A GENERIC FAULT DETECTION AND DIAGNOSIS APPROACH FOR PNEUMATIC AND ELECTRIC DRIVEN RAILWAY ASSETS

by

HAO BAI

A thesis submitted to

The University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

Electronic, Electrical and Computer Engineering
School of Engineering
The University of Birmingham
July 2010

Abstract

The railway assets studied in this project, are those widely distributed pieces of equipment that are critical to the dependable operation of the railway system. A failed asset is likely to cause significant delay to rail services, and may even place the system into an unsafe state. A generic fault detection and diagnosis (FDD) solution for a number of railway assets of different types is therefore desired.

In this thesis, five assets, namely the pneumatic train door, point machine and train-stop, the electric point machine and the electro-hydraulic level crossing barrier, are considered as case studies. Based on their common dynamic characteristics, these assets are also known as Single Throw Mechanical Equipments (STMEs). A generic FDD method is proposed for these STMEs, which consists of sensor inputs and pre-processing, fault detection processes and fault diagnosis processes. A generic model, composed of a series of sub-models, is constructed to describe the behaviour of each asset. The results of fault detection approaches indicate that the proposed method has good performance and is generically applicable to the five assets. Two fault diagnosis methods using fault model and residual analysis are proposed and the fault model based fault diagnosis is preliminarily approached. Finally, a new three level architecture for railway condition monitoring is discussed for practical applications.

Dedicated with love and thanks to

my parents and my wife Wenjing Jia

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, Dr. Clive Roberts, for his strong support and patient guidance over the past years. I benefited significantly from his rich knowledge and experience. Special thanks to Mr. Andy Dunn for his technical support and contribution to my lab work.

Thanks to Mr. Rhys Davies for his support to my working environment in the office. In addition, I would like to thank all the members of the group I have worked with: Dr. Stuart Hillmansen, Dr. Paul Weston, Dr. Hans Chou, Dr. Jiehua Chen, Mr. Fan Bo, Mr. Shaofeng Lu and Mr. Lei Chen. Thank you all for you help and support over the last three years. I would also like to thank Ms. Mary Winkles for taking care of the administrative issues.

Finally, a big thank you is given to my wife, Wenjing Jia, and my family for their constant support and encouragement, which is never less significant for this thesis.

Table of Contents

Chapter 4

A generic fault detection and diagnosis approach for pneumatic train door

Chapter 5

Generic fault detection approaches for other STMEs

Chapter 6

A distributed condition monitoring architecture for railway assets

Chapter 7

Conclusions and further work

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Background

In any industrial process, it is essential that maintenance is provided to ensure that the equipment runs safely and normally. Properly maintained industrial plants have significant benefits, such as higher productivity, equipment which has a longer lifespan and, as a consequence, lower production costs. An effective and efficient maintenance plan requires that information concerning the condition of the equipment can be accessed on a timely basis. In the early 19^{th} century, maintenance was only carried out following a failure as there was a lack of means to understand the status of machinery. Since that time, routine maintenance has been performed in order to find and fix problems before a fault occurs. However, time period based maintenance inspection is still not sufficient, particularly for incipient faults. With the development of electronic technology, a low-cost, on-line condition monitoring system has become realistic for industrial applications. Predictive maintenance is, therefore, achievable via deliberated fault detection and diagnosis (FDD) algorithms.

Railway systems are both safety and time critical. A large number of trackside railway assets, such as point machines and level crossing barriers, contribute to regular train services in a complex operating context. A failed asset is likely to cause a significant delay to rail services, and may even place the system into an unsafe state. Appropriate maintenance for these widely distributed assets is the main concern of infrastructure management, which is currently undertaken by Network Rail in Britain. In 2007/2008,

£1,118 million was spent on maintenance of the UK main line network (Network Rail 2008), mainly through scheduled basis. A time period based regular maintenance regime has been used on the railway systems since the 1950s. A predetermined set of tasks is performed for each asset (Roberts 2007). To date, predictive maintenance has not been fully delivered to the railway system.

In recent years, there have been many studies on predictive (condition-based) maintenance for a wide range of industries, including railway systems, with the aim of increasing the operational reliability of industrial processes (Roberts *et al*. 2001, Lehrasab et al. 2002, Becker and Poste 2006 and Redeker 2006). The most effective maintenance strategy, predictive maintenance, provides continuous condition monitoring of the equipment and any deviation from the desired operating characteristics triggers maintenance requirements. This type of maintenance strategy is especially good at detecting incipient and gradually developing faults, which are difficult to find, even with expert knowledge. Proactive conduct can prevent a total failure of the equipment, which may result in significant costs and unpredictable hazards.

Condition monitoring is the key point of the predictive maintenance strategy. Modern technology, such as sensors, data transmission and computing, has enabled information collection and processing for remote condition monitoring to be carried out at a high speed with an effective cost. For widely distributed railway assets, local networks, i.e. Fieldbus, are capable of organising data from installed sensors and transferring it to a Central Processing Unit (CPU), where the data is processed with FDD algorithms to identify the operating status of the equipment. The classical FDD

process usually consists of analytical models (quantitative or qualitative) and relevant detection and diagnosis algorithms. Data from sensors is validated against the models to check any existing deviation and the results are used to analyse the health status of the equipment. If significant inconsistency is detected, further fault diagnosis will be carried out.

Based on the features of railway assets, a concept of $(RCM)^2$ was proposed to combine the concepts of reliability centred maintenance and remote condition monitoring (Roberts and Fararooy 1998). The underlying intention of this concept is to realise automated condition-based maintenance through remote condition monitoring techniques, whereby the maintenance costs are reduced and the safety of rail services is reinforced. As an initial basis of $(RCM)^2$, a comprehensive Failure Mode and Effects Analysis (FMEA) is usually developed for a specific asset using expert knowledge and accumulated experience.

1.2 Motivation and objectives

In the railway system, there are a large number of safety and dependability critical assets spread over a wide geographical area. It is essential to keep these assets working in a healthy condition by using proper maintenance systems. Currently, a scheduled maintenance regime is carried out; however, defects are commonly found in the following areas:

- It is impossible to be aware of the onset of faults that occur between scheduled maintenance inspections.
- Some faults cannot be found by a visual inspection.

- If a sudden failure occurs, maintainers may be required to work during traffic hours to provide emergency maintenance. This could create a health and safety risk to the personnel.

- The tasks performed by routine maintenance are normally predetermined, which may result in inadequate preparation for unusual maintenance requirements.

- Due to the intermittent nature of some faults, the asset is often found fault-free when tested by the maintenance staff; however, the asset may fail after a random period of time. This number of instance of 'Tested OK on arrival' (TOK) is high in railway infrastructure maintenance.

- Since the assets are distributed over a wide area, a large number of maintainers are often required to take care of them. In the absence of detailed fault information, an asset sometimes has to be inspected repeatedly due to either inadequate preparation or failed fault identification. Consequently, the financial and time costs of maintenance are increased significantly, while the reliability of the railway system decreases.

Failed railway assets can lead train traffic into an unforeseen situation which places passengers in danger. Abrupt asset failures and emergency maintenance often significantly delay train services. A penalty charge of up to £120 per delayed train per minute is currently applied. The financial loss can be huge for the infrastructure owner, particularly when main line traffic is disrupted. It is therefore obvious that a more effective and efficient maintenance methodology is required to ensure the reliability and safety of the railway system.

Previous studies have been carried out on fault detection and diagnosis to railway assets through condition monitoring. Lehrasab (1999) presented Single Throw Mechanical Equipment (STME) concept-based approaches on the modelling and diagnosis of pneumatic train doors, train-stops and point machines. Roberts (2007) discussed the methodology of fault detection and diagnosis on several railway assets outlined in this thesis and presented practical applications and results. As a closed loop controlled asset, the electric train door was studied by Dassanayake (2001). For the electro-hydraulic level crossing barrier, various approaches can be found in Suthasinekul *et. al.* (1976), Brinkmann and Spalmann (1996), Yazdi *et. al.* (1998), Nash and Roberts (1999), Garcia Marquez *et. al.* (2007) and Ishak *et. al.* (2008). Strategies were developed for the health state detection of electric point machines, such as statistics based RCM^2 algorithms (Garcia Marquez and Pedregal 2007), Principal Component Analysis (PCA) (Garcia Marquez 2006), unobserved component models for wear in point machines (Garcia Marquez et. al. 2007) and qualitative presentation of parameter trends (Silmon and Roberts 2006). However, most of these studies were related to a specific asset rather than producing a generic solution to a class of similar railway assets, which is pursued in this study.

The objective of this study is to investigate a generic fault detection and diagnosis method for a number of multi-type simple railway assets. The benefits are:

- Condition monitoring, integrated with a FDD algorithm, is capable of collecting fault information upon the occurrence of a fault and diagnosing the fault for appropriate maintenance.

- Intermittent faults can be detected and diagnosed in time, which reduces the

occurrence of ‘TOK’.

- Incipient faults can be detected and relevant maintenance can thus be carried out before a failure occurs.

- With the ability to monitor multiple types of asset, the condition monitoring system is simplified and therefore the capital cost is reduced.

- The proposed generic FDD solution is able to replace a series of specifically designed algorithms. The computation required can also reduce by integrating the generic FDD software into one centralised computation unit.

1.3 Thesis structure

The work, completed to achieve the above objectives, is presented in this thesis, and the structure is outlined as follows:

Chapter 1 - Introduction to the study. The background, motivation and objectives of the project are presented.

Chapter 2 - Definition and techniques of fault detection and diagnosis. An overview is provided of the development of fault detection and diagnosis methodology over the last few decades. As analytical methods, the model-based mathematical (quantitative) and artificially intelligent (qualitative) fault detection and diagnosis methods are discussed and compared.

Chapter 3 - Five railway assets are considered in this study, namely the pneumatic train door, point machine and train-stop, electric point machine and

electro-hydraulic level crossing. The assets are introduced and the lab-based test rigs for data acquisition are described in detail. Data collected from these assets are presented. The definition of Single Throw Mechanical Equipement (STME) is introduced, and the relevance of this definition to the case studies is explored. As a classification of STME, common features are abstracted from the five railway assets, based on which a generic fault detection and diagnosis method is proposed for condition monitoring. Based on statistics theory, an adaptive thresholding algorithm is proposed for residual generation. Two fault diagnosis approaches are presented and discussed.

Chapter 4 - A generic fault detection and diagnosis approach is presented for the pneumatic train door. As the essential part of FDD, the modelling work is illustrated. The results of fault detection using an adaptive threshold algorithm are presented. The results are analysed and the performance of the proposed generic fault detection and diagnosis method is proved to be good. A fault diagnosis approach using a fault model is presented and discussed.

Chapter 5 - Four case studies for other assets considered in this thesis are provided. The generic fault detection method is applied for the other two pneumatic railway assets, the train-stop and the point machine, as for the pneumatic train door. Fault detection results are presented, by which the fault detection method is proved to be good and generic for the pneumatic assets. Initial fault detection approaches for the electric point machine and

electro-hydraulic level crossing are presented with results. The feasibility of the generic fault detection method to the electric and electro-hydraulic assets is discussed and confirmed.

Chapter 6 - A distributed condition monitoring architecture for simple multiple railway assets is described. A three level condition monitoring architecture for the assets is introduced and discussed. Based on a generic fault detection and diagnosis solution, a more economic three level architecture is proposed. Digital communication networks suitable for local condition monitoring networks are also introduced.

Chapter 7 - Conclusions of this study and discussion of further work.

Chapter 2

Fault detection and diagnosis methods

2.1 Introduction

In order to achieve reliability, maintainability and safety in industrial processes, fault detection and diagnosis (FDD) technology has been rapidly developed and improved over the last four decades. In the 1970s, initial FDD applications in chemical and industrial plants used threshold testing to check system data. Using this method, a fault can be detected when a measured value crosses a given threshold (Vaclavek 1974 and Himmelblau 1978). This classical limit-value-based method is simple and reliable; however, it only responds to a relatively large change to a feature, therefore a detailed fault diagnosis becomes impossible (Isermann 1997).

With increasing system complexity and requirements for reliability, a quantitative model of a practical system was required and many investigations were therefore made using analytical approaches during the 1980s and 1990s. The idea was to generate signals that represent inconsistencies between normal and faulty system operation (Patton, Lopez-Toribio and Uppal 1999). Based on analytical model, the algorithms, such as observers (Chen and Patton 1999), parity equations (Gertler 1998) and parameter estimation (Isermann 1994a), were designed for inconsistency signal generation (also known as residuals generation). These model-based FDD methods have been widely implemented in many industrial fields, such as nuclear power plants (Lee *et al.* 2006), railway vehicles (Li and Goodall 2003 and Li, *et al.* 2007), jet engines (Patton and Chen 1997) and electrical machines (Combastel *et al.* 2002). In

some safety-critical industrial systems, e.g. nuclear reactor, aircraft or fast rail, high cost hardware redundancy integrated with analytical methods is applied to avoid incident when a fault occurs (Patton and Chen 1997).

More recently, modern computing and analysis methods, e.g. neural networks, fuzzy logic and pattern recognition, have been investigated as powerful modelling and decision making tools. A survey of Artificial Intelligence (AI) approaches to FDD was given by Patton (1999). Neural networks can be used for continuous linear and non-linear systems modelling, where the model itself has the potential to be improved by learning from new input and output information from a real system. Thus, some of the difficulties of using mathematic models are overcome, which makes neural network based FDD algorithms more applicable to real systems. However, neural networks are generally treated as black-box models, which indicates that it is not easy to achieve insight into the behaviour of the models. To obtain maximum benefit, neural networks are usually combined with qualitative models or inference methods, e.g. fuzzy logic, to enhance the diagnostic reasoning capabilities.

This chapter provides an introduction to fault detection and diagnosis methodology by means of a literature review. The FDD methods are generally classified into two categories: quantitative FDD methods, including observers, parity equations and parameter estimation; and qualitative methods, including neural networks, fuzzy logic and neuro-fuzzy systems. The comparison of three quantitative methods is also provided.

2.2 Fault detection and diagnosis methodology

A *fault* is defined as an unpermitted deviation of at least one characteristic parameter

of a system from normal (healthy) status (Isermann 1984, 1997). A system which has the capacity of detecting, isolating and identifying faults is called a *fault detection and diagnosis* system (Patton *et. al.* 1999). A *failure* is defined as the state of a permanent invalidation of a system to perform normal functions. The system mentioned above is often a machine/plant/network or a railway asset in this thesis.

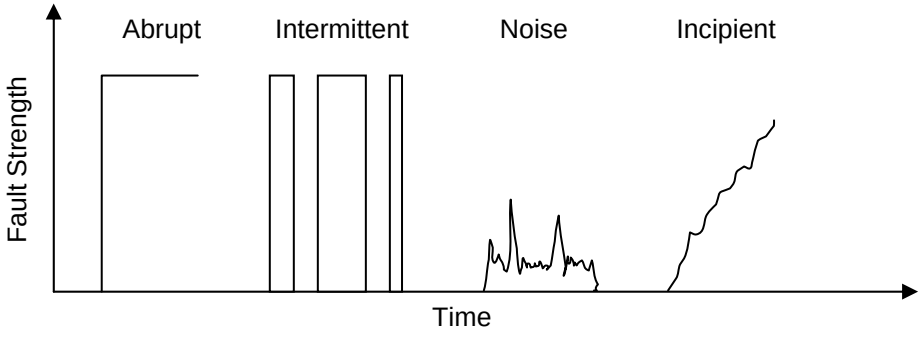


Figure 2.1 Typical temporal faults.

Time-varying faults leading to the failure of a system are mainly classified into four categories as shown in Figure 2.1. When abrupt faults occur, the system jumps from healthy to failure without a sign. This type of fault can not usually be predicted and often happens due to the sudden failure of an important component. Intermittent faults usually happen when electric connections are unstable. A long term observation would help to identify this fault. Noise (disturbance) is defined as an unknown extra input to a system, which may randomly cause malfunctions or failure. This class of fault may be statistically defined; however, it is very difficult or impossible to predict. Therefore, the fault detection and diagnosis methods need to be robust to these normal system variations, as the behaviour of the system changes with the presence of this extra input. Incipient faults are typical in mechanical systems, where faulty elements deviate from their behaviour in a gradual manner (drift). A certain threshold can be set

for detection of this type of fault, and a decision can be made when the threshold is exceeded. The tracking of the deviation can also provide information for early fault prediction.

The fault categories mentioned above are based on the time-varying characteristics of a system. Regarding the effect of faults on a process, they can also be classified as: additive faults and multiplicative faults (Isermann 1997 and Dassanayake 2001). Additive faults are caused by unknown inputs or disturbances. The outputs are changed independently from the known inputs. Sensor and actuator faults are normally modelled as additive faults. When unknown variables are added to a system and the physical parameters change accordingly, the resulting faults are referred to as multiplicative faults. With the occurrence of this type of fault, the outputs of a system change depending on the magnitude of these unknown variables. As subsets of additive faults, sensor faults and actuator faults are common in industrial plants (Frank 1990). Sensor faults generate fake deviations between actual system values and measurements. In the same way, actuator faults cause discrepancies between the actual performance of actuators and the expected performance by the commands.

As mentioned in the definition of fault detection and diagnosis systems, a FDD system theoretically consists of three functions: fault detection, fault isolation and fault identification. Fault detection (FD) is referred to as the capability to observe something wrong and determine a fault which is occurring in a system. The fault can then be classified and located by the isolation function. Based on the knowledge abstracted from the aforementioned two steps, the fault identification function implements the identification of fault strength. In practice, only fault detection and

isolation are normally included since fault strength identification is often too difficult to apply.

A FDD system is often required to have a set of desirable characteristics which offer a reliable, safe and efficient target system. Quick detection and accurate diagnosis are normally required as key points of a good FDD system (Isermann 1997 and Dash and Ventkatasubramanian 2000). However, timely detection is often a trade off between robustness and performance, since high speed execution is always sensitive to high frequency influences. Therefore, desensitisation to disturbances/noises is a must. Furthermore, an accurate diagnostic algorithm also needs to overcome model uncertainties in order to correctly classify multiple faults. In practical FDD applications, a priori estimations on classification error are expected to enhance reliability and provide more information to users (Ventkatasubramanian 2003a). The robustness of a diagnostic system is the key point which can directly affect performance under different practical or environmental conditions.

2.2.1 General model-based FDD algorithm

The aim of model-based fault diagnosis is to generate information about faults which have occurred in target systems using actual measurements, as well as the quantitative, qualitative or combined system model. The model-based method is referred to as an analytical method, which is low-cost compared to hardware redundancy in some safety-critical applications, provided that a model can precisely simulate the behaviour of a real system. Typically, the target system is considered as a continuous-variable dynamic system, which has an input u and an output y, with an unknown fault f. Figure 2.2 illustrates the conceptual structure of a model-based fault detection and

diagnosis system.

The inconsistency between the model and the actual system (also known as residual), *r(t)*, is generated by the following equation:

$$r(t) = Hy(t) + Hy_m(t) \tag{2.1}$$

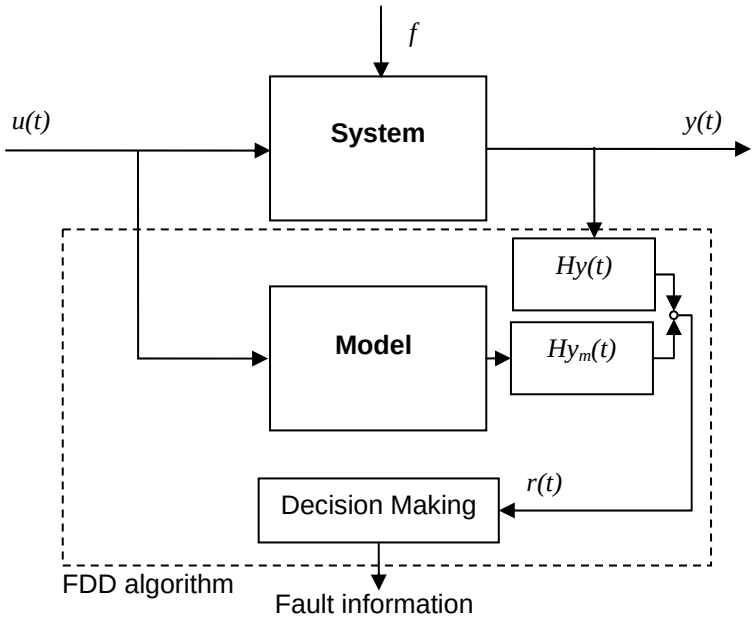


Figure 2.2 Concept of model-based fault detection and diagnosis.

H represents the manipulation function of the system output, *y(t)*, and model output, $y_m(t)$. The status of the system can be observed by *r(t)*:

$r(t) = 0$, fault free;

$r(t) \neq 0$, fault occurs.

In theory, the residuals must be either zero in a fault free case, to indicate that no fault occurs, or non-zero in the case of a fault. However, in practice, deviations normally exist with different magnitudes. A threshold is, therefore, required for sensitivity adjustment. The value at which a threshold is set determines whether the FDD system has enough sensitivity to detect a fault or not. The balance is a trade-off between

detection accuracy and false alarm frequency.

Residuals represent the likelihood of fault occurrence, and the decision making system (Figure 2.2) provides a decision rule to examine whether these residuals are indicating a fault (Patton *et. al.* 1999). In addition to the threshold testing method for fault detection described above, statistical and inference methods can also be used to make a decision (Frank and Ding 1996). As shown in Figure 2.2, decision making system generates fault information by investigating and analysing a set of residuals.

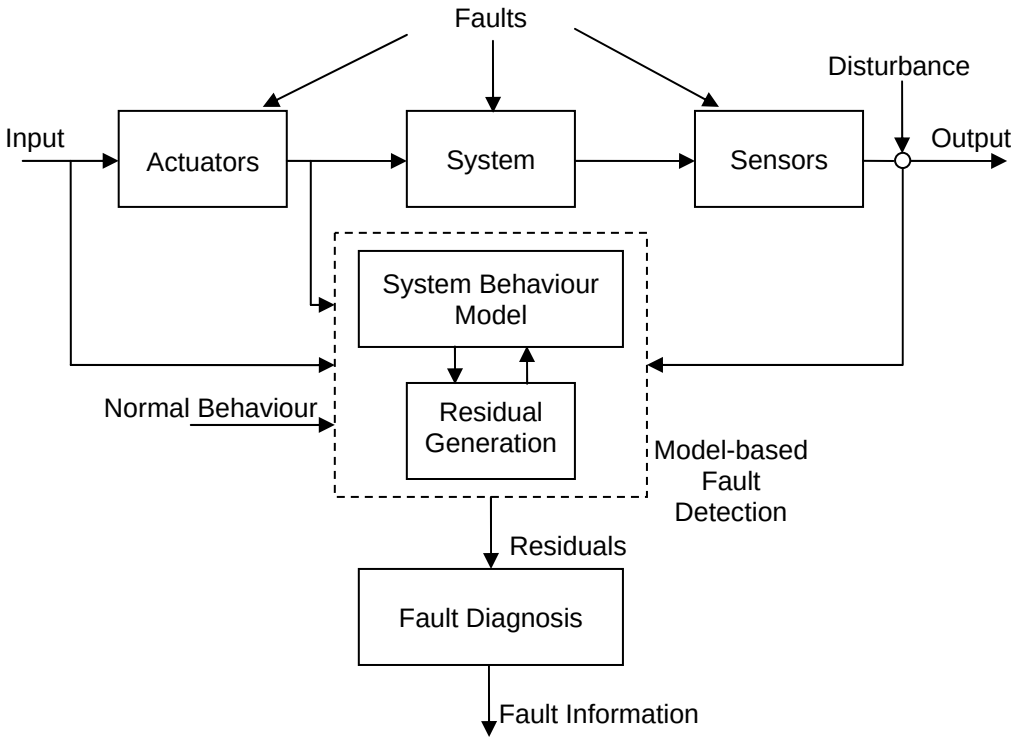


Figure 2.3 General scheme of model-based fault detection and diagnosis (Isermann 2005).

A general scheme of model-based fault detection and diagnosis is shown in Figure 2.3. In this figure, the whole system consists of actuators, the target system and sensors, where the faults can be grouped as actuator faults, system faults or sensor faults. Disturbance (noise) is added on the sensor output. Both input and output are physical

measurements, which are compared with the prediction from the system behaviour model for residual generation. The residual generator aims to produce a set of inconsistencies to indicate whether a fault is present. Normal behaviour information is used as an input to the system behaviour model to detect any change in system features and to produce symptoms to aid further diagnosis. Basically, an intact FDD system includes three stages (procedures) with different functions: system modelling, residual generation and fault diagnosis. Firstly, a precise mathematical model is required to predict system performance. For most systems, precise mathematical models are often very difficult to obtain. The robustness of the FDD scheme is often achieved by designing algorithms where the effects of model uncertainties and unmodelled dynamic disturbances on residuals are minimised and sensitivity to faults is maximised (Patton *et. al.* 1999). Secondly, a set of residuals is generated to represent the deviation between actual and nominal features. Finally, the residuals are evaluated to relate to certain faults and to locate the fault if it is present. The model implementation and residual generation compose the model-based fault detection system.

2.2.2 Quantitative FDD methods

This section reviews quantitative, model-based residual generation methods. The most frequently used FDD approaches, including observers (state estimation), parity equations and parameter estimation, are described and the three methods are compared. The mathematic descriptions of both observers and parity equations can take the form of either continuous or discrete time equations. In this section, the observers are described in continuous and parity equations in discrete time to illustrate the basic principles.

2.2.2.1 Observer based methods

Diagnostic observers appeared in the early 1970s and have been well developed in the last thirty years. The algorithms of residual generation in this model-based method are based upon the calculation of estimation error using observer equations, which are applicable to both continuous and discrete systems (Chen and Patton 1999). A detailed survey of this technique has been given by Frank and Ding (1997).

The concept of the observer-based method is to observe any change to the system states by state estimation. A continuous, time variant linear system with faults is assumed in the form of a State-Space model (Frank and Ding 1997 and Patton 2000).

$$\dot{x}(t) = Ax(t) + Bu(t) + R_1 f(t) \tag{2.2}$$

$$y(t) = Cx(t) + Du(t) + R_2 f(t) \tag{2.3}$$

Where $x(t) \in R^n$, is the state vector; $u(t) \in R^k$, is the system input vector, whilst $y(t) \in R^m$, is the output vector; $f(t) \in R^j$, is the fault vector in the temporal domain ($f_i(t)$ can represent a set of faults with i=1,2,…,j, if they are present); R_1 and R_2 are matrices representing how the fault relates to the system. An observer is then required to estimate a set of states, which can be written as follows:

$$\hat{\dot{x}}(t) = A\hat{x}(t) + Bu(t) + K\left[y(t) - \hat{y}(t)\right] = (A - KC)\hat{x}(t) + (B - KD)u(t) + Ky(t) \tag{2.4}$$

$$\hat{y}(t) = C\hat{x}(t) + Du(t) \tag{2.5}$$

where $\hat{x}(t)$ is the estimated state vector and $\hat{y}(t)$ is the observer output vector; $K \in R^{n \times m}$ is the designed feedback gain matrix of the observer. Therefore, the state estimation error can be described with the following equation.

$$e(t) = \hat{x}(t) - x(t) \tag{2.6}$$

Where $e(t)$ satisfies

$$\dot{e}(t) = (A - KC)e(t) + (KR_2 - R_1)f(t) \tag{2.7}$$

The residuals, $r(t) \in R^p$, can then be obtained.

$$r(t) = W(\hat{y}(t) - y(t)) = WCe(t) - WR_2 f(t) \tag{2.8}$$

Where $W \in R^{n \times m}$ is the residual weighting matrix. This equation shows that the residuals are sensitive to estimation errors and fault signals. In a fault-free case, where $f(t)=0$, the $e(t)$ should converge to zero as $t \to \infty$, if the absolute values of the Eigenvalues of WC are less than 1 (Venkatasubramanian 2003a).

2.2.2.2 Parity equations

The aim of parity equations is to compare the parity (consistency) of the analytical models with the actual outputs (measurements from sensors) of a real system. In theory, the result of parity equations (residuals) is zero under fault-free conditions, where an accurate and robust system model is a must. In the form of a State-Space model, a discrete system is illustrated as (Chow and Willsky 1984):

$$x(k+1) = Ax(k) + Bu(k) \tag{2.9}$$

$$y(k) = Cx(k) + Du(k) \tag{2.10}$$

where $x(k) \in R^n$, is the state vector; $u(k) \in R^p$, is the input vector, whilst $y(k) \in R^m$, is the output vector; $x(k+1)$ represents the state of the system at time $k+1$. A, B, C and D are constant matrices to relate output to input parameters. To produce residual signals temporally, an output, y, of the model at time, k, is deduced. The output at time, $k+1$, is:

$$y(k+1) = CAx(k) + CBu(k) + Du(k+1) \tag{2.11}$$

For any $\Delta t > 0$, the output at $k+\Delta t$ is:

$$y(k+\Delta t) = CA^{\Delta t}x(k) + CA^{\Delta t-1}Bu(k) + CA^{\Delta t-2}Bu(k+1) + ... + CBu(k+\Delta t-1) + Du(k+\Delta t) \tag{2.12}$$

where Δt =0,1,…, n. Writing the above equation in a compact form as:

$$Y(k) = Qx(k-n) + RU(k) \tag{2.13}$$

Therefore, residuals generated at time, k, can be written as (Yu *et al.* 1996 and Ding *et al.* 1999):

$$r(k) = W(Y(k) - Y_m(k)) \tag{2.14}$$

where $Y_m(k)$ is the measurement of system outputs; W is a vector for residual generating manipulation. A well designed residual generation vector is selected to achieve a specific structured residual response to faults and to decouple from unknown disturbances and model uncertainties.

2.2.2.3 Parameter estimation

The parameter estimation approach detects faults by the estimation of parameters within a dynamic system, where the faults are assumed to be reflected by these features (Isermann 1984, Isermann 1994, Chen and Patton 1999 and Patton *et. al.* 1999). The system parameters can be classified as physical and abstract parameters, which directly and indirectly represent the status of a real system, respectively. As, in most practical cases, these parameters are not obtained, parameter estimation methods are applied by measuring the input and output signals, provided that the first principle (physical principle based) model is well known (Isermann 2005). However, due to the difficulty of constructing an accurate model for a complex non-linear system, the application of this method is often restricted to simple linear systems.

A single-input, single-output, time variant, linear model around a steady state operating point (Y_{00}/U_{00}) can be described by a differential equation (Isermann 1997 and Patton 1989).

$$a_0 y(t) + a_1 y^{(1)}(t) + ... + y^{(n)}(t) = b_0 u(t) + b_1 u^{(1)}(t) + ... + b_m u^{(m)}(t) \tag{2.15}$$

$$y(t) = Y(t) - Y_{00}; u(t) = U(t) - U_{00} \tag{2.16}$$

Where *a* and *b* are model parameters and $y^{(n)}(t) = d^n y(t) / dt^n$. In compact form, the equation 2.15 can be written as:

$$y^{(n)}(t) = \psi(t)\theta + e(t) \tag{2.17}$$

where *e(t)* is the equation error, regression $\psi(t)$ and the parameter vector θ are:

$$\psi(t) = [-y(t), -y^{(1)}(t), ..., -y^{(n-1)}(t), u(t), u^{(1)}(t), ..., u^{(m)}(t)] \tag{2.18}$$

$$\theta = \left[a_0, a_1, ..., a_{n-1}, b_0, b_1, ..., b_m\right]^T \tag{2.19}$$

The measurements of input and output are at discrete time $t=k\Delta t$, where k=0,1,…,N, is the number of sampled data. The N+1 equation is as:

$$\hat{y}^{(n)}(K) = \psi(K)\hat{\theta} + e(K) \tag{2.20}$$

where $\hat{\theta}$ is the estimated parameter vector and $K = (N+1)\Delta t$, which can be obtained by the loss function, *J*.

$$J = \sum_{k=0}^{N} e^2(k) = \sum_{k=0}^{N} (\hat{y} - y)^2 = \sum_{k=0}^{N} (\psi(k)\hat{\theta} - y(k))^2 \tag{2.21}$$

The $\hat{\theta}$ is, therefore, estimated using the well-known least-squares (LS) algorithm (Mandel 1964 and Ljung 1987).

$$\begin{aligned} &\hat{\theta} = [\Psi^T \Psi]^{-1} \Psi^T Y \\ &\Psi = [\psi(0), \psi(1), ..., \psi(N)], Y = [y^{(n)}(0), y^{(n)}(1), ..., y^{(n)}(N)] \end{aligned} \tag{2.22}$$

For the linear model as mentioned above, a LS algorithm is normally applicable for minimising the loss function, *J*, and therefore the estimated parameter vector $\hat{\theta}$ can be

obtained. However, this method is generally not valid for non-linear models, except a class of which have the linearity characteristic in parameters and the non-linear functions are known (Gertler 1998). The non-linear model with features of a linear model in the parameters can be described by the following equation.

$$y(t) = a_1 f_1[u(t)] + a_2 f_2[u(t)] + ... + a_r f_r[u(t)] + \varepsilon(t) \quad (2.23)$$

Where $f[u(t)]$ is a known non-linear function with input *u(t)*; $a_1, a_2 \ldots a_r$ are unknown parameters and $\varepsilon(t)$ is the equation error. In this case, the regression $\psi(t)$ and parameter vector θ can be described as:

$$\psi(t) = \left[f_1[u(t)],\ f_2[u(t)], ...,\ f_r[u(t)] \right] \quad (2.24)$$

$$\theta = \left[a_1, a_2, ..., a_r \right]^T \quad (2.25)$$

The output is therefore predicted by the following equation:

$$\hat{y}(t) = \psi(t)\hat{\theta} \quad (2.26)$$

The estimated parameter vector can also be obtained using the LS algorithm as in equation 2.22.

2.2.2.4 Comparison of the three methods

The three model-based residual generation methods have been well developed over several decades, the research and application of which can be found in literature such as observer-based approaches in Frank (1996), Frank and Ding (1997) and Patton (1997), parity equations in Chow and Willsky (1984), Patton (1994) and Kosebalaban and Cinar (2001) and parameter estimations in Isermann (1997, 2003a) and Patton (1999).

In various aspects, these methods have similar and different characteristics, which can

be illustrated as follows:

– The observer-based approach realises fault detection through observing the change in system state, which is also known as a state redundancy. The parity equations method functions by identifying the inconsistencies between the analytical model and the real system. The aim of parameter estimation is to detect the variety in the physical parameters of the system by estimation.

– Fundamentally, the observer-based and parity equation methods have an equivalent structure, which results in the possibility that the parity equation method can be transformed into the form of observer structures in some specific conditions (Chen and Patton 1999), i.e. parity equations can be transformed to observer representation by use of linear filter (Gertler 1991).

– The model structures for the three methods are all required to be well known. The system parameters are also necessary for the first two methods that also require the analytical (mathematic) model to fit the process well (Isermann 1994).

– For non-linear systems, these methods can be applied; however, restrictions exist. Non-linear observers and non-linear parity equations are only applicable to a particular class of non-linear circumstance. Parameter estimation can be applied to non-linear systems, provided that the parameters have a linear feature and the non-linear functions are known.

– The performances of the three fault detection methods highly depend on their designs for different applications. The three methods, with appropriate designs, have their own characteristics. In comparison to parameter estimation, observers

and parity equations have a very fast response to sudden faults. The computation demand of these two methods is moderate, which makes on-line real-time application feasible. Parameter estimation is relatively slow since a large number of data samples are required to make the estimation accurate. However, this method can detect small changes in the system, irrespective of whether they are slow or fast developing faults, since the parameter values directly indicate the fault strength. A pre-condition of applying parameter estimation is that the system must stay dynamic; otherwise the estimation result may drift to unpredictable values.

With the strict requirement of an explicit model for the target system, these traditional methods are often difficult to implement in practical tasks, especially for those complex highly non-linear objects. In recent years, artificial intelligence (AI) techniques have been introduced to the fault detection and diagnosis field either as a system modelling tool or as a residual classifier.

2.2.3 Qualitative FDD methods

In this section, the applications of AI techniques, such as neural networks and fuzzy logic, in the fault detection and diagnosis field are introduced. The combination of these two methods, the neuro-fuzzy technique, is also reviewed.

2.2.3.1 Artificial neural networks

An artificial neural network (ANN) is a processing system with a number of interconnected elements, which are called neurons (Winston 1992 and Patton *et. al.* 1999). This technique was initially developed in the late 1940s in order to model

certain aspects of the function of a human brain. A neuron has only a very simple specific function and structure; however, a large number of connected neurons in parallel provide huge processing power. Each neuron can be thought of as a mathematical function, which connects to other neurons to map the inputs and outputs within a network (Patton *et. al.* 1999).

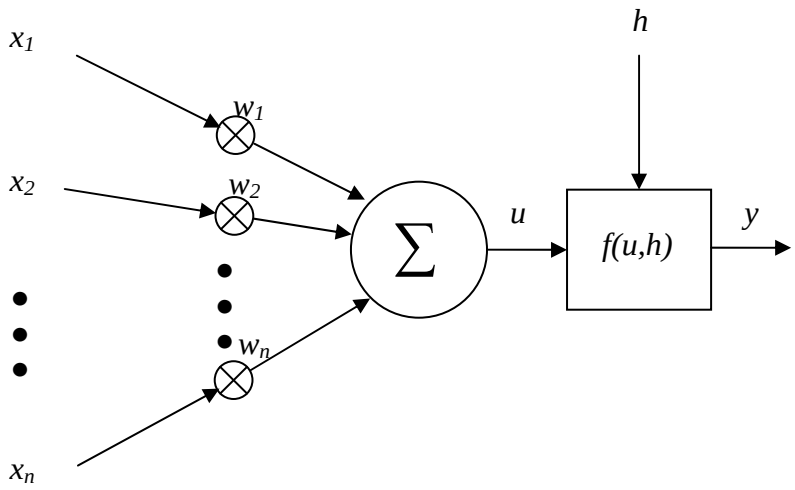


Figure 2.4 Concept of an *n*-input neural network processing unit.

Figure 2.4 illustrates a typical neuron in a neural network with n inputs, where x_i (i=1,2,..,n) is input; y is output; w_i (i=1,2,…,n) is connection weight to corresponding x_i; h is threshold; f is activation function and u is the summation of weighted inputs.

A neuron has a series of input signals which are changed when they travel along the connections by combining with the connection strength (weight). After gathering the inputs from connections, an activation value is then computed using an activation function. The output of this unit, y, is written in mathematical form as:

$$y = f(u,h) = f(\sum_{i=1}^{n} x_i w_i, h) \tag{2.27}$$

The threshold relation for obtaining the output, *y*, can be written as:

$$y = \begin{cases} 1, & if \quad u \geq h \\ 0, & if \quad u < h \end{cases} \tag{2.28}$$

Function 2.28 presents the 'all-or-nothing' characteristic of the processing unit threshold function (Gurney 1997). Apart from the step function (2.28), other activation functions are also popular, for example a sigmoid function. The sigmoid function (2.29) makes the output, *y*, smoothly depend on *u*.

$$y = f(u,h) = \frac{1}{1+e^{(h-u)/\rho}} \tag{2.29}$$

Where $e \approx 2.7183$ is a constant and ρ determines the shape of the function.

A neural network consists of a number of connected neurons. Depending on their functionality, the neurons are organised as three layers: input layer, hidden (middle) layer and output layer as shown in Figure 2.5, where the hidden layer may be more than one layer. A fundamental property of neural networks is the ability to learn. The learning process requires a set of training data, which include desired inputs and outputs that reflect the behaviours of a system. With feedback loops, the network trains itself to be a close approximation of the actual. In the fault detection and diagnosis field, neural networks are normally applied to system modelling and fault classification.

Compared to analytical models, neural network modelling does not require explicit information about the physics of the real system. The advantage is that it can automatically learn the system by extracting the system features from historical training data. The neural network can handle both linear and nonlinear systems, although it is not efficient for linear systems. The capability of simulation of non-

linear behaviours with given suitable weighting factors and network structure is the most important feature of neural networks (Narendra and Parthasarathy 1990 and Narendra 1996).

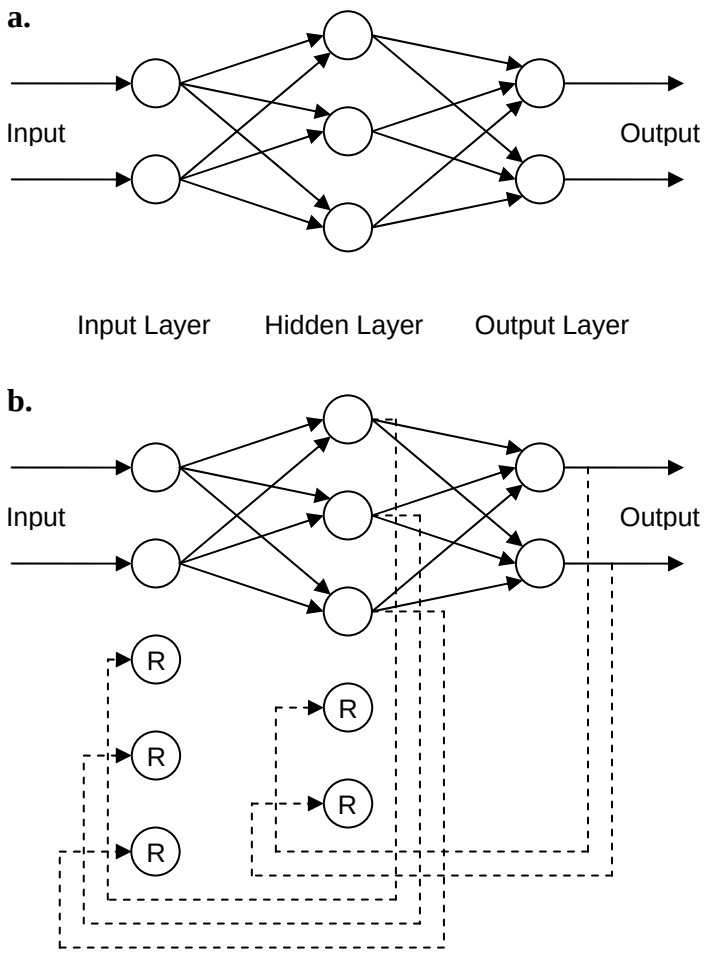


Figure 2.5 Three-layer feed forward (a.) and recurrent (b.) networks.

In Figure 2.5, the architecture of a three-layer neural network is displayed. For system modelling, the two main classes of neural network can be considered: multi-layer feed forward network (Figure 2.5 a) and recurrent network (Figure 2.5 b). The first network maps the linear and non-linear relations with activation functions at each neuron. However, this is static, rather than dynamic, mapping. Narendra and

Parthasarthy (1990) investigated the possibility of representing non-linear dynamic systems using the feed forward network combined with linear dynamic systems. With the feed-back character, the recurrent network is applicable to dynamic non-linear modelling. The deviation from the desired output is propagated back to re-input to neurons at the middle and/or input layers for network behaviour adjustment.

A non-linear dynamic system can be generally described as:

$$y(k) = F(y(k-1), \cdots, y(k-n), u(k), \cdots, u(k-n)) \tag{2.30}$$

where $u(k) \in R^i$ is the input vector, $y(k) \in R^j$ is the output vector, F is the non-linear function and n is the system order. A one-step prediction model of this system using a feed forward neural network can be presented in the form of an equation. Time delay units are also employed for output prediction, which simulates the feed back function to make the feed forward neural network self-adjustable.

$$\hat{y}(k) = NN(W, y(k-1), \cdots, y(k-n), u(k), \cdots, u(k-n)) \tag{2.31}$$

Where W is the weight matrix. Therefore, the residual function is:

$$r(k) = \hat{y}(k) - y(k) \tag{2.32}$$

During training, the weight W is adjusted to reduce the inconsistency from measured outputs. The error function is as the following function.

$$E = \frac{1}{2}\sum_{k=1}^{m}(\hat{y}(k) - y(k))^2 \tag{2.33}$$

Where E is the sum of the squared error and m is the number of training tasks.

Typically, neural network architectures for system modelling include feed forward networks, Radial Basis Function (RBF) networks and dynamic neural networks, because of their powerful approximation and generalising abilities (Patton *et. al.* 1999).

Fault classification is a logic decision-making process to transform quantitative knowledge into qualitative statements that determine fault occurrence and location (Korbicz *et. al.* 2004 and Chen and Patton 1999). For various fault situations, a fault classifier is used to classify the generated symptoms into corresponding distinguishable patterns. The capability of generating arbitrary regions in space makes neural networks feasible for this task (Cybenko 1989). In fault diagnosis, the regions may represent different fault types or locations. Based on the architecture and learning algorithms, probabilistic neural networks and radial basis function networks are the most suitable for fault classification (Winston 1992). On the other hand, unsupervised neural networks, such as self-organising neural networks, can also be used for classification, due to their adaptive structure based on the input to the network. To achieve an accurate classification of a fault, the training data must contain all possible faults that may happen in the process.

2.2.3.2 Fuzzy logic

Fuzzy logic is a technique used to deal with ambiguous rather than precise reasoning problems using multi-valued logics derived from fuzzy set theory (Zadeh 1965 and Klir and Yuan 1996). In classic predicate logic, the degree of statement to truth is defined as ‘crisp’ {true, false}, however, in fuzzy logic, it can range from 0 to 1, where the statement is declared based on the closeness to each value. In recent years, the research and application of fuzzy logic to model-based fault diagnosis has been attempted (Chen *et al.* 1996, Ballè and Fuessel 2000 and Evsukoff *et al.* 2000). The main idea of model-based fault detection and diagnosis is to generate the inconsistency, termed residuals, between the model and the actual system as

mentioned in Section 2.2.1. The robustness of the model-based method can be influenced by the accuracy and uncertainty of the model, especially in the case of unknown disturbances, which may lead to the generation of vague residuals and, therefore, difficulty in fault classification. A possible solution is to tune the parameters in the system state observer and controller by estimating the real system outputs (Schneider and Frank 1994 and Patton *et. al.* 1999). On the other hand, the fuzzy logic method is suitable to deal with this type of uncertain situation with known knowledge. With inference by fuzzy logic, ambiguous residuals or structured residuals can be fuzzily isolated, and the degree of possibility of belonging to a certain fault pattern can be deduced.

As mentioned above, the fuzzy logic method is based on fuzzy set theory. Figure 2.6 illustrates how fuzzy set theory is defined (Terano *et. al.* 1991). The rectangular frame represents the whole set ***X***; the dotted circle is the ambiguous border of fuzzy set ***A***; x is an element (member) in ***X***. ***A*** is a fuzzy subset of ***X***. Fuzzy set theory defines the degree of how x belongs to ***A***, where the function to give the degree is called the membership function. The fuzzy set ***A*** can be written as:

$$A = \left\{ \left(x,\ \mu_A(x) \right) \mid x \in X \right\} \tag{2.34}$$

where $\mu_A(x)$ is the membership function to fuzzy set ***A***. Based on the membership function, a one-to-one correspondence for fuzzy sets over the region [0,1] is achievable. An example of residual amplitude classification is used to explain how the membership function works (Figure 2.7).

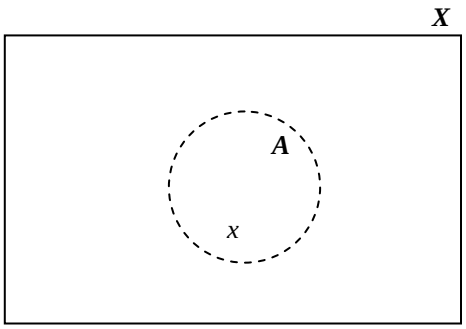


Figure 2.6 Fuzzy set *A*.

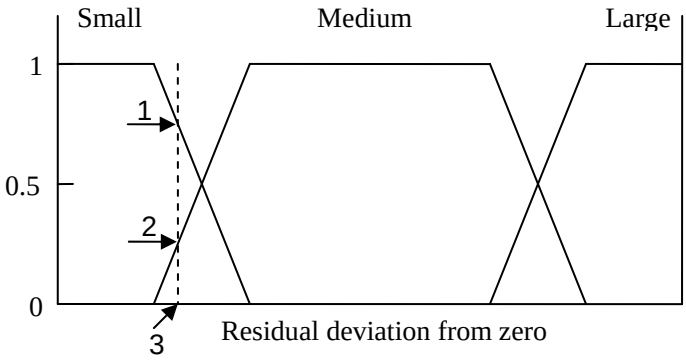


Figure 2.7 Membership functions for residual amplitude classification.

In this figure, the meaning of the expressions, ‘Small’, ‘Medium’ and ‘Large’, is represented by three functions mapping residual amplitude. Each function maps the same residual value into [0, 1]. The dotted vertical line represents a particular value of a residual, which is given three truth values by each membership function correlated to each of the expressions. Three arrows point the truth values obtained from each function: arrow ‘3’ points to zero, which linguistically means ‘not large’; arrow ‘2’ describes the residual amplitude as ‘slightly medium’ and arrow ‘1’ shows ‘fairly small’.

Rule-based mechanisms have been developed for fuzzy inference systems to process fuzzy inputs, which can mainly be classified as: direct and indirect reasoning methods

(Tanaka 1996). The most popular reasoning methods are direct methods. The inference rules to perform fuzzy reasoning are expressed in an IF-THEN format called 'fuzzy IF-THEN rules'.

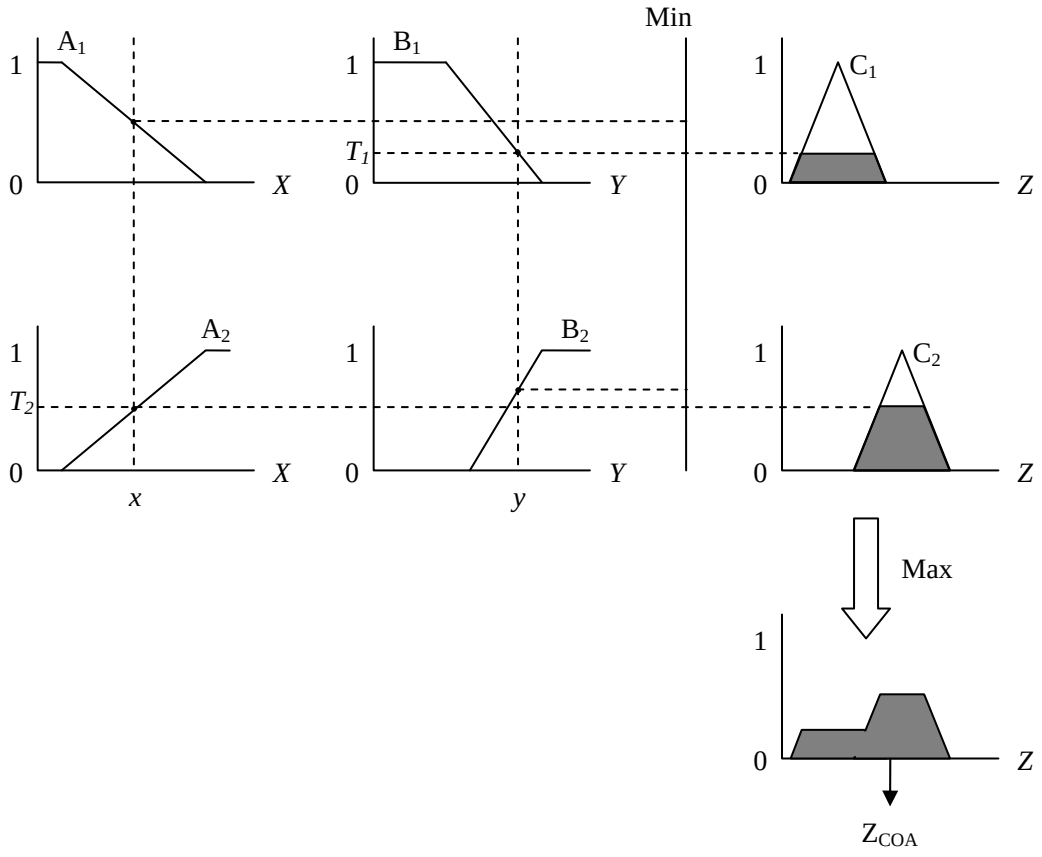


Figure 2.8 Fuzzy inference process using Mamdani's direct method.

An example of a two-input one-output reasoning mechanism using Mamdani's direct method (Mamdani 1975) is presented in Figure 2.8. A_1, A_2, B_1, B_2, C_1 and C_2 are fuzzy sets. *X* and *Y* are inputs and *Z* is the output. *x* and *y* represent two input variables. The two fuzzy inference rules in IF-THEN format are

Rule 1: IF *x* is A_1 and *y* is B_1 THEN *z* is C_1

Rule 2: IF *x* is A_2 and *y* is B_2 THEN *z* is C_2

By the membership functions chosen as trapezoidal shapes, two membership values

for the two crisp inputs *x* and *y* are obtained, where a min-operation, represented by $\wedge$, is applied. The adaptability of rule 1 and 2 is presented as follows.

$$\begin{aligned} T_1 &= \mu_{A_1}(x) \wedge \mu_{B_1}(y) \\ T_2 &= \mu_{A_2}(x) \wedge \mu_{B_2}(y) \end{aligned} \tag{2.35}$$

Apply the above result to the consequence part, C_1 and C_2, to obtain the conclusion of the two rules, which are shown by shaded areas.

$$\begin{aligned} \mu_{C_1'}(x) &= T_1 \wedge \mu_{C_1}(z) \qquad \forall z \in Z \\ \mu_{C_2'}(x) &= T_2 \wedge \mu_{C_2}(z) \qquad \forall z \in Z \end{aligned} \tag{2.36}$$

The final result is then achieved by applying the max-operation, represented by $\vee$, and the reasoning process is completed.

$$\mu_C(z) = \mu_{C_1'}(z) \vee \mu_{C_2'}(z) \tag{2.37}$$

In practice, a defuzzification is applied to convert a fuzzy set to a crisp output in the form of a definite value. For this purpose, there are two common methods (Tanaka 1996). One is to take the centre of gravity of the fuzzy set (Centroid Of Area, COA), which is described by the following function.

$$z_{COA} = \frac{\int_z \mu_C(z) z dz}{\int_z \mu_C(z) dz} \tag{2.38}$$

The other is to take the maximum value of the membership in the fuzzy set.

$$z_{\max} = \left(\max_z \mu_C(z) \right) \tag{2.39}$$

With the ability to handle ambiguous problems by a priori knowledge, the fuzzy logic technique has been widely used in industrial scenarios. For system description, fuzzy logic provides linguistic rules instead of only traditional mathematical methods, which allows heuristic knowledge derived from expert experience to be integrated. As

linguistic expressions are used, the states of a system can be described to engineers usefully and meaningfully. With the fuzzy inference system, fuzzy logic is a convenient method for dealing with imprecise residual data, which is then used for fault detection and diagnosis. For example, fuzzy observers, which combine the fuzzy logic and model-based methods for non-linear dynamic system fault diagnosis, using the Takagi-Sugeno fuzzy model, were described by Patton (1999).

2.2.3.3 Neuro-fuzzy system

A neuro-fuzzy system is a hybrid system combining neural networks and fuzzy logic, where these two methods are complementary to each other (Jang *et al.* 1997). As mentioned in Section 2.2.3.1, neural networks are ideal for non-linear modelling and fault classification, however, they are not good at explaining how decisions are reached, due to their 'black-box' characteristic. Fuzzy logic, which is highly suitable for reasoning with imprecise information, can easily explain the decision making process by linguistic expressions, but the rules can not be obtained automatically for decision making. These limitations, therefore, raise the requirement for hybrid intelligent systems, where two or more techniques are combined together to overcome the drawbacks of each individual method. For instance, a fault detection and diagnosis task normally includes a signal processing task and a reasoning task, where, respectively, neural networks and fuzzy logic can be applied depending on adaptability. The combination of neural networks and fuzzy logic therefore gives the ability to learn and to deal with system uncertainties by using the advantages of each system.

The models composed of neural network and fuzzy logic mainly have two classes

depending on whether neural network or fuzzy logic is defined as the input interface or the decision maker. As a system input interface, fuzzy logic responds to linguistic statements and provides a fuzzy input vector to a multi-layer neural network, while the decision making is performed by the neural network. This type of combined system is normally referred to as a fuzzy neural network system. In this system, the weight and neurons are considered to be fuzzy sets. The fuzzy neurons are designed to process fuzzy inputs with linguistic expressions in the form of IF-THEN rules and the fuzzy weights are updated by backpropagation algorithms to realise the learning procedure. Conversely, when fuzzy logic is employed as the decision maker, the neural network works as an input interface and receives feedback from output decisions. The architecture of this combination, which is called a neuro-fuzzy system, allows the membership functions of fuzzy logic to be automatically tuned. Since the design and tuning of membership functions are often time consuming tasks, this ability based on the neural network learning mechanism can improve system performance and reduce the time cost of system design.

With the combination of quantitative and qualitative information, the neuro-fuzzy technique is capable of handling complex systems. The neuro-fuzzy model also becomes more transparent than the simulation using a neural network, due to the linguistic expression of rules which humans can interpret. These advantages have resulted in the wide application of this method in industrial processes. Fault detection and diagnosis can also be facilitated, and related literature can be seen in the supervision of vehicle tyre pressure (Ayoubi and Isermann 1997), fault detection of fuel injection system (Förstner and Lunze 2000), fault diagnosis in AC motor control (Alexandru 2004) and railway junction cases (Roberts *et. al.* 2002).

2.3 Conclusions

Model-based quantitative and qualitative methods for fault detection and diagnosis have been discussed in this chapter. The quantitative model-based fault detection methods, namely observers, parity equations and parameter estimation, are introduced and compared. The performances of the three methods depend on whether they are well designed for specific applications. With appropriate designs, the first two methods have the advantage of a fast response to sudden faults, provided that explicit system models and adequate disturbance information is available. In comparison, parameter estimation is good at detecting incipient faults by observing changes in system parameters, however, a slow response speed and a high computation cost are expected and the system must keep dynamic. Accurate models and decoupling from disturbances are often required to improve the robustness of these three methods.

Some qualitative methods, neural networks, fuzzy logic and neuro-fuzzy, have been discussed. With input and output data, neural networks can model complex non-linear behaviour and generate residuals for fault diagnosis. This approach can also facilitate fault classification by training residuals to map to different faults. The black-box feature makes it impossible to understand the trained network and abstract rules from the neural network model. In comparison, fuzzy logic systems are more favourable for practical applications, since the model behaviour can be easily explained based on fuzzy rules and the performance can be adjusted by tuning the rules. However, design and tuning of membership functions can often take a lot of time, traditionally fuzzy logic systems are therefore only applicable to fields where prior expert knowledge is available and the number of inputs is small. Neuro-fuzzy systems combined the

strength of both the two techniques, which enable the system to learn and to be transparent with the fuzzy IF-THEN rules. Neuro-fuzzy systems also have the advantage of the ease of the rules and linguistic model design and the ability to apply the method to complex, uncertain and inherently non-linear systems, which makes accurate fault detection and diagnosis more achievable.

Chapter 3

Generic fault detection and diagnosis for STMEs

3.1 Introduction and motivation

Apart from rolling stock and rails, a railway system runs with the coordination of a number of rail-side assets. These assets are simple, low-level and widely distributed along the track; however, they are often critical to operational success. A failed asset is likely to cause a significant delay to rail services, and may even make the system unsafe. Five safety-critical assets, the electro-pneumatic pneumatic train door, train-stop and point machine, the electric point machine and the electro-hydraulic level crossing barrier, are introduced in this chapter. A generic fault detection and diagnosis method is proposed for these assets based on their command features. The laboratory based test rigs for data collection and condition monitoring are also discussed.

In order to develop a generic fault detection and diagnosis solution, an abstract description (model) of a generic class of assets is required, which can be used for multiple asset types of the same class. As the abstract model, the STME is defined according to a series of dynamic features of a class of assets with similar operational characteristics, such as reciprocating movement and nonlinear load. The detailed definition of the STME will be presented in this chapter. The definition of STME also provides a set of common features which the devices should posses if they fall into this group. A physical modelling approach for the assets is discussed to indicate why a physical model is very difficult to obtain. Based on the above knowledge, a generic fault detection and diagnosis method is illustrated in the following aspects:

- Common feature extraction at the parameter and system levels;
- Generic fault detection and diagnosis methodology and processes;
- Adaptive thresholds design with statistics theory;
- Fault codes and initial fault diagnosis.

3.1.1 Single Throw Mechanical Equipments

The railway assets studied in this thesis are different in design, functionality and power source. However, based on similar operational characteristics, these assets are classified as Single Throw Mechanical Equipments (STMEs). The STME is a class of electro-mechanically operated equipment, which shares the properties of mechanical switches and reciprocating systems but differs from them due to the following features: slow throw speed, long throw time, non-periodic operation and large varying load.

According to the definition given by N. Lehrasab and S. Fararooy (1998), an STME operates with the following characteristics:

- An STME has two stable states. Whenever activated, it physically moves from one state to another. This transition is known as a *throw*;
- During the transition, the load to be moved is large and nonlinear;
- Compared with reciprocating systems, such as relays and switches, the time taken from one state to another, the *throw time*, is correspondingly longer and the *throw speed* is lower. Due to the large nonlinear load, the throw dynamic is much more

important than the throw time and transients;

– The operation of STMEs is not periodic; it starts only upon the receipt of a control command;

– In the open-loop configuration, for safety reasons, the speed-limiting mechanisms, such as dampers, are often employed.

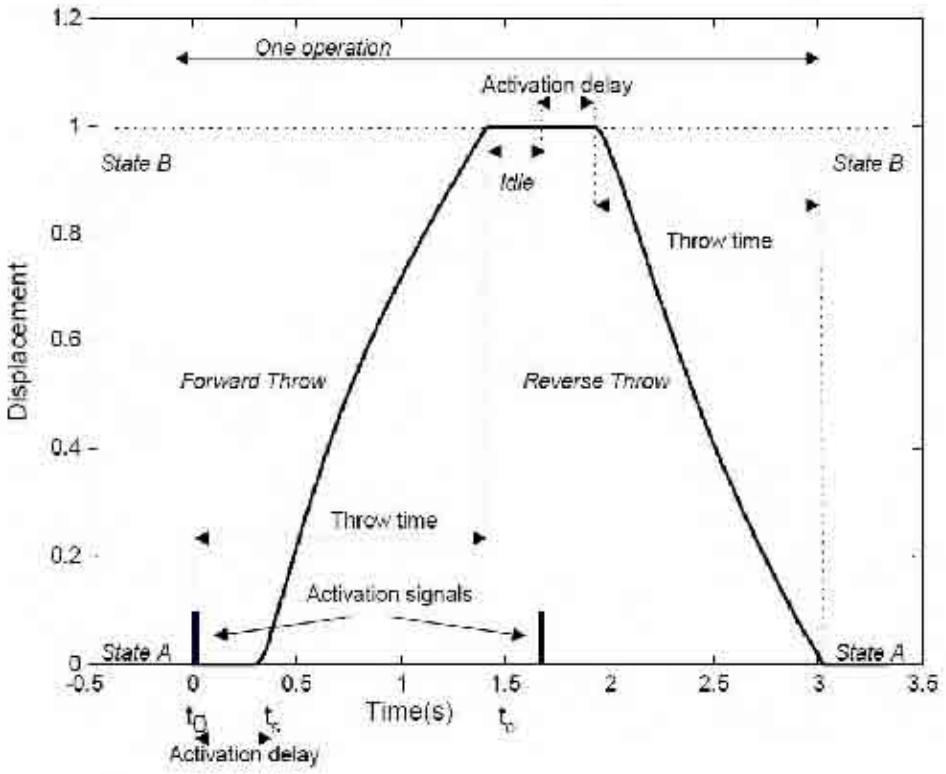


Figure 3.1 A typical STME displacement profile.

A typical STME displacement profile is shown in Figure 3.1. Before an operation command is given, the system is at State A. The transition starts from State A, S_A, to State B, S_B, on the receipt of an activation pulse. This is referred to as a *forward throw*. Then the system stops and stays at rest at S_B. After an arbitrary time, another control signal is given to make the STME move from S_B back to S_A, which is referred to as a *reverse throw*. Forward throws and reverse throws are commonly referred to as

operations. From Figure 3.1, we can see an obvious time difference between the activation signal (t_0) and actual movement (t_s), which is known as the *activation delay*, $t_a = (t_s - t_0)$. At time t_e, the throw completes at S_B. Therefore, the *throw time* (t_t) can be defined as $t_t = (t_e - t_0)$. In most cases $t_0 = 0$, then $t_t = t_e$. The distance between the initial position and the end position is defined as the *throw distance*, $S_t = (S_B - S_A)$.

In most STMEs, throw time, t_t, is dependent upon the input force, ξ; if the force increases, the throw time reduces, and vice versa. Due to the physical limitation of system mechanics, with the force increase, t_t reduces to a minimum value t_{min}. Throw time is inversely proportional to the force. It was found experimentally that the rate of change of t_t with respect to the force applied is proportional to the difference between the actual and minimum throw time. The relation function can be expressed as:

$$\begin{aligned} &\frac{dt_t}{d\xi} = -k_1(t_t - t_{\min}), \\ &t_t = k_2 e^{-k_1\xi} + t_{\min}. \end{aligned} \tag{3.1}$$

Where, k_1, k_2 are the STME constants.

For the purpose of fault detection and diagnosis, a STME definition, as a high level model, is essential to generically describe the basic operation of the STME system. The common features, e.g. throw time and activation delay, can therefore be abstracted from all such equipment, and a generic fault detection and diagnosis method becomes achievable.

3.1.2 Involved railway assets and data visualisation

Five STME railway assets, the electro-pneumatic train door, the train-stop and point machine, the electric point machine and the electro-hydraulic level crossing barrier,

are discussed in this section. For the purpose of visualisation, some of the measured profiles are normalised into the region of [0, 1] to present the dynamics.

3.1.2.1 Electro-pneumatic train door

An electro-pneumatic train rotary door test rig developed by Vapor (UK) was used for data collection. The mechanical configuration is shown in Figure 3.2. A master and a slave actuator are used to control the 'open' and 'close' operation of a pair of door panels via two rotary arms connected to each panel.

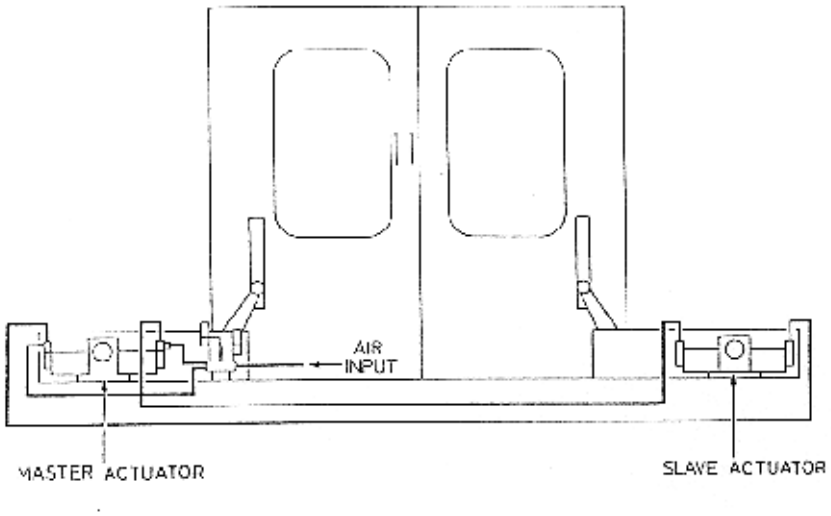


Figure 3.2 Mechanical configuration of pneumatic train door.

Figure 3.3 details the mechanism of the actuator. This type of train door actuator is similar to that used on London Underground's Central Line rolling stock. This actuator consists of two cylinders which operate the 'open' and 'close' respectively. The cylinders are controlled by four 52 V DC activated solenoids and share one piston head. With the compressed air as power source, the 'open cylinder' is filled by the train air supply to push the piston to the desired direction in which the rotary arm is driven to open the panels. Meanwhile, the air in the 'close cylinder' is exhausted and

the airflow is measured by a sensor for condition monitoring. Conversely, the compressed air is filled in the ‘close cylinder’ to perform a ‘close’ operation. A hydraulic damper is installed in the actuator to provide damping force, which reduces the speed of the door panel movement and makes it smooth, to avoid any potential harm to passengers. Two springs are also attached to the door arm to enable it to be pushed back whenever an obstruction happens. The trajectories of the door panel are different on ‘open’ and ‘close’ operations due to the design of the system, where these two operations are defined as normal and reverse throws respectively. Door displacement, compressed air pressure and exhausted airflow are normally monitored for fault detection and diagnosis.

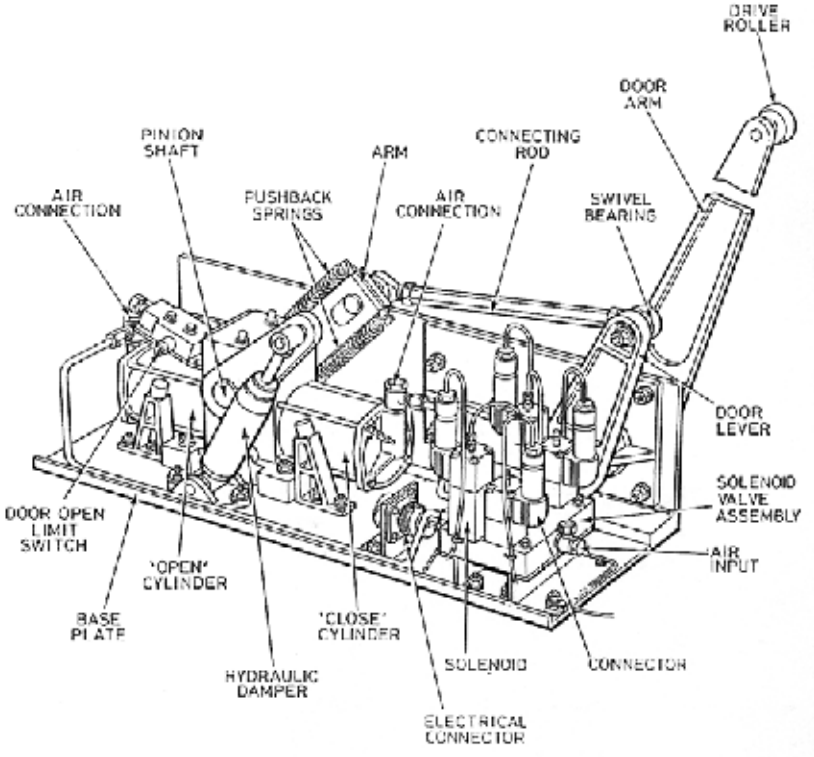


Figure 3.3 Electro-pneumatic train door actuator.

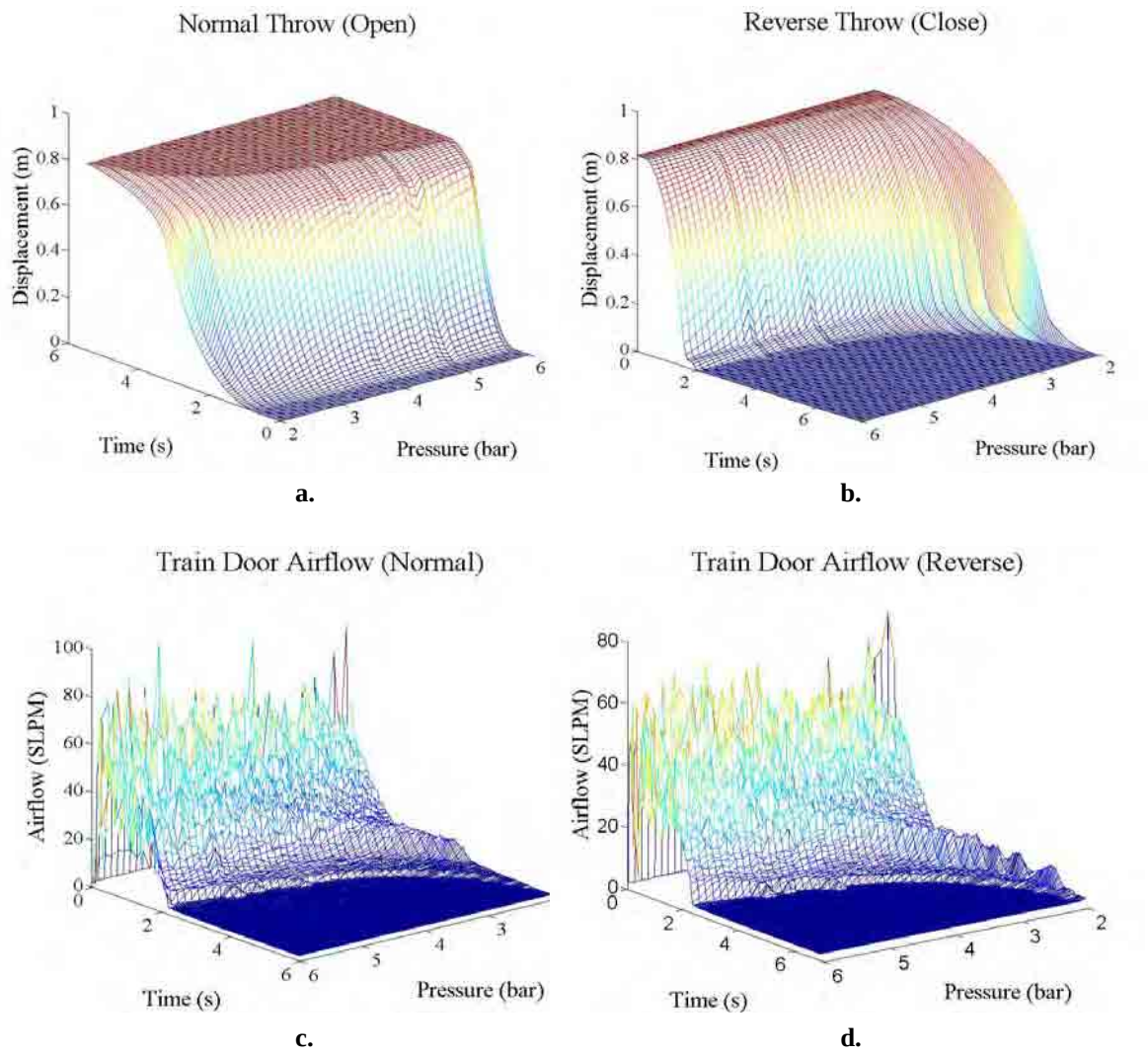


Figure 3.4 Data visualisation for the displacement and airflow of pneumatic train door.

The train door displacement profiles are displayed in Figures 3.4 a and b. In the 3-D figures, the air pressure increased from 2 to 6 bar with a 0.1 bar step. With the air pressure increasing, the activation delay and throw time decreases correspondingly. The throw time falls in the region of 2 to 6 s for a normal throw (opening) and 2 to 8 s for a reverse throw (closing). Under the same conditions, the airflow was also measured and shown in Figures 3.4 c and d.

The unit of airflow is standard litre per minute (SLPM), which means the number of

litres per minute at standard condition (0 °C and 1 standard atmosphere). In the originally measured airflow data, the noise may exist due to the high response frequency of the sensor. In the process of modelling the noise can be reduced by using a low-pass filter.

3.1.2.2 Electro-pneumatic train-stop

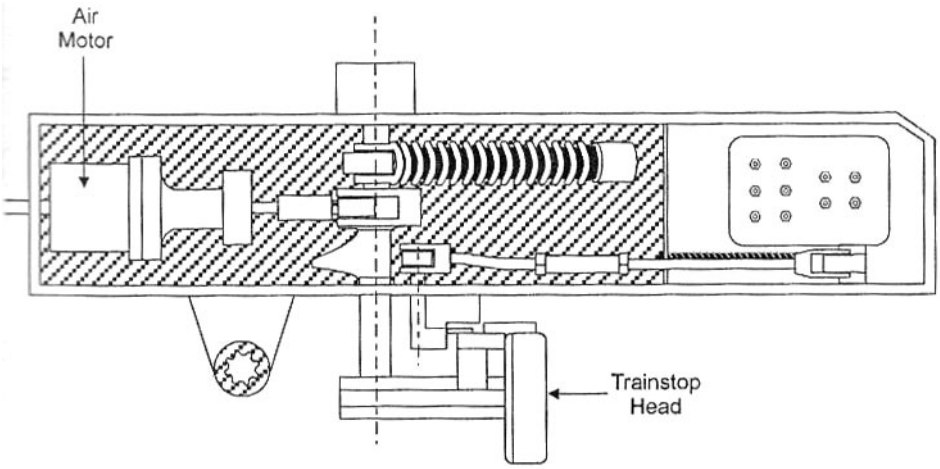
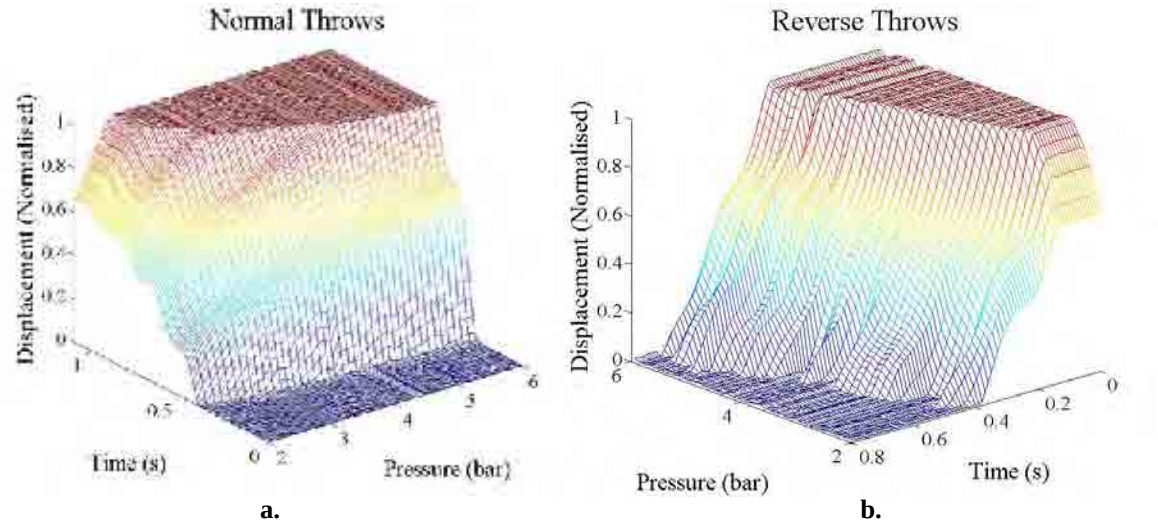


Figure 3.5 Oil-filled J-type Train-stop.

The train-stop is a tripping mechanism, which is fitted on the railway track in front of the signals to apply emergency brake to trains that pass through red signals by mistake (Allan 1993). This mechanism is usually required by some rapid transit systems such as London Underground. An example of an oil-filled J-type train-stop is shown in Figure 3.5. The operation of this train-stop is electro-pneumatic. It is designed to fail in the safe position, raised, which means that the train would be stopped for safety reasons when the train-stop fails to operate. When the green signal is illuminated, the train-stop arm is lowered by the pressure of compressed air (nominal 4.1 bar, critical 2.8 bar) via the electrically activated solenoid valve. This operation compresses the main spring, which forces the arm back to the up/danger position in the absence of air

pressure. When the train-stop operates, the angular movement of the arm is turned into a lateral one by a conversion mechanism to move the indication/detection rod, which shorts/opens the circuit of contact switches in the proving box to indicate the train-stop down/up state to the central control.

The measured data of train-stop displacement is normalised and visualised in Figure 3.6. As a member of the classification of STMEs, the train-stop also has two throws: normal (the train-stop head is pressed down) and reverse (the train-stop head is released up). The train-stop head is designed to be in the 'up' position by the spring in case of a lack of compressed air. The design of this mechanism makes the reverse throw different from the normal throw. In the figures, the throw time falls in the range of 0.45 to 1.2 s for a normal throw and of 0.35 to 0.73 s for a reverse throw when the compressed air pressure changes from 2 to 6 bar.



a. **b.**

Figure 3.6 Train-stop displacement visualisation in 3-D.

3.1.2.3 Electro-pneumatic point machine

Points are a mechanism by which the train can continue along its current tracks or be

guided onto the adjacent ones. The points are driven by point machines with different actuators (e.g. electric, electro-hydraulic or electro-pneumatic). The original point machines were operated manually. This approach was superseded by rod driven points where compressed air was used to drive the points via air motors. This mechanism was known as the 6-foot pneumatic point machine. However, this type of point machine has difficulty of working in tunnels due to its requirement for space outside the tracks. To solve this problem, a new type of point machine located between the running rails was designed and named the 4-foot pneumatic point machine. The schematic of this type of point mechanism is shown in Figure 3.7. The 4-foot pneumatic point machine is typically driven by compressed air of around 3 bar pressure. Two electrically activated valves work alternately to determine the normal or reverse movement.

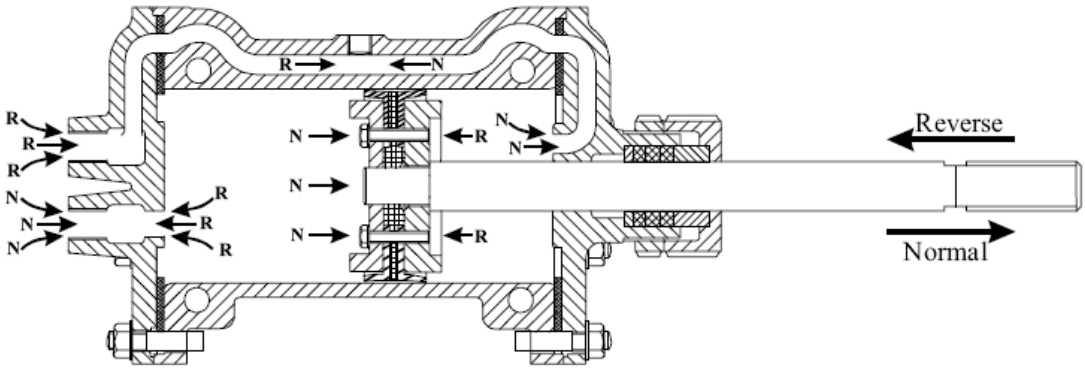


Figure 3.7 Schematic of a 4-foot electro-pneumatic point machine.

The 3-D displacement profiles of an unloaded 4-foot pneumatic point machine are displayed in Figure 3.8. Here the compressed air pressure has been changed from 0.5 to 6 bar with 0.1 bar increments. In this pressure range, the throws for both directions can be completed in less than 2 seconds.

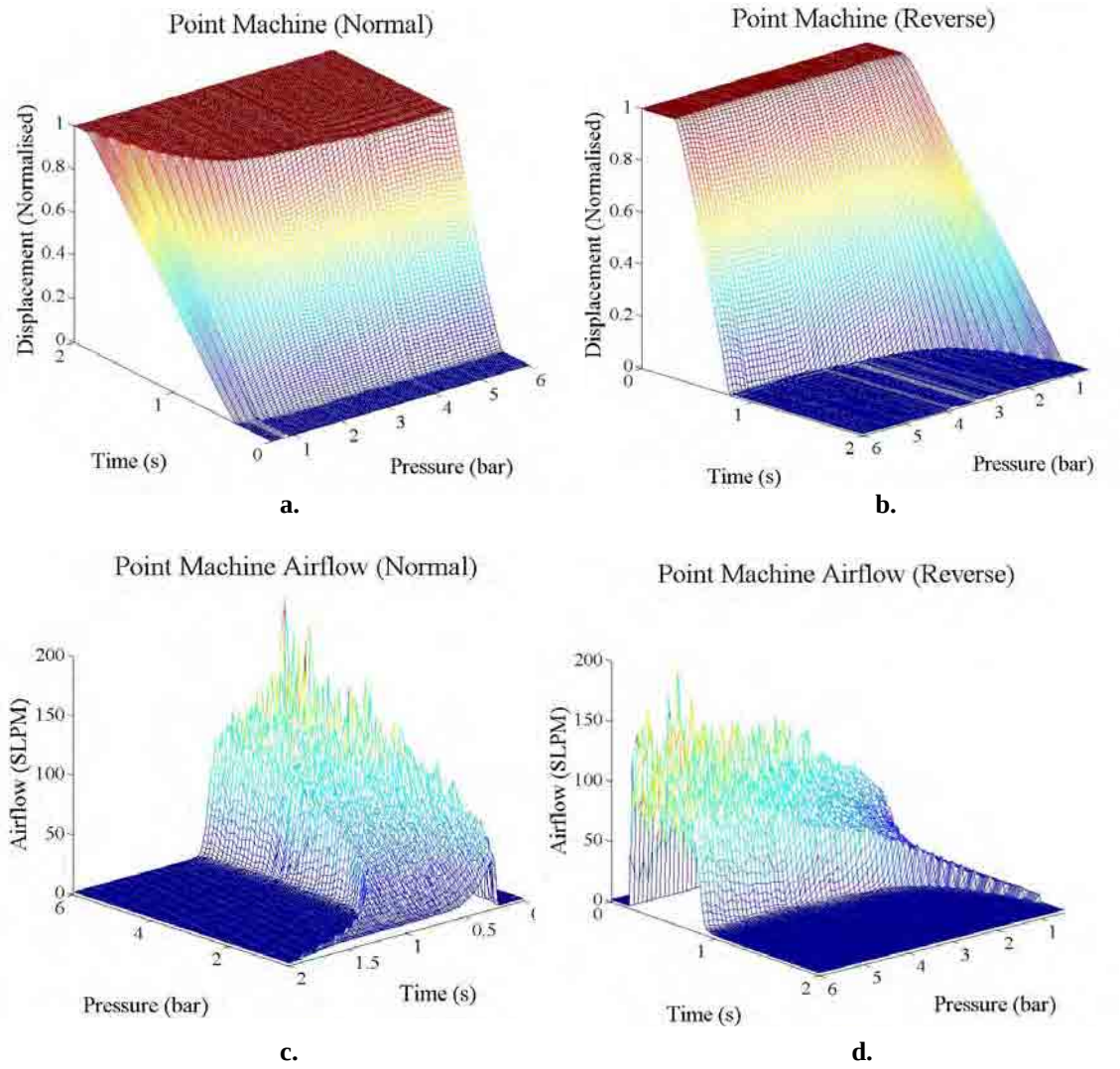


Figure 3.8 Data visualisation for point machine displacement and airflow.

3.1.2.4 Electric point machine

A Westinghouse M63 electric point machine, which was set up in the laboratory, is introduced here as a case study. The electric point machine has the same function as the pneumatic one discussed above, but it performs the route switching by different means. The M63 electric point machine consists of several subsystems as shown in Figure 3.9. A 110 V DC motor provides the driving force where the rotary force from the motor is transferred to a lateral force on a drive bar by a gear system. In turn, the

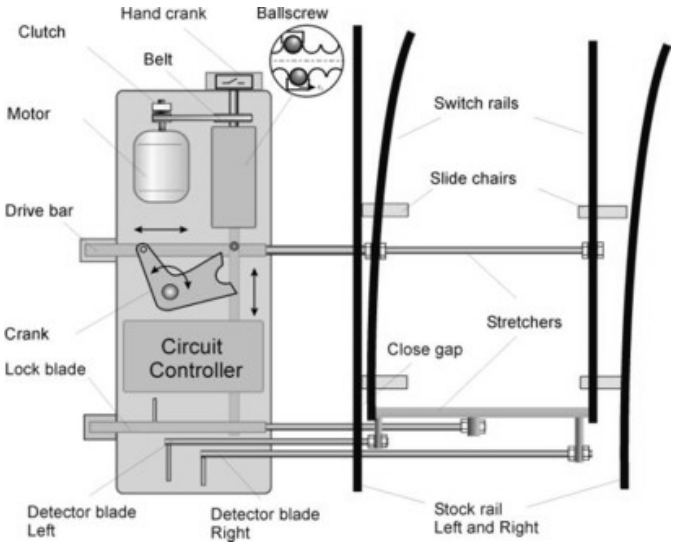


Figure 3.9 Diagram of an electric point machine (Zhou *et. al.* 2002).

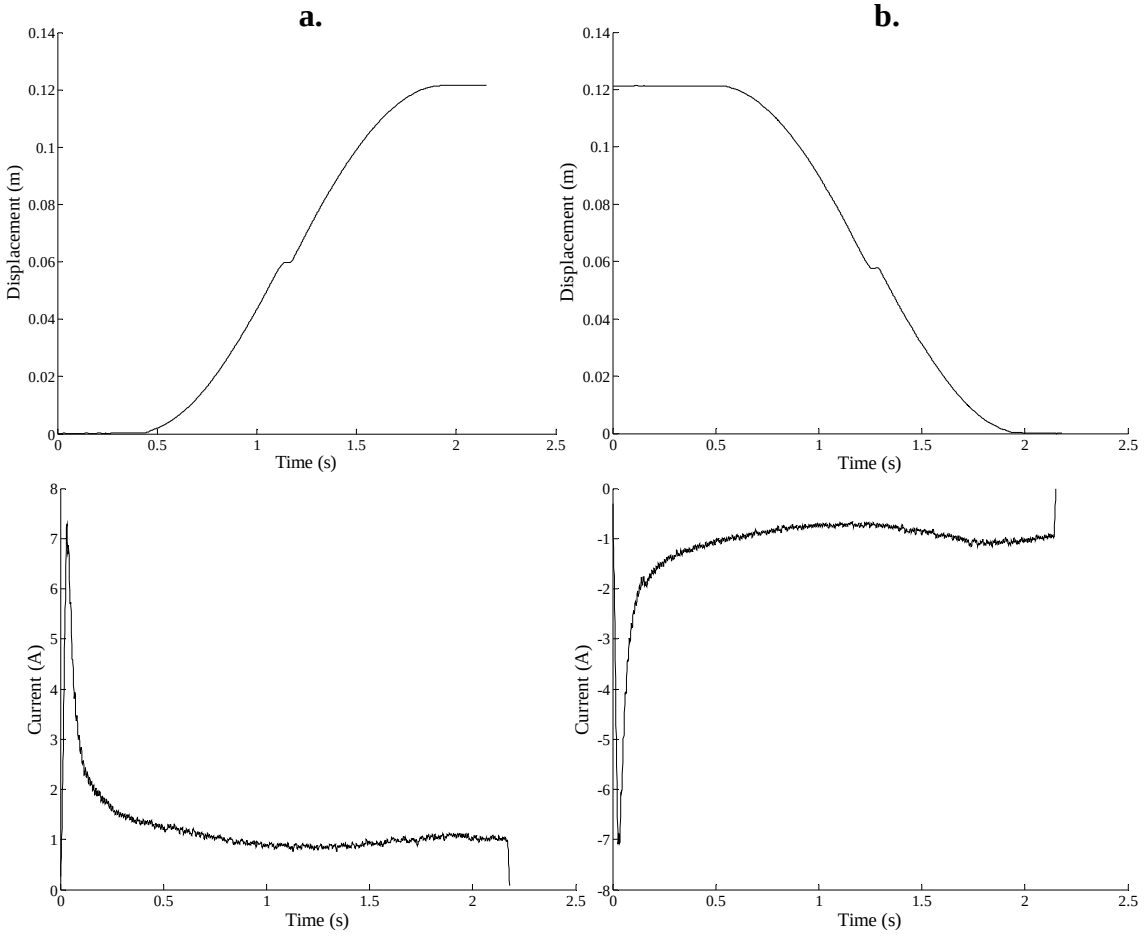


Figure 3.10 Data visualisation for the displacement and current of electric point machine (Loaded).

force is transferred through the drive bar to the switch blade via the drive arm. To avoid any damage when the point is obstructed, a clutch is employed to slip at a predetermined load. For the same purpose, a snubbing device slows the motor down rapidly when the switch blade reaches the end. To 'lock' the point at rest, a mechanical 'lock dog' is used and the status is detected by a set of contacts. A set of motor contacts is designed to allow the direction of movement to be set. In case of emergency, the point machine can be operated manually. The displacement and current data of the loaded electric point machine is visualised in Figure 3.10.

3.1.2.5 Electro-hydraulic level crossing barrier

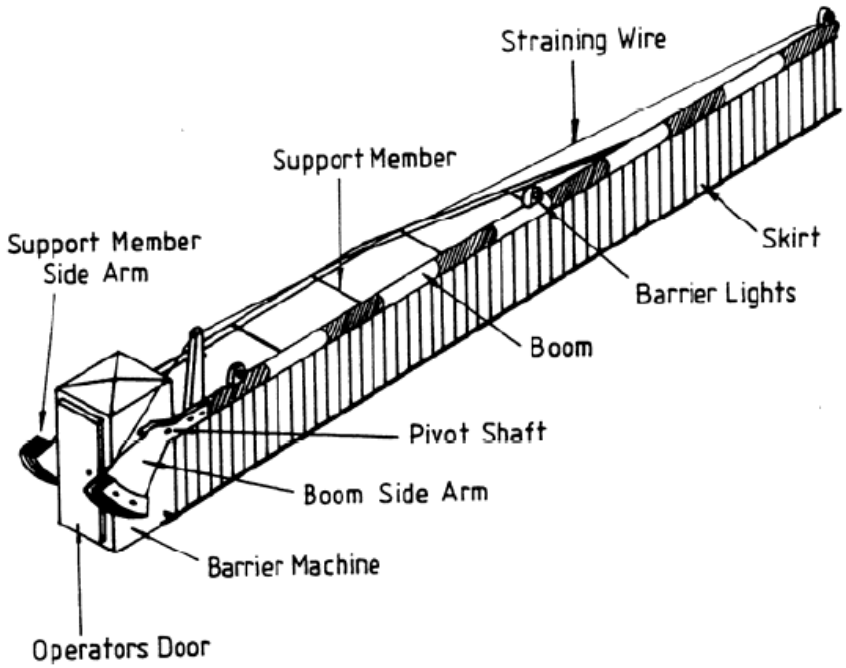


Figure 3.11 Level crossing barrier.

A typical illustration of an electro-hydraulic level crossing barrier BR843 is shown in Figure 3.11. This type of barrier is widely used in the UK. This equipment can be

operated either manually or by an electrically driven hydraulic system. These two operation systems are accommodated in a steel weatherproof cabinet, called a 'barrier machine'. The normal length of the boom measured from pivot to tip is in the range of 3.6 m to 9.1 m. The boom is balanced by adding a variety of balance weights on the boom side arm. The balance weights are chosen according to both the length of the boom and the practical effects of wind. For booms longer than 6.6 m, support members and a straining wire are used to strengthen the boom.

In the laboratory, a BR843 level crossing barrier was set up with a 1.63 m boom. The boom is shorter than would be found on an operational railway junction. To simulate a practical operation, a brass rod was inserted into the hollow boom to provide the equivalent load to 4.1 m. A 24 V DC power supply with maximum 30 A output current was deployed to drive the DC motor (nominal 24 V DC and 15-29 A DC). An oil hydraulic pump driven by the motor generates force to raise the barrier. The working hydraulic oil pressure was 34 bars/500 psi and the current of the control solenoids was around 260 mA.

As an STME, the operation of a level crossing barrier includes two throws: the normal throw, which is defined as the movement from the lowered to the raised position powered by the hydraulic force; and the reverse throw, with movement from the raised to the lowered position, where the force is provided by gravity and the boom freely falls. In Figure 3.12, the parameters measured from the level crossing barrier are displayed. The angular displacement, voltage, current and oil pressure of a normal throw are shown in Figures 3.12. The voltage and current are zero for the reverse

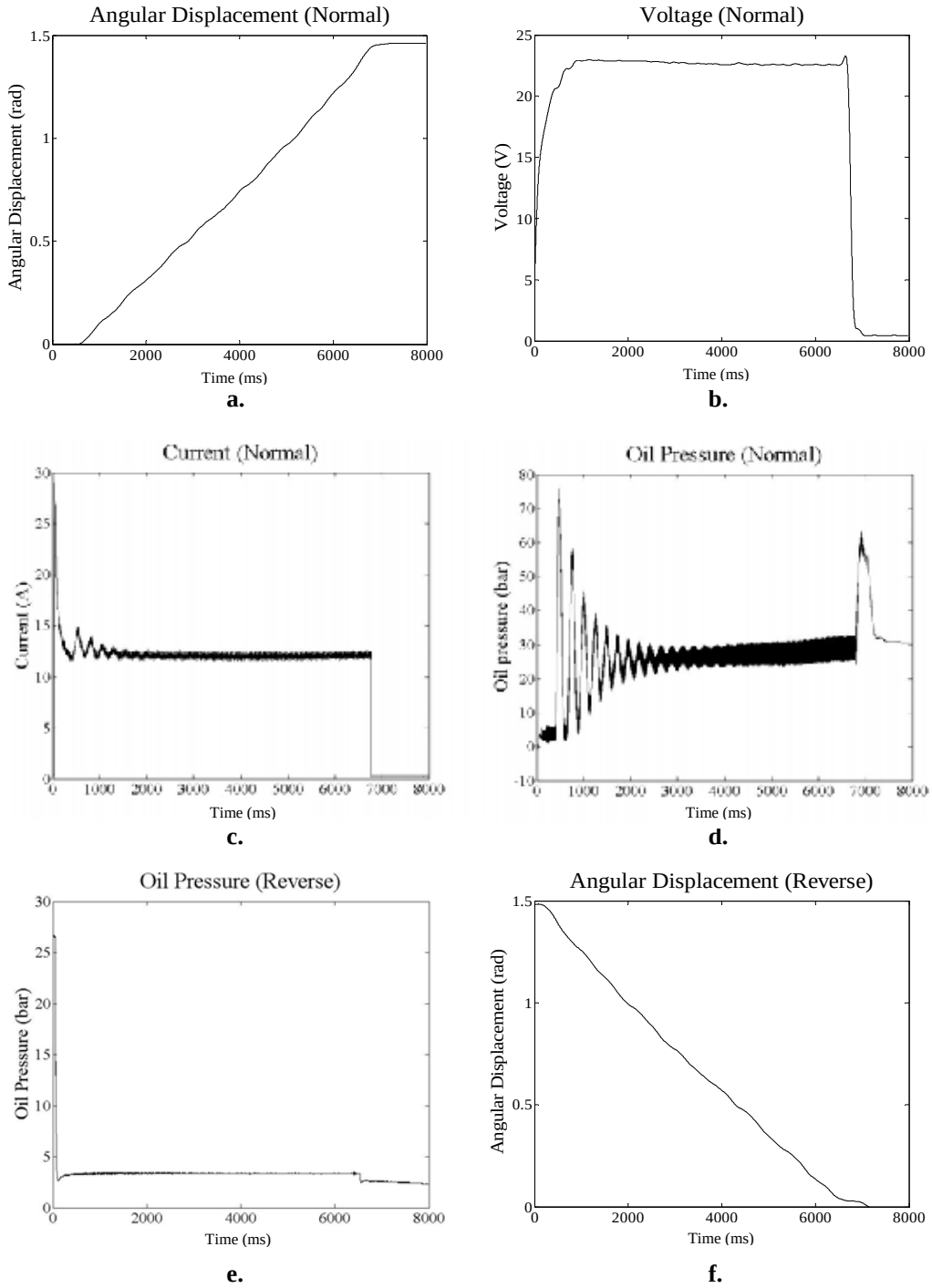


Figure 3.12 Level crossing barrier data visualisation.

throw, therefore only oil pressure and angular displacement are shown in Figure 3.12 (e and f). Figure 3.12 d indicates that the oil pressure inside the pump vibrates due to the interactional forces of both the DC motor and barrier gravity, before it reaches

constant.

3.1.3 A physical modelling approach

Using the data collected from the assets discussed above, a physical modelling (State-space model) approach is presented to indicate that it is very difficult to accurately describe STME assets with one physical model. The pneumatic train door is used as an example in this section.

As proposed by N. Lehrasab (1999), the STME may be modelled as a single-input (pressure) single-output (displacement) (SISO), linear, lumped, time-invariant dynamical system, which can be represented by a state-space model. Due to the process parameters which are needed for state-space modelling, a system simulator should be designed first, where a 'Spring-Damper-Mass' model is used to represent the simple STME system.

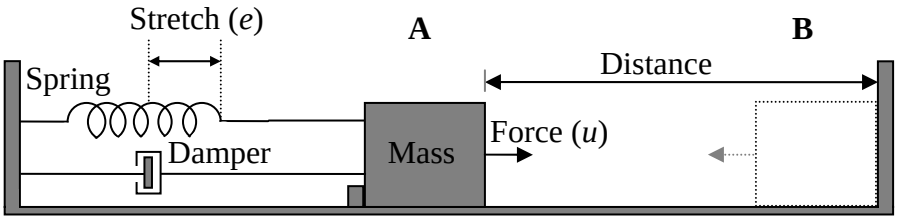


Figure 3.13 'Spring-Damper-Mass' model of STME.

Pneumatic STME equipment usually has a piston-based configuration to convert the pneumatic pressure into useful work. A piston with surface area A is exposed to pressure $u(t)$ at time t during the throw (t_e-t_0), where t_0 refers to the activation time for the throw (from state **A**) and t_e refers to the time when the moving part reaches another stable state (state **B**). A schematic view of this 'Spring-Damper-Mass' system

is displayed in Figure 3.13.

Using Newton's law, the modelling equation for this system when force *u* is applied to the moving part with mass, *M*, the displacement, *y*, can be expressed as:

$$u - Dy' - Ke = My'' \Rightarrow My'' + Dy' + Ke = u \tag{3.2}$$

where, *K* is the spring constant (Hookes Law); *D* is the damping effect (all the frictional forces may be modelled as the damping effect); *M* is the mass of the moving part and load of the equipment; *u* in the pneumatic case represents the pressure of compressed air; *e* is the stretch length of the spring, where *Ke* represents the force produced by the spring; A and B are two states of the moving part.

This second order system was thought not to be accurate enough for residual generation due to large modelling errors by N. Lehrasab (1999). A state-space model was therefore used to represent the system dynamics of the train door panel in the following form:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{3.3}$$

where, *u(t)* is the input of this model, which, here, is referred to as the pressure of compressed air; *y(t)* is the output, which, here, is referred to as the displacement of the moving part of the STME; *x(t)* is the state vector. From equation 3.2,

$$My'' = -Dy' - Ke + u \tag{3.4}$$

Due to $e' = y'$ (the spring stretch and moving part have same speed),

$$\begin{bmatrix} y'' \\ e' \end{bmatrix} = \begin{bmatrix} -\frac{D}{M} & -\frac{K}{M} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y' \\ e \end{bmatrix} + \begin{bmatrix} \frac{1}{M} \\ 0 \end{bmatrix} u \tag{3.5}$$

The state matrix *x(t)* can be written as,

$$x = \begin{bmatrix} y' \\ e \end{bmatrix} \tag{3.6}$$

From equation 3.5,

$$\begin{aligned} \dot{x} &= \begin{bmatrix} -\frac{D}{M} & -\frac{K}{M} \\ 1 & 0 \end{bmatrix} x + \begin{bmatrix} \frac{1}{M} \\ 0 \end{bmatrix} u \\ y &= \begin{bmatrix} 1 & 0 \end{bmatrix} x \end{aligned} \tag{3.7}$$

Thus,

$$\begin{aligned} A &= \begin{bmatrix} -\frac{D}{M} & -\frac{K}{M} \\ 1 & 0 \end{bmatrix} \\ B &= \begin{bmatrix} \frac{1}{M} \\ 0 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \end{bmatrix} \end{aligned} \tag{3.8}$$

Using the estimated parameters suggested by N. Lehrasab (1999), the coefficients of the state-space model, A, B, C and D, are:

$$A = \begin{bmatrix} -25 & -0.1 \\ 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 50 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \end{bmatrix}, D = 0 \tag{3.9}$$

The output of the state-space model and the measured displacement profile are compared in Figure 3.14. In the figure, it can be observed that the state-space model cannot simulate the system performance (represented by the displacement) accurately. By using state-space modelling, the system can be described as a physical principle model, which may give an insight into STME systems. However, this physical model is not precise enough to be used for residual generation. A physical model of the whole system is, therefore, considered to be very difficult to obtain for fault detection and diagnosis application.

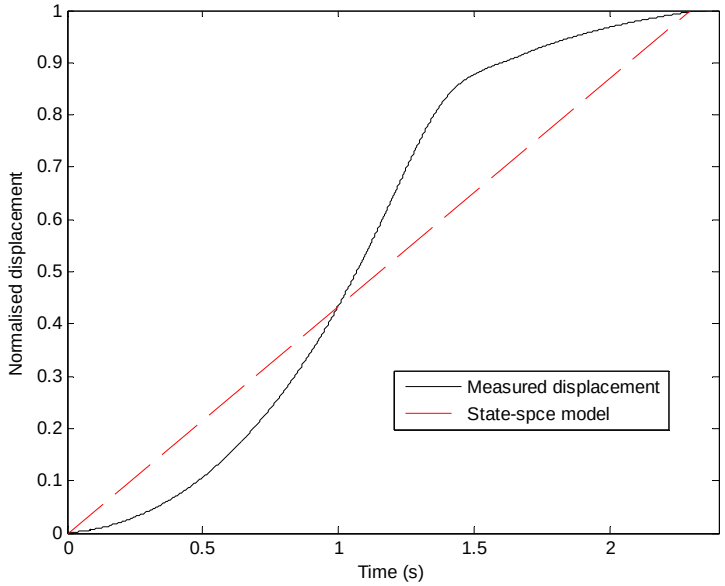


Figure 3.14 Comparison of measured displacement and state-space model estimation (train door normal throw) (Lehrasab 1999).

3.2 A generic FDD method for STMEs

Based on the previous introduction and discussion, a generic fault detection and diagnosis method is proposed in this section based on common features extracted from the STMEs.

3.2.1 Feature extraction

As discussed in section 3.1.1, the assets, considered in this study, can be classified as STMEs, and therefore exhibit the characteristics of non-periodicity and relatively slow reciprocating operation with large and non-linear loads. However, these assets also have their own specific features, the aspects of which can be described under the two headings: parameter features and system features.

Due to the difference in the driven principle between the pneumatic and the electric

and electro-hydraulic assets, the features of electro-pneumatic equipment are illustrated in detail and comparisons are made between the features of the different assets for a clear illustration of the similarities between their parameters and systems.

3.2.1.1 Parameter features

The parameters of the five assets, which are monitored for the purpose of generic fault detection and diagnosis, are listed in Table 3.1. As discussed previously, the electro-pneumatic assets are driven by the potential force of compressed air. The electric part of these assets is only the electrical activated solenoids. In order to distinguish from the electrically driven assets, the electro-pneumatic assets will be called pneumatic assets in the following sections and chapters.

Assets	Monitored Parameters	
Pneumatic Train Door	Linear displacement	Airflow Air pressure
Pneumatic Point Machine	Linear displacement	Airflow Air pressure
Pneumatic Train-stop	Angular displacement	*Airflow** Air pressure
Electric point machine	Linear displacement	*Voltage** Current
Electro-hydraulic level crossing	Angular displacement	Voltage Current Oil pressure

***** *Data was not collected for this parameter, as no appropriate sensor was available.*

Table 3.1 Monitored parameters of the STME assets.

3.2.1.1.1 Pneumatic assets

The parameter features of pneumatic assets are presented in this section. The relations between the three parameters are analysed.

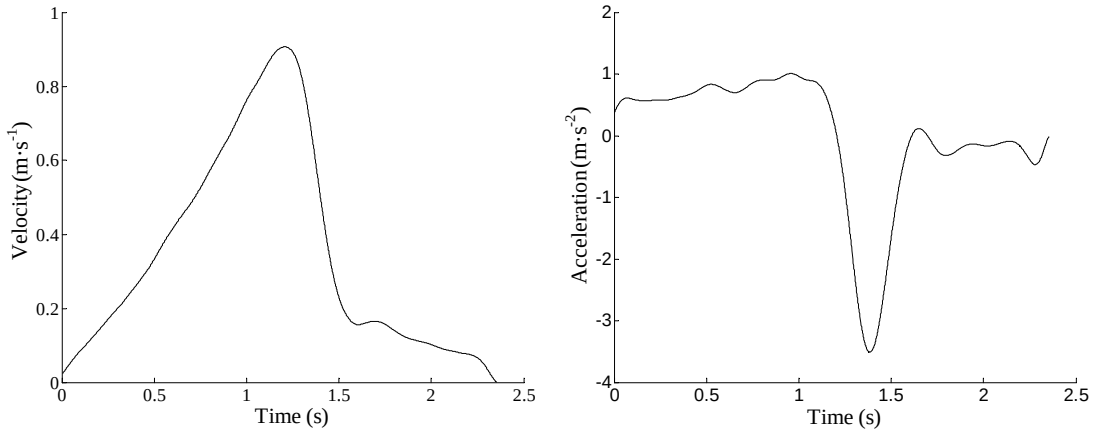
Displacement

Displacement is a common parameter which can be monitored on all five assets considered in this thesis. As the velocity and acceleration can be calculated using the displacement data and corresponding sampling frequency, the displacement is very important for the analysis of operation dynamics. As a second order differential of the displacement, the acceleration maxima indicate the dynamic changes of the movement. As an example, the velocity and acceleration profiles of the forward throw of the train door panel are displayed in Figure 3.15. The velocity and acceleration are calculated by the following equations.

$$v = \frac{y(t+\Delta t) - y(t)}{\Delta t} \tag{3.10}$$

$$a = \frac{v(t+\Delta t) - v(t)}{\Delta t} \tag{3.11}$$

Where v is the velocity and a is the acceleration. Δt is the sampling period which is fixed during the measurement.



a) Velocity vs. time **b) Acceleration vs. time**

Figure 3.15 Velocity and acceleration profiles of pneumatic train door normal throw at 3.5 bar.

In Figure 3.15, the velocity and the acceleration can be observed to vary with time.

The variation is affected by both the specifics of the design, for example the damper (reducing the door panel speed to avoid harm to passengers), and the supplied compressed air pressure. A higher pressure leads to a higher velocity and acceleration. For the particular case considered in this thesis, with a 3.5 bar air supply the highest velocity reaches approximately 0.9 ms^{-1} and the highest acceleration values are approximately 1 ms^{-2} for acceleration and 3.5 ms^{-2} for deceleration.

When the displacement was measured using a draw-wire displacement sensor (Micro-epsilon WPS-1250 MK46), the sampling frequency was set at 1000 Hz. The measured discrete data resulted in an unsmoothed displacement profile. In addition, noise was also observed in the collected data. In order to get smoothed velocity and acceleration profiles and to reduce the noise, the displacement profile was filtered using a Butterworth low-pass filter. As a compromise between distortion and fidelity, the order of the filter and the normalised cut-off frequency were chosen as 2 and 0.01 $rad \cdot s^{-1}$ respectively. The filtered results retained the throw dynamics and, simultaneously, reduced the distortion caused by noise or sampling resolution.

In the definition of an STME, the dynamic characteristics are the most important criterion to judge whether an item of equipment is an STME or not. Consequently, the dynamics of an STME are also the most useful tools for operation status identification. In order to precisely analyse the performance of an asset, the acceleration features are used to divide the whole throw period into several temporal regions, where each region responds to a dynamic status, acceleration or deceleration. In Figure 3.16, 8 temporal regions, a_{t1}, a_{t2}, …, a_{t8}, for acceleration and 7 temporal regions, d_{t1}, d_{t2}, …, d_{t7}, for deceleration are displayed.

For fault detection and diagnosis, a comparison between a model output and measured practical data is usually used to detect occurrence of a fault. In the case of fault detection for STMEs, it is considered that it is more applicable and accurate to carry out the comparisons within the defined regions. In the regions, which correspond to certain dynamic status, the dynamic changes caused by faults are more easily observed.

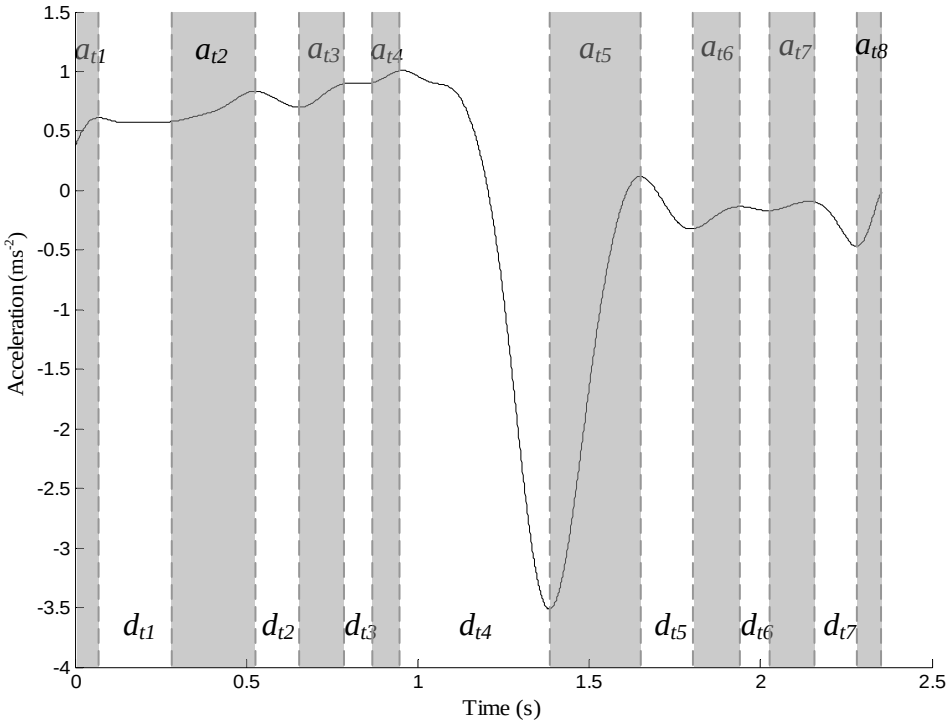


Figure 3.16 Temporal regions division using acceleration feature.

In order to relate a fault to the position of the train door panel on its trajectory, the spatial regions must therefore be studied. As shown in Figure 3.17, 6 spatial regions, a_{s1}, a_{s2}, …, a_{s6}, for acceleration and 6 spatial regions, d_{s1}, d_{s2}, …, d_{s6}, for deceleration are divided using acceleration maxima.

By using the above two region division methods, the dynamic changes within a throw

are clearly illustrated. However, these two methods produced two different region division results. When a fault, such as an excessive friction or a leakage, occurs, the actuator normally requires a longer time to finish a throw. A temporal model of the displacement with a fixed throw time will therefore be unavailable to be used for a comparison with the measured displacement. Since the throw distance is constant when the throw is finished, the spatial model is considered more appropriate.

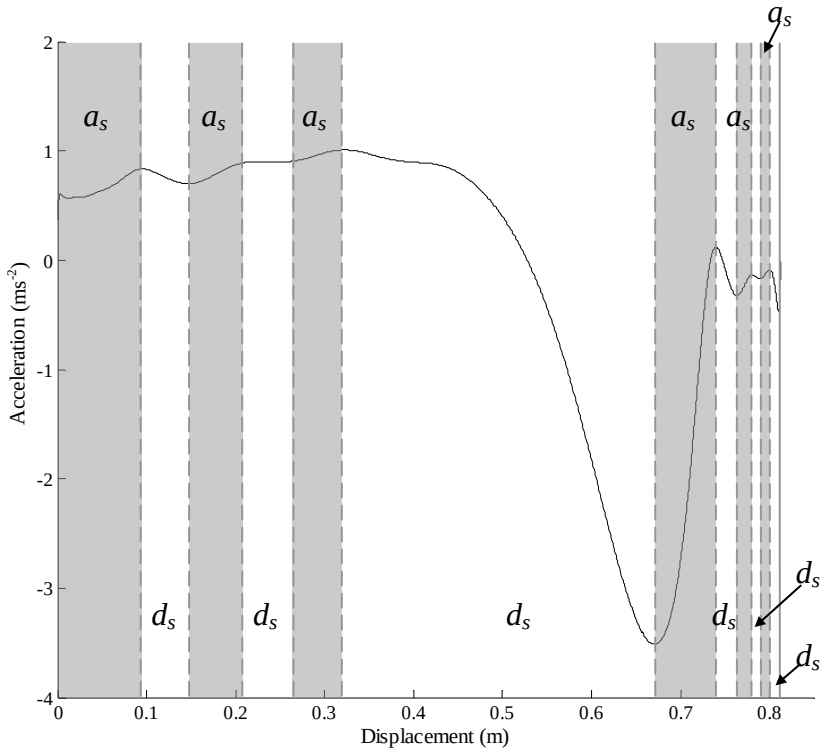


Figure 3.17 Spatial regions division using acceleration feature.

In Figure 3.17, the size of the regions is not even, where regions, a_{s5}, $a_{s6,}$ d_{s5} and d_{s6}, are small and region, d_{s3}, is large. In addition, a logic understanding is also required to mark each of the regions, which is meaningful for fault location. In this situation, the regions are improved by combining both the acceleration and the velocity profile features.

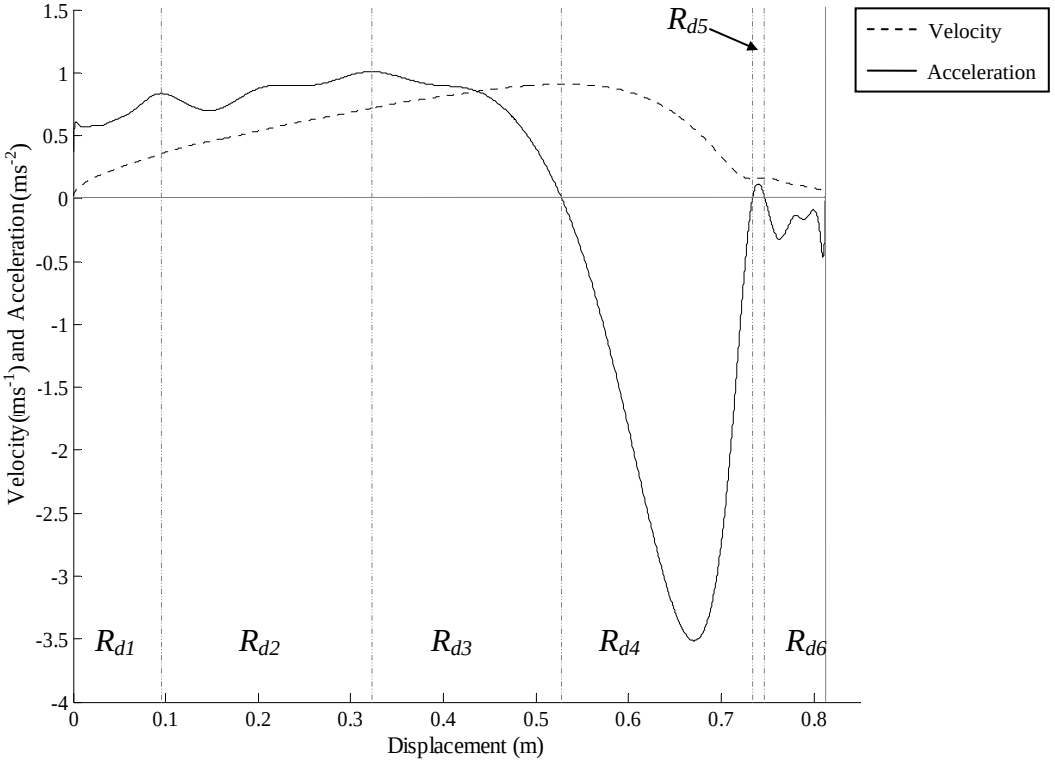


Figure 3.18 Spatial regions division using acceleration and velocity features.

Regions	Boundaries (metre)	Boundaries (second)	Description
R_{d1}	0 – 0.095	0 – 0.53	Throw start
R_{d2}	0.095 – 0.32	0.53 – 0.96	Intermediate 1
R_{d3}	0.32- 0.53	0.96 – 1.21	Intermediate 2
R_{d4}	0.53 – 0.73	1.21 – 1.61	Intermediate 3
R_{d5}	0.73 – 0.75	1.61 – 1.7	Intermediate 4
R_{d6}	0.75 – 0.812	1.7 – 2.35	Throw stall

Table 3.2 Boundaries of spatial regions.

As shown in Figure 3.18, a healthy normal throw of a train door can be divided into 6 regions and each region represents a status of the throw process. R_{d1} and R_{d6} indicate the start (velocity increasing from zero) and end (velocity decreasing to zero) of the train door throw respectively. R_{d2} and R_{d3} are both in an area of velocity increasing,

however, a maximum value of acceleration at 0.322 m is used to divide the regions finely and therefore enhance the degree of precision. R_{d4} is a deceleration region while R_{d5} has a small level of acceleration. The boundaries and descriptions of logic meanings are shown in Table 3.2. The regions definition method presented for the pneumatic train door is also applicable to other STMEs considered in this thesis.

Airflow

An air cylinder is often used as a force generator for pneumatic STMEs. The control of the cylinder is normally operated by an electrically activated solenoid. Once a solenoid is activated, the valves are switched on or off to allow compressed air into or out of the cylinder chambers. The compressed airflow expelled from the chamber is considered in this study. This feature is useful in indicating the working conditions within a machine. Before extracting the characteristics of this feature, a mathematical analysis of cylinders is given.

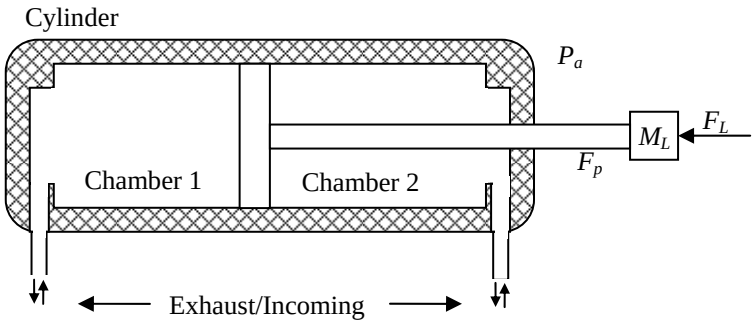


Figure 3.19 Schematic representation of the pneumatic cylinder.

In Figure 3.19, a schematic of a pneumatic cylinder is shown. M_L is the external mass; M_{pr} is the assembly mass of the piston and rod; F_L is the external force; F_p is the output force provided by the rod; P_1 and P_2 are the absolute pressure in the two

chambers; P_a is the absolute ambient pressure; A_1 and A_2 are the effective areas of piston, and A_r is the cross-sectional area of the rod.

Under ideal conditions, where the friction, the mass of the piston and rod and the ambient pressure are ignored, the output force is described as:

$$F_p = P_1 A_1 - P_2 A_2 \tag{3.12}$$

$$A_1 = A_2 + A_r \tag{3.13}$$

where equation 3.13 describes the relationship of the piston effective areas between chamber 1 and 2. The output force is determined by the difference in pressures in the two chambers. Practically, when all the factors are considered, the dynamic process of an operating cylinder is presented with a differential function (Barber 1986 and Richer and Urmuzlu 2000).

$$(M_L + M_{pr})\ddot{x} + \beta\dot{x} + F_f + F_L = P_1 A_1 - P_2 A_2 - P_a A_r \tag{3.14}$$

Where x represents the piston position; β is the viscous friction coefficient and F_f is the Coulomb frictional force.

The mass flow of compressed air can be illustrated as (Warring 1969 and Daugherty *et. al.* 1985):

$$\dot{m} = \begin{cases} \dfrac{A_0 P_1}{\sqrt{T_1}} \sqrt{\dfrac{2k}{R(k-1)}} \sqrt{\left(\dfrac{P_2}{P_1}\right)^{2/k} - \left(\dfrac{P_2}{P_1}\right)^{k+1/k}} & : \dfrac{P_2}{P_1} > \left(\dfrac{2}{k+1}\right)^{k/k-1} \quad (unchoked) \\ \dfrac{A_0 P_1}{\sqrt{T_1}} \sqrt{\dfrac{k}{R}\left(\dfrac{2}{k+1}\right)^{k+1/k-1}} & : \dfrac{P_2}{P_1} \le \left(\dfrac{2}{k+1}\right)^{k/k-1} \quad (choked) \end{cases} \tag{3.15}$$

where A_0 is the effective valve orifice area; P_1 is the upstream pressure; P_2 is the downstream pressure; T_1 is the upstream temperature and the constant $k=c_v/c_p=1.4$ for air (c_v and c_p are the constant volume and constant pressure specific heats of air); R is

the ideal air constant, where $R = \frac{pV}{nT}$ (p is the absolute pressure; V is the volume of gas; n is the amount of gas and T is thermodynamic temperature) in the ideal gas law. The '*choked*' and '*unchoked*' are two limiting conditions for two mass flow rate descriptions, where a choked flow indicates a condition that the mass flow rate will not increase with a further decrease in the downstream pressure environment while upstream pressure is fixed. The state of '*choked*' is reached when the flow velocity approaches the speed of sound; otherwise, the flow is considered as '*unchoked*' (Miller 1996).

A relationship between mass flow and airflow velocity can be described by the following equation.

$$\dot{m} = \rho A_0 u \tag{3.16}$$

Where ρ is the air density and u is the airflow velocity.

If we substitute equation 3.16 into equation 3.15, velocity of airflow is described as:

$$u = \begin{cases} \frac{P_1}{\rho\sqrt{T_1}}\sqrt{\frac{2k}{R(k-1)}}\sqrt{\left(\frac{P_2}{P_1}\right)^{2/k} - \left(\frac{P_2}{P_1}\right)^{k+1/k}} & : \frac{P_2}{P_1} > \left(\frac{2}{k+1}\right)^{k/k-1} \quad (unchoked) \\ \frac{P_1}{\rho\sqrt{T_1}}\sqrt{\frac{k}{R}\left(\frac{2}{k+1}\right)^{k+1/k-1}} & : \frac{P_2}{P_1} \le \left(\frac{2}{k+1}\right)^{k/k-1} \quad (choked) \end{cases} \tag{3.17}$$

In this equation, the value of airflow velocity is influenced by the pressures in the two chambers, the absolute temperature and the density of the air expelled through the orifice. The size of the open area of the orifice does not affect the airflow velocity.

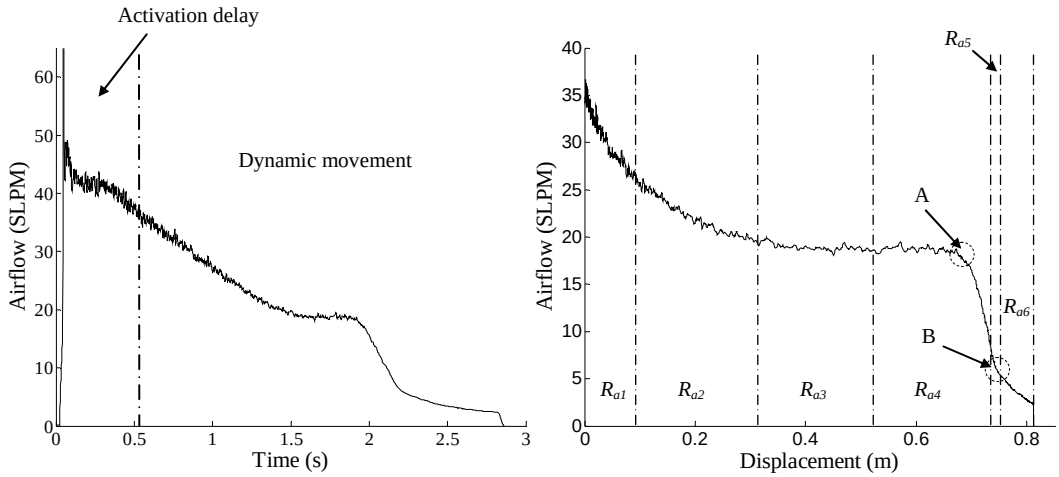
In the laboratory the airflow of the pneumatic STMEs was measured by a mass airflow sensor, Honeywell AWM720P1, which has a larger measurement range, up to 200 SLPM (Standard Litre per Minute), than other airflow sensors. However, the

actual airflow is still over the measurement range of this sensor in the case of the point machine. According to equation 3.15, the mass airflow is proportional to the orifice area. The orifice area of the point machine was therefore increased by diverting the main airflow into several pipes, one of which was measured. The branched airflow cannot indicate the value of the total mass airflow, however, it represents the airflow dynamics, which is important for condition monitoring.

Since the variation of density of the air flowing through the orifice and of the pressure in the two chambers shows non-linear characteristics, it is hard to accurately model the mass airflow for the cylinder operating process. From equations 3.14 and 3.15, the displacement of the piston can be observed to be differentially related to the mass airflow; furthermore, it can be deduced that the displacement of the load driven by the pneumatic cylinder is also non-linearly related to the mass airflow. In this situation, a precise mathematical model to relate displacement to mass airflow becomes difficult. Instead of a mathematical description, the mass airflow can be modelled using artificial neural networks by which non-linear relations can be studied to achieve a better simulation performance (Patton *et. al.* 1999).

A throw of the STMEs is normally composed of two parts: activation delay and dynamic movement. The displacement of the load remains as zero in the activation delay period, however, contrarily the airflow produces a large output. Based on the design of a pneumatic cylinder, the piston can only move when the air in one idle chamber is partly expelled and the pressure in this chamber drops to a certain value. The value of the pressure drop depends on the pressure in the working chamber and the force required to drive the load. As shown in Figure 3.20 a, a large amount of air

is expelled within the activation delay period. The plotted airflow curve is the mean of 100 airflow measurements of the train door normal throws.



a) Two sections of airflow. **b) Spatial regions for airflow.**

Figure 3.20 Analysis of the airflow data of train door normal throw.

With the boundaries defined in Table 3.2, the dynamic part of the airflow is divided into 6 spatial regions, as shown in Figure 3.20 b, where the airflow in R_{a1} corresponds to the start of the throw, the regions, R_{a2}, …, R_{a5}, correspond to the intermediate sections of the throw and R_{a6} is the end of the throw. In this figure, two turning points, A (0.682 m) and B (0.745 m), can be observed in regions R_{a4} and R_{a5} respectively. By comparing to Figure 3.18, a feature can be observed that turning points A and B respectively respond to the low and high points of the acceleration profile within regions R_{d4} and R_{d5}. This feature relates the airflow to the acceleration on these two points, which could be useful to identify the synchronisation of the piston and the load. Apart from this correlation, the airflow still shows a non-linear relationship to the displacement based features.

Air pressure

Other than displacement and airflow, there is another parameter, air pressure, which is available to be measured and monitored. As a part of the test rig for pneumatic STMEs, the compressed air was supplied by an air compressor, an Airmaster Tiger 8/50. This compressor generates enough compressed air to provide constant pressure for most of the experiments, which performs a good simulation of a practical compressed air supply. In practical circumstances, the air pressure may vary within a permitted range, however, as a constant value, this parameter can be monitored by an air pressure sensor and a pre-defined threshold could be used for fault detection.

3.2.1.1.2 Electric and electro-hydraulic assets

In this section, the parameters of the electric point machine and the electro-hydraulic level crossing barrier, monitored for fault detection and diagnosis, are compared with the pneumatic assets. The parameters monitored on the pneumatic assets, air pressure, airflow and displacement, represent the working status of the pneumatic driven assets. The variation of these parameters reflects the performance changes of the assets and, therefore, the faults could be detected by the monitoring system. In the case of the electric and electro-hydraulic assets, the parameters for monitoring are different.

Electric and electro-hydraulic STMEs have measurable parameters for condition monitoring. In the electric point machine, the voltage supplied for the electromotor and the current through the inside coil show the performance of the motor. The displacement can also be monitored to describe the movement profile of the drive bar and mechanically connected switch blades. In previous studies (Marquez 2007 and 2008), a force transducer was employed to the electric point machine for the

generation of the output force data. However, this measurement was not included in this study, since the force output is a specific quantity for the point machine. Using this parameter conflicts with the aim of a generic solution in which the displacement of all five STMEs is used as a common feature. On the other hand, the measurement of the displacement and the output force are considered to represent the same dynamic characteristics; in other words, these two parameters are intrinsically linked and reflect the performance in two different ways. In the electro-hydraulic level crossing barrier, the voltage and current of the electric motor and the oil pressure in the hydraulic pump provide the information to show how the machine works. Unlike the output force for the electric point machine, the motor current and oil pressure are both monitored since they respectively represent two devices within the machine. Importantly, as the parameter presenting the throw dynamic, the rotary displacement of the axis is also monitored.

In comparison with the pneumatic STMEs, the parameters of the electric point machine also present the condition of the system. The voltage and displacement show the condition of the machine input and output. The current, which gives a better description of the motor condition than the voltage, directly reflects the force output and represents the health status of the whole machine.

The input and output of the electro-hydraulic level crossing barrier are figured by the motor voltage and the rotary displacement of the axis in the machine box. The current indicates the health condition of the electromotor. The oil pressure profile is very useful to detect a leakage or power failure of the hydraulic pump. These two parameters are also intrinsically linked to the performance of the barrier raising up or

falling down. The health condition of the level crossing machine is well described by a combination of these two parameters.

As members of the classification of STMEs, the electric point machine and electro-hydraulic level crossing barrier have similar performance, known as throws (an operation includes one normal and one reverse throw), as the pneumatic assets, where the throw dynamic is generically presented by the displacement profiles. For each asset, the status can be presented by a specific parameter for input, machine and output, based on which the fault detection is applicable by monitoring the parameter changes.

3.2.1.2 System features

Due to the similar mechanisms and power source of the pneumatic assets, their system features are discussed together. In order to extract common system features from all the assets considered in this study, the system features of the electric and electro-hydraulic assets are analysed by comparing them with the pneumatic assets. Although the electric and electro-hydraulic assets are very different in both mechanism and power source from the pneumatic assets, common features can still be found at system level. In Figure 3.21, the system structures of three types of assets are displayed.

3.2.1.2.1 Pneumatic assets

As shown in Figure 3.21, the whole process of the operation of pneumatic STMEs consists of three main sections: input, machine and output. In each section, at least one parameter can be used to represent or indicate the condition. This structure is not specific to these pneumatic assets, as it can be found widely in various systems, such

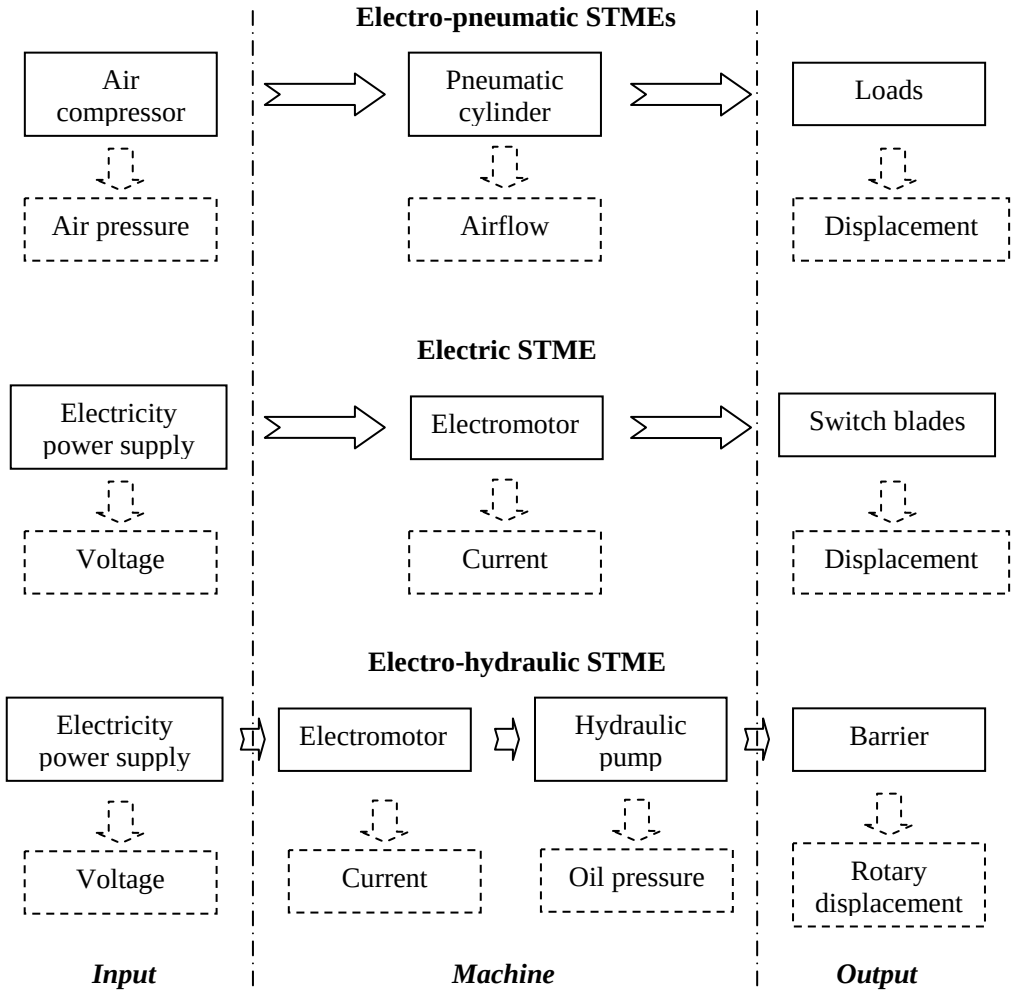


Figure 3.21 Comparison of three groups of STMEs.

as elevators and electric windows in cars. However, if a generic solution is required for the three items of differently designed equipments, the generic features would be an essential base.

Input

The input includes power supply and control command signals. The status of the input can be ascertained by monitoring the air pressure and the corresponding command signals.

Machine

A pneumatically driven machine can be generally split into two functions: actuation and force transmission. The actuation is performed by an actuator to generate force output upon the receipt of a command signal. The control unit of the actuator, the solenoid, is electrically activated to magnetically control the valves to enable the air to flow into or out of the cylinder. With the pressure difference increasing in the chambers of the cylinder, the piston is pushed to generate an output force to drive the load. The force is transferred to the load through the transmission system, which includes the driving shaft, gearing system, etc. The load is mechanically connected to the transmission system, such as the train door panel, and it is therefore classified as a part of the machine. During the operation of the machine, airflow can be observed, which reflects the internal working conditions.

Output

The output is mainly referred to as the movement of the driven load which is represented by the measured displacement profile. This parameter is a presentation of the machine dynamics, which is very useful for condition monitoring.

3.2.1.2.2 Electric and electro-hydraulic assets

The mechanism design determines the fundamental principle of the assets. The principal features of the three groups of STMEs were introduced previously. For comparison purposes, the features are briefly listed as follows:

- The pneumatic assets operate using the power of compressed air. The energised solenoid valves control the air in/out of the chambers of the cylinder to drive the piston and consequently the mechanically connected load.

- The electric point machine is powered by electricity. A motor provides a rotary force to drive the switch blades through a gear system and a drive bar.

- The electro-hydraulic level crossing barrier is powered by electricity. A hydraulic pump, driven by an electric motor, delivers oil under pressure to extend the actuator and thus raise the barrier.

In Figure 3.21, the system structures of assets are displayed schematically in the absence of the signalling and self-protection systems. These STMEs differ in the following respects: 1) different power supplies are used as the driving forces; 2) the loads are different and have different characteristics; 3) the mechanisms are significantly different for each of the three types of STMEs. Despite these differences, a generic systematic feature can still be observed, as shown by the vertical dotted lines. These assets are open-loop controlled machines composed of three parts: input, machine and output. This type of structure could be found in many systems, whether big or small, complex or simple. However, for the purpose of generating a generic FDD solution, it is still a useful basis for the algorithm development.

With the structure defined at the system level, the electric and electro-hydraulic assets are considered to have similar system features as the pneumatic assets. Thus there is potential that the structure of the generic FDD solution designed for pneumatic assets could be applied to these two railway assets.

3.2.2 Generic FDD methodology

As the aim of this study, a generic FDD solution should cover a large number of railway assets as an STME group and provide an efficient and low-cost condition

monitoring system. Based on feature extraction and analysis, a generic view of the measurable parameters and system structure was presented. Consequently, a generic FDD solution will be proposed and explained in this section.

3.2.2.1 Principle of generic FDD method

The model-based fault detection and diagnosis solution normally relies on an accurate system model. The more accurate the model is, the better the condition of the system can be identified. For the five railway assets, more modelling work and more sensors could be applied to improve the level of response to the FDD method. However, these efforts would limit condition monitoring to each single machine, instead of being applicable to a group of assets with different mechanical designs. Furthermore, this more precise FDD solution would greatly increase the economic cost and require longer development time, which deviates from the original intention of this study. The generic FDD solution therefore emphasises on the common features extracted from these STMEs and proposes to monitor a large number of them by using a relatively small number of sensors and relatively simple mathematical models. The generic solution will rely on the common features of the mechanisms and the integration of models. Fault detection will then be implemented by using several sub-models and the fault diagnosis will be based on the logical analysis of mechanical common features in the form of model combination.

The operation process of the STME can be divided into three parts: input, machine and output. At each part, a parameter represents the status. Based on the mechanical design, the three parts are dependent upon each other. The performance of the machine is influenced by the input. In other words, the input can determine the

performance to different extents. The machine also affects the output and in the same way the output is indirectly affected by the variation of the input.

I_f	M_f	O_f	O => M	$O \not\Rightarrow M$	**Fault location**
0	0	0	N/A	N/A	No fault
1	1	1	1	0	Input fault
1	0	0	1	0	Pressure/Voltage sensor
0	0	1	0	1	Displacement sensor
0	1	0	0	1	Airflow/Current/Oil pressure sensor
0	1	1	1	0	External fault (friction etc.)
0	1	1	0	1	Internal fault (cylinder etc.)

Table 3.3 Basic rules of fault diagnosis for STMEs.

Based on the connections between the three parts, a logical fault diagnosis regulation can be constructed. The input is defined as I_f for a faulty status. In the same form, the machine is defined as M_f and the output is defined as O_f. O => M represents that the output keeps consistent with the machine variables in fault mode, which can be identified by the synchronisation between the displacement and airflow (or current, oil pressure); meanwhile, the inconsistence, represented by $O \not\Rightarrow M$, indicates that mechanical faults might exist inside the machine. The basic rules for fault diagnosis on STMEs are displayed in Table 3.3. In the table, the '1' represents 'true' and the '0' represents 'false'.

When applying the rules to a logical analysis for the condition monitoring, the fault model O => M could be a key point of the solution. To apply this model, it is important that a distinguishable change exists in the synchronisation between machine variables and the output (displacement) when the machine itself operates under normal and faulty conditions. This requirement may lead to a failed diagnosis for small faults inside the machine which only slightly affect the machine variables or

which do not affect at all. In this case, the fault could only be located by long term observation.

3.2.2.2 Process of generic FDD method

Based on the previous analysis, a generic fault detection and diagnosis diagram was drawn to illustrate the FDD process in detail, as shown in Figure 3.22. In this figure, the whole FDD process was divided into three stages: sensor inputs and pre-processing, fault detection processes, and fault diagnosis processes. In each of the processes, the tasks are generically defined for all the railway assets considered in this study.

The generic fault detection and diagnosis method proposed in this study is based on a generic model of the STMEs, where the generic model is considered to be a combination of a series of sub-models. The sub-models are the models of the variables presenting the working status of the assets. Before applying this generic FDD method to the assets, the modelling work for these sub-models was carried out and the details will be presented in Chapter 4. During the process of fault detection, these variables need to be monitored by specific sensors and the data are collected. The acquired data is then used for observing deviations between the measured profiles and estimated profiles produced by sub-models to determine whether the system is working normally or not. Any detected change of the measured data would potentially indicate a fault. The change is, therefore, processed to residuals by a pre-designed algorithm to finally determine a fault alarm, to be passed to the next procedure, fault diagnosis. Using the fault information obtained during the fault detection procedure, the fault diagnosis algorithm works to identify the fault type and location.

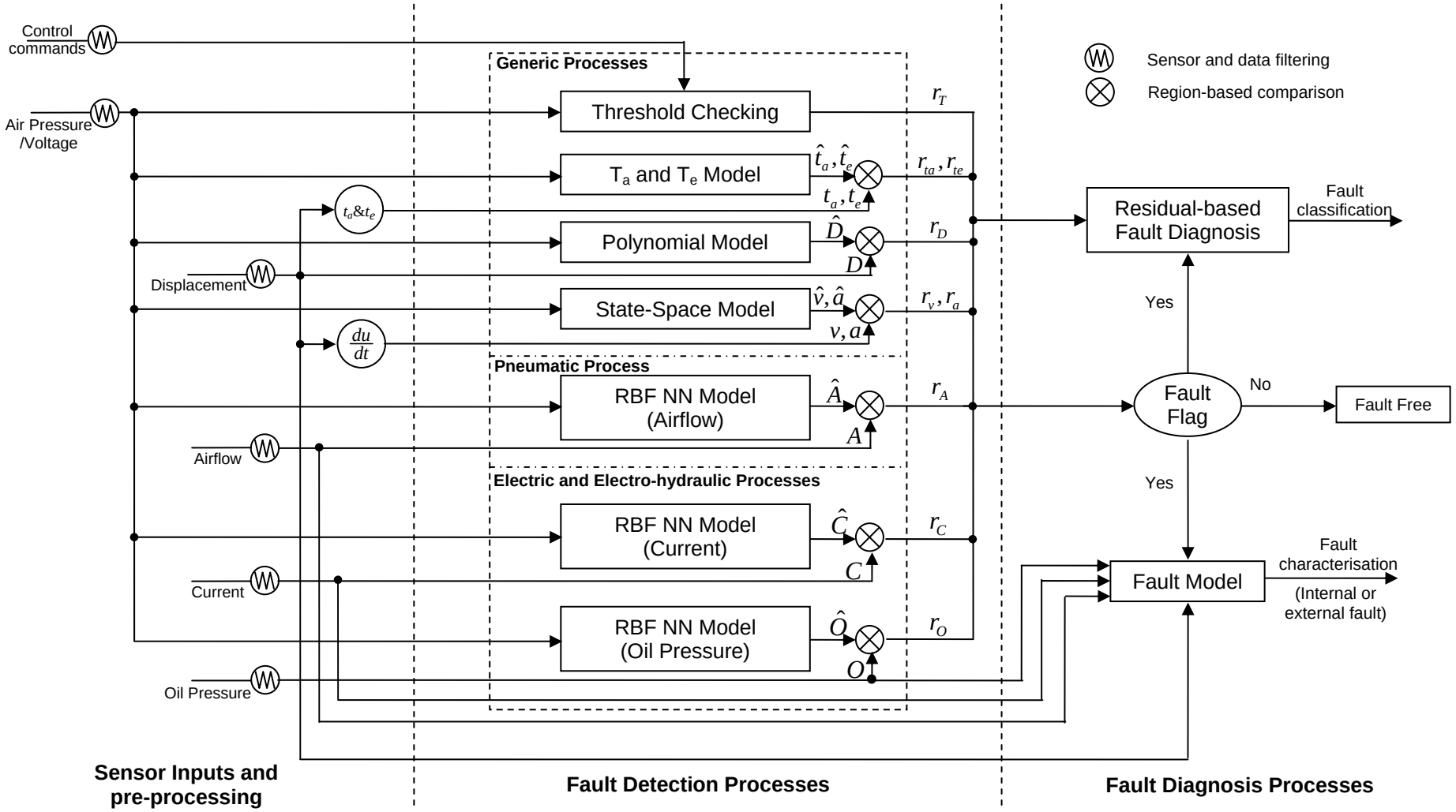


Figure 3.22 Diagram of generic fault detection and diagnosis for STMEs.

The final result of this generic FDD method would either be that the asset works under fault-free conditions or that the fault(s) occurred with relevant fault information, such as the fault is on the machine or caused externally.

Sensor inputs and pre-processing

At the first stage of the FDD process, the data of the variables in each section of the system (input, machine and output) are collected by sensors. As displayed in Figure 3.22, the variables monitored in this generic FDD process include the control command signals, air pressure/voltage, angular or linear displacement, current and oil pressure. Due to unavoidable noise during data collection and the sensor resolution, the data requires pre-processing by a low-pass filter to smooth the profile at a predetermined level. Furthermore, the variables, velocity and acceleration, which represent the dynamic of the system performance, are calculated using the processed displacement data. For the T_a and T_e models, the activation delay, t_a, and the throw time, t_e, are abstracted from the measured displacement.

Fault detection processes

At this stage the model based fault detection is carried out with the pre-processed data. The measured data is compared with the data estimated by the pre-designed sub-models to detect changes. As discussed previously, the whole profile of variables was divided into pre-defined spatial regions in order to increase the accuracy of comparison. The comparison of the measured and estimated data profiles is, therefore, performed within each spatial region. To determine whether an inconsistency represents an actual fault, an adaptive thresholding algorithm was developed, which will be explained in detail in the next section. Based on the thresholds, the

inconsistency would be considered either as a fault or tolerable. Consequentially, the inconsistency is then processed and passed to the fault diagnosis procedure as residuals.

The fault detection process contains three sections: generic processes, pneumatic process, and electric and electro-hydraulic processes. Each of the sections has different functions during the fault detection process.

The generic processes have three models and one threshold checking algorithm. The variables considered in this section exist among all the five assets.

- Threshold check for compressed air pressure and command signals:

  $$V_{Th} = f_{Threshold}(P, T_P, S, T_S) \tag{3.18}$$

  where P and S are the pressure and the amplitude of the command signal for monitoring. T_p and T_s are the given thresholds for the pressure and command signal respectively. V_{Th} is the comparison result which indicates normal or faulty status.

  The threshold checking produces two residuals r_P and r_S to indicate whether the voltage of the control command and the initial air pressure/voltage are in a normal range.

- The T_a and T_e model for activation delay and throw time:

  $$T_{a\&e} = f(\exp(k_{1\cdots n}, P_{1\cdots m})) \tag{3.19}$$

  where $k_{1\cdots n}$ is the coefficient of the exponential model; $P_{1\cdots m}$ is the air pressure. Since this model is built to respond to m pressure inputs, one single T_a (activation delay) or T_e (throw time) value is achievable at a certain pressure

value. $T_{a\&e}$ is an activation delay or throw time value at a given pressure.

Two residuals, r_a and r_e, are produced to show the difference between the model estimated and measured activation delay and throw time.

- The polynomial model for displacements:

$$Disp_{poly} = f(poly(r_{1\cdots i}, t), P_{1\cdots m}) \tag{3.20}$$

where $r_{1\cdots i}$ is the coefficient of the $(i\text{-}1)^{th}$ polynomials. The model generates a 3-dimensional output responding to a time sequence, t, and a pressure range, $P_{1\cdots m}$. The output $Disp_{poly}$ is a displacement data set at a selected pressure P.

The residual, r_D, represents a series of residuals, r_{d1}, r_{d2}, … , r_{dk}, where k is the number of spatial regions. These residuals indicate the comparison result in each spatial region of the displacement.

- The State-Space model for the throw dynamic including velocity and acceleration:

$$Dyn_{ss} = f(SS_{1\cdots j}((a,b,c,d), p), P_{1\cdots m}) \tag{3.21}$$

where $SS_{1\cdots j}$ is the State-Space model for j spatial regions and the a, b, c and d are parameters for each of the models. The p is the abscissa of spatial plotting. Responding to every pressure, $P_{1\cdots m}$, the spatial regions are defined and the State-Space model is set up for each of the regions. The calculation result, Dyn_{ss}, is therefore a velocity or an acceleration data set.

Two residuals, r_v and r_a, are produced for the velocity and acceleration respectively. r_v contains r_{v1}, r_{v2}, … ,r_{vk}, and r_a contains r_{a1}, r_{a2}, … , r_{ak} for the spatial regions, where k is the number of the spatial regions.

The pneumatic process has one model for airflow that is available for the three pneumatic assets (airflow data is not available from the train-stop, since no proper sensor was found).

- The neural network model for airflow estimation:

  $$A_{NN} = NN(p, A_{1\cdots m}) \tag{3.22}$$

  where p is the spatial X-axis and $A_{1\cdots m}$ is the airflow profile responding to the pressure range, $1\cdots m$. This neural network model is trained using healthy airflow data sets in a spatial scale. The A_{NN} is an airflow data set estimated by the model at a certain air pressure.

  The residual, r_A, representing a series of residuals r_{A1}, r_{A2}, … , r_{Ak}, where k is the number of spatial regions, indicates the inconsistency of the estimated and measured airflows.

The electric and electro-hydraulic processes have two models working for the electric current and hydraulic oil pressure respectively.

- The neural network models for current estimation of the electro-hydraulic level crossing barrier and the electric point machine:

  $$C_{NN} = NN(p, C_{1\cdots m}) \tag{3.23}$$

  where p is the spatial X-axis and $C_{1\cdots m}$ is the current profile responding to the voltage range, $1\cdots m$. This neural network model is trained using healthy current data sets in a spatial scale. The C_{NN} is a current data set estimated by the model at a certain voltage.

  The residual generated by using this model, r_C, is composed of r_{C1}, r_{C2}, … , r_{Ck},

where k is the number of spatial regions. r_C indicates the deviation of the measured current from the model estimated values.

- The neural network model for oil pressure estimation:

$$OP_{NN} = NN(p, OP_{1\cdots m}) \tag{3.24}$$

where p is the spatial X-axis and $OP_{1\cdots m}$ is the oil pressure profile responding to the voltage range, $1\cdots m$. This neural network model is trained using healthy oil pressure data sets in a spatial scale. The OP_{NN} is an oil pressure data set estimated by the model at a certain voltage.

The residual, r_O, containing r_{O1}, r_{O2}, … r_{Ok}, represents the inconsistency of the measured and estimated oil pressure and k is the number of spatial regions.

These models mentioned above were constructed and applied for the assets considered in this study. The modelling work will be presented in Chapter 4.

Fault diagnosis processes

The residuals produced in the fault detection processes are used in the process of 'fault flag' to determine whether a fault occurred. Since the number of residuals could be large with an increase in the number of spatial regions, the residuals used for residual-based fault diagnosis processes are re-organised into another form based on the characteristics of the variables.

Another fault diagnosis method using fault models is also proposed for initial fault characterisation. The fault occurred during the performance can be classified as either an external fault (fault caused by external factors, such as friction or obstruction) or an internal fault (fault caused by the machine mechanism, such as broken force

transmission). This initial fault characterisation would be very helpful for maintenance.

The fault model relating faulty displacement to correlated faulty airflow, current or oil pressure:

$$X_T = NN(D_{f_{1\cdots m}}, X_{f_{1\cdots m}}) \tag{3.25}$$

where $D_{f_{1\cdots m}}$ is the faulty displacements and $X_{f_{1\cdots m}}$ is the faulty airflow, current or oil pressure at *m* pressure/voltage values. This neural network model is trained using faulty displacement data as an input and the target is the correlated faulty airflow, current or oil pressure. The detail of the fault diagnosis processes will be introduced in section 3.2.4.

3.2.3 Residual generation

In model-based fault detection and diagnosis, the generation of residuals is a central issue for inconsistency identification and fault information collection (Chen and Patton *et. al.* 1999). In Chapter 2, three common methods for residual generation were introduced. In this section parity equations are employed for FDD of the pneumatic STMEs.

3.2.3.1 Parity equations approach

As discussed in Chapter 2, the parity equations method compares the parity (consistency) of the analytical models with the measurements of a real system. The experiment-based models can be considered as analytical approaches. Here the models are applied in parallel with the system process and the results are used to generate parity vectors.

In the spatial scale of throw distance, the models can be written in a general form as:

$$y(k) = f(P, u(k_0)) \tag{3.26}$$

where the function f can take both linear or non-linear formats; P is the model coefficient vector for linear models. In neural network models, P represents the parameters, such as the number of neurons, goal and spread and $u(k_0)$ is the model input. When the displacement is taken as the input, $k_0=k$ where k is the spatial scale of throw distance. When the input is air pressure, $k_0=0$ and $u(0)$ represents the initial pressure of the air supply. According to equation 2.14 in Chapter 2, the residuals can be described as:

$$r(k) = W(f(k) - f_m(k)) \tag{3.27}$$

where $f_m(k)$ is the measurement at position k and W is the residual generation vector.

3.2.3.2 Adaptive threshold

With the models constructed under fault-free conditions, it is possible to generate healthy STME performance data. The residual generation method also provides the methodology for producing inconsistencies in a mathematic form.

A thresholding process is normally involved in the decision-making stage of model-based FDD (Chen and Patton 1999). The choice of threshold is not a straightforward issue, as it determines the robustness of the whole FDD algorithm. If a fixed threshold is used, the sensitivity of faults will vary depending on the limit selected. When the threshold is chosen to be too high, the sensitivity to faults will reduce; on the contrary, the false alarm rate will increase if a threshold is chosen to be too low. In addition, a fixed threshold is often not sufficient with the presence of model uncertainty and measurement noise.

In order to enhance the robustness of the FDD system, an adaptive threshold approach is proposed, which will take model uncertainty and measurement noise into account. This approach improves the monitoring of performance with increased detection sensitivity and few false alarms (Patton, Frank and Clark 1989). The methodology and procedures for designing adaptive threshold were summarised theoretically by Ding and Frank (1991) and Emami-Naeini *et. al.* (1988).

3.2.3.2.1 Model analysis using statistical theory

In this section, the train door displacement of a normal throw is taken as an example for analysis. The polynomial model (model 3.20) discussed in section 3.2.2.2 is used to provide healthy displacement data at 3.5 bar. This polynomial model was constructed using a mean displacement value of 200 throws (normalised in the region of [0,1]) under fault-free conditions. Even though the 200 train door displacements were measured under the same conditions, the data profiles were not identical due to the variation in lubrication level. The maximum deviations between the model output and the 200 measured displacements are displayed in Figure 3.23. The residual amplitude varies from -0.015 to 0.015 through the whole throw, however, it can be seen that the distribution of residuals is symmetrical from the standard.

The deviations displayed in Figure 3.23 indicate that the standard displacement generated by the polynomial model does not precisely match the actual displacements due to inevitable model uncertainty. In a practical environment, this model would cause a high false alarm rate or failure of fault detection, if a fixed threshold were to be applied. Therefore, the threshold needs to be adaptively designed with consideration for the model uncertainty and noise which may exist.

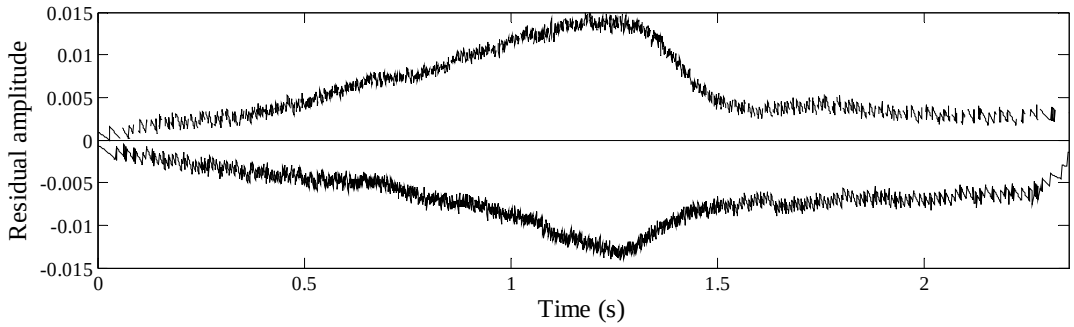


Figure 3.23 The residuals under fault-free conditions.

Further investigation of the residuals was conducted by statistically counting the residuals at different amplitude levels for all the 200 experimental displacement data sets. Based on the counting result, a probability distribution function (PDF) could be estimated.

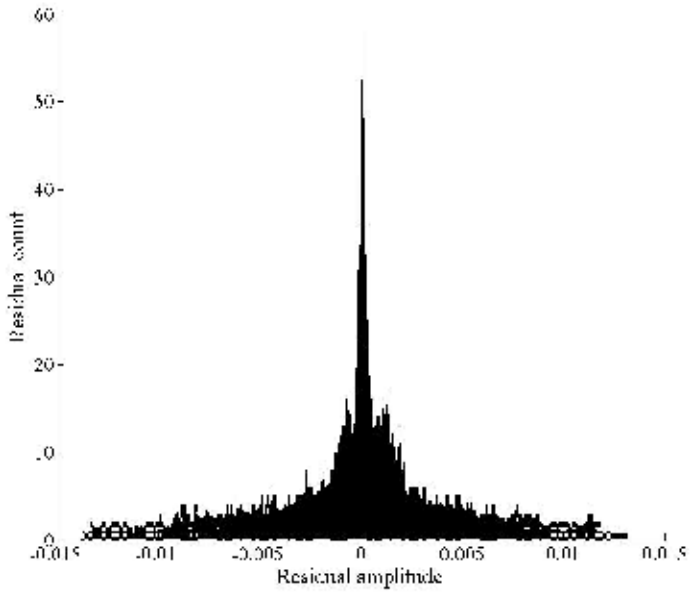


Figure 3.24 Distribution of residuals *vs.* throw times.

As shown in Figure 3.24, the counting result of 200 throws indicates that the count value is high when a residual approximates to zero, and the count value drops gradually to zero when the residual amplitude increases. This histogram therefore

shows a typical feature of a Gaussian distribution. The Gaussian distribution, also known as normal distribution, is a continuous probability distribution used to describe the data that clusters around a mean (Bishop 2006). As shown in Figure 3.25, the residuals were counted according to their amplitudes and the density distribution was fitted by a Gaussian function with a 95 % confidence level.

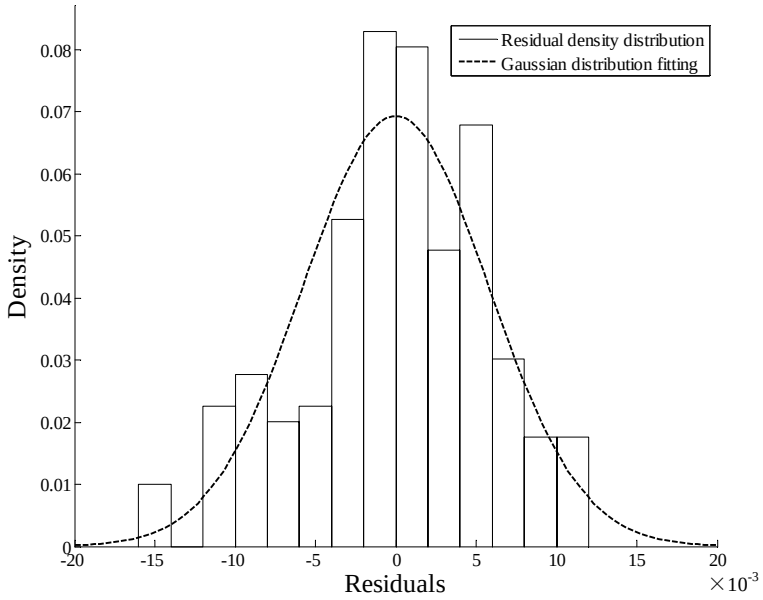


Figure 3.25 A histogram of residual density distribution and Gaussian distribution fitting.

The probability density function for Gaussian distribution can be illustrated by an exponential function (Dougherty 1990):

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{3.28}$$

where μ is the mean; σ is the standard deviation. The estimated value of μ is 1.7×10^{-14} and σ is 5.75. Based on the assumption of Gaussian distribution, a Lilliefors test was carried out for confirmation.

The Lilliefors test, as an adaption of the Kolmogorov-Smirnov test, is used to test the

null hypothesis that data comes from a normal (Gaussian) distribution (Lilliefors 1967 and Gonzalez *et al.* 1977). The Lilliefors test statistic can be described as:

$$\max_x \left| SCDF(x) - CDF(x) \right| \tag{3.29}$$

which has the same form as the Kolmogorov-Smirnov test. In the function, x represents the sample vector; *SCDF* is the empirical cumulative distribution function (CDF) estimated from the sample and *CDF* is the normal CDF with mean and standard deviation equal to the mean and standard deviation of the sample. At a significance level of 0.01, the distribution of residuals under fault-free conditions is confirmed to be close to Gaussian distribution. A Gaussian distribution probability plot of the residuals is shown in Figure 3.26.

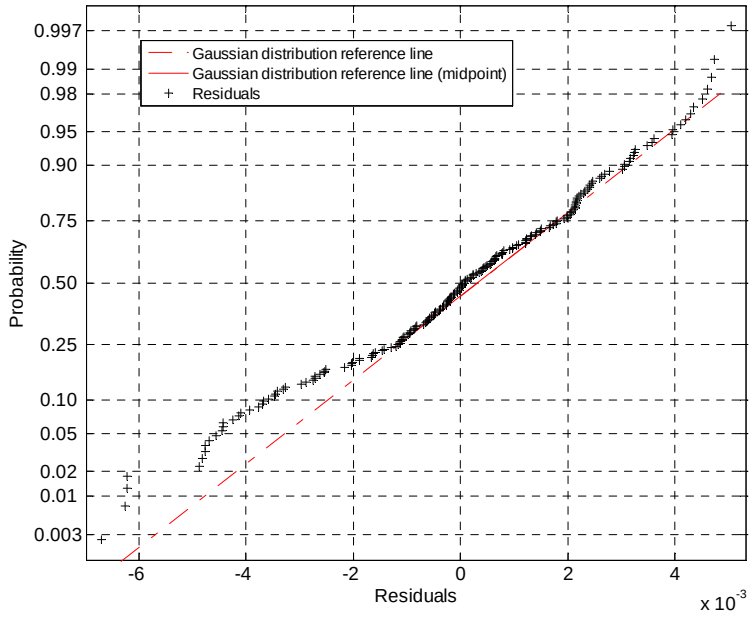


Figure 3.26 Gaussian distribution probability of residuals.

When the displacement profile generated by the polynomial model was compared with 200 measured displacement profiles, the measured profiles were found to be in

parallel with the model output without crossing. In this case, if the displacement profile generated by the model is taken as a standard, a mean value of the deviations of a measured displacement in time series could be used to represent its average deviation from the standard. In Figure 3.26, 200 average deviations obtained from 200 healthy displacement data sets are displayed. The linear distribution of the data indicates that the distribution of the data is close to a Gaussian distribution.

Based on the results of the Lilliefors test and Gaussian probability plotting, a conclusion can be drawn that the deviations of measured displacement profiles from the model output (the displacement profile generated by the polynomial model) obey the Gaussian distribution. The deviations at any time instant are also Gaussian distributed.

3.2.3.2.2 Adaptive threshold design

The displacement profile is a set of discrete data, where the number of data equals the number of samples. In order to set up an adaptive threshold, at each time instant the deviations between measured healthy displacement and the displacement generated by the model need to be statistically analysed. If the number of data is assumed to be *n*, the distribution of the deviations at a time instant, t_k, can be described as:

$$P(t_k) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(t_k-\mu)^2}{2\sigma^2}} \tag{3.30}$$

where $k = 1 \cdots n$; μ and σ are the mean and the standard deviation at time, t_k. In the train door case, the normal (forward) throw of the pneumatic train door at the air pressure of 3.5 bar is employed as an example. In a practical environment, the air pressure might vary in a range, and different supplied pressures result in

corresponding displacements. In this case, it is useful to introduce the air pressure, *p*, to the threshold design. Therefore, the μ and σ at time, t_k, with air pressure, p_j, can be illustrated by the following functions.

$$\hat{\mu}(p_j, t_k) = \frac{1}{m}\sum_{i=1}^{m} r_i(p_j, t_k) \tag{3.31}$$

$$\sigma(p_j, t_k) = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(r_i(p_j, t_k) - \hat{\mu}(p_i, t_k))^2} \tag{3.32}$$

where $j = 1 \cdots q$, q is the number of predetermined air pressure values; $\hat{\mu}$ is the estimated mean value. In this study, the air pressures supplied for the pneumatic train door are at an increment of 0.1 bar from 2 to 6 bar, where, for example, q is 41; m is the number of healthy experimental displacement data sets at pressure, p_j; r is the deviation at time, t_k.

From statistical theory, a confidence level, represented by 1-α, is normally used to indicate the reliability of an estimation, where α is a number between 0 and 1. In practical applications, the confidence level for Gaussian distribution is often selected to be between 95 % and 99 % (Shi et. al. 2005). With a given confidence level, the probability function can be described as follows using the mean, μ, and the standard deviation, σ (Crowder and Hand 1990).

$$P\{\hat{\mu} - z\sigma < \mu < \hat{\mu} + z\sigma\} = 1 - \alpha \tag{3.33}$$

where z is a coefficient related to the confidence level. This equation presents the confidence level by the confidence limits of the mean. In the case of thresholding, this equation indicates that (1-α)·100 % deviations lie in the interval between $\hat{\mu}(p_j, t_k) - z\sigma(p_j, t_k)$ and $\hat{\mu}(p_j, t_k) + z\sigma(p_j, t_k)$. In equation 3.33, the estimated mean

and standard deviation are usually known parameters, but *z* needs to be calculated. The Gauss error function is therefore employed, which can be written as:

$$erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \tag{3.34}$$

This function presents the probability that the deviations have a distance less than *z* from the mean value at the centre. The *z* can therefore be calculated using the reverse error function as follows.

$$z = \sqrt{2} erf^{-1}(1-\alpha) \tag{3.35}$$

When a confidence level is given as 97 %, α is then 0.03 and the coefficient $z = 2.17$. If the standard displacement profile generated from the polynomial model is represented by $D(p_j)$, the two thresholds at a confidence level of 97 % can be described as ($k = 1 \cdots n$):

$$\xi(p_j) = D(p_j) + \hat{\mu}(p_j, t_k) \pm 2.17\sigma(p_j, t_k) \tag{3.36}$$

With the consideration of the variation of air pressure, this design of adaptive threshold is capable of producing a series of thresholds for measured displacements under the air pressures, p_j.

In Figure 3.27 and 3.28, the adaptive thresholds with 97 % confidence for the displacement and airflow of a train door normal throw are displayed for 6 temporal regions (the air supply was provided at 3.5 bar). The displacement and airflow data used for model fitting were the average of 200 sets of experimentally collected data under healthy conditions. Meanwhile, the 200 sets of data were used for the adaptive threshold generation, where the possible varieties of healthy performance of the train door were supposed to be included. In the two figures, the threshold is represented by the dashed red lines and the dynamics of displacement and airflow at each time point

are outlined, which indicates that the level of the variation of healthy data would be tolerated by these thresholds. The false alarm rate would therefore be reduced.

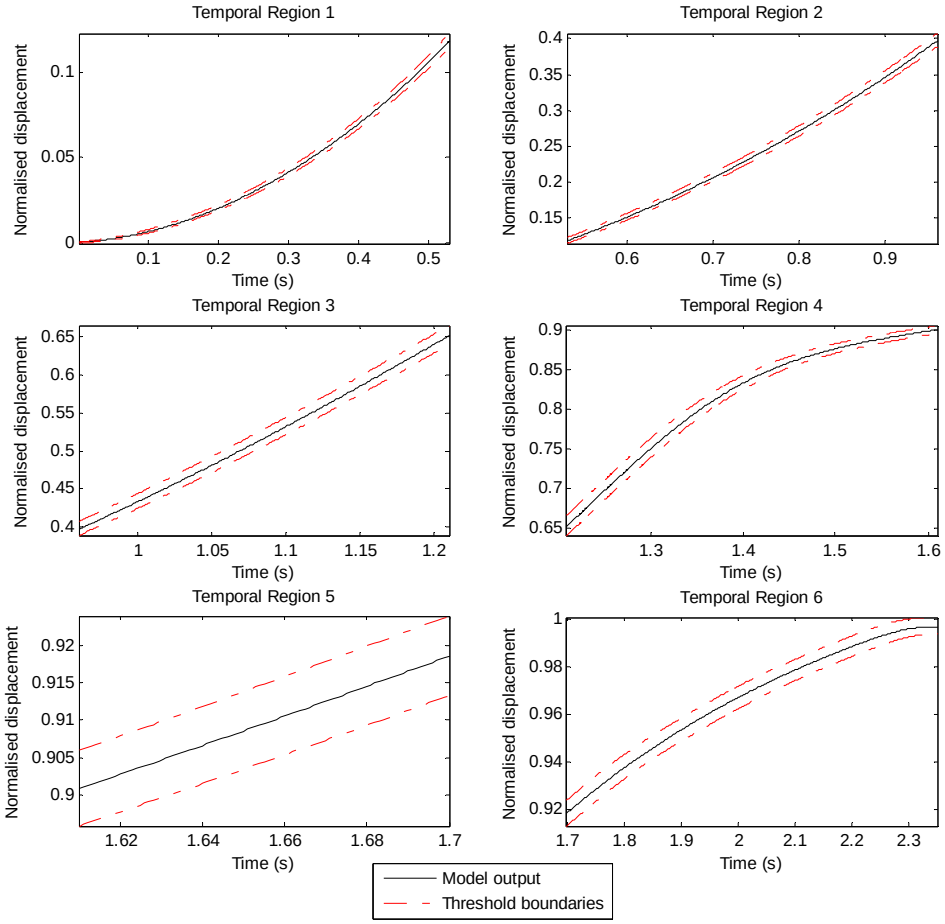


Figure 3.27 Adaptive thresholds for displacement of a train door normal throw generated by a polynomial model.

The confidence level selected for threshold design directly influences the fault detection accuracy. A higher confidence level increases the threshold to tolerate larger deviations from the modelled prediction. The detection sensitivity for minor faults is therefore lowered. However, the false alarm rate is also reduced. The selection of confidence level is therefore a compromise between the fault detection accuracy and the false alarm rate.

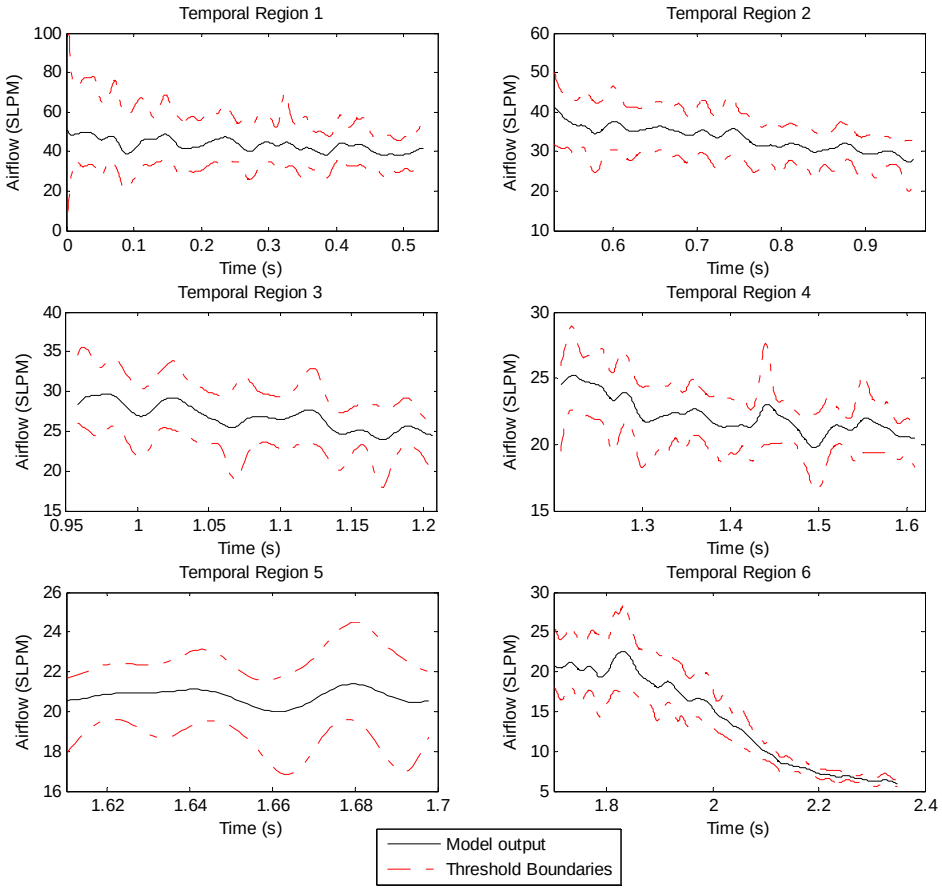


Figure 3.28 Adaptive thresholds for airflow of a train door normal throw generated by a state space model.

3.2.4 Fault diagnosis

As explained in Chapter 2, a fault detection and diagnosis system involves both detection and diagnosis, where the diagnosis of faults is also known as identifying the classification, location and strength of the faults. Previous sections introduced the generic fault detection method for the railway assets considered in this study. So far, two fault diagnosis methods have been designed for the pneumatic assets, which are illustrated in this section. For the purpose of fault diagnosis, the fault codes are also defined for all five assets.

When a fault occurs on an asset and prevents it from performing a normal operation, three items of information are normally required to identify the fault: what the fault is, where/how it occurs and the size or strength of the fault (Collacott 1977 and Patton *et. al.* 1989). The first question requires gathering and analysing the fault information from residuals generated when the fault is detected. The residuals could be mapped to certain faults providing that the residual patterns are unique to each other. The answer to the second question can be generally said to be that the fault is either inside the machine or caused by external factors. The last question is the most difficult to answer. In this section, a solution is initially proposed by combining the parameter estimation and inference system to identify the fault type and strength.

3.2.4.1 Fault codes

In order to identify the failure modes for the five assets considered in this study, the faults, which may occur during operation, are classified and coded. In practice, the faults could usually be located to different mechanical or electrical components. However, in order to apply the generic fault detection and diagnosis method for the assets with different mechanisms, these various faults are categorised in a similar format.

Pneumatic train door

Table 3.4 lists the failure modes of a pneumatic train door. These faults were also simulated on the test rig and corresponding data were collected for FDD method evaluation. The data collection in healthy and fault modes was carried out at a range of air pressure values with 0.1 bar increments from 2 to 6 bar.

Code	Failure mode	Details
TD-F0	Healthy operation	Open and close profiles with nominal air pressure of 3.5 bar
TD-F1	Friction	Significant friction resulting in a larger activation delay and throw time
TD-F2	Obstruction	The train door panel is stuck and failed to complete an operation
TD-F3[1]	Mechanical faults	Poor sealed cylinder piston or broken linkage between piston and driven load
TD-F4	Non-critical leakage	A slight leakage on the air pipe or cylinder, but complete operation can be performed
TD-F5	Critical leakage	A severe leakage on the air pipe or cylinder resulting in operational failure
TD-F6	Non-critical air supply failure	Air pressure is in the range of 2 to 5 bar, but not at 3.5 bar (experimental definition)
TD-F7	Critical air supply failure	Air pressure is lower than 2 bar or higher than 5 bar (experimental definition)[2]
TD-F8	Pressure sensor failure	Intermittent signal transmission or sensor failure
TD-F9	Displacement sensor failure	Intermittent signal transmission or sensor failure
TD-F10	Airflow sensor failure	Intermittent signal transmission or sensor failure

[1] *Fault was not simulated. Other faults were simulated and data were collected.*
[2] *High air pressure (> 5 bar) may cause hazard to passengers due to the high speed of train door movement.*

Table 3.4 Description of faults considered in train door test rig.

Although the performances under non-nominal pressures can be considered as a failure mode, they can also be seen as healthy for those where faults do exist under the same pressures. Taking this into consideration, the experimental operations include 2250 operations in healthy mode and 410 operations in a failure mode.

Pneumatic train-stop

Fault simulation and data collection were also performed for a train-stop. Since no airflow sensor was available for the train-stop, the mechanical fault and airflow sensor fault were not considered. The pressure range defined in experiments for train-stop operation is 0.1 bar increments from 2 to 6 bar. In total, 1230 operations were performed for healthy conditions with 205 faulty operations. In Table 3.5, one healthy mode and eight fault modes are listed.

Code	Failure mode	Details
TS-F0	Healthy operation	Normal and reverse throws with nominal air pressure of 4.1 bar (60 psi)
TS-F1	Friction	Significant friction resulting in a larger activation delay and throw time
TS-F2	Obstruction	The train-stop head is stuck and failed to complete an operation
TS-F3	Non-critical leakage	A slight leakage in the air pipe or cylinder, but complete operation can be performed
TS-F4	Critical leakage	A severe leakage in the air pipe or cylinder resulting in operation failure
TS-F5	Non-critical air supply failure	Air pressure is lower than 4.1 bar, but higher than 2 bar (experimental definition).
TS-F6	Critical air supply failure	Air pressure is lower than 2 bar (experimental definition).
TS-F7	Pressure sensor failure	Intermittent signal transmission or sensor failure
TS-F8	Angular-displacement sensor failure	Intermittent signal transmission or sensor failure

Table 3.5 Description of faults considered in the train-stop test rig.

Pneumatic point machine

Code	Failure mode	Details
PM-F0	Healthy operation	Normal and reverse throws with a nominal air pressure of 3 bar
PM-F1	Friction	Significant friction resulting in a larger activation delay and throw time
PM-F2	Obstruction	The point machine rod is stuck and failed to complete an operation
TD-F3*	Mechanical faults	Poorly sealed cylinder piston or broken linkage between piston and driven load
PM-F4	Non-critical leakage	A slight leakage on an air pipe or cylinder, but complete operation can be performed
PM-F5	Critical leakage	A severe leakage on an air pipe or cylinder resulting in a failure of operation
PM-F6	Non-critical air supply failure	Air pressure is lower than 3 bar, but higher than 0.5 bar (experimental definition for unloaded case)
PM-F7	Critical air supply failure	Air pressure is lower than 0.5 bar (experimental definition for unloaded case)
PM-F8	Pressure sensor failure	Intermittent signal transmission or sensor failure
PM-F9	Displacement sensor failure	Intermittent signal transmission or sensor failure
PM-F10	Airflow sensor failure	Intermittent signal transmission or sensor failure

** Fault was not simulated. Other faults were simulated and data were collected.*

Table 3.6 Description of faults considered in point machine test rig.

Two point machines, one healthy and the other with a leaky air cylinder, were used to generate data in both healthy and failure modes. The air pressure range was set up at

0.1 bar increments from 0.5 to 6 bar. The unloaded healthy point machine was used to generate healthy data for model construction. The total operations of the point machine in healthy mode were 1680, and 560 operations were performed for each failure mode. In Table 3.6, one healthy and ten failure modes are listed.

Fault code	Failure mode	Practical faults
LCB-F0	Healthy	Fault-free condition
LCB-F1	Friction	Main shaft bearing tight Boom up/down slowed Incorrect balance weight
LCB-F2	Obstruction	Main shaft seized Counter balance weight obstructed Hydraulic system blockage Maintenance door obstructing boom
LCB-F3	Motor fault	Pump motor failure
LCB-F4	Hydraulic pump fault	Piston rod broken Hydraulic fault in pack Hydraulic oil leakage
LCB-F5	Power failure	Power pack disconnected Power pack on incorrect boom length
LCB-F6	Displacement sensor fault	N/A
LCB-F7	Oil pressure sensor fault	N/A
LCB-F8	Current sensor fault	N/A
LCB-F9	Voltage sensor fault	N/A
	Other failures	Main shaft moved Solenoid valve fails Circuit controller seized/general wear Down stop missing Detecting relays disconnected High resistance in electric circuits Circuit controller band dirty Boom resting > 90 degree Circuit controller spring high resistance

Table 3.7 Faults considered for the electro-hydraulic level crossing barrier.

Electro-hydraulic level crossing barrier

Based on a Failure Mode and Effects Analysis (FMEA) (Roberts 2007), the faults were analysed and sorted into categories in the same form as the fault types for the pneumatic assets. The reason this work was undertaken is that the generic FDD solution is designed only to the same type of faults for different assets, and the faults thus cannot be accurately located to single components.

The fault types considered in this study and the corresponding practical faults are listed in Table 3.7. The practical faults included are only those which could result in a degraded or failed operation. Since the FMEA focuses on the faults of the machine mechanism, no practical faults were suggested for sensor faults. For the development of generic FDD algorithms, the data in fault modes is required, however, it is difficult to simulate the mechanical faults by damaging the components. The simulation of an obstruction fault would also raise a high risk to both the operator and the machine. Therefore, the machine has only been operated with a simulated friction fault and the data has been collected.

Fault code	Failure mode	Description
EPM-F0	Healthy	Fault-free condition
EPM-F1	Friction	Throw time is longer than normal, but the throw is completed.
EPM-F2	Obstruction	Throw fails to complete
EPM-F3	Motor fault	Faulty motor performance reflected by current
EPM-F4	Power failure	Voltage failure
EPM-F5	Displacement sensor fault	Sensor failure or bad connection
EPM-F6	Current sensor fault	Sensor failure or bad connection
EPM-F7	Voltage sensor fault	Sensor failure or bad connection

Table 3.8 Faults considered for electric point machine.

Electric point machine

Since a Fault Mode and Effects Analysis (FMEA) from experts was not available for the electric point machine, the fault modes for consideration are initially outlined as a generic view, as shown in Table 3.8.

3.2.4.2 Fault model approach

The fault model in this study is understood as a generic model of the assets to describe performance in fault mode and to generate fault information at the same time. The construction of a fault model normally requires in-depth knowledge of how the system physically works and how it could go wrong. For a generic FDD solution, however, specific physical descriptions of the assets are not considered and the possible faults are only considered generically as listed in section 3.2.4.1. In this situation, a simple fault model was designed. How it works on the pneumatic assets is explained in this section.

In the case of pneumatic assets, displacement and airflow are two important parameters which reflect the performance of the assets. In this study, displacement is considered as the final presentation of the operation of the asset, which determines whether the operation is healthy or faulty. In parallel with displacement, the airflow is generated from the working cylinder as a by-product, which also presents the process of the operation. For example, the speed (acceleration or deceleration) of a vehicle represents the performance, however, the exhaust of a gas-based vehicle can also show the working status and, furthermore, the exhaust indicates the condition of the engine, of which the speed may not be able to give any information. Airflow is

therefore an important parameter which provides information about the cylinder and the mechanical system of the machine.

When a fault occurs on a pneumatic asset, the displacement is always affected. The change in displacement is often similar no matter whether the fault is caused by a faulty machine or an external factor, i.e. friction. For airflow, the situation is different in that airflow is also influenced when a friction or obstruction fault occurs, but it correlates with the displacement profile (as mentioned in section 3.2.1.1.1) only if the pneumatic system is healthy. When a fault occurs on the machine, e.g. badly sealed cylinder piston or broken linkage from piston to driven load, the airflow will lose synchronisation with the displacement. Another issue to point out is that the fault of leakage does not influence the synchronisation between the airflow and the displacement. From the working principle of the air cylinder introduced in section 3.2.1.1.1, the mass airflow only relates to the air pressure and the movement of the cylinder piston. This phenomenon was also observed during the analysis of collected data.

A fault model (function 3.25) was designed to identify whether a fault was caused externally or internally, where the air leakage was classified as an external fault. This fault model involves a radial basis function neural network (RBFNN) to predict an airflow profile using an actual measured displacement profile and compares it with measured airflow. This type of neural network has the capability of accurate approximating with adequate training data (Patton *et. al. 1999*). The data used to train the network includes the data collected from fault simulation, where the faults are only caused by external factors such as friction, obstruction and air leakage. The result

of comparison between the fault model output and measured airflow indicates that the fault is considered to be caused externally and the asset is in a healthy condition, if the two airflow profiles are similar. Otherwise, the fault is likely to be caused by an internal machine fault. External faults were used for model training because these faults are easier to simulate on the test rig than mechanical faults, which would need to be simulated by damaging a machine.

3.2.4.3 Residual analysis approach

The residuals produced by the fault detection procedure are essential information for fault diagnosis, especially when the system is taken as a black box without knowledge of its internal workings. In order to make the FDD method generic to all the assets considered in this study, their physical details were not taken into account and the assets were considered as black boxes only with inputs and outputs. The residuals are therefore the unique information for fault diagnosis. Compared with the method using a fault model introduced in the last section, the analysis of residuals is expected to generate more accurate classification results by relating the residual patterns to certain faults.

The residuals were defined for three pneumatic assets, which are displayed in Table 3.9, 3.10 and 3.11, where '1' represents true and '0' represents false. Twelve residuals were designed for the pneumatic train door and the point machine; and 9 residuals were designed for the pneumatic train-stop with the absence of airflow information. With the same parameters monitored on these three assets, the residuals are also of the same design, which facilitates the generic consideration for fault diagnosis.

Fault Code	**Residuals**											
TD-F0	0	0	0	0	0	0	0	0	0	0	0	N/A
	1	1	1	1	1	1	1	1	0	1	0	0
TD-F1	0	1	1	0	1	1	0	1	0	1	0	0
	0	0	1	0	0	1	0	1	0	1	0	0
	1	1	1	1	1	1	1	1	0	1	1	0
TD-F2	0	1	1	0	1	1	0	1	0	1	1	0
	0	0	1	0	0	1	0	1	0	1	0	0
TD-F3	1	1	1	1	1	1	1	1	0	1	1	1
TD-F4	1/0	1/0	1	1	1	1	1/0	1	0	1	0	0
TD-F5	1	1	1	1	1	1	1	1	1	1	1	0
TD-F6	1/0	1/0	1	1	1	1	1/0	1	1	1	0	0
TD-F7	1	1	1	1	1	1	1	1	1	1	1	0
TD-F8	0	0	0	0	0	0	0	0	1	0	0	0
TD-F9	1	1	1	1	1	1	1	1	0	0	1	1
TD-F10	0	0	0	0	0	0	0	0	0	1	1	1
Description	Start velocity	Intermediate velocity	Stall velocity	Start acceleration	Intermediate acceleration	Stall acceleration	Activation delay	Throw time	Air pressure	Airflow	Airflow synchronisation	Fault mode airflow
Residuals	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	r_{12}

Table 3.9 Residuals for pneumatic train door case.

Fault Code	**Residuals**								
TS-F0	0	0	0	0	0	0	0	0	0
	1	1	1	1	1	1	1	1	0
TS-F1	0	1	1	0	1	1	0	1	0
	0	0	1	0	0	1	0	1	0
	1	1	1	1	1	1	1	1	0
TS-F2	0	1	1	0	1	1	0	1	0
	0	0	1	0	0	1	0	1	0
TS-F3	1/0	1/0	1	1	1	1	1/0	1	0
TS-F4	1	1	1	1	1	1	1	1	1
TS-F5	1/0	1/0	1	1	1	1	1/0	1	1
TS-F6	1	1	1	1	1	1	1	1	1
TS-F7	0	0	0	0	0	0	0	0	1
TS-F8	1	1	1	1	1	1	1	1	0
Description	Start velocity	Intermediate velocity	Stall velocity	Start acceleration	Intermediate acceleration	Stall acceleration	Activation delay	Throw time	Air pressure
Residuals	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9

Table 3.10 Residuals for pneumatic train-stop case.

Fault Code	**Residuals**											
PM-F0	0	0	0	0	0	0	0	0	0	0	0	N/A
	1	1	1	1	1	1	1	1	0	1	0	0
PM-F1	0	1	1	0	1	1	0	1	0	1	0	0
	0	0	1	0	0	1	0	1	0	1	0	0
	1	1	1	1	1	1	1	1	0	1	1	0
PM-F2	0	1	1	0	1	1	0	1	0	1	1	0
	0	0	1	0	0	1	0	1	0	1	0	0
PM-F3	1	1	1	1	1	1	1	1	0	1	1	1
PM-F4	1/0	1/0	1	1	1	1	1/0	1	0	1	0	0
PM-F5	1	1	1	1	1	1	1	1	1	1	1	0
PM-F6	1/0	1/0	1	1	1	1	1/0	1	1	1	0	0
PM-F7	1	1	1	1	1	1	1	1	1	1	1	0
PM-F8	0	0	0	0	0	0	0	0	1	0	0	0
PM-F9	1	1	1	1	1	1	1	1	0	0	1	1
PM-F10	0	0	0	0	0	0	0	0	0	1	1	1
Description	Start velocity	Intermediate velocity	Stall velocity	Start acceleration	Intermediate acceleration	Stall acceleration	Activation delay	Throw time	Air pressure	Airflow	Airflow synchronisation	Fault mode airflow
Residuals	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	r_{12}

Table 3.11 Residuals for pneumatic point machine case.

- The residuals, $r_1 \cdots r_6$, are based on the displacement measurement. The three residuals generated from velocity data in a spatial scale could provide the information as to whether and where a fault occurred. As the dynamic changes are amplified by acceleration, the residuals generated from acceleration are useful for the detection of small incipient faults.

- The residuals of activation delay (r_7) and throw time (r_8) are time related, which are used to check whether an operation starts and ends within the nominal time range. These two residuals are sensitive to the faults that influence the operation time.

- The monitoring of air pressure provides a deviation from the nominal air pressure and identifies whether this deviation is within or outside an allowed pressure range.

- The airflow related residuals, $r_{10} \cdots r_{12}$ (for the train door and point machine), give information on the operation of the assets in parallel with the displacement. The airflow and fault mode airflow are used to check the dynamic change of the airflow profile. The monitoring of synchronisation proves whether the two sections of an asset, cylinder piston and driven load, are working in step.

The normal working sensors are critical for the application of the generic FDD method. Among the three sensors, it is considered that only one sensor is allowed to fail in each operation. The dependency of the residuals on sensors was analysed under the assumption that no other faults were involved.

Displacement sensor failure

The consequential residuals of a failed displacement sensor are provided in Tables 3.9, 3.10 and 3.11. In this situation, the initial fault diagnosis can still be carried out by comparing the output of the airflow sensor with the modelled prediction of airflow with respect to the reading of the air pressure sensor. The result could identify either the displacement sensor fault or that the asset failed to operate upon receiving the command signal.

Airflow sensor failure

The consequential residuals of the airflow sensor failure are also presented in the tables. The FDD could still be performed using the displacement based residuals and air pressure value. In the case where no other faults are present, the airflow sensor fault could be identified.

Air pressure sensor failure

The models, 3.20 to 3.22, depend on the air pressure value to generate a corresponding prediction of the displacement, velocity, acceleration and airflow. A faulty air pressure sensor could lead to the failure of these models. In this case, the fault detection and diagnosis would totally collapse. In order to avoid this situation, two methods could be used to decouple the sensitivity of these models to an air pressure sensor fault.

The first solution was mentioned in previous literature by Roberts (2002). The decoupling of the displacement related residuals can be realised through the calculation of secondary residuals. A precondition is that the regions need to be more finely divided to enable the first order approximations to replace the state-space

models for velocity and acceleration (3.21). The primary residuals are achieved directly from the comparison between actual measurements and the modelled prediction. The secondary residual could be calculated by eliminating the first order variable (pressure) using subtraction for two primary residuals. The secondary residuals will be independent of air pressure.

The other solution proposed is to predict the air pressure value during the failure of the sensor. The model 3.20 predicts the displacement with a given pressure value and stored polynomial coefficients. In the healthy mode, the displacement profiles at a pressure range of 0.1 bar increments from 2 to 6 bar were used to train the polynomial model. When the pressure reading is unavailable, this model could also be able to inversely predict the pressure using the measured displacement profile. The application of the inverse prediction of air pressure requires that the displacement sensor works normally to provide data. The inverse function of model 3.20 can be illustrated as follows.

$$P = f^{-1}(poly(r_{1\cdots i}, t), Disp) \tag{3.37}$$

Where the *Disp* is the measured displacement and *P* is the predicted air pressure. The function of this model is to compare the measured displacement profile with the predicted displacement automatically using the air pressure at 0.1 bar increments from 2 to 6 bar. Once the most similar one is found, the corresponding air pressure value will be the predicted air pressure.

3.3 Test rig development

A generic test rig for the five STME assets was set up in the laboratory (Appendix A).

The hardware overview of the test rig is shown in Figure 3.29. In this system, a computer was used as the control centre to send control instructions and receive data from sensors via two NI-DAQ USB-6008 data acquisition cards. The USB-6008 provides a maximum of 8 analogue input connections in single-ended mode or 4 analogue input connections while in differential mode. These analogue channels are used to interface with the sensors for data collection. The two analogue output channels of the USB-6008 can provide up to 5 V DC (5 mA). These are used to send control signals to the control units of the STMEs. Apart from the features described above, this DAQ card also has 12 TTL/CMOS digital channels, a 2.5 V DC (1 mA) reference voltage and a 5 V DC (200 mA) power supply.

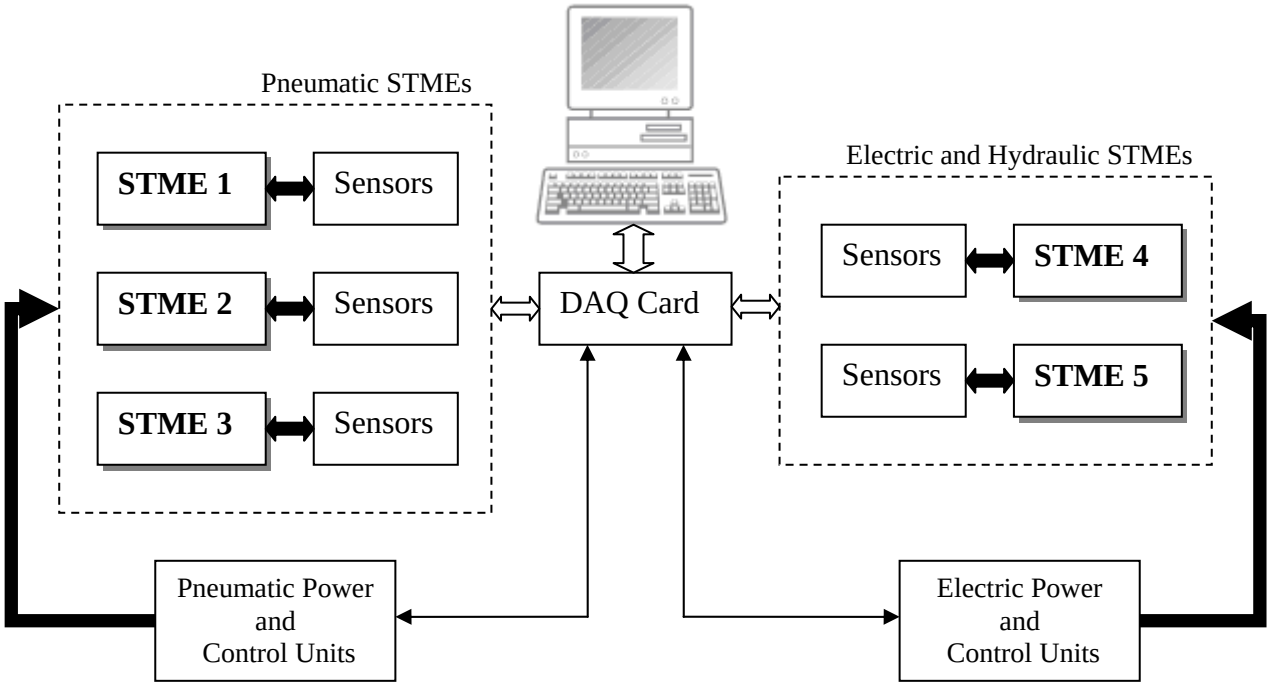


Figure 3.29 Overview of STME test rig hardware.

For the pneumatic STMEs, an air compressor (Airmaster Tiger 8/50 with maximum pressure of 8 bar and 50 litre capacity) was used to provide the power. In order to control the pressure of the output of compressed air, a digital air regulator (SMC

ITV2050-31F2BS3-Q) was employed. A 5 μm air filter was also used to dry and filter the air from the compressor to make the air regulator operate normally.

A 24 V DC heavy-duty electric power supply drove the level crossing barrier. The operation of the barrier was controlled by relays to start, stop and reverse the motor. In order to drive the motor of the electric point machine, a 1.5 kVA transformer was used to convert 240 V AC to a 110 V AC supply.

3.3.1 Sensors for data collection

STMEs	Parameters	Sensors*
Pneumatic Train Door	Linear displacement	Micro-epsilon WPS-1250 MK46
	Airflow	Honeywell AWM720P1
	Air Pressure	SMC ITV2050-31
Pneumatic Train-stop	Angular displacement	106-degree rotary position sensor
	Air pressure	SMC ITV2050-31
Pneumatic Point Machine	Linear displacement	Penny & Giles HLP190 LVDT
	Airflow	Honeywell AWM720P1
	Air pressure	SMC ITV2050-31
Electric Point Machine	Voltage	LEM CV 3-500
	Current	LEM PointSenz PCM 30-P
	Linear displacement	Micro-Epsilon WPS-250 MK30
Electro-hydraulic Level Crossing Barrier	Angular displacement	Micro-Epsilon WPS-250 MK30
	Voltage	LEM LV 25-P
	Current	LEM LTA 50P/SP1
	Oil pressure	GENSPEC GS4200

** All the sensors listed in the table were installed by the author for experimental purposes.*

Table 3.12 Sensors installed on STME assets.

Various sensors were installed on the assets for data collection. Several electric power supplies, providing 24 V DC, 10 V DC, 5 V DC and ±15 V DC power**,** were used to enable the sensors to work properly. A list of the used sensors is given in Table 3.12.

Displacement sensors

Two draw-wire displacement sensors, Micro-Epsilon WPS-1250 MK46 and WPS-250 MK30, were used to measure linear displacement up to 1250 mm and 250 mm respectively. The travel distances of the train door panel and the drive bar of the electric point machine, 812 mm and 122 mm, were measured using these sensors. The draw-wire displacement sensor measures movement using a highly flexible steel cable. The 'cable drum' is attached to a sensor element which provides a proportional output signal. This type of displacement sensor has the advantages of having a high dynamic response speed and good linearity. The resolution of WPS-1250 MK46 is 0.4 mm and that of WPS-250 MK30 is 0.1 mm. The WPS-250 was used to measure the angular displacement of the level crossing barrier. When the axis of the barrier machine rotates, the cable attached to the surface of the axis is drawn out vertically and the angular displacement is therefore calculated using the displacement reading and known axis radius.

A 106-degree rotary position sensor was attached to the pneumatic train-stop to measure of rotary displacement of the train-stop head.

Airflow sensor

The airflow exhaust of a pneumatically powered device can be measured. As a dynamic parameter reflecting the practical action of the cylinder, the mass airflow velocity was measured using a Honeywell AWM720P1 sensor on pneumatic STMEs.

This sensor provides in-line flow measurement with a specially designed flow housing and measures mass velocity up to 200 standard litres per minutes (SLPM) with a response time of 6 ms. This sensor was chosen because the maximum airflow of the point machine (160 SLPM) and the train door (80 SLPM) are higher than the measurement range of other sensors. A 10 V power supply is required by this airflow sensor and the output ranges from 1 to 5 V corresponding to 0 to 200 SLPM. The output signal is not linear with respect to the varying airflow, however, the response curve can be fitted by a second-order Gaussian function.

$$f(x) = a_1 e^{-((x-b_1)/c_1)^2} + a_2 e^{-((x-b_2)/c_2)^2} \tag{3.38}$$

Where x is the voltage signal and $f(x)$ is the corresponding airflow value. The parameters were calculated using the typical input-output relations provided by the datasheet of the airflow sensor and are listed in Table 3.13.

a_1	b_1	c_1	a_2	b_2	c_2
2.972×10^{16}	24.34	3.359	81.64	5.46	2.337

Table 3.13 Gaussian curve fitting parameters for airflow sensor.

Air regulator/pressure sensor

In order to make the compressed air pressure adjustable, an air regulator, SMC ITV2050-31F2BS3-Q, was employed. This pneumatic regulator allows the setting of the output pressure up to 9 bar with 0.01 bar adjustment resolution, which satisfies the experiments in which a maximum of 6 bar pressure is required. The power supply used was a 24 V DC supply and the input signal ranged from 0 to 10 V DC. The pressure in 0.1 bar resolution can be displayed by an LED indicator and can also be read from the output signal (1 to 5 V DC) to achieve a more accurate value.

Voltage sensor

Two voltage sensors, LEM CV 3-500 and LEM LV 25-P, were used to monitor the electrical power supply. The CV 3-500 measures the primary voltage up to 500 V (r.m.s. 350 V). With a ±15 V DC power supply, the conversion ratio is 500 V / 10 V and the accuracy is at ±0.2 % at 25 ºC. The LV 25-P performs the same function as the CV 3-500, but the accuracy is lower at ±0.8 % at 25 ºC. The nominal electrical supplies for the electric point machine and the level crossing barrier are 110 V DC and 24 V DC respectively, which both fall in the measurement range of these two sensors.

Current sensor

In order to measure the current of the motors in the electric point machine and the level crossing barrier, two current sensors, LEM PointSenz PCM 30-P and LEM LTA 50P/SP1, were employed respectively. The PCM 30-P requires a 24 V DC power supply and measures current up to 30 A nominally. With a ±15 V DC power supply, the LTA 50P/SP1 measures nominal current of 50 A and gives output at the ratio of 100 mV / A.

Pressure sensor

A pressure sensor, GENSPEC GS4200, was used to measure the oil pressure in the hydraulic system of the level crossing barrier. This sensor provides a measurement range of 0 to 700 bar. With a 15 V DC power supply the output span is 0 to 10 V DC to indicate the pressure reading, i.e. 700 bar pressure gives 10 V output voltage.

Software

To enable the functionality of the hardware, the software was designed to control and

read the sensor via the NI DAQ card. National Instrument's LabVIEW was chosen as the developing tool for software programming. LabVIEW provides a graphical development environment to interface with the DAQ card for machine control and data acquisition (Appendix B).

3.3.2 Load simulation for point machines

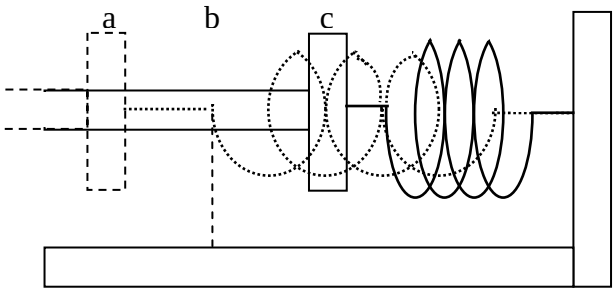
As a test rig, a M63 electric point machine was installed on a machine base in the laboratory. With a 110 V DC power supply, the point machine operated normally and the data was collected through a computer based data acquisition system and installed sensors, including a voltage, a current and a draw-wire displacement sensor.

In order to create a practical application environment, a load was designed to simulate the dynamic force generated by the switch blades. The load device used a strong spring, of which the design parameters are listed in Table 3.14.

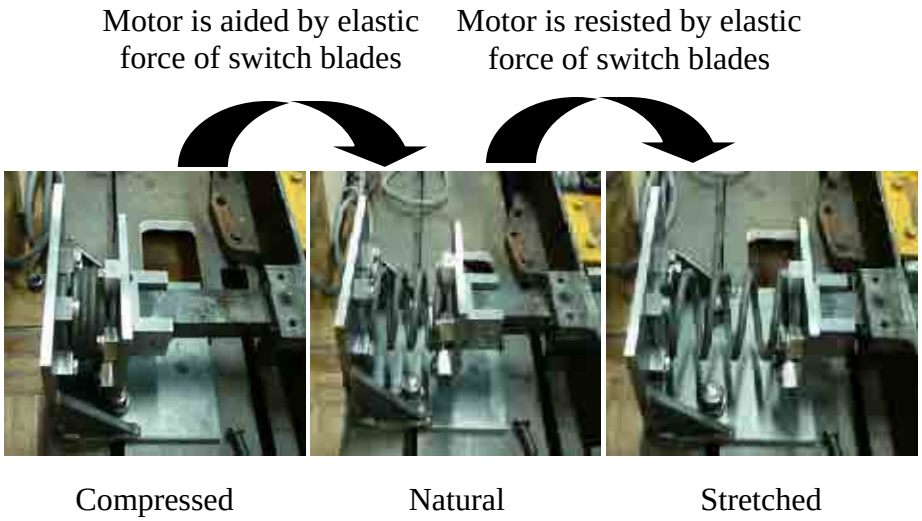
Natural length	Deformed length (Stretched)	Deformed length (Compressed)	Spring rate	Max load
14 cm	20.25 cm	7.75 cm	43200 N/m	2700 N

Table 3.14 Parameters of load simulated for electric point machine.

In practical application, the drive bar of the point machine is connected to the switch blades through the stretchers, as shown in Figure 3.9. Once the point machine is activated, the drive bar moves vertically to the switch blades and, therefore, the switch blades are pushed or pulled to one side of the stock rails, via which two routes are switched. In the design of the load, the force output by the drive bar was calculated and the dynamic changes were considered.



a.



b.

a) Schematic structure of simulated load b) Images of the load responses

Figure 3.30 Load simulation using single spring.

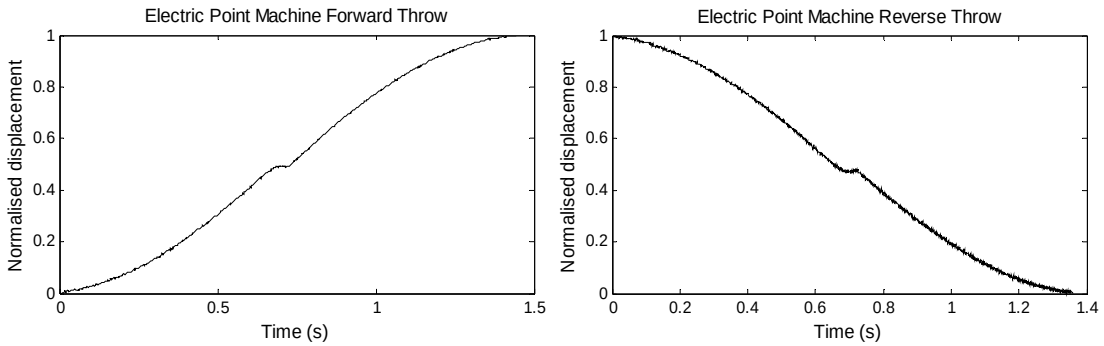


Figure 3.31 Displacement profiles for electric point machine with load.

The point machine moves a variable load due to the bending of the switch blades. Initially, the motor of the point machine is aided by the released force of the switch blades. As the blades traverse to their final position, an increased force is exerted as the switch blades bend in the opposite direction (Oyebande 2002). At the end of the throw, the motor stops and the switch blades are locked. To simulate these working statuses, a single spring load device was developed as shown in Figure 3.30 a. The elastic forces of the spring on compression and stretching are used to simulate the similar force experienced when the switch blades of a point machine are bent. When the drive bar of the point machine travels from *a* to *c* and from *c* to *a*, it experiences forces similar to those experienced during a real throw. The load was fitted at the other end of the drive bar instead of the one connected to switch blades. As shown in Figure 3.30 b, for the normal (forward) throw (the switch blades are pushed to the farther side of stock rails), the travel starts from the point the spring is fully compressed to the end where the spring is fully stretched. The reverse throw repeats the process inversely. The test results (shown in Figure 3.31) proved that the simulation basically reflects the practical dynamic process; however, this one spring system is still too simple to accurately present the force changes of the practical stock rails, and the force generated by the spring is still weaker than expected.

3.4 Conclusions

Five railway assets, the pneumatic train door, train-stop and point machine, the electric point machine and the electro-hydraulic level crossing barrier, are introduced in this chapter. A generic fault detection and diagnosis method using common features was proposed. A method of adaptive thresholding was also proposed and explained.

Two fault diagnosis approaches for the pneumatic assets, fault model and residual analysis, were also discussed. For the development of the generic FDD method, test rigs were set up in the laboratory and the details were also presented.

The assets considered in this study are different in both function and mechanism; however, they have similar dynamic characteristics during their operations. Based on these similarities, such as non-periodical and reciprocating operations with large nonlinear load, these assets are classified into a group called Single Throw Mechanical Equipments (STMEs).

As the aim of this study is to design a generic FDD method for these STMEs, the common features of the assets are the emphasis for consideration and utilisation. Two aspects of the features were extracted: the parameter features, including the displacement, airflow/current/oil pressure and air pressure/voltage; and the mechanism features, including the input, machine and output.

Based on the information monitored on parameter features and the structure outlined by system features, a generic FDD method was proposed. A common system structure for each asset is composed of input, machine and output with dependency relations. Each section in the structure can be represented by one corresponding parameter. Based on this system structure, a fault detection and diagnosis process diagram was provided for illustration of the FDD procedures. The processes of application of the generic FDD method include three stages: sensor inputs and pre-processing, fault detection processes, and fault diagnosis processes. In each of the processes, the tasks are generically defined. At the first stage, the performance data of the asset is measured by a series of sensors and pre-processed. At the stage of fault detection,

seven different types of sub-models are used for different variables, where the sub-models include the T_a and T_e model for activation delay and throw time, the polynomial model for displacement, the state-space models for velocity and acceleration, the radius basis function neural network model for airflow, current and oil pressure. Each of these sub-models works for one dependent task and the combination of them composes a generic model to describe the whole system. At the fault diagnosis stage, two processes, fault model and residual analysis, are included. The faulty model, built using faulty mode data, has the function of initial fault characterisation (external faults or internal faults). The residual analysis is to generate fault information by using reformatted residuals generated from the fault detection procedure. The diagram of generic fault detection and diagnosis (as shown in Figure 3.22) describes the procedures of fault detection and diagnosis and will be tested on the assets in the following chapters.

A parity equation approach for residual generation was discussed and the design of an adaptive threshold using statistical theory for fault detection and residual generation was introduced in detail. The threshold has the capability to tolerate normal dynamic variation for healthy operation data and thus minimises the false alarm rate. With confidence levels, the threshold could be adjusted to adaptively change the detection sensitivity for faults.

Two approaches to fault diagnosis for pneumatic assets were discussed. One approach uses the fault model, which benefits from a rough fault classification. Mechanical faults, such as leakage at the cylinder piston and failed linkage between the cylinder and the driven load, and external faults, such as friction or obstruction, could be

classified into different categories. The precondition of applying the fault model is that the sensors of displacement and airflow are in a normal working condition. Another approach is based upon the analysis of residuals and pointing the residual patterns to certain faults. Since the residuals come from the comparison between the prediction by the model and actual measurements by sensors, the decoupling of the residuals from sensor faults is important. Two methods were introduced to decouple residuals from the air pressure sensor fault. The fault diagnosis methods have not been tested on the test rig, which will be left as a future work.

The test rig for data collection from these assets was built in the laboratory. The hardware, including sensors, control units and power supplies, was introduced in detail. The software programme, using the LabVIEW graphical language for controlling and data acquisition, was also introduced. Automatic data collection and storage was achieved by the combination of both hardware and software.

By the information provided above, the idea of generic fault detection and diagnosis was illustrated and clearly explained. For the five simple STMEs, the fault detection and diagnosis method and an adaptive threshold were deliberately designed using their common systematic and parametrical features. The results of applying this method are presented in Chapters 4 and 5.

Chapter 4

A generic fault detection and diagnosis approach for pneumatic train door

4.1 Introduction and motivation

This chapter presents a generic fault detection and diagnosis (FDD) approach for the pneumatic train door. Several modelling methods discussed in Chapter 3, such as exponential modelling (T_a and T_e modelling), polynomial modelling, state-space modelling and neural networks modelling, are introduced and combined to compose a practical generic STME model appropriate for FDD. The failure modes of the train door, derived from the failure mode and effects analysis (FMEA), were provided in Chapter 3, based on which the fault detection results are presented. An initial approach to fault diagnosis using the fault model is also discussed.

The motivation for research into safety-critical railway equipment is both to improve the reliability of asset performance and to enhance the quality of service to passengers. This study aims to develop a practical FDD method which will allow a large range of different assets to be effectively monitored and, at the same time, reduce the development cost of monitoring systems. In addition to the above benefits, detection of incipient faults should also be achievable, which is very useful for efficient infrastructure management.

The approach presented in this chapter is to illustrate how the proposed generic fault detection and diagnosis method works on an STME asset (the pneumatic train door)

and to test the effectiveness of this method.

4.2 Modelling for STMEs

As an essential part of fault detection and diagnosis, the modelling methods for the STMEs were studied and several modelling approaches are presented in this section. For the convenience of modelling work, the measured data was normalised into the region of [0,1].

4.2.1 Exponential model

The compressed air supply of the pneumatic STMEs may vary within a range. In this study, the performance of pneumatic assets was considered within the pressure range from the minimum value, less than that at which the asset will absolutely fail to operate, to the maximum value, the maximum pressure that the compressor can continuously supply, i.e. 2 to 6 bar for the pneumatic train door. These two pressure limits were set based on the laboratory environment rather than practical applications in which the threshold for failure pressure supply is likely to be stricter.

In the case of the train door, the 3-D profiles of normal and reverse throws are displayed in Figures 4.1 a and b respectively. The top views of these two 3-D profiles are shown in Figures 4.1 c and d with contour lines, in which the throw time can be clearly observed to be exponentially related to the air pressure. The activation delay (delay between the solenoid activation and the physical movement of the door) of the train door panel also has an exponential relationship with respect to the air pressure.

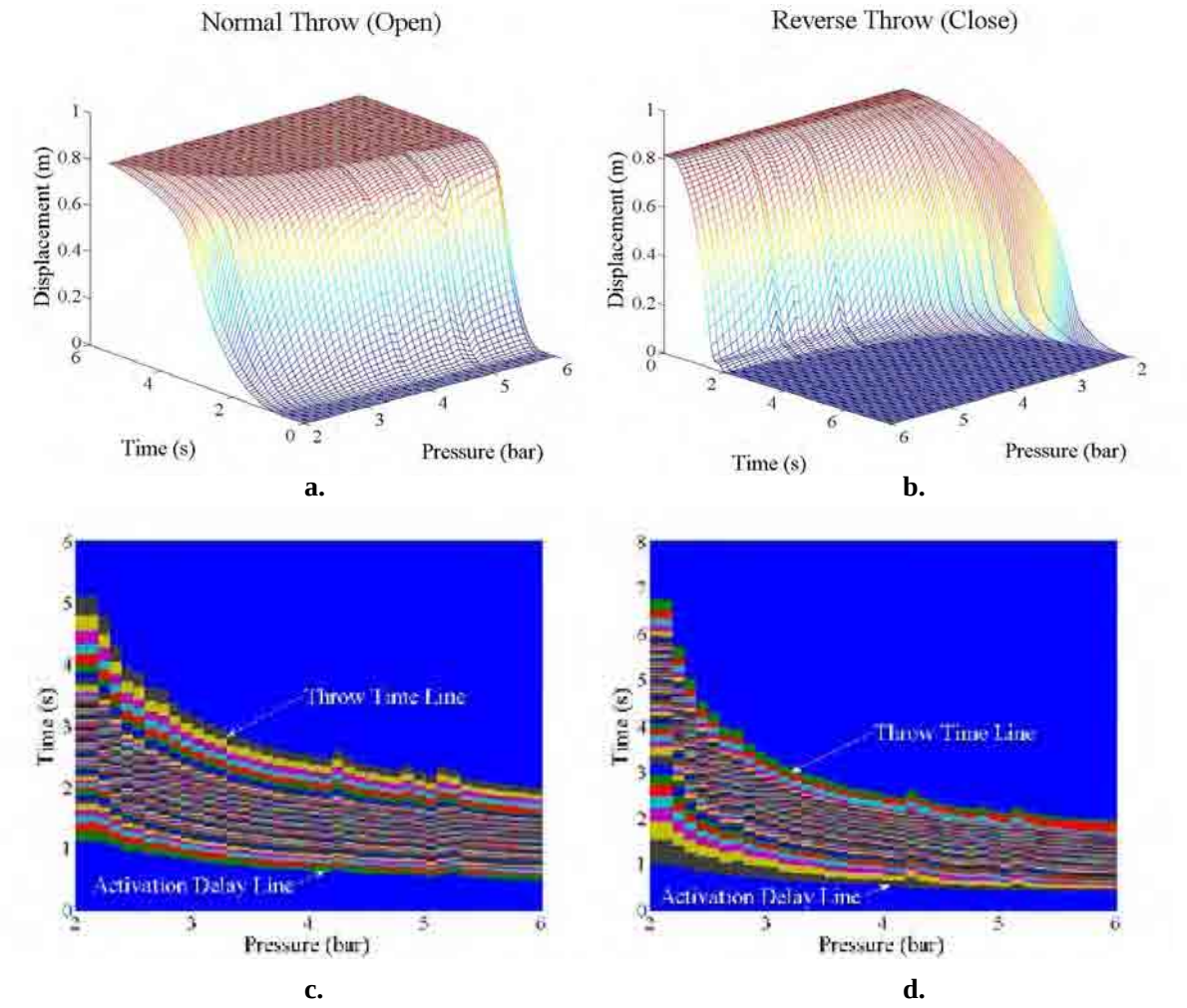


Figure 4.1 Exponential feature in the train door throws.

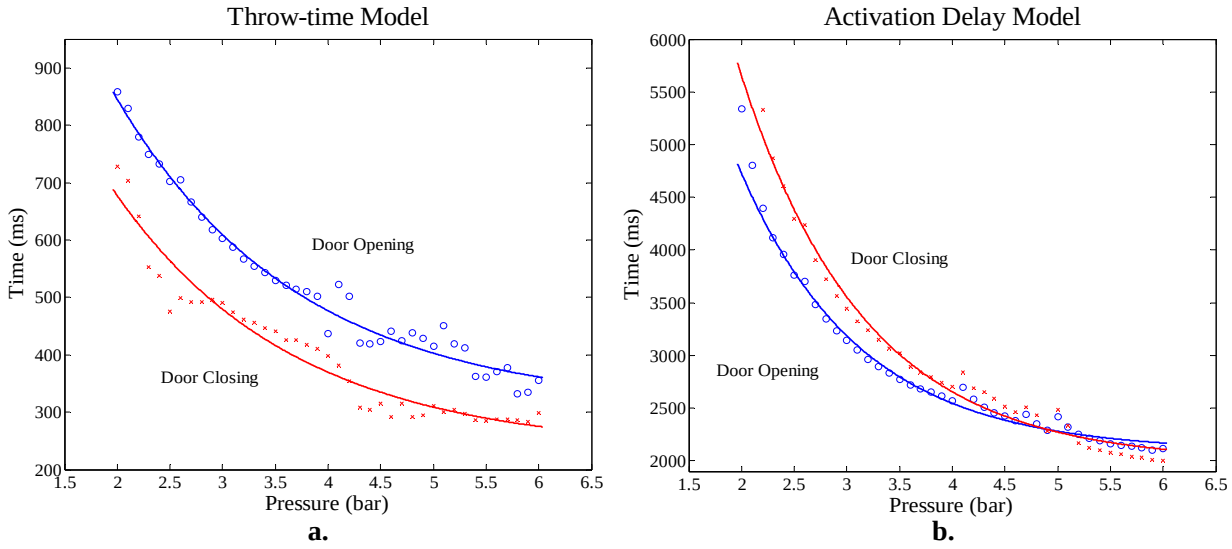


Figure 4.2 Throw time and activation delay *vs.* pressure for the pneumatic train door.

In Figure 4.2, the throw time and activation delay obtained from experimental data are displayed. To model these data, the curve was fitted using a general exponential equation.

$$t(u) = x_1 e^{-x_2 u} + x_3 \tag{4.1}$$

Where t is either the throw time or activation delay; u is the input pressure and x_1, x_2 and x_3 are the designed coefficients. The curve fitting process aims to select appropriate coefficients to minimise the misfit between the function and the training set data points, where the misfit is usually represented by an error function. A widely used error function is given by the sum of the squares of the error between the predicted and actual data points, which can be simply described as the follows (Bunday 1984 and Bishop 2007).

$$E = \sum_{i=1}^{n} \left[(x_1 e^{-x_2 u_i} + x_3) - t_i \right]^2 \tag{4.2}$$

Where i=1, 2, 3, … , n. n is the number of measurements. u_i is the i^{th} pressure value and t_i is the i^{th} measured throw time or activation delay.

Type	State	Model	Accuracy (R-square)
Throw-time	Opening	$t(u) = 454.4e^{-1.053u} + 2091$	99.25%
	Closing	$t(u) = 667.6e^{-1.02u} + 1987$	98.64%
Activation Delay	Opening	$t(u) = 168.4e^{-0.6942u} + 308.5$	97.73%
	Closing	$t(u) = 136.6e^{-0.7062u} + 233.2$	93.87%

Table 4.1 Temporal models of throw time and activation delay.

In Table 4.1, the optimised model parameters for healthy train door operation are shown. The data in Figure 4.2 for modelling are mean values obtained from 100

randomly selected throw data sets, thus the model can be thought to reliably reflect the nominal operation characteristics.

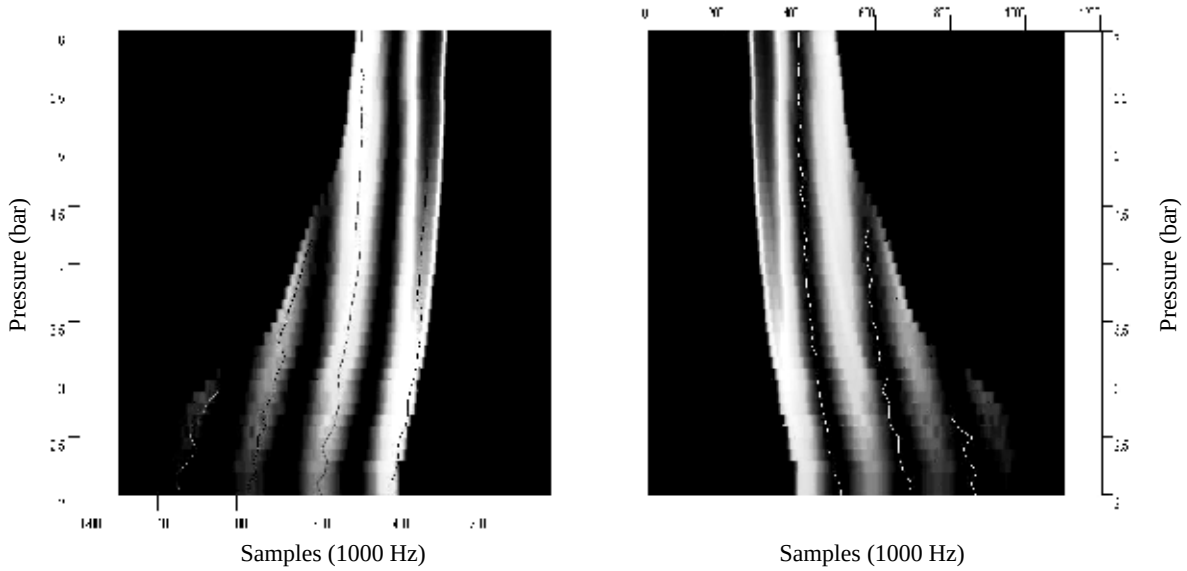
As shown in Table 4.1, the R-square was calculated to indicate the accuracy of models. This statistic indicates the goodness level of fit in explaining the variation of the training set data. The R-square is the square of the correlation between the predicted and actual values, which is defined as the ratio of the sum of squares of the regression and the total sum of squares and the function can be written as (Everitt 1998):

$$R-square = 1 - \frac{\sum_{i=1}^{n} w_i (t_i - \hat{t}_i)^2}{\sum_{i=1}^{n} w_i (t_i - \bar{t})^2} \tag{4.3}$$

Where w_i is the selected weight; $\hat{t}_i$ is the predicted value; $\bar{t}$ is the average measured value and t_i is the i^{th} measured throw time or activation delay. The R-square values in Table 4.1, which indicates accuracy by being close to 1 and poor accuracy by being close to 0, show that the models achieved a high accuracy, i.e. 99.25% of data could be explained by the throw time opening model.

Apart from the throw time and activation delay, the exponential feature can also be observed in the throw dynamics. This feature is more clearly presented by the throw velocity and acceleration. As an example, Figure 4.3 displays the top view of acceleration 3-D profiles of the train-stop normal throw, where the pressure applied ranges from 2 to 6 bar. In this figure, the light areas indicate the wave crests (the acceleration maxima) and the dark areas between the light areas are the wave valleys (the deceleration maxima). The maxima and minima of acceleration at different air

pressures were linked up and displayed by lines in the figure. It can be observed that these lines also exhibit exponential features between the time and pressures.



a) Trend lines of wave crests. **b) Trend lines of wave valleys.**

Figure 4.3 Peak values of acceleration (train-stop).

4.2.2 Polynomial model

The polynomial model is one of the most frequently used models for curve fitting. It has many advantages such as low computation cost and well-understood mathematic forms (Wolberg 1967 and Hunt 1993). The main aim of designing a polynomial model is to select a suitable model order to map the input and output of a given data sets covering all possible system operation conditions. In this study, the polynomial model is used to build up a functional relationship between the air pressure and displacement. As the displacement profiles of the pneumatic train-stop are the most complex for modelling among the five assets, the train-stop is selected as an example of polynomial modelling.

An m^{th} order polynomial function can be described in the form:

$$y(x,W) = w_0 + w_1 x + w_2 x^2 + \cdots + w_m x^m = \sum_{i=0}^{m} w_i x^i \tag{4.4}$$

where m is the order of the polynomial and x is the scalar input variable. $W=\{w_0, w_1, w_2,\ldots,w_m\}$ is a vector of designed model parameter. $y(x,W)$ is a nonlinear function of x and is a linear function of coefficients, W, which presents the model output.

The polynomial modelling process aims to select a polynomial function with an order that fits the given training data. The coefficients are determined by inverse calculations. The accuracy of fitting is determined by minimising the error function, which is in the same form as equation 4.2 and written as:

$$E(W) = \sum_{j=1}^{n} \left[y(x_j, W) - t_j \right]^2 \tag{4.5}$$

where t_j is the j^{th} target value.

The polynomial fitting in the train-stop case by the CFTool of MATLAB is taken as an example. Two methods can be considered to fit the displacement profile using polynomial functions. The first one is to map the input, air pressure, and the output, displacement, where the pressure varies in the range of 2 to 6 bar, in the case of the train-stop. However, due to the relatively large variation of the data, the fitting results, especially at 2 to 3.5 bar, were proved to be not good enough, even with the high 9^{th} order polynomial. Furthermore, the computation cost using high order polynomial functions is high. The second method is to model the displacement profiles using the time point as input and the position of the train-stop head at each time point as the target output. Compared with the first method, this second method benefits from a low computation power requirement and the ability to predict the displacement which is not included in the training data. For example, the training data for the displacement

profile was selected at every 0.1 bar increment from 2 to 6 bar, however, the model has the ability to generate the displacement in 0.05 bar increment.

In order to apply polynomial fitting on the training data, a discrete time interval needs to be selected. A small time interval will result in an accurate model, conversely, a large time interval makes the model less accurate. Lehrasab (1999) gave an analysis on time interval selection and suggested that 10 Hz (100 ms) was sufficient for dynamic presentation. However, for experimental purposes, a 10 ms time interval was chosen to provide higher accuracy and better prediction capability. On the other hand, a conclusion drawn from previous trials is that one polynomial function with fixed order is not sufficient for modelling due to the large variations among the datasets. Therefore, the data was grouped into four temporal regions (as shown in Figure 4.4) for appropriate fitting methods and polynomial orders.

Apart from polynomials, rational fitting was also used to model the data sets in temporal regions (290-390 ms and 590-1100 ms) as shown in Table 4.2. The rational fitting can be considered as an extension of polynomial functions in the form:

$$f(x) = \frac{p_1 x + p_2 x^2 + \cdots + p_{m-1} x^{m-1} + p_m x^m}{x^n + q_1 x^{n-1} + q_2 x^{n-2} + \cdots + q_{n-1} x + q_n} \tag{4.6}$$

Where p_1, p_2, … , p_m are the coefficients of numerator polynomial and q_1, q_2, … , q_n are the coefficients of denominator polynomial.

At a frequency of 100 Hz, 111 sets of data were extracted as the training data. As illustrated in Table 4.2, 3rd and 4th order polynomial and ‘cubic/quadratic’ and ‘cubic/cubic’ rational models were trained in four temporal regions. To ascertain the reliability of the model, the model output of training data is shown in Figure 4.5 a. In

comparison with the measured train-stop normal throw (Figure 3.6 a.), the performance of the model is good. With the function of prediction, the model also generates predicted displacement data at every 0.2 bar from 1.95 to 5.95 bar, as shown in Figure 4.5 b. The coefficients of polynomials are given in Appendix C.

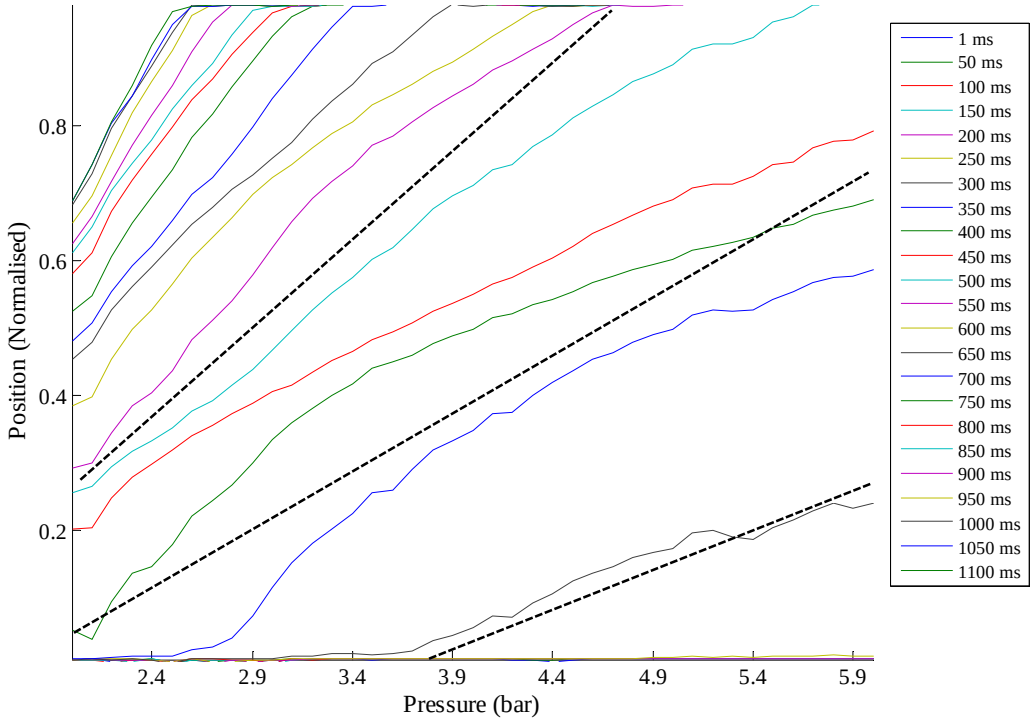
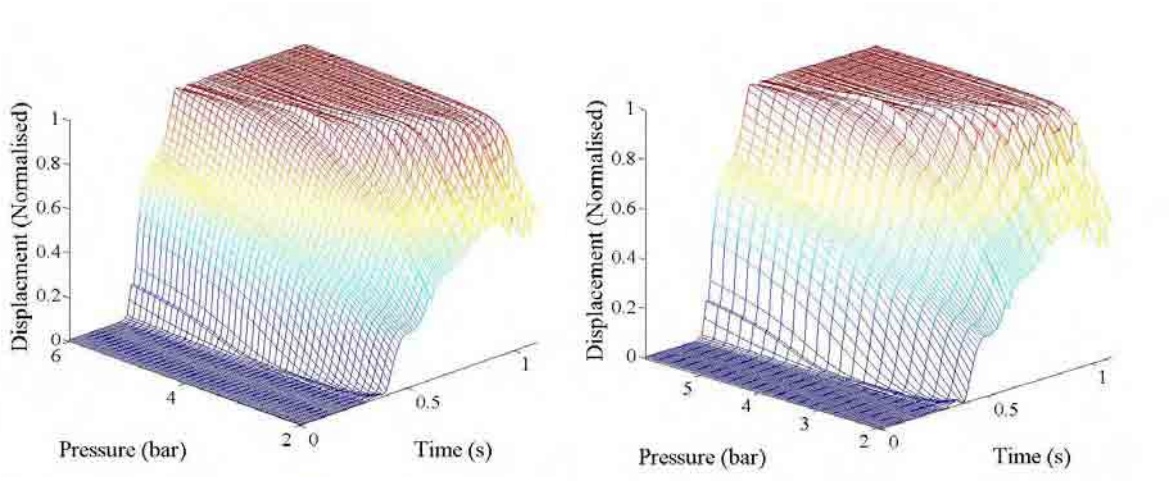


Figure 4.4 Temporal regions for polynomial model.

Temporal Regions	Polynomials
0-280 ms	3rd order polynomial
290-390 ms	Cubic/quadratic rational
400-580 ms	4th order polynomial
590-1100 ms	Cubic/cubic rational

Table 4.2 Temporal regions for polynomial fitting.



a) Model validation. **b) Prediction by model.**

Figure 4.5 Polynomial model of train-stop normal throw.

4.2.3 State space model

Two methods of modelling STME features have been discussed in the last two sections, the exponential model for throw time and activation delay and the polynomial model for displacement dynamics. However, another type of parameter, which is very important to identify the health status of the asset, is also monitored. This is the airflow (for pneumatic assets) or the current and oil pressure (for electric or/and hydraulic assets). This type of parameter is normally too complex to model, however, it directly reflects the operation dynamic within the equipment. In other words, these parameters can indicate the working conditions of the mechanisms.

The aim of the state space modelling approach discussed in this section is to build up a simple and low computation cost model for the airflow, current and oil pressure. The difficulties of modelling these parameters are the large variations and the disturbances of measured data. As an example, a pneumatic train door airflow profile at 3.5 bar is shown in Figure 4.6.

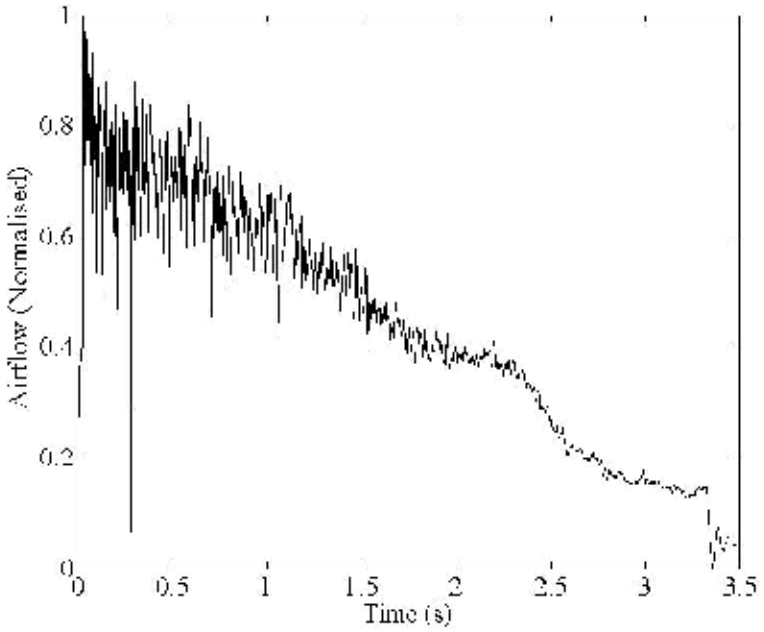
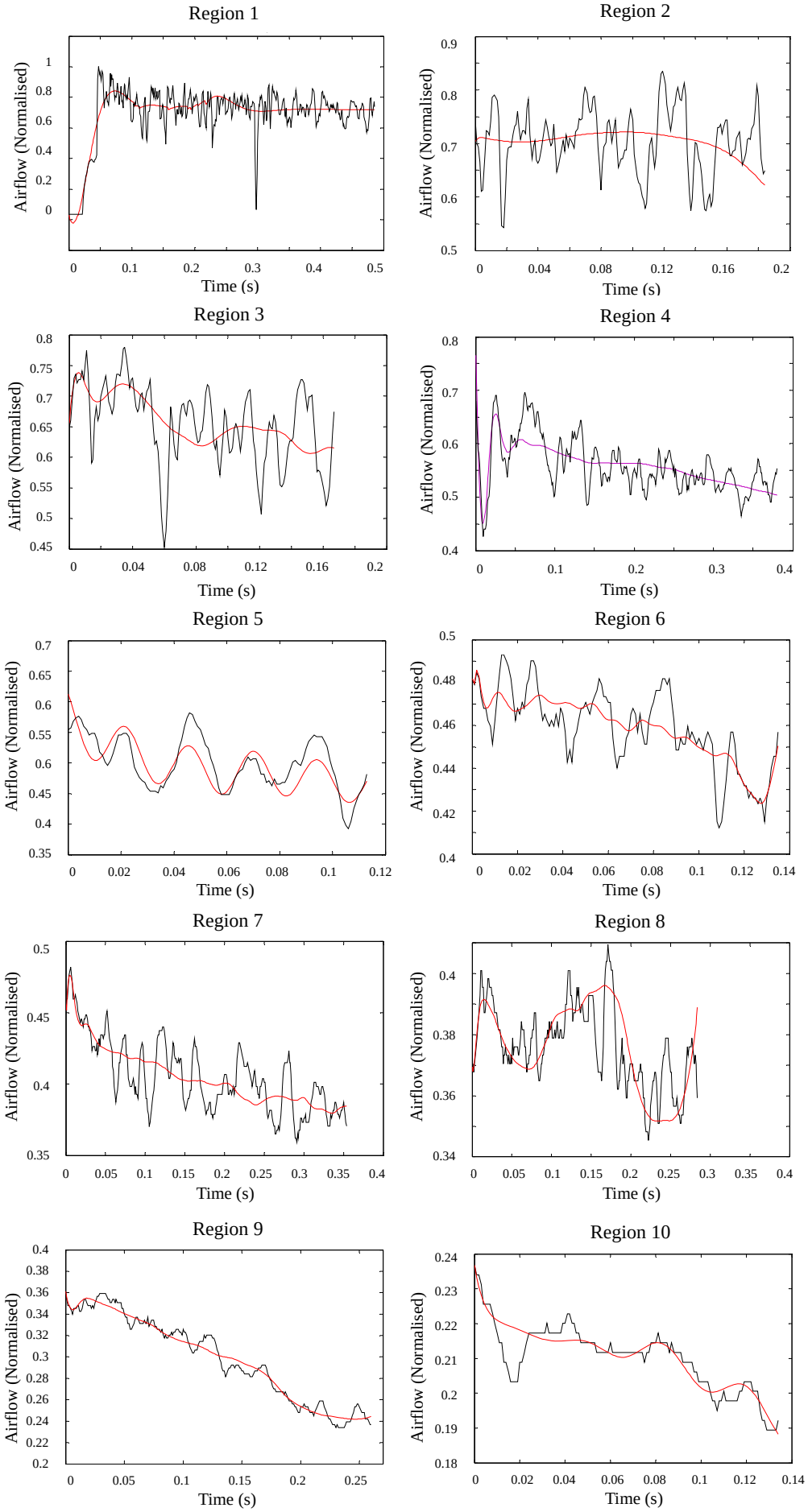


Figure 4.6 Airflow profile of train door normal throw at 3.5 bar.

A state space model of the airflow of the train door normal throw is presented in Figure 4.7. To reduce the complexity of modelling, the airflow data was divided into 17 temporal regions based on the acceleration feature of train door movement. The data in each region was modelled with the state space model and the predicted outputs are compared with the original data. The discrete time-variant state space model used in this study can be described as the following:

$$\begin{aligned} x(t+Ts) &= Ax(t) + Bu(t) + Ke(t) \\ y(t) &= Cx(t) + Du(t) + e(t) \end{aligned} \tag{4.7}$$

where $x(t) \in \mathbb{R}^n$, is the state vector; $y(t) \in \mathbb{R}^q$, is the output vector, and $u(t) \in \mathbb{R}^p$, is the input vector. *A* is the state matrix in $n \times n$ dimension; *B* is the input matrix in $n \times p$ dimension; *C* is output matrix in $q \times n$ dimension and *D* is feedforward matrix in $q \times p$ dimension. *Ts* is the sampling period and *e(t)* is the estimation error.


Region 1
Airflow (Normalised)
Time (s)
Region 2
Airflow (Normalised)
Time (s)
Region 3
Airflow (Normalised)
Time (s)
Region 4
Airflow (Normalised)
Time (s)
Region 5
Airflow (Normalised)
Time (s)
Region 6
Airflow (Normalised)
Time (s)
Region 7
Airflow (Normalised)
Time (s)
Region 8
Airflow (Normalised)
Time (s)
Region 9
Airflow (Normalised)
Time (s)
Region 10
Airflow (Normalised)
Time (s)

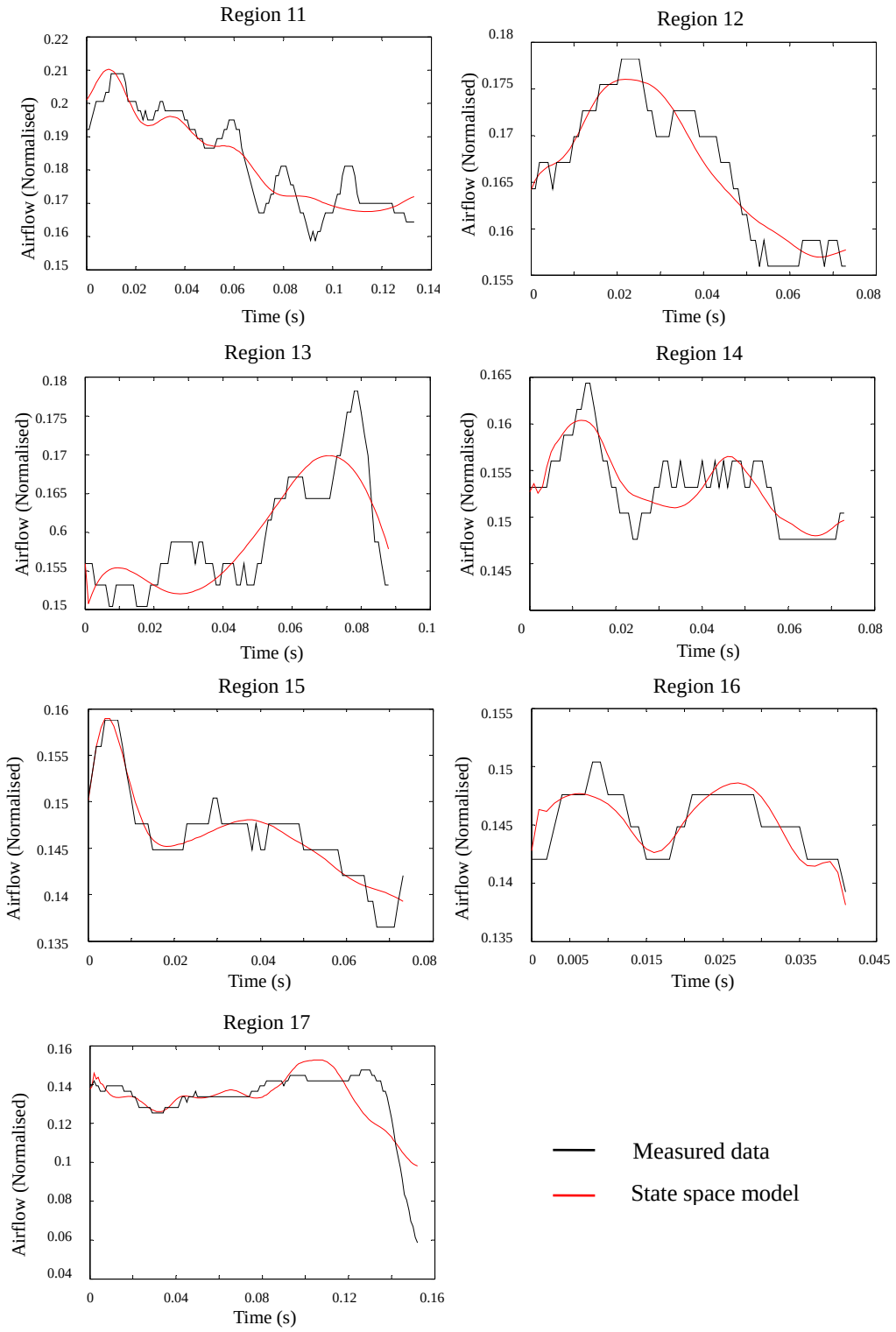


Figure 4.7 State space model of the airflow of the train door normal throw.

In the figures, the state space model provides good simulation results. Although the

division of temporal regions increases the number of models, the accuracy of the model is improved and the model size for each region is small. The low request of computation power for the state space model makes it an economic solution for practical applications.

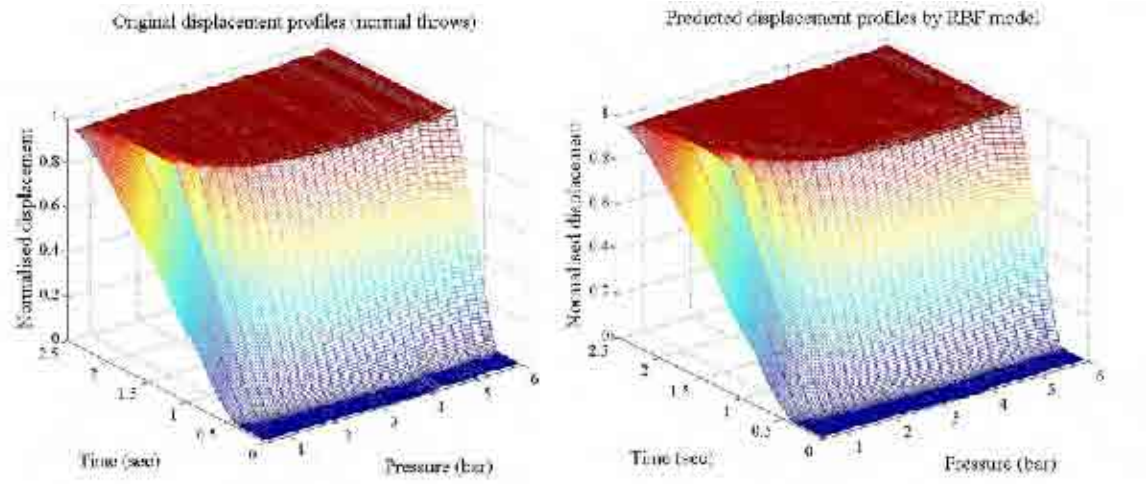
4.2.4 Neural network model

Due to the limits of using mathematical models in complex modelling and to make fault detection and diagnosis algorithms practical for real systems, an approach to the simulation of the dynamics of STMEs was applied using ANN modelling techniques, such as radial basis function neural networks (RBF). However, sufficient system processing power is required for the necessary mathematics.

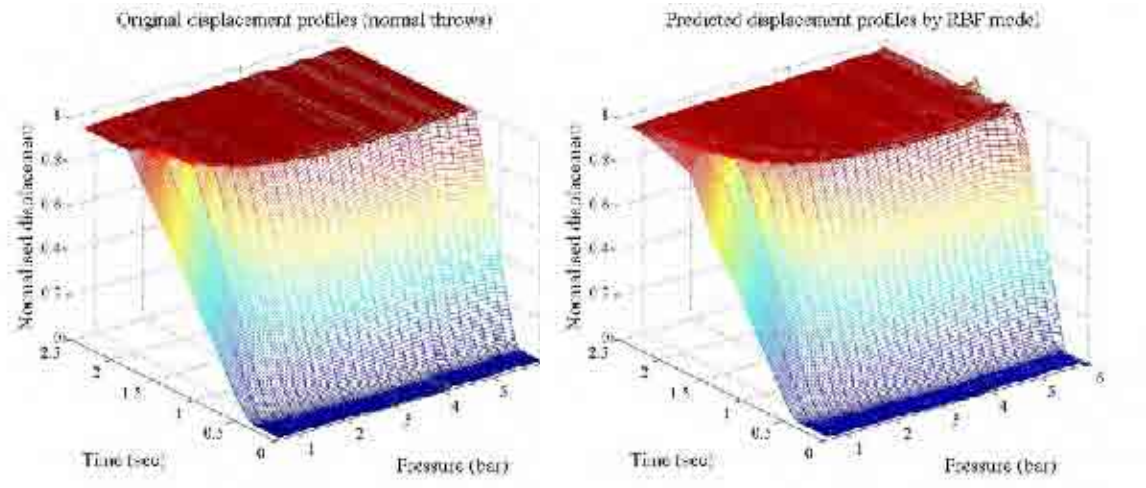
RBF neural networks are single-hidden-layer feed-forward networks, which can be presented as:

$$f_i(x) = \sum_{j=1}^{n_h} \eta_{ji} \varphi(\|x - c_j\|, \rho_j) \quad 1 \le i \le m \tag{4.8}$$

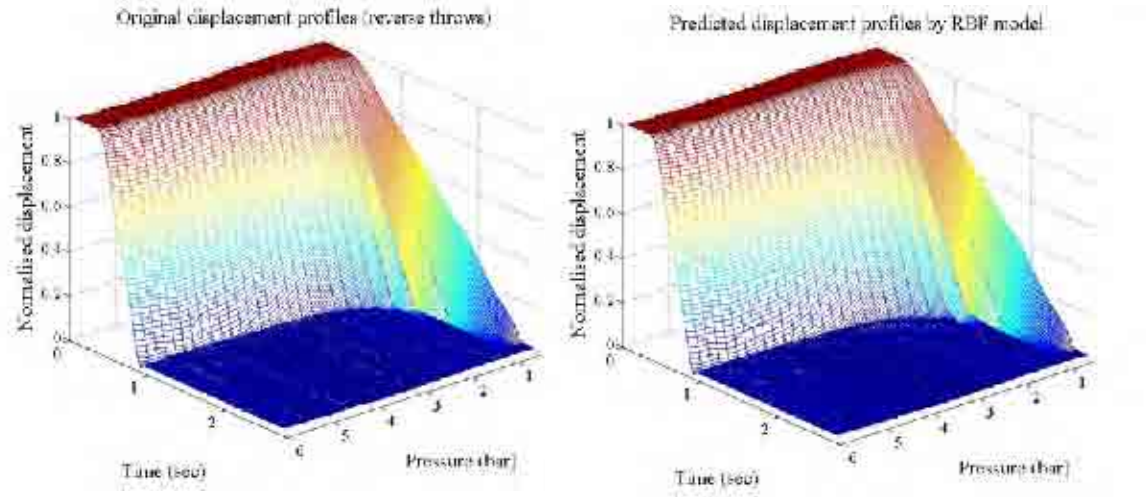
where η_{ji} are the weights of the linear combiner; $\|.\|$ denotes the Euclidean norm; c_j is the RBF centre; ρ_j is the positive scalars width; n_h is the number of nodes in the hidden layer; m is the number of inputs and $\varphi(z, \rho) = e^{-(z,\rho)^2}$. RBF neural networks are capable of modelling any non-linear behaviour with arbitrary accuracy. With higher accuracy requirements, the number of neurons increases as do the model complexity and size. The RBF model trained by the data of the unloaded point machine is illustrated in Figure 4.8.



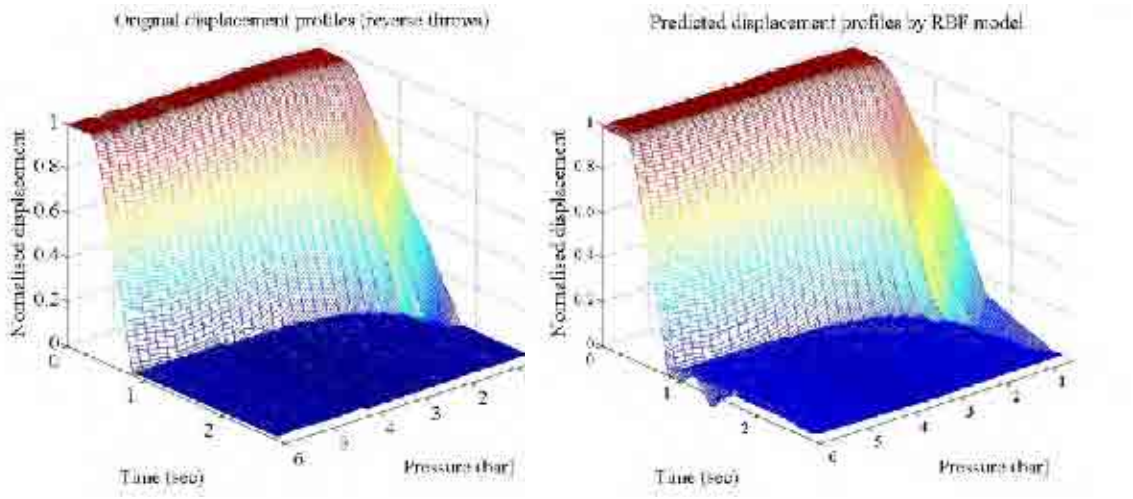
a) Original displacement profiles (training data) vs. outputs of RBF model for normal throws.



b) Original displacement profiles (test data) vs. predictions by RBF model for normal throws.



c) Original displacement profiles (training data) vs. outputs of RBF model for reverse throws.



d) Original displacement profiles (test data) vs. predictions by RBF model for reverse throws.

Figure 4.8 Prediction results of RBF model of the pneumatic point machine.

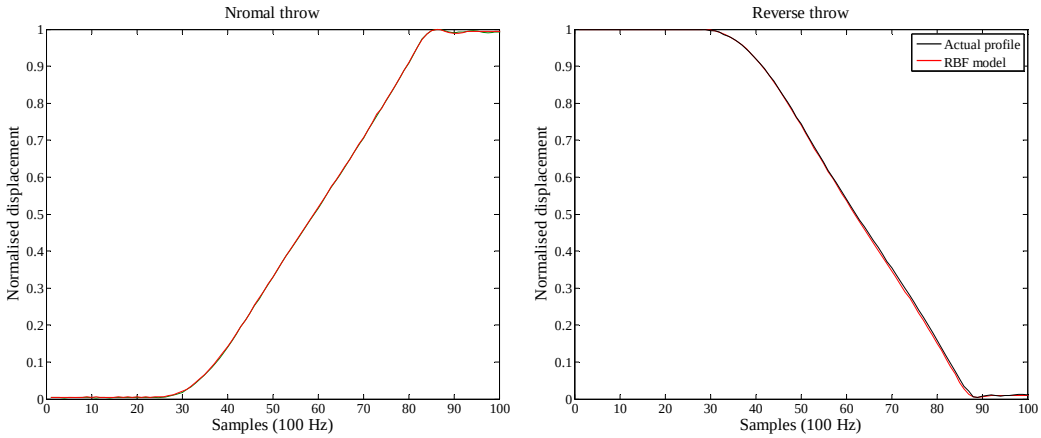
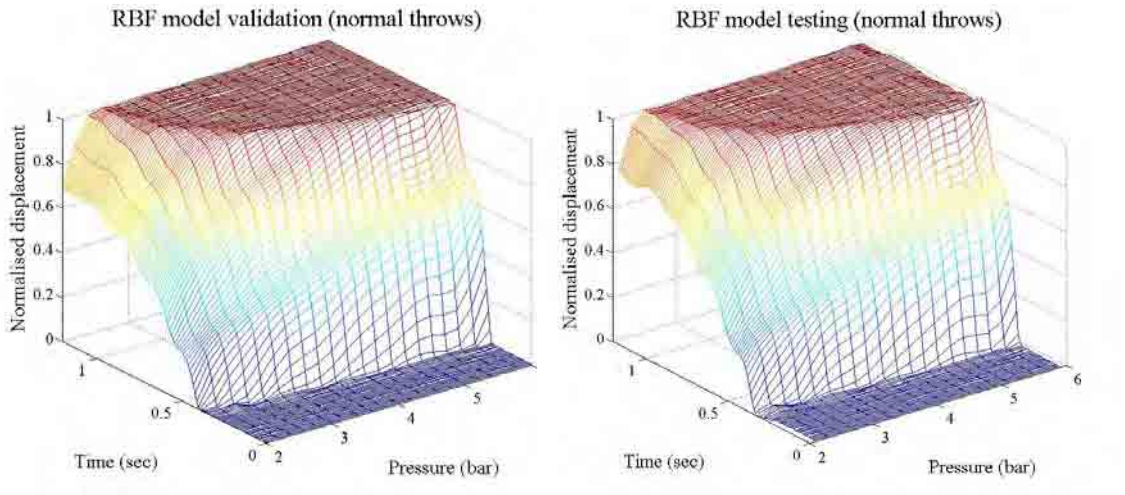


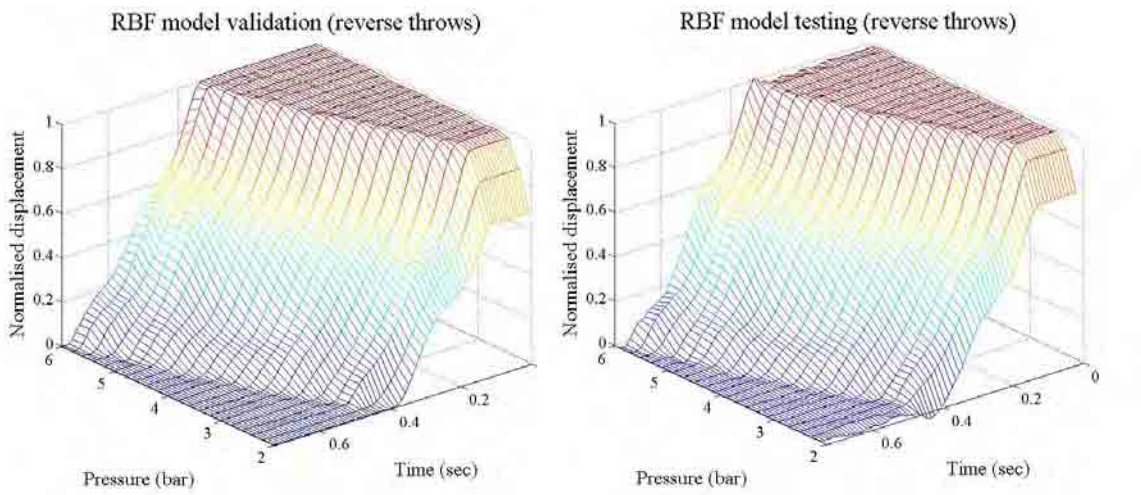
Figure 4.9 Measured data *vs.* RBF model output of point machine at 3 bar.

The pneumatic air pressures are used as the input of the RBF model and the output is the displacement of the rod of the point machine. To train this network, 56 sets of data at an increment of 0.1 bar from 0.5 to 6 bar were collected from the asset as training data. In order to test the accuracy of the prediction of the model, the data were divided into two groups. The first group included the data at pressures of 0.5, 0.7, …, 5.7, 5.9 bar, and the data at pressure of 0.6, 0.8, …, 5.8, 6.0 bar were used for model testing. The RBF model was created and trained by the first data group and tested by the second

group. The comparison between the actual data and the RBF model output at the nominal pressure of 3 bar is presented in Figure 4.9.



a) Validation by training data and test of RBF model for normal throws.



b) Validation by training data and test of RBF model for reverse throws.

Figure 4.10 Prediction results of RBF models of the train-stop.

The same method was also applied to the train-stop case study. Another RBF model was trained using half of the train-stop displacement data at every 0.2 bar from 2 to 6 bar, and tested by the data from 2.1 to 5.9 bar in 0.2 bar increments. Results of model are shown in Figure 4.10. The comparisons between the measurement and the

RBF model output at the pressure of 3.1 bar are displayed in Figure 4.11.

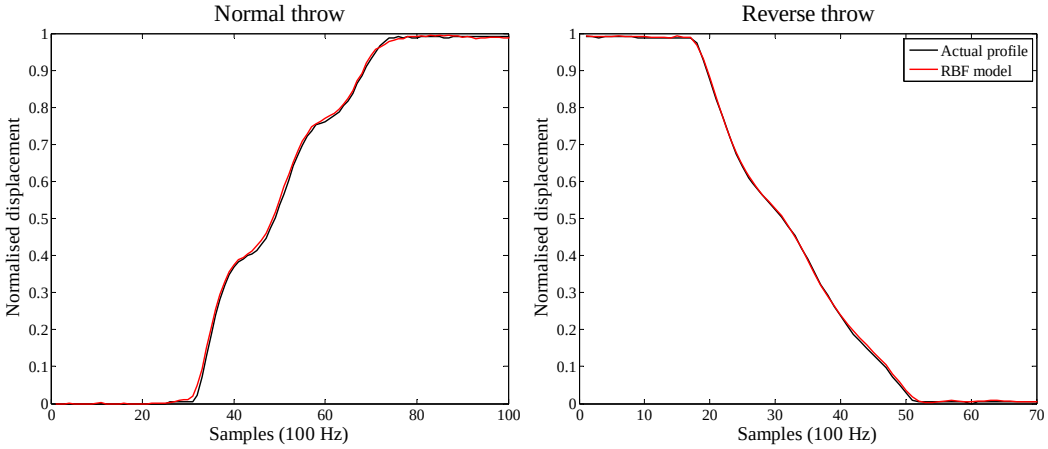


Figure 4.11 Measured data *vs.* RBF model output of train-stop at 3.1 bar.

The results presented above indicate that the RBF neural network model provides good simulation and prediction performance. In practice, the computation cost for the neural network models is relatively high and special integrated circuits could be required to handle the complex algorithms.

4.3 Fault detection for pneumatic train door

Based on the modelling work discussed above, a fault detection approach for the pneumatic train door is presented in this section. According to the diagram of generic fault detection and diagnosis in Chapter 3, the diagram for the pneumatic train door is revised and shown in Figure 4.12. In this figure, the generic and pneumatic processes remain and are displayed in black. The processes for electric and electro-hydraulic assets are displayed in grey, which means that they are not involved in the FDD procedures.

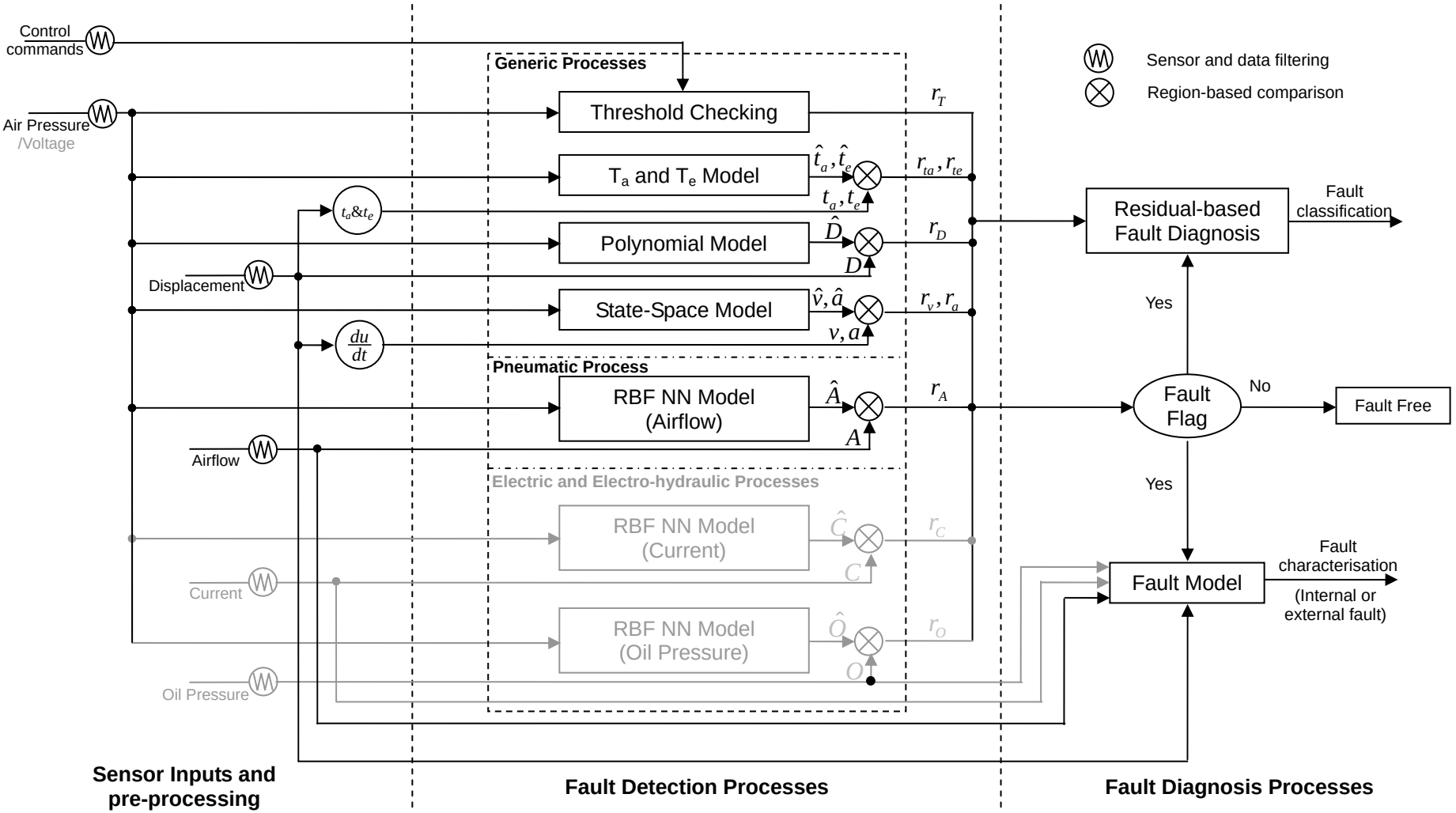


Figure 4.12 Diagram of generic fault detection and diagnosis for pneumatic STMEs.

During the fault detection process, the actual measured and model estimated variables are compared within a series of pre-designed spatial regions. The dividing of spatial regions for a pneumatic train door normal throw was introduced in Chapter 3 (shown in Figure 3.18 and Table 3.2). The spatial regions of the reverse throw are displayed in Figure 4.13 and the boundaries are listed in Table 4.3.

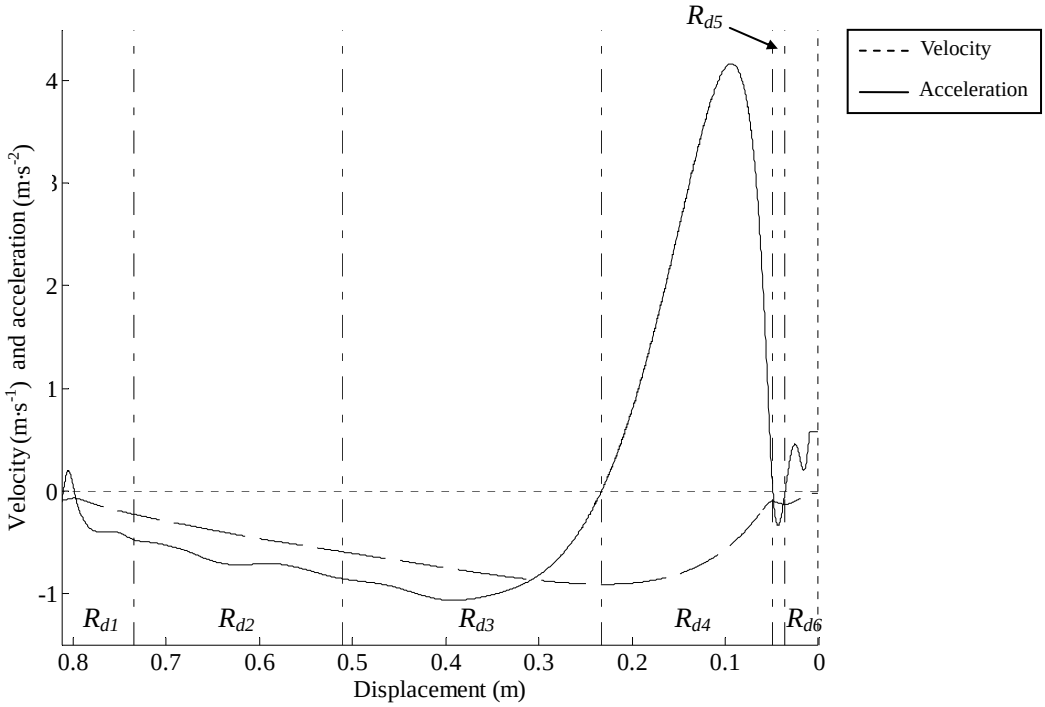


Figure 4.13 Definition of spatial regions for the reverse throw of pneumatic train door (3.5 bar).

Regions	Boundaries (m)	Description
R_{d1}	0.812 – 0.735	Throw start
R_{d2}	0.735 – 0.51	Intermediate 1
R_{d3}	0.51 – 0.233	Intermediate 2
R_{d4}	0.233 – 0.049	Intermediate 3
R_{d5}	0.049 – 0.036	Intermediate 4
R_{d6}	0.036 – 0	Throw end

Table 4.3 Boundaries of spatial regions for the reverse throw of pneumatic train door (3.5 bar).

The fault detection method with an adaptive threshold was evaluated using the

collected data and the results are displayed in Table 4.4. Five single models, the exponential model, polynomial model, state-space model (velocity), state-space model

Model	Fault detection (%)	False alarm (%)	Confidence level (1- α)%
Exponential model (activation & throw-time)	65.57	5.5	99
	71.07	11.13	97
	73.28	15.63	95
Polynomial model (displacement)	84.37	2.88	99
	87.83	7.63	97
	88.95	11.93	95
State-space model (velocity)	94.46	1.5	99
	95.46	7	97
	96	10.25	95
State-space model (acceleration)	92.46	8	99
	93.91	12.5	97
	95.09	16.63	95
Neural network model (airflow)	78.5	4	99
	82.41	10.25	97
	85.14	17.25	95
Model combination	96.92	11.82 (0.78)	99
	98.59	17.74 (4.1)	97
	99.25	24.76 (5.79)	95

Table 4.4 Fault detection results for pneumatic train door.

(acceleration) and neural network model, were tested independently with three confidence levels, 99 %, 97 % and 95 %, for threshold design. Flowcharts of generic modelling and evaluation programmes are presented in Appendix D. The fault detection results and false alarm rates are listed in the form of percentages which can be described as:

$$\text{fault detection rate} = \frac{\text{number of faults detected in faulty profiles}}{\text{number of operations with failure mode}} \times 100\% \quad (4.9)$$

$$\text{false alarm rate} = \frac{\text{number of faults detected in healthy profiles}}{\text{number of healthy operations}} \times 100\% \quad (4.10)$$

Under ideal conditions, a rate of 100 % is required for fault detection with a 0 % false alarm rate, which means that all the faults are successfully detected and no fault alarm is triggered during healthy operations. However, due to the presence of noise and model uncertainty, the ideal results can never be achieved by a single model.

In the table, it can be observed that the accuracy of fault detection is different for every single model. As discussed in section 3.2.2.1, the models should be combined in order to achieve a better fault detection result. As shown at the bottom of the table, a combined model approach significantly improves the performance of fault detection. However, the false alarm triggered by any one of the models was counted, the false alarm rate is high. To reduce the false alarm rate, an alarm management system could be introduced to logically analyse the generated alarms. A false alarm supported by multi models could be proved to be a true alarm; otherwise, the alarm could be ignored. For example, the values in brackets at the bottom of the table are the percentages of false alarm rate that were supported by all the models. It can be seen that the false alarm rate is highly reduced.

4.4 Preliminary fault diagnosis approach

In this section, a preliminary fault diagnosis approach for the pneumatic train door using a fault model is discussed. Another fault diagnosis method, residual analysis,

was also introduced in Chapter 3 and the residuals were defined for different failure modes.

In Figure 4.14, a displacement profile of a train door normal throw in healthy mode is shown by a dotted line with a 3.5 bar air pressure. The solid line is a faulty displacement profile, where the fault is created by a restricting force applied by hand in the opposite direction to which the train door is moving, to simulate random friction. The fault occurring during the throw is circled. Due to the resistive force, the throw time was longer.

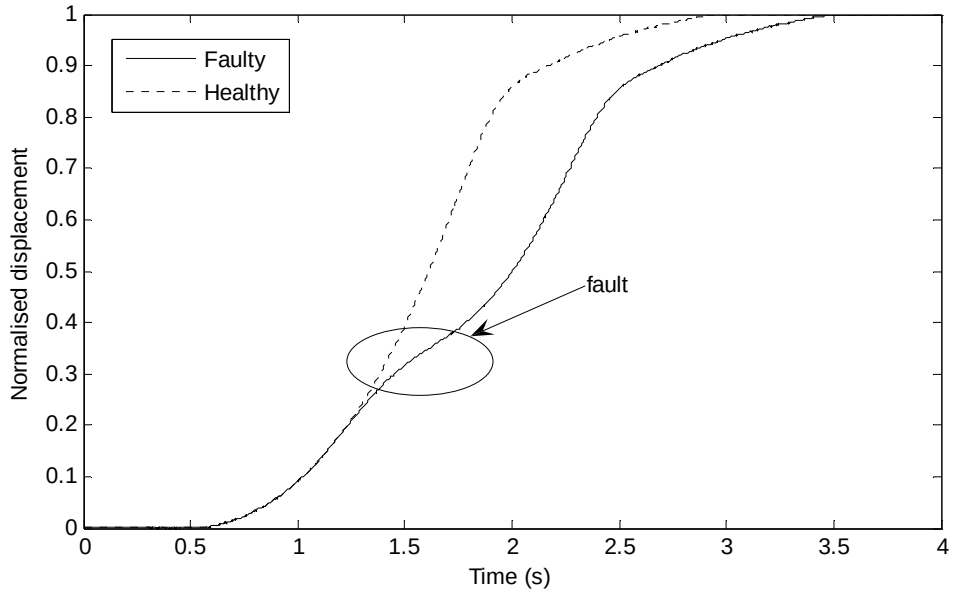


Figure 4.14 Healthy and faulty pneumatic train door displacement profiles.

The corresponding healthy and faulty airflow profiles with this disturbance are displayed in Figure 4.15. It can be observed that the airflow sharply decreased when the disturbance was applied and then increased after the disturbance was removed.

In order to identify whether the asset is in a healthy or a faulty condition, the fault

model was applied to predict an airflow profile from the faulty displacement. The predicted and actual airflow profiles are compared in Figure 4.16.

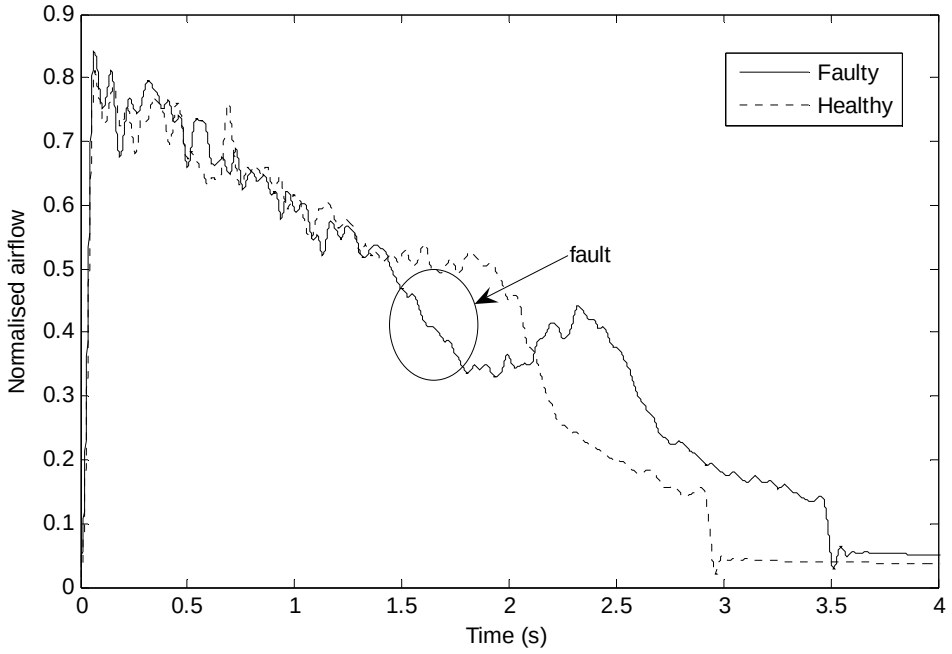


Figure 4.15 Healthy and faulty pneumatic train door airflow profiles.

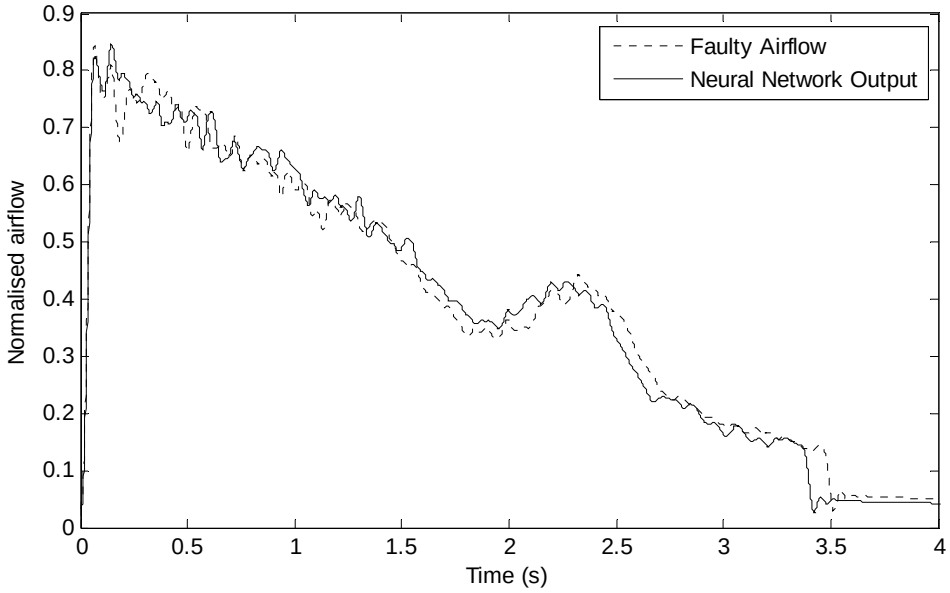


Figure 4.16 Airflow prediction using fault model.

In the figure, the solid and dotted lines respectively represent the airflow predicted by RBF network and the actual airflow. The predicted airflow is considered to be consistent with the actual airflow and a conclusion can be drawn that the asset is in a healthy condition and the fault was caused by an external factor (a friction simulated by a hand-push in this case). Assuming that they do not match, it could be considered that the machine is in a faulty condition.

From the results shown above, it is proved that this fault model is capable of roughly identifying external faults. The accuracy of the classification depends on how many fault mode data could be achieved from the assets to train the network. Another requirement of applying this method is that the displacement and airflow sensors are available and working normally. The sensor faults are therefore not tolerable for this fault model. This method did not apply to the train-stop since no airflow sensor is available for this asset. Due to the lack of data of mechanical faults, the diagnosis of machine faults is left for future work.

4.5 Conclusions

A generic fault detection and diagnosis approach for the pneumatic train door was presented in this chapter. As the essential part of the fault detection and diagnosis method, the modelling work was also illustrated. Based on the constructed models, the proposed generic FDD method was applied on the pneumatic train door by following the procedures designed in the generic FDD diagram. The results of the fault detection were provided. A preliminary fault diagnosis approach using a fault model was also discussed.

Data analysis for modelling was discussed and four modelling approaches and results were illustrated. For each of these modelling methods, the economic application and practical accuracy were considered. These two factors have to be achieved in compromise. The exponential model for throw time and activation delay is an efficient way to provide a healthy value for fault detection. In comparison with the neural network method, the polynomial model has the advantages of straightforward model structure and low computation requirements. However, when accuracy and prediction capability are emphasised, the neural network model becomes more suitable. Therefore, the polynomial model method is better for low cost applications and the neural network model is more often used in experimental cases. The state space model was discussed for the modelling of some parameters, such as airflow, current and oil pressure. The spatial region division provides the possibility of application of this simple model. This model can also be used for dynamic simulation in either temporal or special regions.

The results of fault detection were provided for the pneumatic train door. The results suggested that the performance of fault detection using a single sub-model varies; however, a combination of sub-models has a good fault detection capability. As well as improving the sensitivity of fault detection, the false alarm rate was also enlarged by model combination, since the false alarms generated by each sub-model were taken into account. A mechanism of managing the false alarms is introduced to only count the false alarms supported by all the sub-models, by which the false alarm rate is highly reduced.

Overall, the generic FDD method has a good response on laboratory-based test rigs. A limitation of this method, however, still exists which may reduce the applicability of

employing the method in practice. This limitation is caused by the individuality of multiple instances of the same system. The models and FDD methods in this study are based on assets in the laboratory. Other assets in the practical environment may exhibit different characteristics, although they are of the same type and performing the same operations. This limitation could be overcome by the construction of a self learning mechanism, which would collect the data for model construction from the first several operations, then the features of each asset would be automatically modelled and adapted by the FDD method.

Chapter 5

Generic fault detection approaches for other STMEs

5.1 Introduction and motivation

As introduced in Chapter 3, five railway assets, known as STMEs, were studied in this project. These assets can be further grouped into three categories by the type of power supply: pneumatic, electric and electro-hydraulic. The pneumatic train door has been studied previously, for which a generic fault detection and diagnosis method has been proposed and applied. This chapter considers the case studies for the other four assets including the pneumatic train-stop, point machine, the electric point machine and the electro-hydraulic level crossing barrier.

The case studies for the four assets in this chapter are to test and analyse whether the proposed FDD method is applicable for all these assets. For the two pneumatic assets, the train-stop and the point machine, the same FDD processes as for the pneumatic train door are applied. Although the electric and electro-hydraulic assets are grouped as a class of Single Throw Mechanical Equipments (STMEs), based on the manner of reciprocating operations, they have significant differences in mechanical designs due to the different power sources. Preliminary fault detection approaches for the electric and electro-hydraulic assets are, therefore, applied with the emphasis on analysing the feasibility of this method. The results are also presented and concluded.

5.2 Generic fault detection for pneumatic assets

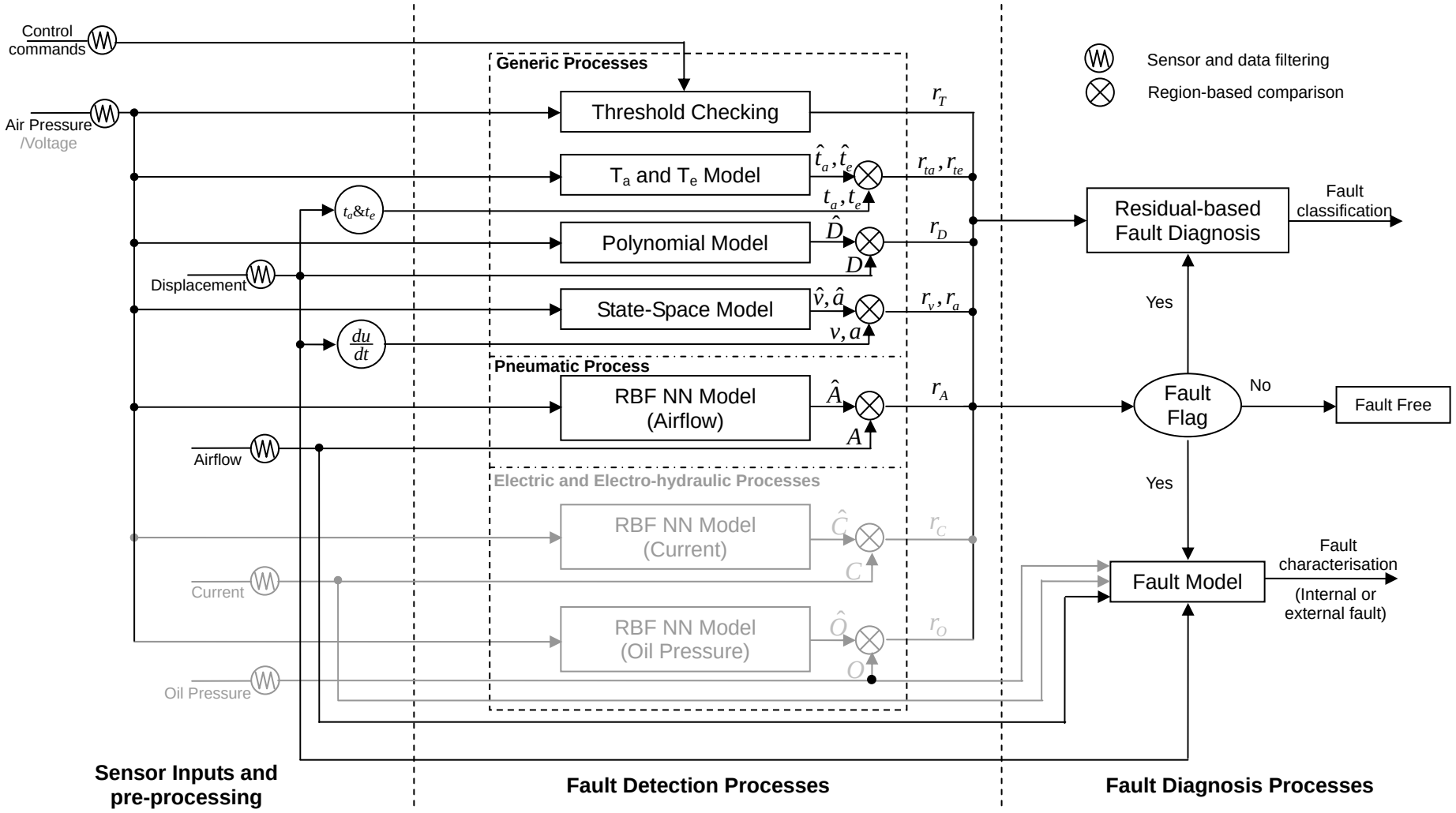


Figure 5.1 Diagram of generic fault detection and diagnosis for pneumatic train-stop and point machine.

This section presents the generic fault detection approaches for the two pneumatic assets: the train-stop and the point machine. In order to illustrate the generic FDD method, the generic fault detection and diagnosis diagram for the pneumatic train door is also used here and shown in Figure 5.1. Since the three machines are all pneumatic and have similar mechanisms, the diagram is the same as the one for the pneumatic train door. In the diagram, the electric and electro-hydraulic processes are displayed in grey to indicate that these processes are not functional.

5.2.1 Fault detection for pneumatic train-stop

Before applying the FDD method on the pneumatic train-stop, the spatial regions dividing method is exhibited by the spatial regions figures and the table of boundaries. The spatial regions, divided using velocity and acceleration features, are displayed in Figures 5.2 and 5.3 for normal and reverse throws respectively. The region boundaries

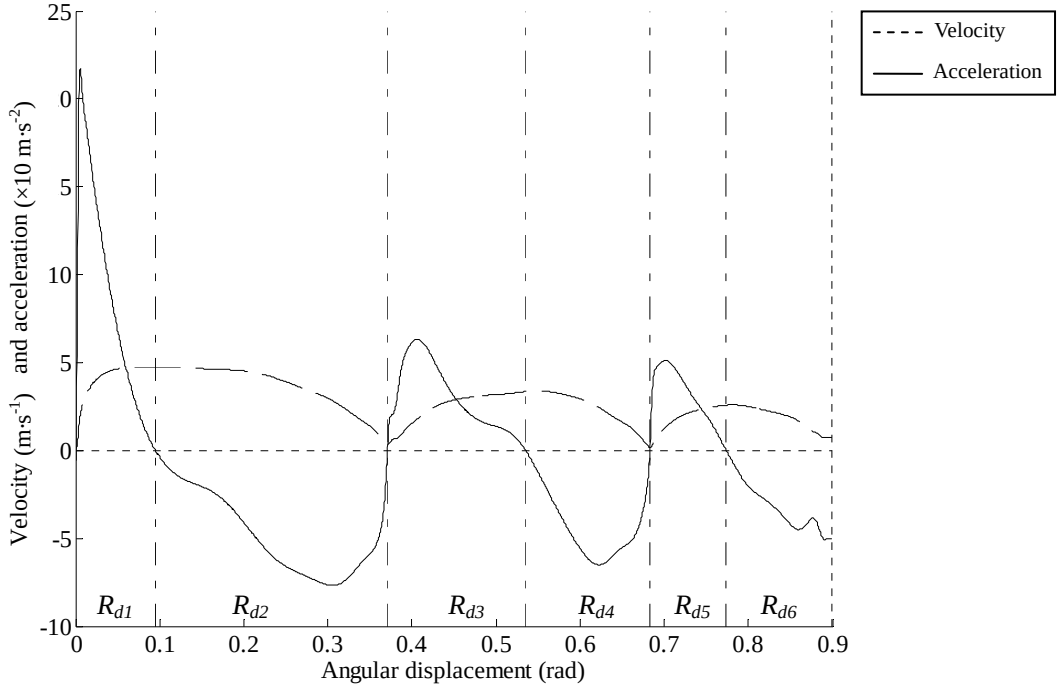


Figure 5.2 Definition of spatial regions for the normal throw of pneumatic train-stop (3.1 bar).

Regions	Boundaries (rad)	Description
R_{d1}	0 – 0.095	Throw start
R_{d2}	0.095 – 0.371	Intermediate 1
R_{d3}	0.371 – 0.535	Intermediate 2
R_{d4}	0.535 – 0.683	Intermediate 3
R_{d5}	0.683 – 0.774	Intermediate 4
R_{d6}	0.774 – 0.9	Throw end

Table 5.1 Boundaries of spatial regions for the normal throw of pneumatic train-stop (3.1 bar).

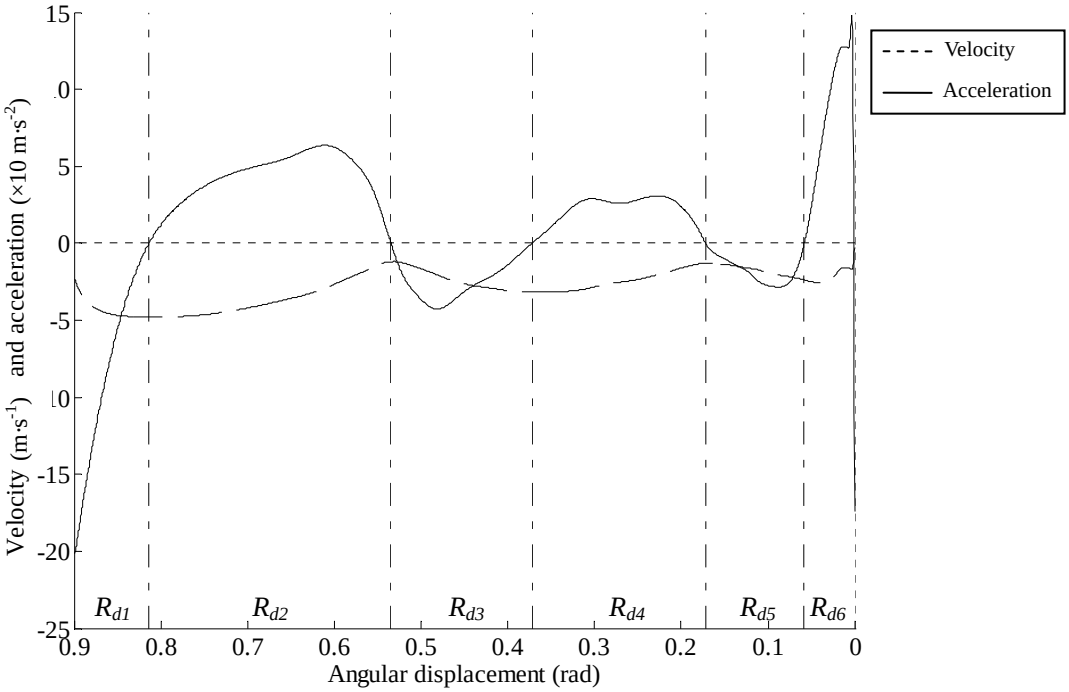


Figure 5.3 Definition of spatial regions for the reverse throw of pneumatic train-stop (3.1 bar).

Regions	Boundaries (rad)	Description
R_{d1}	0.9 – 0.814	Throw start
R_{d2}	0.814 – 0.535	Intermediate 1
R_{d3}	0.535 – 0.372	Intermediate 2
R_{d4}	0.372 – 0.171	Intermediate 3
R_{d5}	0.171 – 0.058	Intermediate 4
R_{d6}	0.058 – 0	Throw end

Table 5.2 Boundaries of spatial regions for the reverse throw of pneumatic train-stop (3.1 bar).

are listed in Tables 5.1 and 5.2. As mentioned previously, the regions would increase in accuracy, when the comparison between actual measured and model estimated variables is carried out within proper defined regions.

Model	Fault detection (%)	False alarm (%)	Confidence level (1- α)%
Exponential model (activation & throw-time)	65.51	1.38	99
	68.84	2.89	97
	69.35	3.84	95
Polynomial model (displacement)	90.74	10.22	99
	92.41	13.73	97
	96.11	16.29	95
State-space model (velocity)	87.78	8.73	99
	94.26	11.71	97
	96.94	14.93	95
State-space model (acceleration)	87.78	9.76	99
	91.85	13.32	97
	93.43	16.41	95
Model combination	98	16.8 (0.79)	99
	98.33	24.6 (2.73)	97
	98.89	29 (3.75)	95

Table 5.3 Fault detection results for pneumatic train-stop.

Based on the deliberately designed models and the adaptive thresholds, the fault detection processes were applied and the results are listed in Table 5.3. Due to the absence of an airflow sensor, the neural network model for airflow was not included. From the results, it can be observed that the polynomial model gives the best detection rate; however, the corresponding false alarm rate is also higher than other models. This phenomenon is considered to be caused by the large dynamic changes in the

displacement profiles. The large dynamic changes can help the detection of faults; meanwhile, they can also trigger more false alarms due to the existing model uncertainties. The solution to guarantee good capability of fault detection could be to use a combination of models. The results of the combination of models can be observed to have much better sensitivity to the simulated faults in this case study. Under the condition that the false alarms are only counted when they are triggered by all the sub-models, the false alarm rates (shown in the brackets) are also lower than individual models.

5.2.2 Fault detection for pneumatic point machine

The spatial regions, based on the velocity and acceleration features, are displayed in Figures 5.4 and 5.5 for pneumatic point machine normal and reverse throws respectively. The region boundaries are listed in Tables 5.4 and 5.5.

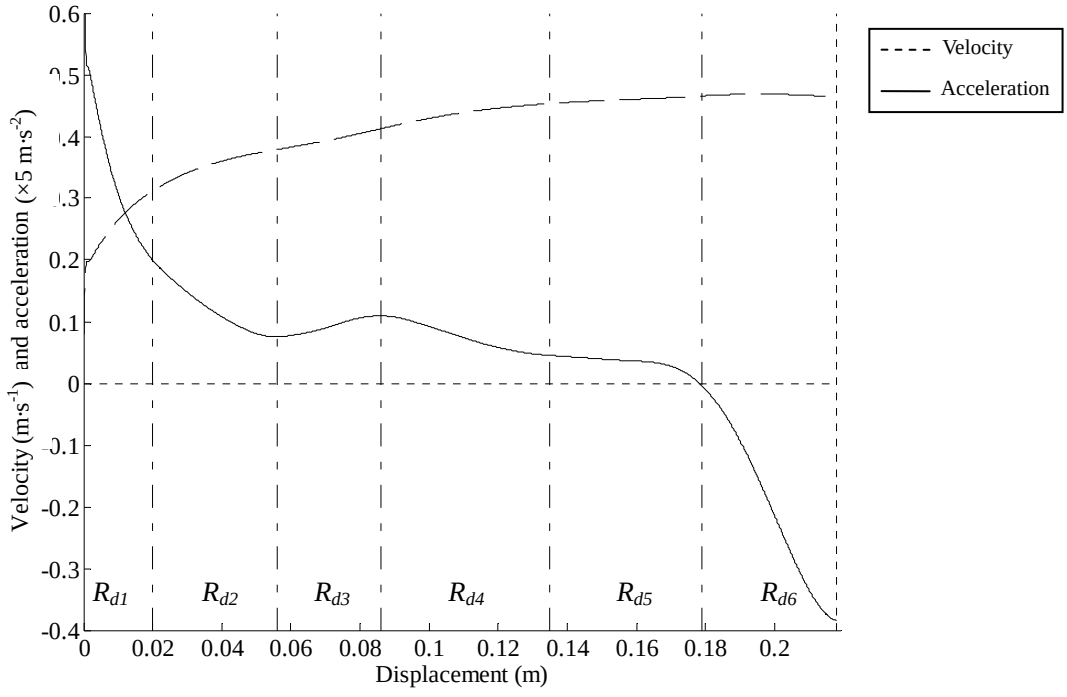


Figure 5.4 Definition of spatial regions for the normal throw of pneumatic point machine (3 bar).

Regions	Boundaries (m)	Description
R_{d1}	0 – 0.02	Throw start
R_{d2}	0.02 – 0.056	Intermediate 1
R_{d3}	0.056 – 0.086	Intermediate 2
R_{d4}	0.086 – 0.135	Intermediate 3
R_{d5}	0.135 – 0.179	Intermediate 4
R_{d6}	0.179 – 0.218	Throw end

Table 5.4 Boundaries of spatial regions for the normal throw of pneumatic point machine (3 bar).

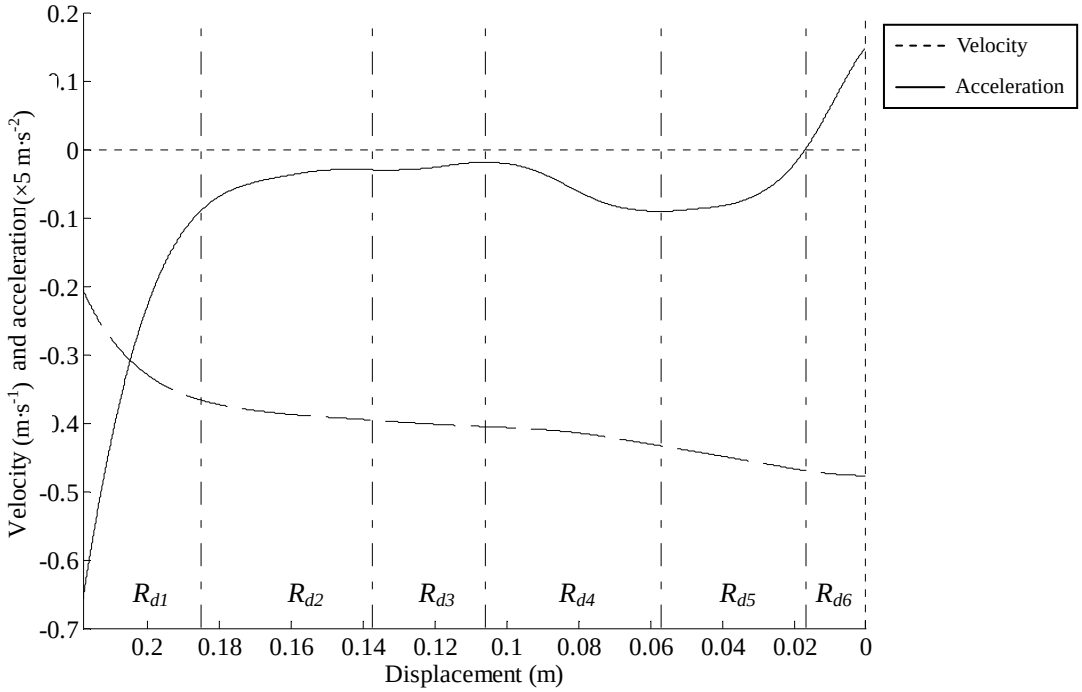


Figure 5.5 Definition of spatial regions for the normal throw of pneumatic point machine (3 bar).

Regions	Boundaries (m)	Description
R_{d1}	0.218 – 0.185	Throw start
R_{d2}	0.185 – 0.138	Intermediate 1
R_{d3}	0.138 – 0.106	Intermediate 2
R_{d4}	0.106 – 0.057	Intermediate 3
R_{d5}	0.057 – 0.017	Intermediate 4
R_{d6}	0.17 – 0	Throw end

Table 5.5 Boundaries of spatial regions for the normal throw of pneumatic point machine (3 bar).

Model	Fault detection (%)	False alarm (%)	Confidence level (1- α)%
Exponential model (activation & throw-time)	69.1	1.07	99
	71.88	2.95	97
	73.73	4.55	95
Polynomial model (displacement)	97.92	2.95	99
	99.77	4.73	97
	99.77	7.05	95
State-space model (velocity)	97.92	2.68	99
	98.84	5.09	97
	99.07	7.95	95
State-space model (acceleration)	97.69	0.54	99
	97.92	5.45	97
	100	10.18	95
Neural network model (airflow)	90.05	0	99
	91.9	0.45	97
	92.13	3.93	95
Model combination	100	4.64 (0)	99
	100	11.07 (0.45)	97
	100	21.16 (3.87)	95

Table 5.6 Fault detection results for pneumatic point machine.

In Table 5.6, the fault detection results for the pneumatic point machine are listed. Except for the exponential model, the models present good detection sensitivity for simulated faults at each confidence level. In comparison with the train door and train-stop, the displacement profiles of the unloaded point machine used for model construction have less dynamic variations. The accuracy of the model is therefore highly enhanced. The model combination provides fault detection rates of 100 % at all confidence levels, which is an ideal result for a fault detection system. However, these

good results would also lead to the difficulty of transplanting this method onto other pneumatic point machines. This issue will be discussed in the conclusion section of this chapter.

5.3 Preliminary generic fault detection for electro-hydraulic and electric assets

The application of the generic FDD solution for the electro-hydraulic and electric assets will be analysed in this section. Based on the work completed so far for these two assets, initial generic fault detection approaches are presented, which focus on the feasibility analysis of the proposed generic fault detection method.

5.3.1 Fault detection for electro-hydraulic level crossing barrier

In order to illustrate the generic FDD process, the generic fault detection and diagnosis diagram for the electro-hydraulic level crossing barrier is displayed in Figure 5.6. In the figure, the process for pneumatic assets is shown in grey to indicate it is not functional. Two sub-models, neural network models for current and oil pressure, are involved in the electro-hydraulic process. The current model in the figure has two representations: the current of the motor of the level crossing machine or the current profiles of the electric point machine motor. In this case study, the current model simulates the current of the level crossing barrier. A lab-based BR843 electro-hydraulic level crossing barrier was used as a test rig, the detail of which was illustrated in section 3.1.2.5.

In order to simulate a friction fault, a weight was fitted at the end of the boom to simply simulate an external resistance during the operation. Since the centre of gravity of the weight is changing when the boom goes up and down, this simulated friction varies

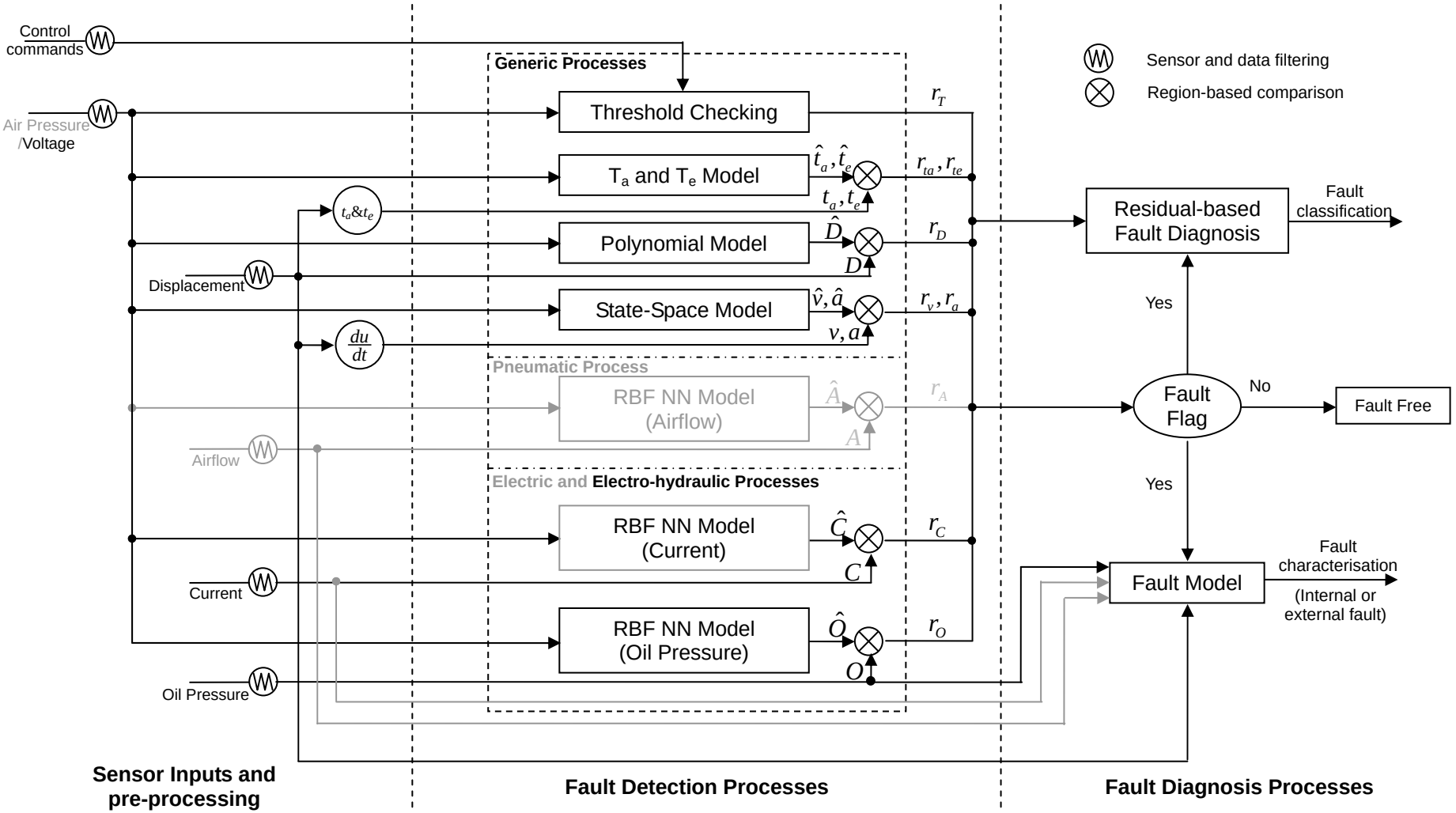


Figure 5.6 Diagram of generic fault detection and diagnosis for electro-hydraulic level crossing barrier.

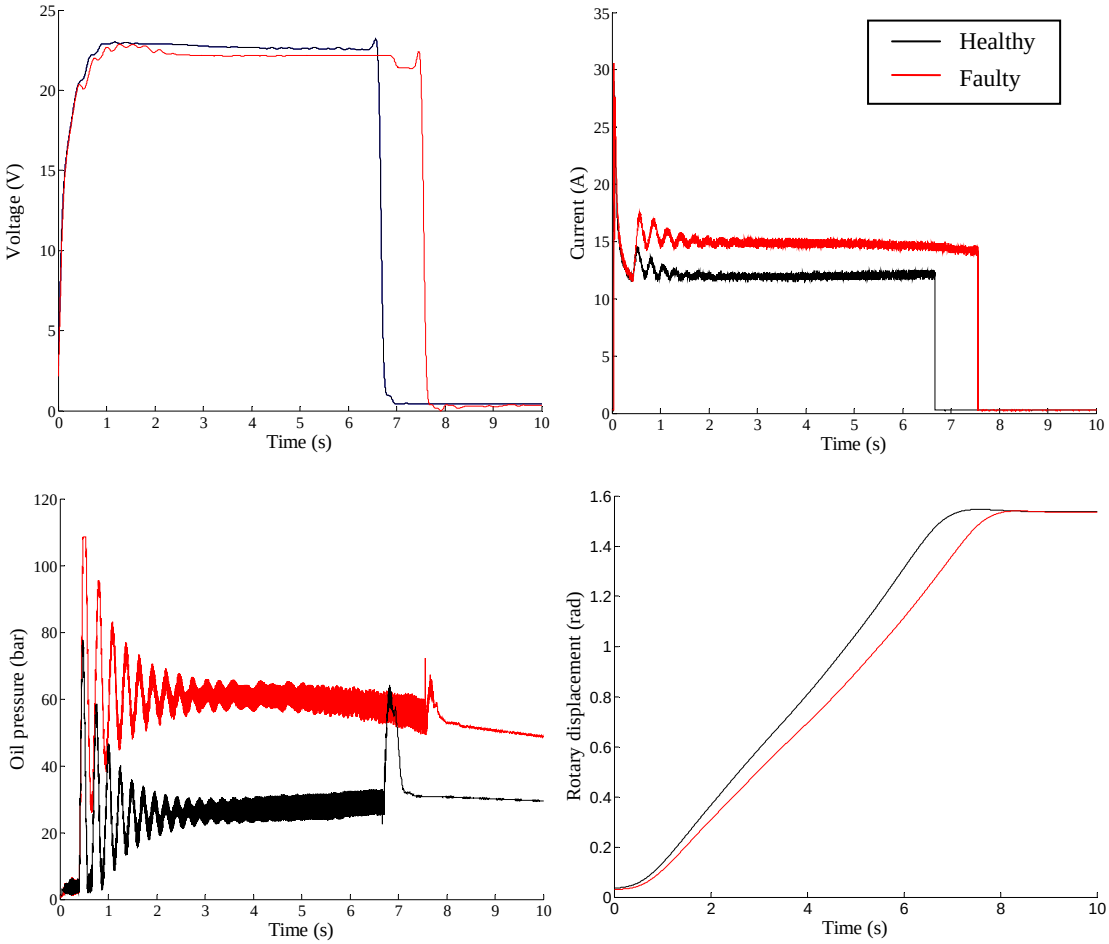


Figure 5.7 Healthy *vs.* faulty profiles of the electro-hydraulic level crossing barrier normal throw.

Operating characteristic	Healthy	With simulated fault
Operating voltage-Average (Volts)	22.8	22.1
Motor current-Max (Amps)	29	30.5
Motor current-Working (Amps)	11.5	14.7
Oil pressure-Max (bar)	77.7	108.6
Oil pressure-Average (bar)	27.7	60.5
Activation delay (sec)	0.7	0.79
Throw time (sec)	6.8	7.7
Throw distance (rad)	1.48	1.48

Table 5.7 Operating characteristics of level crossing barrier normal throw.

with respect to the height it is raised. This simulation is only valid for the normal throw (raising) of the level crossing, since in the reverse throw (lowering) the boom falls by gravity and the weight therefore makes the speed of falling faster instead of slower.

The fault of friction was simulated on the normal throws of the level crossing barrier and the data was collected. In Figure 5.7, the voltage, current, oil pressure and rotary displacement of a normal throw (from lowered to raised position) in both healthy and fault modes are profiled and compared against time. The operating characteristics are listed in Table 5.7.

Due to resistance of the simulated friction, the activation delay and throw time became longer. To drive a larger load, the motor current had an increase of 3.2 Amps; consequently, the oil pressure in the hydraulic pump was 32.8 bar higher than the nominal working condition in average. By the angular displacement of the rotary axis, it can also be observed that the barrier rose at a lower speed and reached the end at approximately 0.9 second later than the healthy throw. The throw distance was measured by degrees of an arc as 1.48 rad (around 85° angle). With the increasing of current output, the power supply had a small drop of 0.7 V.

The changes shown in the figure indicate that the performance of the level crossing can be directly reflected by the measured variables, which, as expected, indicates that the FDD could be realised by condition monitoring. The parameter monitoring based generic FDD solution proposed in Chapter 3 is, therefore, applicable for this asset.

The strategy of the generic fault detection method is to use a single model for each parameter to generate the predicted profile. With the thresholds generated by the adaptive thresholding algorithm, the actual measured data is then compared with the

prediction to give fault information and to generate residuals for further diagnosis. As the adaptive threshold algorithm was developed by a statistical analysis of training data, this threshold method is, therefore, applicable for the models based on the mean of a number of profiles.

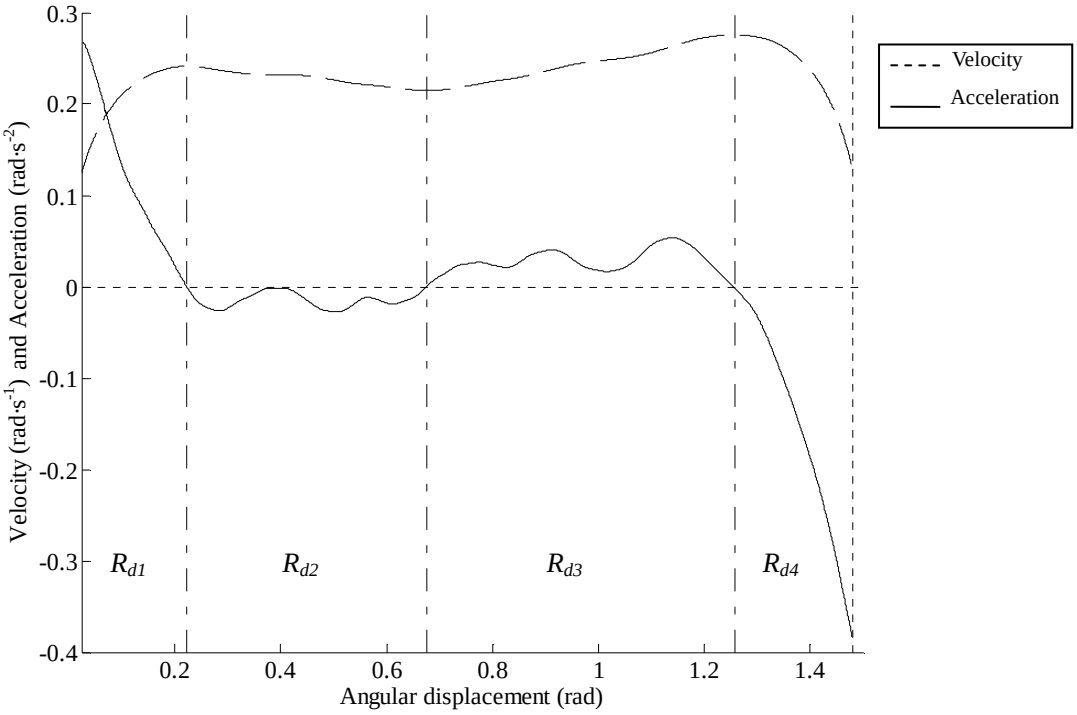


Figure 5.8 Spatial regions division using acceleration and velocity features (level crossing barrier normal throw).

As shown in Figure 5.8, the velocity and acceleration profiles of a normal throw of the level crossing barrier are displayed on a spatial scale. These two profiles were calculated using the mean of 100 displacement profiles to represent a typical normal throw. Low-pass Butterworth filters were used to smooth the profiles and abstract the inherent dynamic features by reducing the noise and the distortion caused by sensor resolution. Four spatial regions are divided according to the zero-crossing points of the acceleration profile. In each of the regions, the throw has a dynamic status marked

as either acceleration or deceleration. The spatial boundaries are listed in Table 5.8. Since the level crossing barrier has a relatively slow operation (6.1 s for a normal throw), the intermediate regions could be further divided in practice. For the purpose of analysis, the 4 regions are individually used for fault detection in this section.

Regions	Boundaries (rad)	Description
R_{d1}	0 – 0.223	Throw start
R_{d2}	0.223 – 0.675	Intermediate 1
R_{d3}	0.675 – 1.26	Intermediate 2
R_{d4}	1.26 – 1.48	Throw end

Table 5.8 Spatial regions of level crossing barrier normal throw.

The modelling methods introduced to pneumatic STMEs in Chapter 4 were applied to parameters of the level crossing barrier. A polynomial model was used to model the displacement profile under fault-free conditions. In Figure 5.9 and 5.10, the velocity and acceleration derived from the polynomial model are displayed with thresholds. The thresholds were designed at a confidence level of 0.97 (α = 0.03). The profiles generated from faulty operations are also displayed red for comparison.

Due to the simulated friction fault, the velocity deviates from the typical profile at the start of the throw and remains low in the following process, which makes the fault very obvious. Since the fault was simulated by adding weight at the end of the boom, the affect to the throw dynamics is therefore constant and results in a faulty velocity profile in parallel with the typical one in the intermediate regions as shown in Figure 5.9. Consequently, the acceleration profile for the faulty throw has clear differences at the start and end of throw but little difference in the intermediate regions as shown in Figure 5.10. A comparison between the two figures suggests that the velocity is more

sensitive to this type of fault.

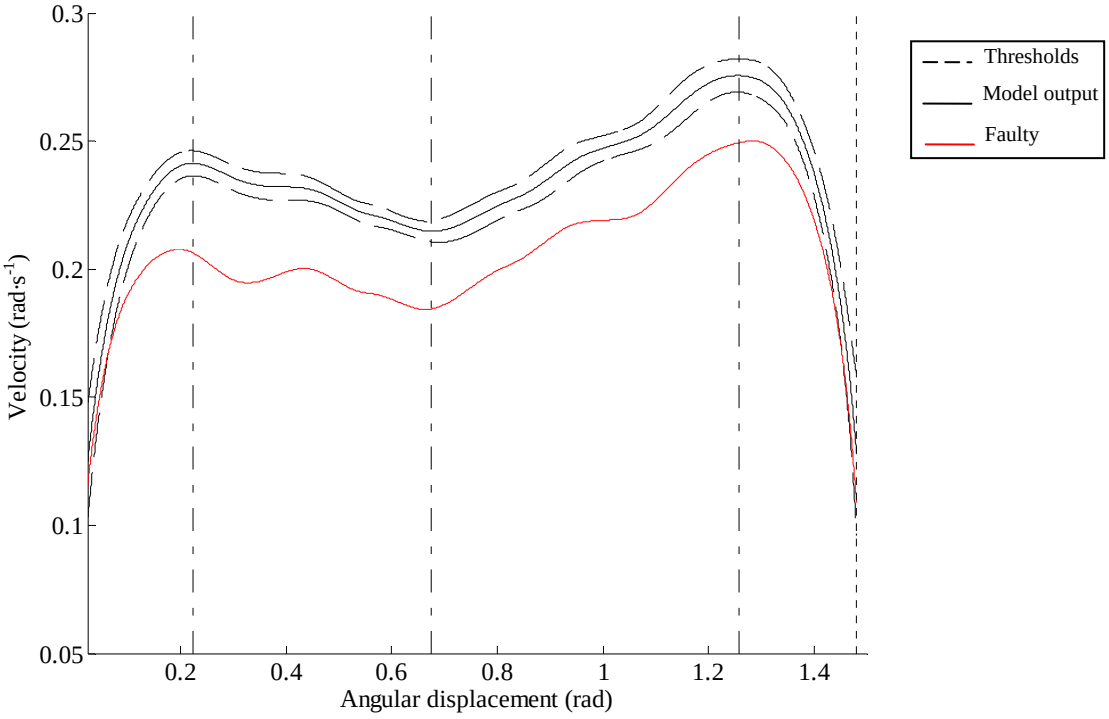


Figure 5.9 Velocity profiles of level crossing barrier (polynomial model *vs.* faulty).

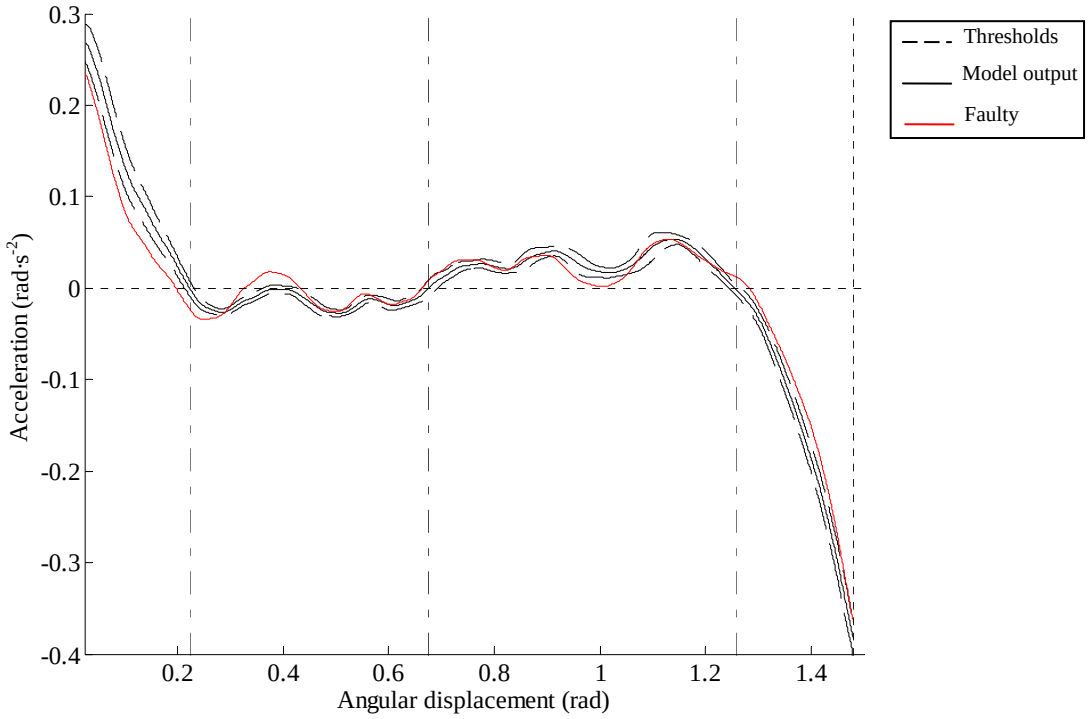
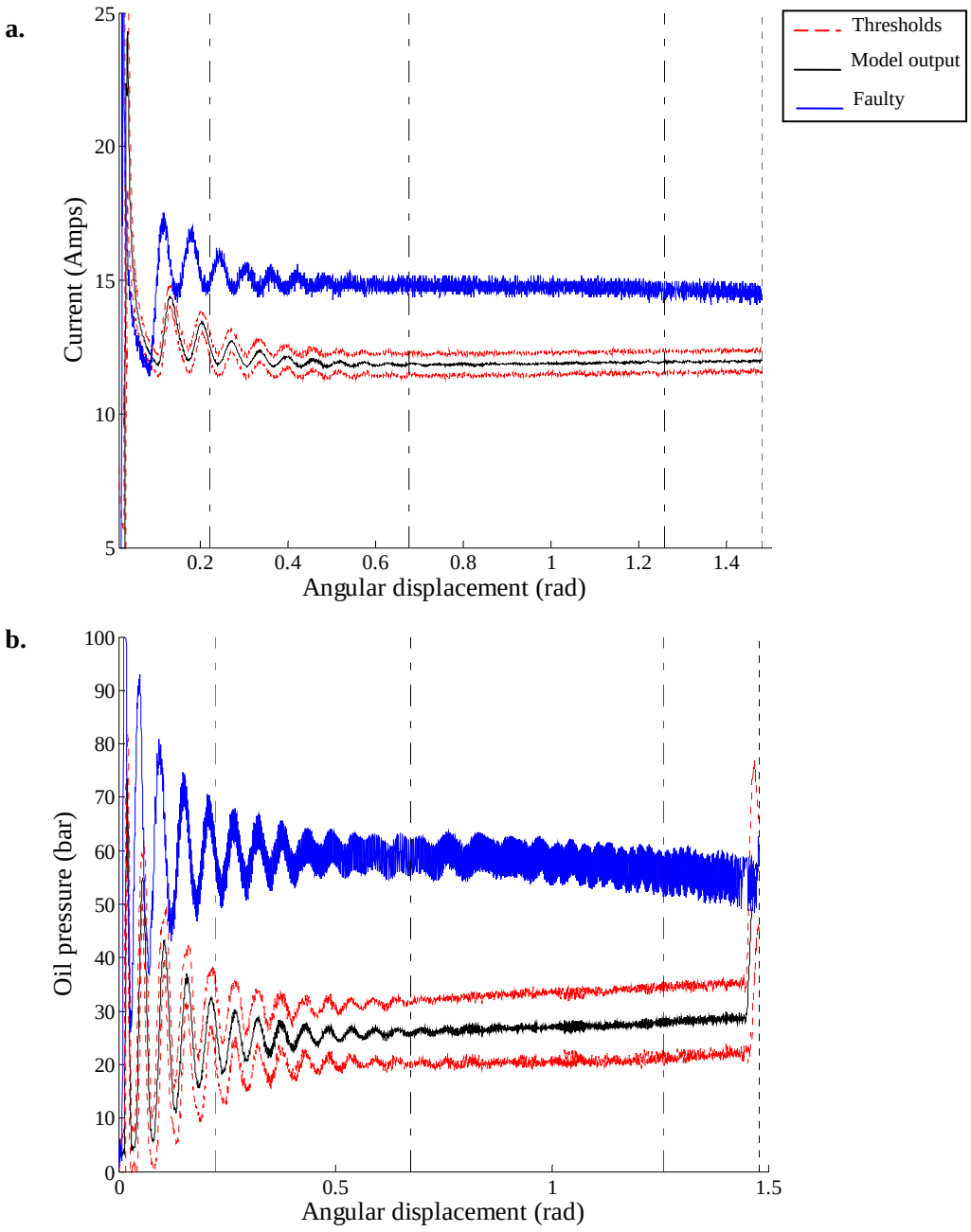


Figure 5.10 Acceleration profiles of level crossing barrier (polynomial model *vs.* faulty).



a) Current profiles **b) Oil pressure profiles**

Figure 5.11 RBF neural network models with thresholds *vs.* faulty profiles.

The current and oil pressure of the level crossing barrier have the features of complex non-linearity. In order to model these two variables, the RBF neural network was deployed. In Figure 5.11 a and b, the black solid line is the predicted output of the

neural network model. The thresholds were also calculated using the previously introduced adaptive thresholding method, shown by red dashed lines in the figure. The faulty profiles are presented to compare with the typical ones. It is clear that the simulated fault causes significant deviations from the model predictions.

Based on these results, the models have a good performance for parameter prediction and the thresholds work well for detecting the simulated fault.

5.3.2 Fault detection for electric point machine

In Figure 5.12, the generic fault detection and diagnosis diagram for the electric point machine is displayed. In this figure, the processes of pneumatic and electro-hydraulic assets are shown in grey, which means that these processes do not function. The current model in the electric process represents current profiles of the electric point machine motor.

Before considering the fault detection, the responses of the electric point machine parameters to the simulated load are illustrated by the current and displacement profiles shown in Figure 5.13. In the figure, the displacement and current profiles under both unloaded and loaded conditions are displayed for comparison. The red lines are the profiles when the point machine is loaded and the black lines are for the unloaded machine. In Figure 5.13 a, it can be observed that, when loaded, the throw started earlier, which indicates that the motor is aided by the released elastic force of the spring. When the drive bar keeps moving, the spring reaches its natural length and the force on the drive bar changes to a resistance force, where the displacement profile shows a step in the middle of travelling. In the second half of the distance, the drive bar pulls the spring and it takes longer to reach the end. The electric current profiles also indicate the

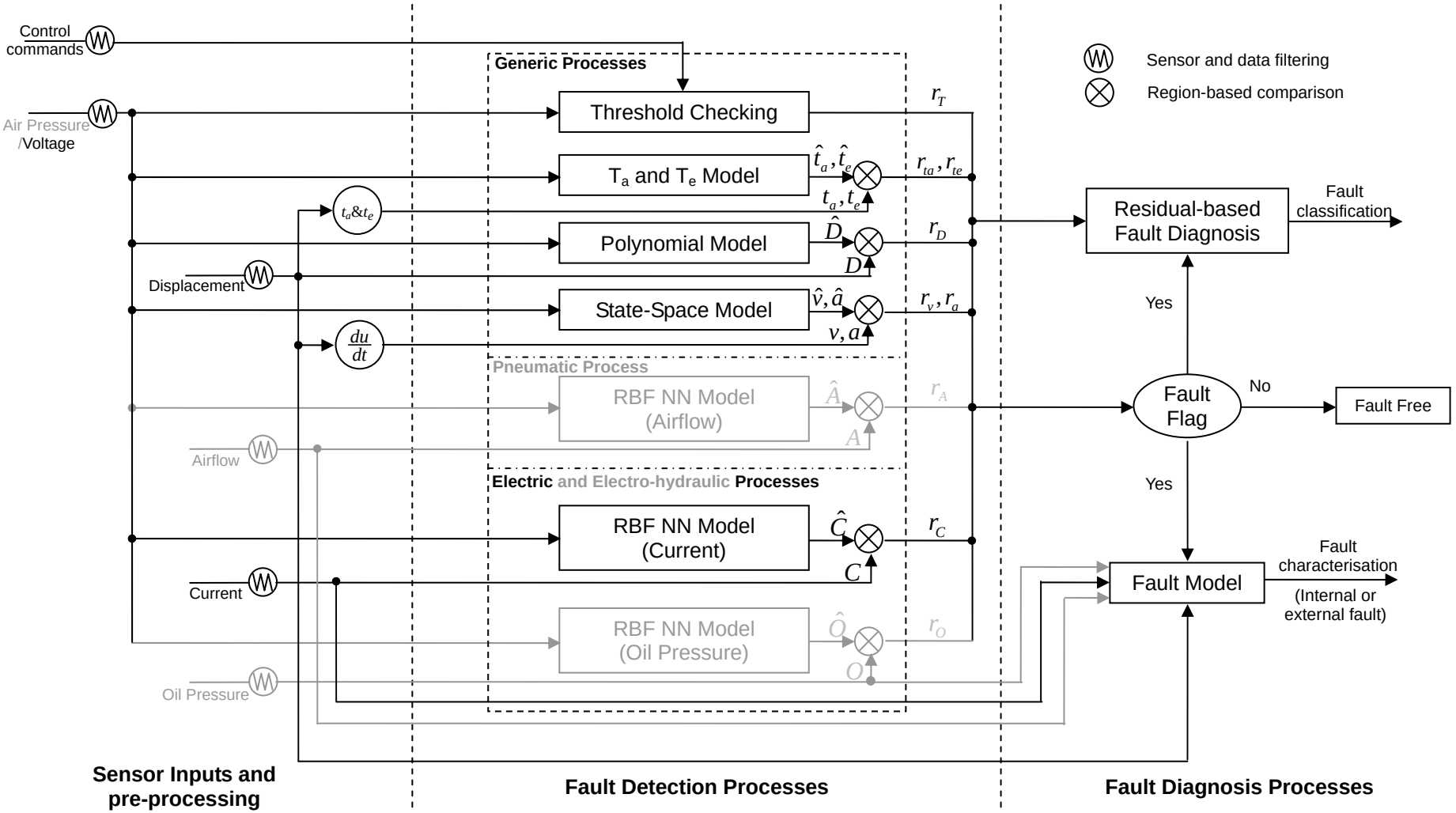
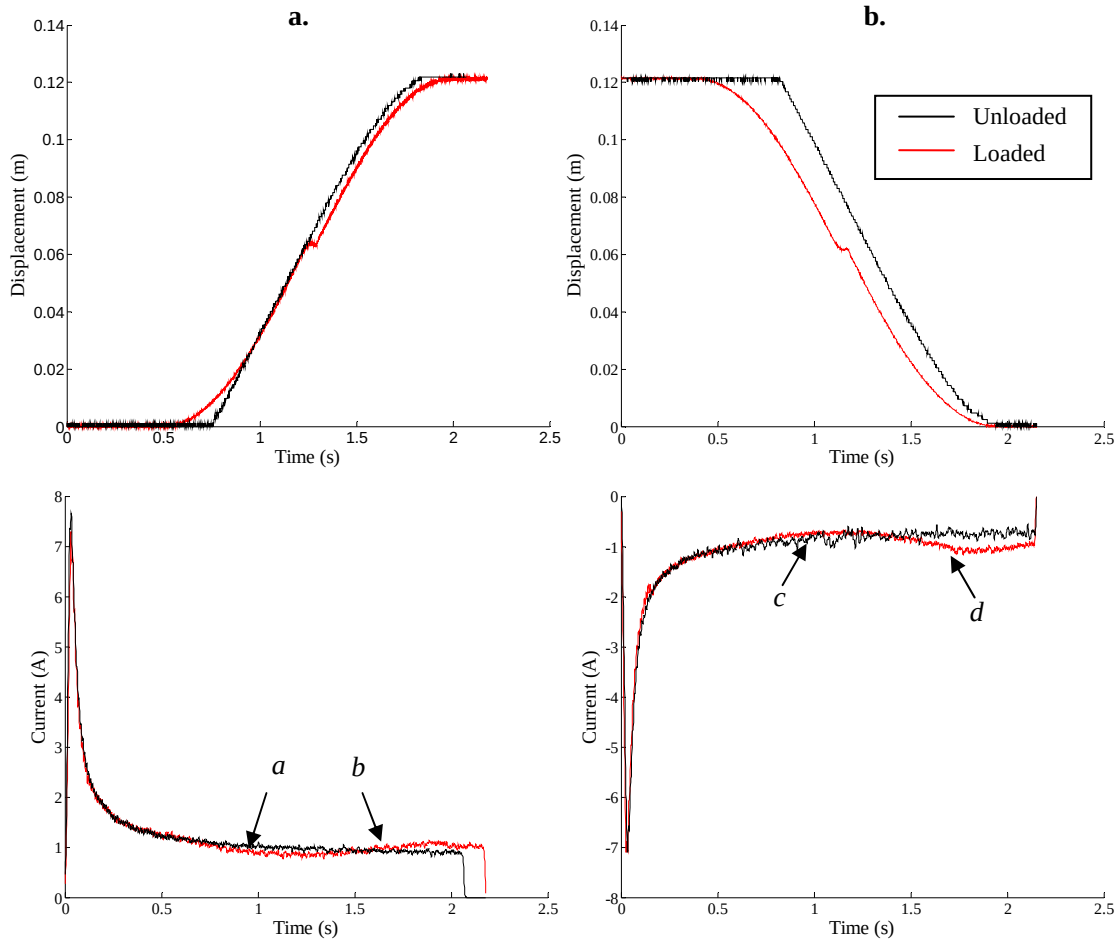


Figure 5.12 Diagram of generic fault detection and diagnosis for electric point machine.

process. In the first half of the distance, as shown by arrow point *a*, the loaded current is lower than the unloaded current due to the assistance of the spring released force. In the last half of the distance, the loaded current becomes higher, as shown by arrow point *b*, since the motor overcomes the resistance to move the drive bar to the end. For the reverse throw, the process is inversely performed and the two points, *c* and *d*, show the same situations. From this analysis, it can be observed that the displacement and the current reflect the dynamic changes in different ways. The correlation between these two parameters could be used to identify the health of the machine.



a) Normal throw **b) Reverse throw**

Figure 5.13 The displacement and current profiles of electric point machine (Unloaded *vs.* Loaded).

Operating characteristic	Unloaded		Loaded	
	Normal	Reverse	Normal	Reverse
Operating voltage-Average (Volts)	110 DC			
Motor current-Max (Amps)	7.6	7.1	7.3	7.1
Motor current-Working (Amps)	0.96	0.79	0.97	0.9
Activation delay (sec)	0.76	0.82	0.53	0.44
Throw time (sec)	1.82	1.94	1.92	1.9
Throw distance (m)	0.122			

Table 5.9 Operating characteristics of electric point machine.

The operating characteristics under both unloaded and loaded conditions are listed in Table 5.9. When the electric point machine is loaded, the activation delays of both normal and reverse throws are shorter and the throw time is longer than the unloaded operations. The spring load basically simulates the dynamic features of the stock rails which are driven by the point machine in practical terms.

In order to analyse the dynamic process of the throws, the velocity and acceleration profiles were calculated using the displacement data. Figure 5.14 and 5.15 respectively show the velocity and acceleration profiles for both the normal and reverse throws of the loaded electric point machine.

In Figure 5.14, the throw is divided into six spatial regions using the features of acceleration profile and the region boundaries are listed in Table 5.10. The regions of the start and end of the normal throw are defined by the maxima of acceleration close to both sides. The intermediate regions are divided using the zero-crossing points. Responding to the changes of the load, regions R_{d2} and R_{d5} represent the processes when the motor is aided and resisted respectively by the elastic force of the spring. Regions R_{d3} and R_{d4} present the dynamic change when the spring moves from

compression to stretch.

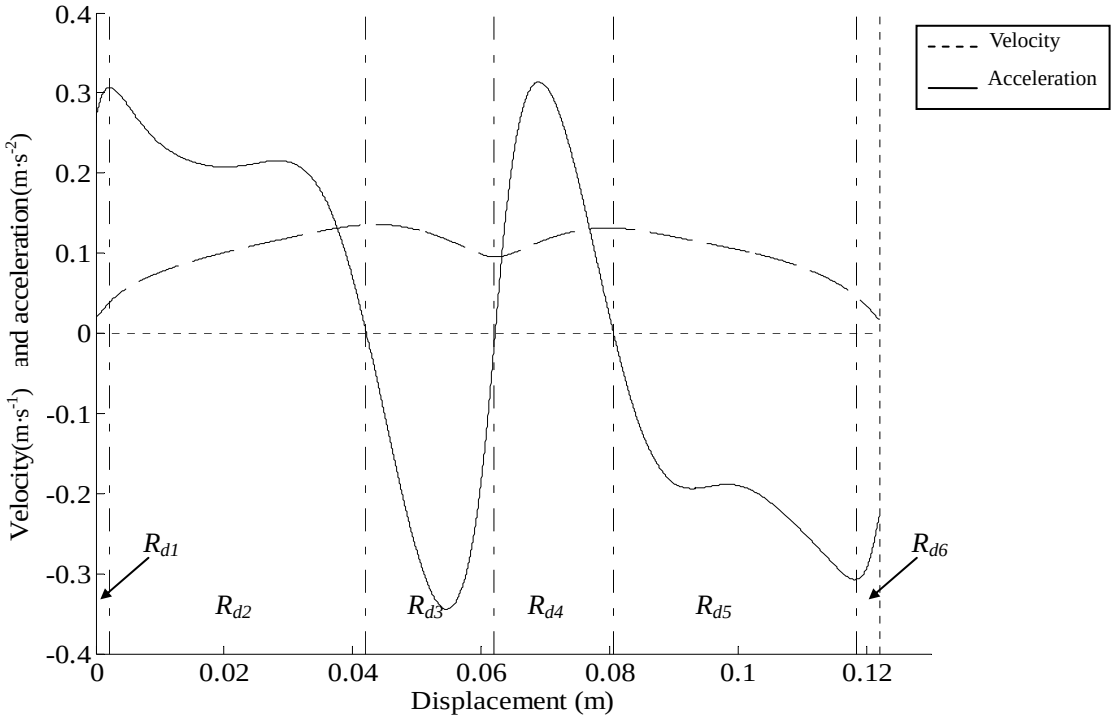


Figure 5.14 Division of spatial regions using acceleration and velocity features (electric point machine normal throw).

Regions	Boundaries (m)	Description
R_{d1}	0 – 0.002	Throw start
R_{d2}	0.002 – 0.042	Intermediate 1
R_{d3}	0.042 – 0.062	Intermediate 2
R_{d4}	0.062 – 0.08	Intermediate 3
R_{d5}	0.08 – 0.119	Intermediate 4
R_{d6}	0.119 – 0.122	Throw end

Table 5.10 Spatial region boundaries of a normal throw of electric point machine.

In Figure 5.15, the profiles for reverse throw are displayed and the region boundaries are listed in Table 5.11. There are also six spatial regions divided according to the dynamic changes of the load. Since the direction of the velocity is reverse to that in normal throw, the velocity is shown by minus values. The dynamic process of reverse

throw is inverse to the normal throw.

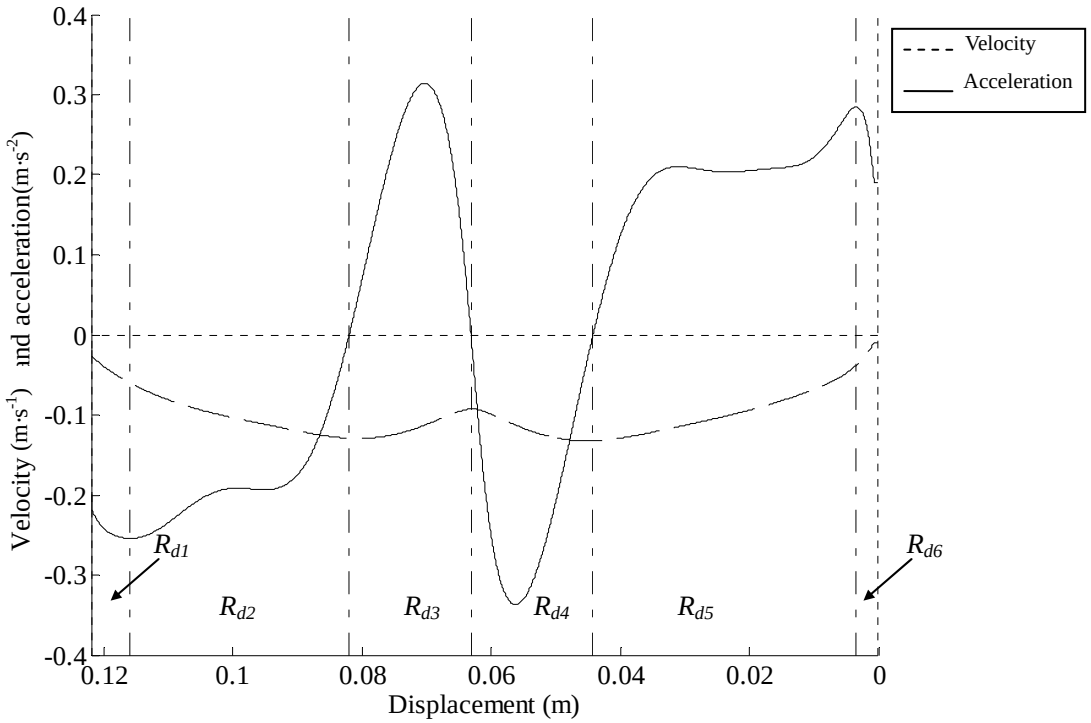
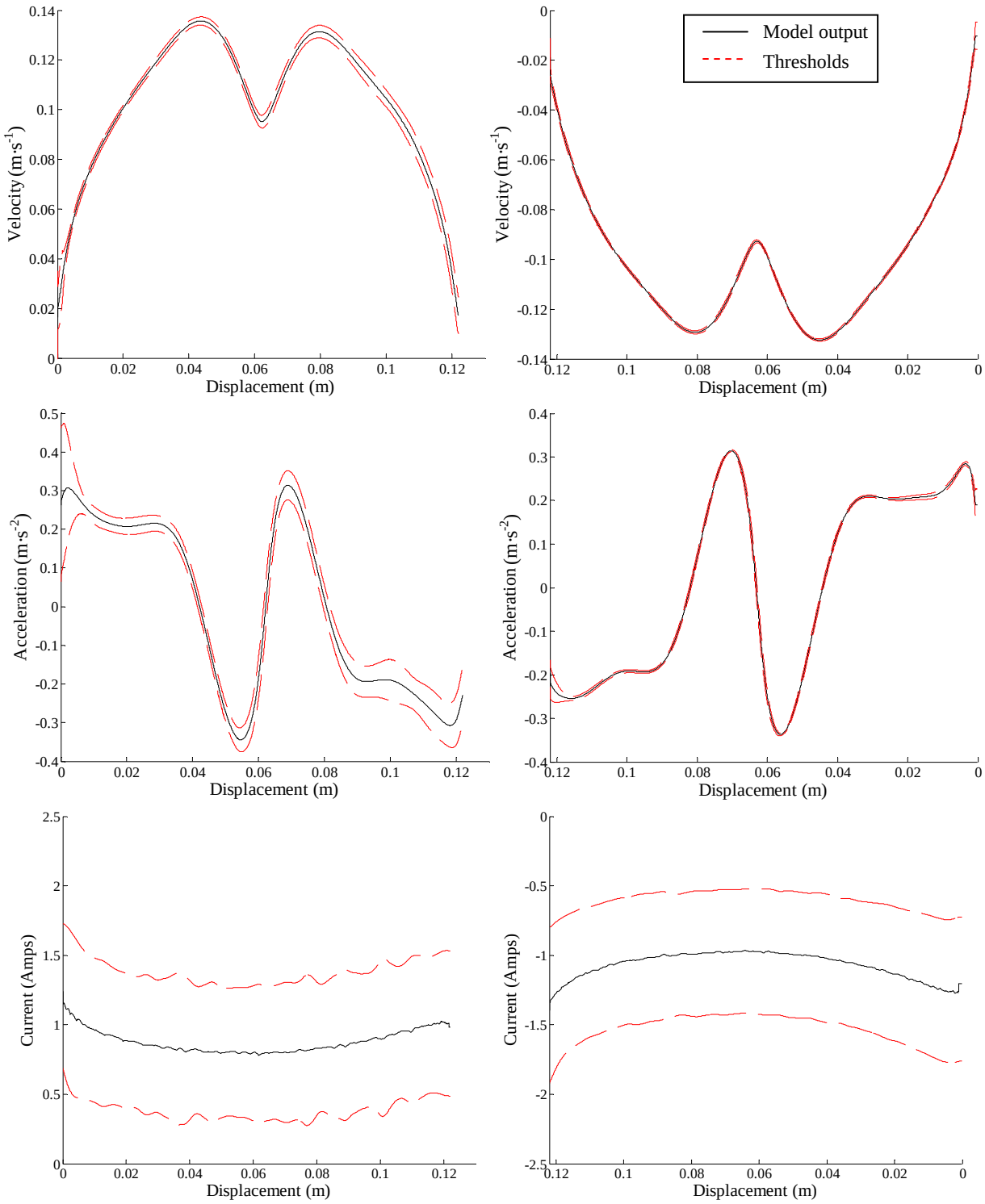


Figure 5.15 Spatial regions division using acceleration and velocity features (electric point machine reverse throw).

Regions	Boundaries (m)	Description
R_{d1}	0.122 – 0.116	Throw start
R_{d2}	0.116 – 0.082	Intermediate 1
R_{d3}	0.082 – 0.063	Intermediate 2
R_{d4}	0.063 – 0.044	Intermediate 3
R_{d5}	0.044 – 0.003	Intermediate 4
R_{d6}	0.003 – 0	Throw end

Table 5.11 Spatial regions of electric point machine reverse throw.

Models for the electric point machine were designed and the model outputs are displayed in Figure 5.16. Polynomial models were deployed for the prediction of



a. **b.**

a) Normal throw **b) Reverse throw**

Figure 5.16 Models of electric point machine with thresholds.

displacement profiles under fault-free conditions. The velocity and acceleration displayed in the figure are calculated using the output of the polynomial model. The RBF neural network was used for the simulation of the motor current, the results of which are also presented in the figure. At a confidence level of 0.97, the thresholds are generated and shown in Figure 5.16. Further work will be the simulation of faults and data collection under different fault modes. It would be interesting to evaluate the models using fault mode data and, meanwhile, enhance the robustness and accuracy of the designed models.

Based on the preliminary fault detection results, it can be concluded that the generic FDD method is applicable to the electric point machine. Monitored parameters, the displacement and motor current, can directly reflect the dynamic process of the asset. Changes of performance caused by faults are capable of being detected when their deviations from the model predictions become larger than the adjustable predefined thresholds.

5.4 Conclusions

This chapter presented four case studies for the pneumatic train door and point machine, the electro-hydraulic level crossing barrier and the electric point machine, using the proposed generic fault detection and diagnosis method. For each of the assets, the diagram of FDD processes was provided for the purpose of illustrating how the generic FDD method can be applied.

In the case studies of the two pneumatic assets, the fault detection, based on the pre-designed sub-models and the adaptive thresholding method, was applied and the

results were listed. The fault detection results indicate that the generic fault detection method performs as well as for the pneumatic train door discussed in Chapter 4. A conclusion can be drawn that the generic fault detection method is good for the pneumatic assets considered in this study. The results also suggested that the accuracy of fault detection by individual sub-model is relatively low compared to the combined model (combination of relative sub-models). The combined model is also considered as a generic model for the assets, which is able to describe their performance and status. The false alarm rates, under the condition that a false alarm is only counted when it is supported by all the sub-models, are in a reasonable range.

The case studies for the electro-hydraulic level crossing barrier and the electric point machine were also presented, where the emphasis was on the analysis of the feasibility of applying the generic FDD method on these two assets. As discussed in Chapter 3, the structure for the electro-hydraulic level crossing barrier and the electric point machine is similar to the mechanism features of the pneumatic assets, which can also be systematically divided as: input, machine and output. The parameter features were similar to the pneumatic assets. For each section of the system, there are one or two parameters which are valid to be monitored and to present the dynamic of the asset. For instance, in the case of the electric point machine, the voltage, current and displacement respectively represent the input, machine and output of the system and the dynamic changes are reflected by these parameters. Preliminary results of generic fault detection were provided and analysed. Based on the results and above analysis, a conclusion can be drawn that the generic solution would be effective and efficient for the two cases. As future work, it would be interesting to collect more fault mode data and evaluate the proposed method properly.

Chapter 6

A distributed condition monitoring architecture for railway assets

6.1 Introduction and motivation

As a safety strategy used to help avoid financial losses and unpredictable hazards, condition monitoring has been widely applied to industry processes. For safety-critical processes, such as nuclear power generation (Weerasinghe *et al*. 1998, Gomm *et al*. 2000), aircraft control and power systems (Patton and Chen 1997, Marcos *et al*. 2004) and chemical processes (Basila *et al*. 1990, Ruiz *et al*. 2001), the application of fault detection and diagnosis takes an essential role. In these cases, stand-alone, dedicated, expensive monitoring systems are utilised. However, the assets considered in this study are in a different situation. In railway systems, there is a large number of these assets, which are simple but safety-critical, thus a generic condition monitoring system in a distributed architecture is desired to reduce the maintenance cost while improving reliability at the same time.

In this chapter, a condition monitoring architecture for a large number of widely distributed multiple railway assets is discussed. Previous studies have considered a three-level distributed architecture that was successfully implemented to condition monitoring, where Embedded Processors (EP) were used to perform the data transmission and preliminary fault detection and diagnosis (FDD) for assets of the same type at close geographic locations (Roberts 2007, Dassanayake *et al*. 2001). An algorithm based distributed fault diagnosis architecture for electric train doors was

also proposed (Dassanayake 2001). As digital communication solutions, the Fieldbus and Ethernet were employed in the two studies and proved to be both effective and economical. These two networks are therefore also utilised in this study. The technical options and the protocol selection for the Fieldbus are also introduced. Finally, a new distributed architecture benefitting from the generic FDD solution will be discussed and a conclusion will be drawn to this approach.

6.2 Condition monitoring architectures

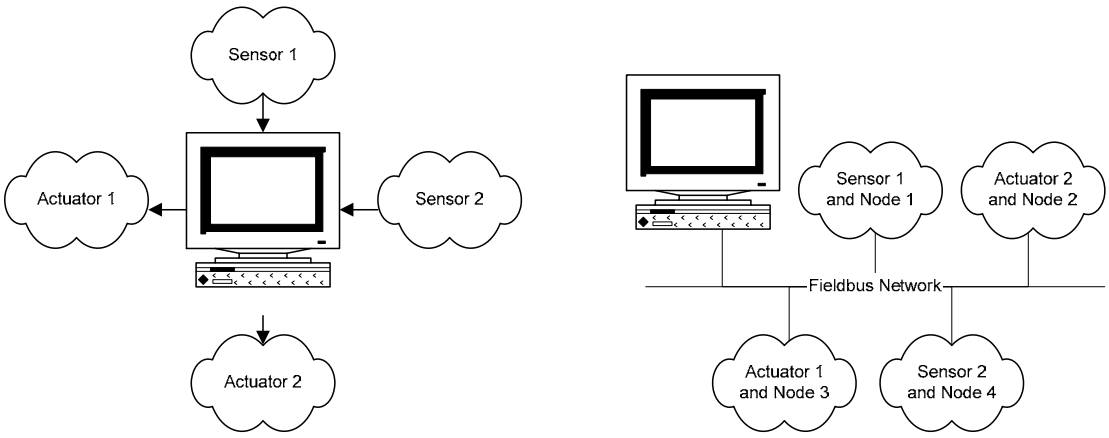
The purpose of developing a condition monitoring architecture is to achieve three system functions: data acquisition, computation and storage (Dassanayake 2001). In this section, these issues are discussed and possible solutions are proposed.

6.2.1 Condition monitoring architectures for multiple assets

In railway systems, multiple assets are distributed over a wide area. However, they are often found clustered together in a small local area, such as around a railway station or a junction. This geographical feature determines that the monitoring system cannot be highly centralised, otherwise data acquisition would be difficult. Consequently, the monitoring system should be geographically distributed to an asset or a cluster of assets. The localisation of the condition monitoring system is a distributed architecture, in which the functionality found in centralised systems is located throughout the architecture.

In the distributed condition monitoring architecture, data acquisition becomes viable. The computation for preliminary fault detection and diagnosis is also carried out close to the assets in order to reduce the amount of data communication. However, the

computation units should not be distributed to each individual asset, since the cost would be increased. As railway assets are often clustered in a local area, a local centralisation could reduce the cost for both data acquisition and computation. In principle, the distributed condition monitoring architecture should be a systematically distributed (the system level referred to as the Maintenance Information System (MIS) of the railway system) but locally centralised (at asset level including multiple assets) system.



a) Centralised monitoring system. **b) Fieldbus based distributed monitoring system.**

Figure 6.1 Two types of condition monitoring systems (Roberts 2007).

Two schematics of condition monitoring systems, centralised and distributed, are displayed in Figure 6.1. In the centralised monitoring system (shown in Figure 6.1 a), the control instructions and the data from sensors are sent out from and sent back to the central management computer. In the distributed system (shown in Figure 6.1 b), the sensors and assets (actuators) connect to a Fieldbus network which manages the communication. From the point of view of the financial cost, these two types of system are compared.

- The centralised monitoring system usually consists of a central management unit

which communicates with assets through point-to-point connections to sensors or control units. For widely distributed railway assets, a condition monitoring system based on this type of topology would require a large amount of cables to operate over long distances and the large number of assets would mean that this would be a complex process. As a result, the cost of data acquisition and computation is high, whereas the reliability and flexibility decrease.

- As discussed above, the distributed monitoring system distributes the data acquisition and preliminary computation to the assets in a local area. The monitoring data would not have to be sent through a long distance to the management centre and the initial FDD process could be carried out via a local network, such as Fieldbus. The financial cost is therefore reduced due to fewer requirements for cables and powerful computation devices.

Compared with the centralised architecture, the distributed condition monitoring architecture is, technically and financially, more suitable for multiple railway assets.

Previous studies proposed and implemented a three level (component level, asset level and system level) condition monitoring architecture for railway assets, including the train door, electro-hydraulic level crossing barrier and point machine (Figures 6.3 and 6.4, Roberts 2007).

In Figure 6.2, a configuration of train door distributed condition monitoring architecture is displayed. In passenger carriages, the train doors are connected to the Embedded Processor (EP) via a Fieldbus network. Each train door contains fieldbus nodes (FN) which link the transducers to the network for communication. The EP processes the preliminary FDD and reports to the Train Central Computer (TCC)

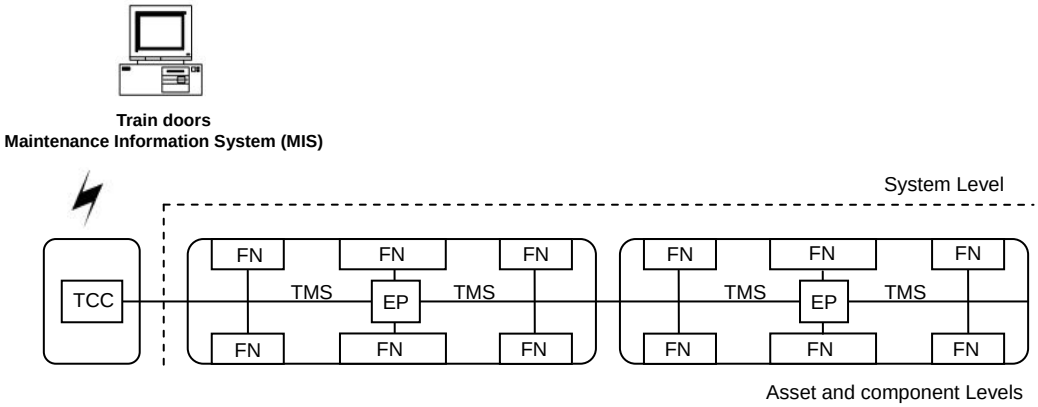


Figure 6.2 Distributed train door condition monitoring architecture.

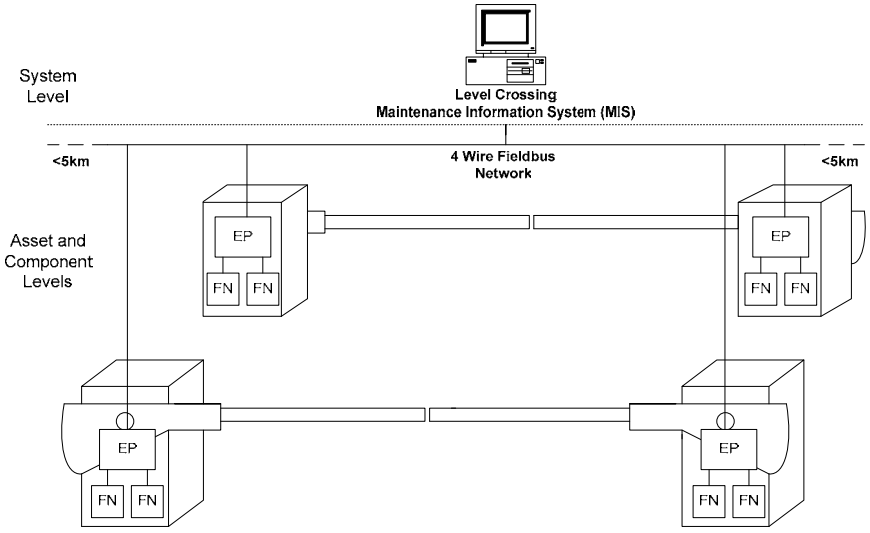


Figure 6.3 Distributed level crossing condition monitoring architecture.

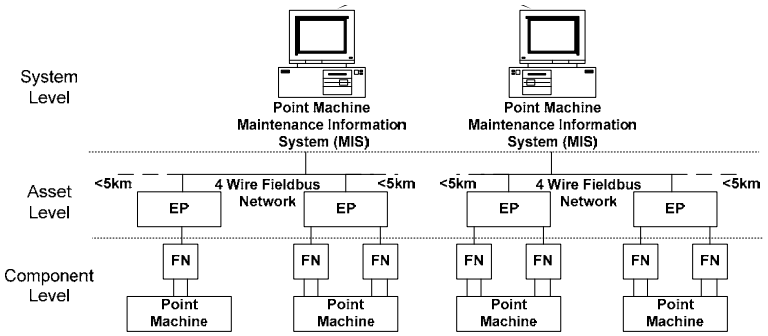


Figure 6.4 Distributed point machine condition monitoring architecture.

through the Train Management System (TMS). In the circumstance of a fault occurring, the TCC sends the fault information to the Maintenance Information System (MIS). The MIS contains the functionality of fault diagnosis which will identify the faults and arrange corresponding maintenance on the schedule.

The distributed condition monitoring architectures are also shown in Figures 6.3 and 6.4 respectively for the level crossing and the point machine. As for the train door, EP is fitted on the assets to process the data from the sensors. The Fieldbus network is also used for data transmission to connect the transducers to the EP via the FNs. The MIS collects the condition information from the EPs and manages maintenance for faulty assets.

This type of distributed architecture splits the data acquisition, computation and storage functions into three levels, via which the cost of the condition monitoring system decreases and the reliability increases. Based on this topology, another distributed architecture using the generic FDD method, which is more economic and powerful, is proposed and will be discussed in section 6.3.

6.2.2 Communication networks

In the distributed condition monitoring architecture, data acquisition and computation need to be localised to a cluster of assets which are in a geographically small area. A demand for a data communication solution in the local area is therefore raised.

In a conventional industrial control system, data communication often uses standards such as 4-20 mA (analogue signal) or RS232 (digital signal) (Thompson 1997). The topology of the network is centralised, where the transducers are connected directly to

the control centre by cables. As discussed previously, this type of topology is not suitable for railway asset condition monitoring. Due to the low reliability of the analogue signal transmission, the 4-20 mA standard is not suitable for critical applications or for long distances. As a replacement, the fieldbus system is introduced to the local network construction.

A number of digital communication networks developed under the International Electromechanical Commission (IEC) 1158 standard are collectively known as Fieldbuses (International Electrotechnical Commission 1993, Patzke 1998, Roberts *et al*. 1999). The advantages of the fieldbus digital communication network can be summarised in three main areas: installation, maintenance and performance.

- Instead of point-to-point, as in conventional systems, the fieldbus network is a multi-drop system. The sensors or assets and their controllers (CPU) are connected to a serial bus through the fieldbus nodes which manage all the information transmitted serially. This network topology offers a significant reduction in cable requirement. The fieldbus is also capable of supplying power to the transducers and transmitting the signals up to a 5 km distance. With an A/D (analogue to digital) conversion, the signals from the analogue sensors can be transmitted digitally through the fieldbus, which avoids the noise or other influences making analogue transmission unreliable. There are many fieldbus protocols available which can provide a high level of noise immunity and reduce the electromagnetic emissions during data transmission.

- The topology of a single bus network is much less complex than conventional systems, which implies that there would be less demand for maintenance. This

multi-drop bus system also provides a high level of flexibility for adding a new device or transducer by the addition of an extra node, where no more cable needs to be laid. Due to the simplification of the network, the operators can easily scan all the devices included in the system and debug any fault, once occurring, via the online diagnostics supported by the fieldbus. With these functionalities, both the financial cost and the time consumed by maintenance are much reduced.

- The protocols of the fieldbus are designed to transfer small data packets between the CPU and the fieldbus node (or between the nodes) with the minimum of time delay, which is ideal for time critical systems, such as controlling and monitoring. The capability of direct communication between the fieldbus nodes greatly increases flexibility when the data is required at different places in the system. System performance is also enhanced by freeing the CPU for other more important tasks. With fieldbus technology, the system can be initialised, operated and repaired faster than conventional systems.

With the advantages introduced, the fieldbus can be cost-effective, reliable and appropriate for the applications within a local condition monitoring system. Meanwhile, the technology and topology of the fieldbus enables the computation to be distributed into a set of smaller processing units, which reduces the data amount for transmission.

The data communication from the local system to a higher level (MIS) can be undertaken by the Ethernet. The ethernet is a frame-based computer networking technology for local area networks (LAN). In comparison with the fieldbus, the ethernet transmits a large amount of data with no time criticality in either wired or

wireless mode, which is thus not suitable for the local network where timed data transferring is required.

6.2.3 Fieldbus for distributed condition monitoring

Under the same standard, various fieldbus network protocols, such as WorldFIP, Profibus, CAN and Interbus, were designed for applications in distinct industrial scenarios (Kolla 2003). Focusing on the industrial application domain, the communication protocols can be grouped into three categories: sensor networks, device networks and control networks (Pratt 2003). In this study, the fieldbus network for asset condition monitoring crosses all the three domains. The selection of appropriate fieldbus protocol is therefore challenging but critical.

Each of the fieldbus networks has its own strengths and weaknesses. In order to select a suitable fieldbus, the following points need to be considered (Noel 2002, Kolla 2003).

- The selection should focus on the characteristics of the application and the specific benefits the designer expects. For serious condition monitoring applications, data transmission needs to be fast and over a long distance. For instance, the WorldFIP is capable of transferring at 1 Mbit/s rates over a distance of up to 5 km. As a number of multiple assets are connected to the bus via field nodes, the variables should be identical and sharable via broadcasting by other nodes. Depending on the number of assets involved, the maximum amount of field nodes, which can be added to the bus, needs to be considered. The power supply for the transducers must be included in the network functionalities.

- As discussed previously, an important criterion is cost. Under the precondition of providing the required performance, cabling and chipsets should be affordable. If new nodes are needed in the future, it might be necessary to re-assess the cost, implied by the compatibility and supportability of the network hardware and protocol.

- The network connectivity or, in other words, how the data can be obtained, needs to be considered when choosing a fieldbus protocol, especially when multiple assets are connected to one fieldbus. The proper network topology must be assigned and each type of asset is with specific identifiers.

- Flexibility of the network is also an important issue if future changes are expected. A flexible network protocol is capable of supporting the potential changes and impacts with an affordable cost. The interface of the network to both other networks (interoperability) and the operators (supported software) must be friendly.

6.3 A distributed architecture for generic FDD based condition monitoring

The main topic of this study focuses on the approach of a generic fault detection and diagnosis solution for multiple railway assets, including the pneumatic train door, train-stop and point machine, electric point machine and electro-hydraulic level crossing barrier. The benefits of this generic method are to allow a number of multiple assets to be monitored simultaneously and to ensure relatively simple algorithm development. Based on the three level distributed condition monitoring architecture

proposed in previous studies, a distributed architecture for generic FDD based condition monitoring is proposed (as shown in Figure 6.5).

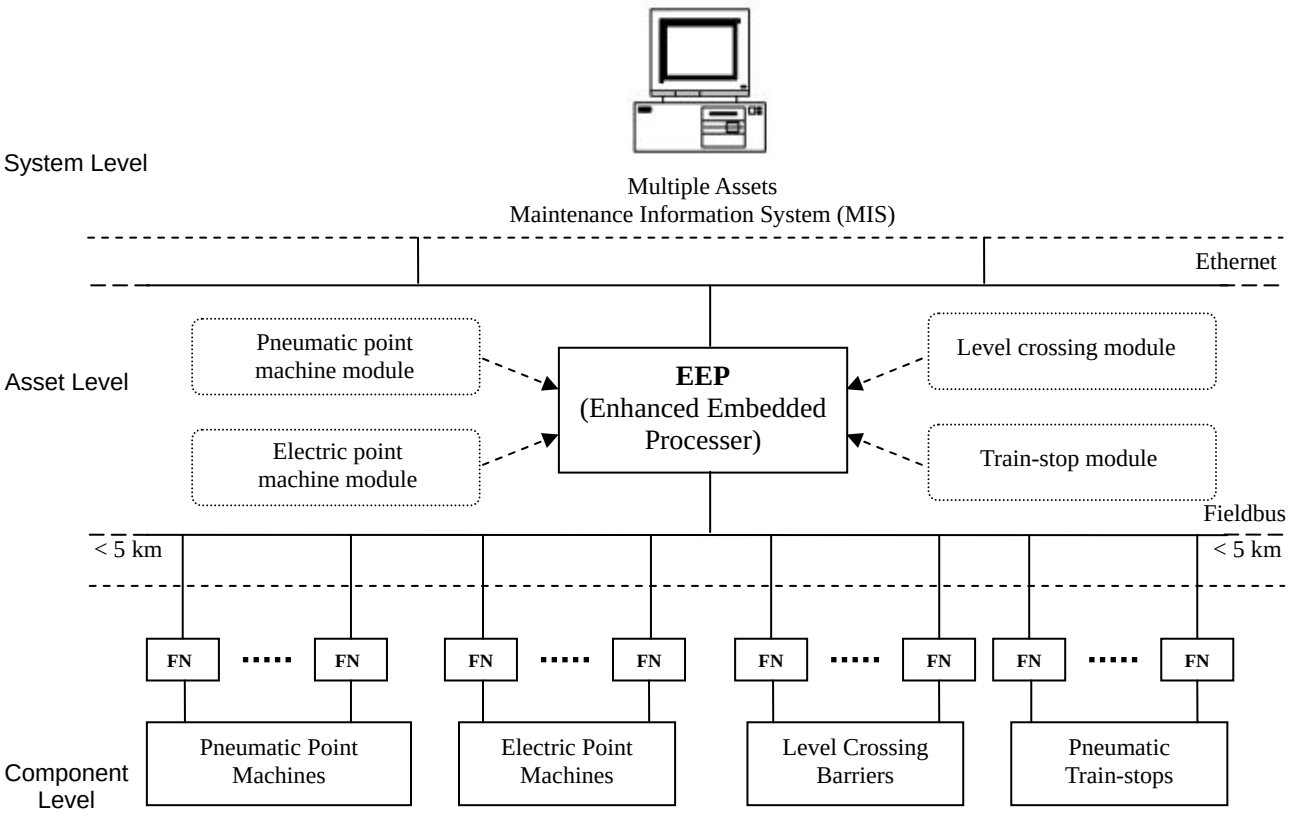


Figure 6.5 A three level distributed architecture for generic FDD based condition monitoring.

In previous architecture systems, preliminary FDD is processed by the Embedded Processors (EP) for railway assets. Since the FDD algorithms developed for each type of asset differ from the others, the EP works specifically on the assets of same type. The computation undertaken by the EPs is, therefore, further distributed to groups of same assets. The generic FDD method proposed in this thesis has same algorithm structure, even though the parameter models are different for the assets with distinct electromechanical principles. It is consequently possible to integrate the algorithms for the multiple assets considered in this study into one processing unit to monitor their conditions at the same time. Based upon this feature, an Enhanced Embedded

Process (EEP) is proposed to replace the EPs at the asset level to further centralise the computation in a local network. Although the EEP requires a more powerful computation capability and a larger data storage capacity for the preliminary FDD processing using multi models, the computation cost could be reduced compared with the cost of a number of EPs.

Of the assets involved in this study, the pneumatic train door is not included in this architecture since it works on a moving train without geographic constraint. Furthermore, with the assistance of distributed EPs, the train central computer is able to independently perform FDD for the doors via the train management system. The proposed architecture will only be suitable for multiple assets in a geographically local area.

In Figure 6.5, the distributed architecture still consists of three levels: component level, asset level and system level. At the lowest level, the fieldbus node links between the transducers and the bus, and controls the data communication. Low-level control alarms can be produced by the node when it is connected to a control signal monitoring component. At the asset level, the EEP unit, which contains 4 modules for 4 types of assets, is capable of interfacing to manage the traffic on the local fieldbus network and processing the fault detection and preliminary diagnosis. Once a fault is detected by the EEP, the information is passed to the Maintenance Information System (MIS) at the system level for further fault diagnosis. The link to the MIS from the EEP could either be a wired or wireless ethernet. The MIS will generate a failure code after isolating the fault and send the information to infrastructure maintainers for maintenance scheduling.

6.4 Conclusions

In this chapter, the architecture for condition monitoring of the railway assets has been introduced and discussed. The feature of geographical distribution of the railway assets determines that a widely distributed and locally centralised condition monitoring architecture is appropriate for condition monitoring of multiple assets. As a data communication tool, the fieldbus was introduced and its advantages for industrial applications were discussed. The principles of selecting a suitable fieldbus protocol were listed and the requests for condition monitoring of railway assets were described.

In a fieldbus based three level distributed architecture, the assets are efficiently managed via a flexible network topology. The cost of this distributed network is lower compared with the traditional centralised network and maintenance is less time consuming and more economical. As a digital communication network, the fieldbus is reliable and the interface to the operator is friendly and supports third-party software.

In the final section of this chapter, an Enhanced Embedded Processor (EEP) based three level architecture is proposed. The computation distributed from centralised system functionality is locally centralised. The EEP requires more computation and data storage costs; however, the total computation cost is still reduced by replacing specifically functioned EPs. A number of multiple assets can be monitored simultaneously with the integration of different FDD algorithm modules. With the ethernet link to the MIS, the modularised condition monitoring software could be changed or upgraded remotely.

Chapter 7
Conclusions and further work

7.1 Introduction

This chapter presents the conclusions of a generic fault detection and diagnosis approach (FDD) for simple multiple railway assets, known as a classification of Single Throw Mechanical Equipments (STMEs). Further work required in some areas is also suggested in order to improve the functionality and reliability of the proposed generic FDD solution.

The case studies in this thesis are on the low-level, simple trackside railway assets, including the pneumatic train door, the point machine and train-stop, the electric point machine and the electro-hydraulic level crossing barrier. In contrast to their low capital cost, these safety-critical assets have significant value for the operation of whole railway system. A cost effective FDD method focusing on generic application, with the aim of improving the reliability of these assets, was studied and presented in this thesis.

In the literature review (Chapter 2), an overview of current FDD techniques was illustrated. A description of the railway assets involved in this study and the development of test rigs for data acquisition and modelling were presented in Chapter 3. According to the common features of performance and load, these assets were classified as STMEs. A generic FDD method was proposed for the railway assets and the development of this method was illustrated. For residual generation, an

adaptive thresholding algorithm was also developed and presented. Two fault diagnosis methods were proposed and discussed. In Chapter 4, a generic fault detection and diagnosis approach was carried out for the pneumatic train door and the modelling work was described in detail. Case studies for the other four assets were provided in Chapter 5. The results of case studies in Chapters 4 and 5 indicated good performance and generic applicability of the proposed FDD method. This common feature based generic solution benefits from the low cost of a generic condition monitoring system. A distributed condition monitoring architecture using the Fieldbus technique was discussed in Chapter 6. The detailed conclusions of this thesis and further work which may be pursued are presented in the next sections.

7.2 Conclusions

This section summarises the findings from this work in the following areas.

7.2.1 Model-based fault detection and diagnosis methods

In previous studies, quantitative model-based residual generation methods, namely observers, parity equations and parameter estimations, were well developed. With adequate information of the system and probable disturbances, the first two methods, when well designed, respond to sudden faults rapidly. Parameter estimation is good at detecting incipient faults by observing changes in system parameters, but the response speed is relatively slow because a large amount of computation is involved. A limitation of parameter estimation is that this method only works for dynamic systems, since the estimated values would drift when the system is in a static status.

Compared with the quantitative methods, the qualitative methods, such as neural

networks, fuzzy logic and neuro-fuzzy, do not depend on the first principle information of the target system. Trained by the input and output data, these methods are capable of simulating the behaviour of the system, where the system can be either linear or nonlinear. The neural network model is known as a black-box model, which indicates that it is impossible to understand the trained network. Inversely, the fuzzy logic model can be easily explained by its fuzzy rules and the model performance can be adjusted by tuning the rules. However, prior expert knowledge is required for the fuzzy system design. A combination of these two qualitative methods, the neuro-fuzzy system, benefits from both the self-learning of the neural network and the transparent model structure of fuzzy logic, which makes the complex system modelling more robust and flexible.

7.2.2 Involved assets and test rig development

Five railway assets involved in this study were described. With similar dynamic characteristics, these assets were classified into a group called Single Throw Mechanical Equipments (STMEs). For the purpose of data acquisition and modelling, lab-based test rigs were developed for these assets. With the interface using LabVIEW software, the asset control, data acquisition and storage were automatically performed by the computer. The details of hardware utilised in this system were provided in Chapter 3.

7.2.3 Generic fault detection and diagnosis method

Among the five assets considered in this study, the three pneumatic assets have similar dynamic features, mechanical designs and parameters, which facilitates the development of a generic FDD method. Common features of their mechanism and

parameters were abstracted, based on which a systematic view of these assets was presented as input, machine and output. The performances of three sections of the asset depend on each other. The logic relations of these sections were used for logic analysis in the circumstance of a fault occurring. The pneumatic assets and the electric and electro-hydraulic assets were compared and a conclusion was drawn that these assets have similar system structure and parameter features and the proposed generic fault detection and diagnosis method could also be applicable to these two assets.

Based on the common features extracted from these STMEs, a generic fault detection and diagnosis method was developed and illustrated by the diagram of processes. The whole FDD process consists of three stages: sensor inputs and pre-processing, fault detection processes and fault diagnosis processes. In each of the processes, the tasks are generically defined for all the railway assets considered in this study. At the stage of sensor inputs and pre-processing, the data of the variables is collected by a set of sensors and pre-processed for fault detection. The fault detection processes include three sections, generic processes, pneumatic process, and electro-hydraulic and electric processes. A series of sub-models are designed for these processes, where a combination of these sub-models is considered to be the generic model for the STME assets. For residual generation, a statistics based adaptive thresholding method was used to automatically generate thresholds. At the fault diagnosis stage, two methods were introduced. The fault model was used to roughly classify the faults into two categories: external faults (e.g. friction or obstruction) and internal faults (e.g. mechanical faults). Another method, using residuals patterns to accurately classify faults, was also proposed, in which two methods were introduced to decouple the residuals from the sensor faults.

7.2.4 Railway assets modelling

Four modelling approaches using data collected in the laboratory were illustrated. In consideration of both the computation cost and accuracy, these modelling methods were discussed. The exponential model, polynomial model and state-space model benefit from a low computation cost at a certain level of accuracy. The neural network model provided an accurate simulation with a larger amount of computation. These modelling methods were selected according to the characteristics of the modelled parameters. The modelling results indicate that the model outputs fit the measured profiles well and the selection of modelling methods is correct.

7.2.5 Fault detection and diagnosis approaches

Fault detection approaches were applied to the three pneumatic assets and the results were provided. From the fault detection results, it can be observed that the performance of a single sub-model is usually not good enough, due to unavoidable model uncertainty and disturbances; whereas a combination of the sub-models presented good sensitivity to faults and the false alarm rate was effectively controlled.

The feasibility of applying the generic FDD method to the electric point machine and electro-hydraulic level crossing barrier was analysed. Preliminary fault detection approaches were carried out and the results indicate that the generic fault detection and diagnosis method is feasible for these two STMEs with an electrical power source.

Based on the results of five case studies, the performance of the proposed generic fault detection and diagnosis method was proved to be good for the pneumatic assets and applicable for all five assets considered in this thesis.

7.2.6 Condition monitoring architecture

Condition monitoring architecture for the railway assets was discussed. Based on the geographical distribution of the simple railway assets, a three level distributed condition monitoring architecture, proposed in a previous study, was introduced. The monitoring system, from assets to the Maintenance Information System (MIS), is divided into three levels: component level, asset level and system level. At the component and asset levels, the fieldbus, a serial-bus digital communication network, was employed due to its flexible network topology and low cost achieved by reducing the amount of cabling. Embedded Processors (EPs) were used to communicate between the components and the fieldbus. From the asset level to the system level, ethernet was used for data transmission.

With the benefits of a generic FDD solution, an Enhanced Embedded Processor (EEP) based, widely distributed and locally centralised condition monitoring architecture was proposed. The generic FDD solution allows simultaneous condition monitoring of a large number of multiple assets, thus the distributed computation of preliminary FDD can be centralised to an EEP in a relatively small local area. This methodology would further reduce the cost of computation by replacing many EPs with one EEP. The modularisation of the FDD algorithm in EEP would allow the operator to change or upgrade the software remotely.

7.3 Further work

Work completed so far has been presented in this thesis. However, it would be interesting to carry out some further work.

7.3.1 Fault detection and diagnosis model improvement

In this study, the work is focused on finding a generic fault detection and diagnosis method for all STME railway assets. Work was also completed on the development of test rigs for data acquisition and modelling. Some interesting points found in the data collection and analysis may be studied further.

7.3.1.1 Improvement by system features

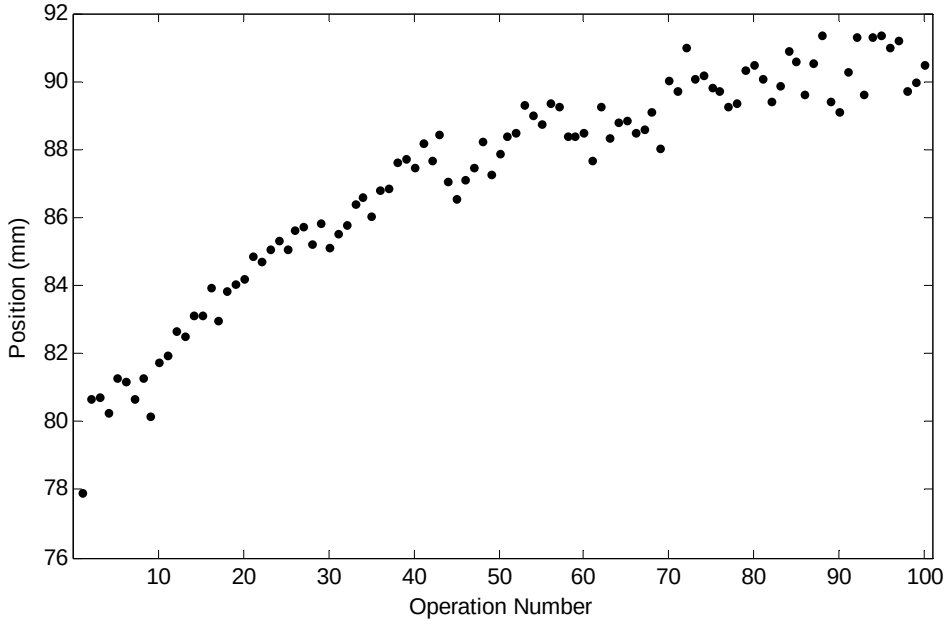


Figure 7.1 Train door positions at 1 sec for 100 normal throws.

When data was collected from the pneumatic train door, it was found that the displacement profiles did not simply repeat under the same air pressure, where the activation delay and throw time decreased with the increase in continuous operation time. Figure 7.1 shows the train door positions at 1 sec for 100 normal throws. With the increase in operation time, the door positions became farther than previous ones at the same time point, in other words, the door travelled faster. The data sets obtained

were therefore a band instead of a single line. This situation was considered in modelling by using an adaptive threshold for tolerance; however, it is still meaningful to study the effects of lubrication on the train door trajectory.

The points shown in the figure could be modelled using exponential or polynomial methods. The model of the train door integrated with this feature could identify the lubrication condition of the door trajectory. This feature exists in the operation of all the STME assets considered in this study and it is more obvious in the relatively weakly powered assets, such as pneumatic assets.

In the case of the train-stop, the peak values of acceleration were shown in Figure 3.15. It was proposed that the modelling of the peak values would benefit from an automatic regions division. This would make the model more flexible and enable a self-learning based preliminary fault detection algorithm for newly produced assets. This feature mainly exists in the case of the train-stop, but would also be useful for other STME assets when they are loaded.

Other factors which would influence the asset performance include temperature and humidity. Currently, temperature is not involved in the modelling work, since the test rigs of assets are lab-based with little temperature change. In practice, the temperature may change significantly which means that this feature would become important. Humidity may also affect the condition of assets as erosion can develop on moving components. Environmental information could be considered in data collection by using intelligent sensors that are able to pre-process the sensor data to allow for the environmental conditions (Tian, Zhao & Baines 2000).

7.3.1.2 Improvement by failure mode data

In the proposed generic FDD method, fault models, trained by failure mode data, are able to identify the faults occurring externally or internally. The accuracy of the model performance relies on the amount of faailure mode data for training. So far, the fault simulation has been carried out on pneumatic assets; however, more failure mode data is required to improve the model accuracy. As a future work, it is necessary to simulate faults on the electric and electro-hydraulic assets to collect failure mode data. As well as training the fault model, the failure mode data is also required for testing the robustness and reliability of parameter models.

7.3.2 Laboratory based online condition monitoring

The test rigs for assets have been set up in the laboratory. With well designed and trained models, a computer based preliminary online condition monitoring system for testing purposes is expected. The online test would include the following procedures:

- The computer sends operation instructions to controlled assets, where the operation could be either continuous or random.
- The parameters of assets are monitored simultaneously by the computer via installed sensors.
- Once a throw is finished, the data would be processed for FDD.
- Faults could be simulated randomly on the assets.
- Testing the FDD software responses to the simulated faults and fault-free conditions.

The test would be very useful for detecting defects in the FDD algorithm and improving the reliability of condition monitoring software.

7.3.3 Real trackside data

Once the lab-based FDD test is accomplished, the practical application of the generic FDD method would be considered. The limitation of the proposed method is that the performance of assets in the laboratory may differ from practical instances. To overcome this limitation, real trackside data would be essential for testing the algorithm.

Since the same examples of an asset may vary in performance, it may be difficult to fit the generic FDD method to every asset. A solution is that a self-learning algorithm is achievable based on the ease of modelling the parameters. The generic FDD method integrated with a self-learning feature would be able to adapt to each individual asset using the initial operation data when the asset is newly fitted. After a period of time, the asset performance may degrade and an alarm would be triggered. If the degradation is caused by normal wear or by environmental influences, the self-learning mechanism could allow re-adaption to the asset with the configuration of the infrastructure maintainer.

7.3.4 Neuro-fuzzy decision making

Neuro-fuzzy is a useful tool to solve nonlinear problems. In fault diagnosis, the residuals, generated in fault detection, need to be mapped to certain faults by which the fault can then be isolated. Furthermore, the weighted residuals would also be able to provide the information on the fault strength and location. The nonlinear mapping from residual patterns to corresponding faults could be processed by the neural

network. With well designed fuzzy rules, the magnitude of residuals could be transferred to the strength of faults and the faults could also be located. It would therefore be interesting to explore a neuro-fuzzy method for residual processing and fault isolation.

Appendix A

Illustration of lab-based STME test rigs

A-1 Pneumatic train door test rig

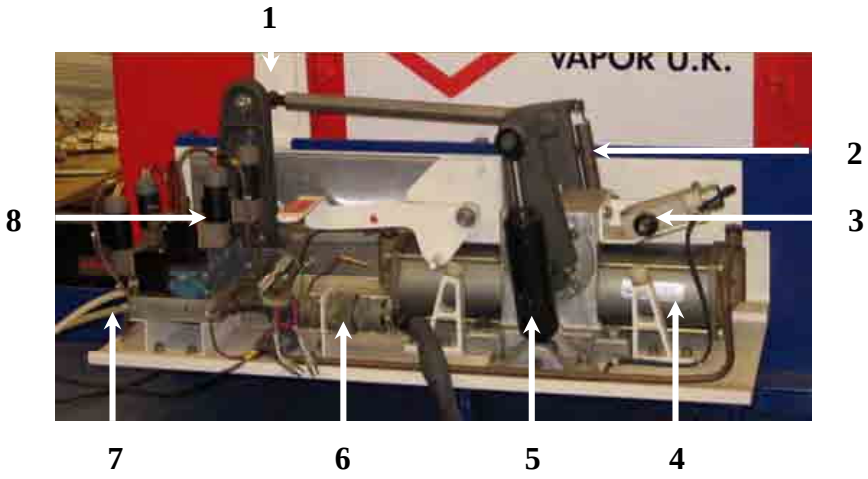


Figure A.1 Illustration of pneumatic train door test rig and actuator.

Description of components			
1	Door arm	**5**	Hydraulic damper
2	Pushback springs	**6**	Electric connector
3	Door open limit switch	**7**	Compressed air input
4	Pneumatic cylinders	**8**	Solenoids

Table A.1 Description of components of pneumatic train door test rig.

A-2 Pneumatic train-stop test rig

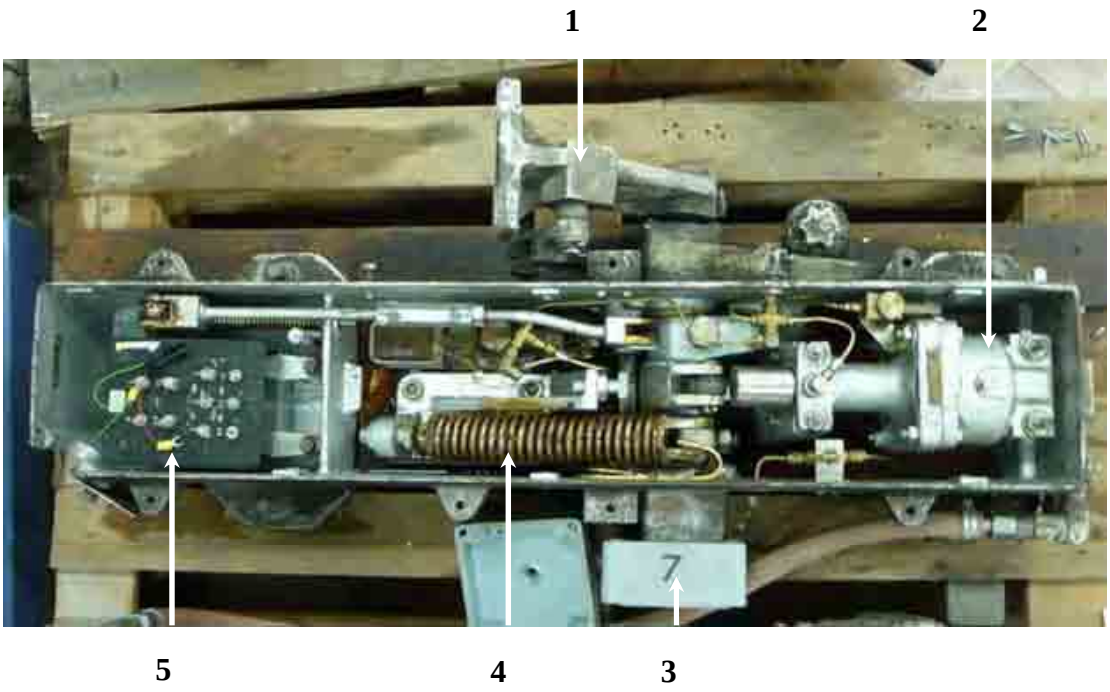


Figure A.2 Illustration of pneumatic train-stop test rig.

Description of components			
1	Train-stop head	**4**	Main spring
2	Air motor	**5**	Proving box
3	Angular displacement sensor		

Table A.2 Description of components of pneumatic train-stop test rig.

A-3 Pneumatic point machine test rig

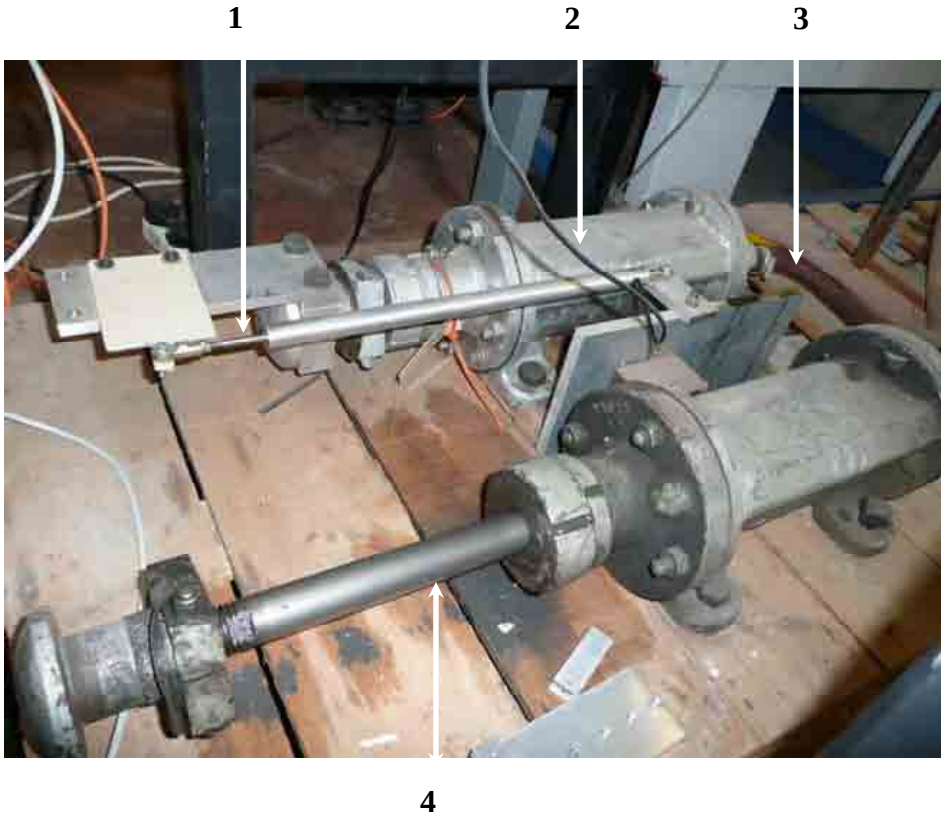


Figure A.3 Illustration of pneumatic point machines test rig.

Description of components			
1	LVDT displacement sensor	**3**	Compressed air supply pipe
2	Pneumatic cylinder	**4**	Drive rad

Table A.3 Description of components of pneumatic point machine test rig.

A-4 Electric point machine test rig

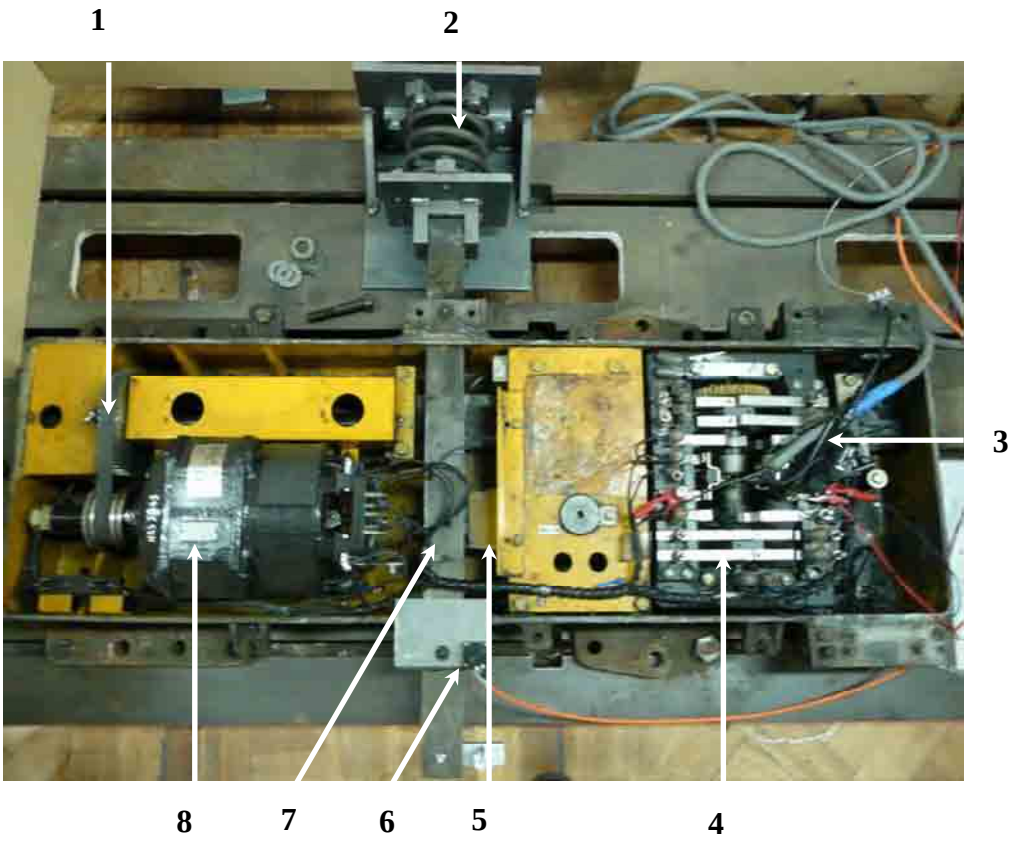


Figure A.4 Illustration of electric point machine test rig.

Description of components			
1	Transmission belt	**5**	Crank
2	Spring load	**6**	Draw-wire displacement sensor
3	Current transducer	**7**	Drive bar
4	Circuit controller	**8**	Electric motor

Table A.4 Description of components of electric point machine test rig.

A-5 Electro-hydraulic level crossing barrier test rig

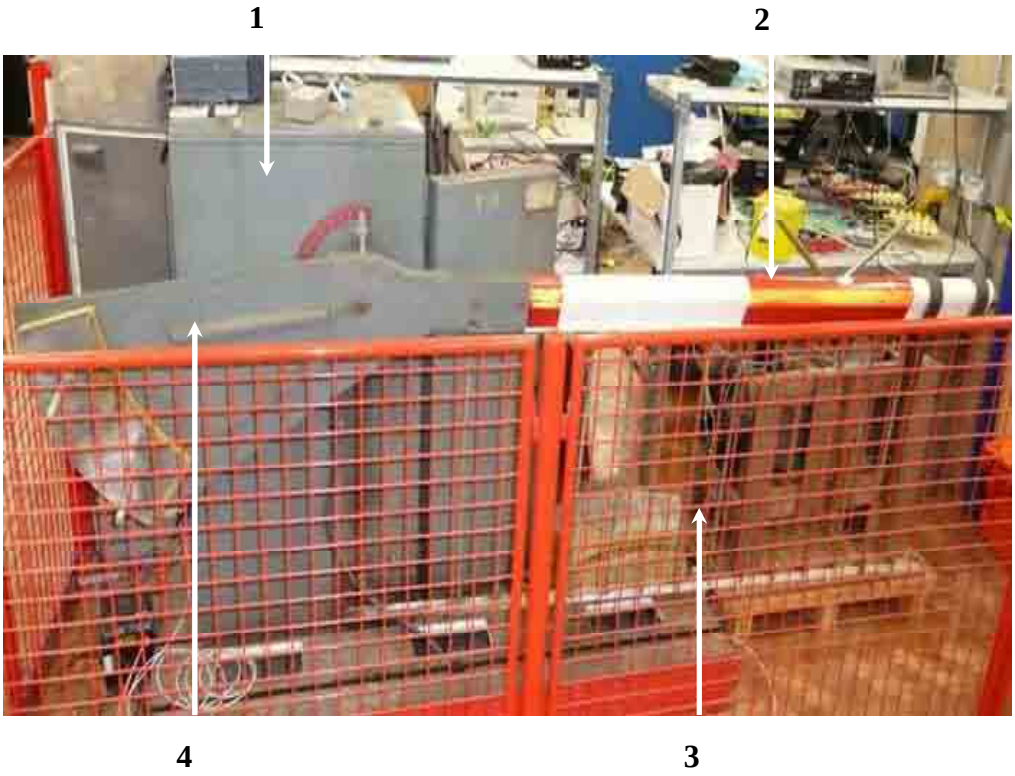


Figure A.5 Illustration of electro-hydraulic level crossing barrier test rig.

Description of components			
1	Barrier machine box	**3**	Skirts
2	Boom	**4**	Boom side arm

Table A.5 Description of components of electric-hydraulic level crossing test rig.

Appendix B

LabVIEW based test rig control and data collection software

As a tool for developing the test rig control and data collection programs, the LabVIEW software was employed to realise various functions via a USB 6008 DAQ. Friendly user interfaces were designed for the pneumatic and electric STMEs respectively, which are shown in Figure B.1 and B.2.

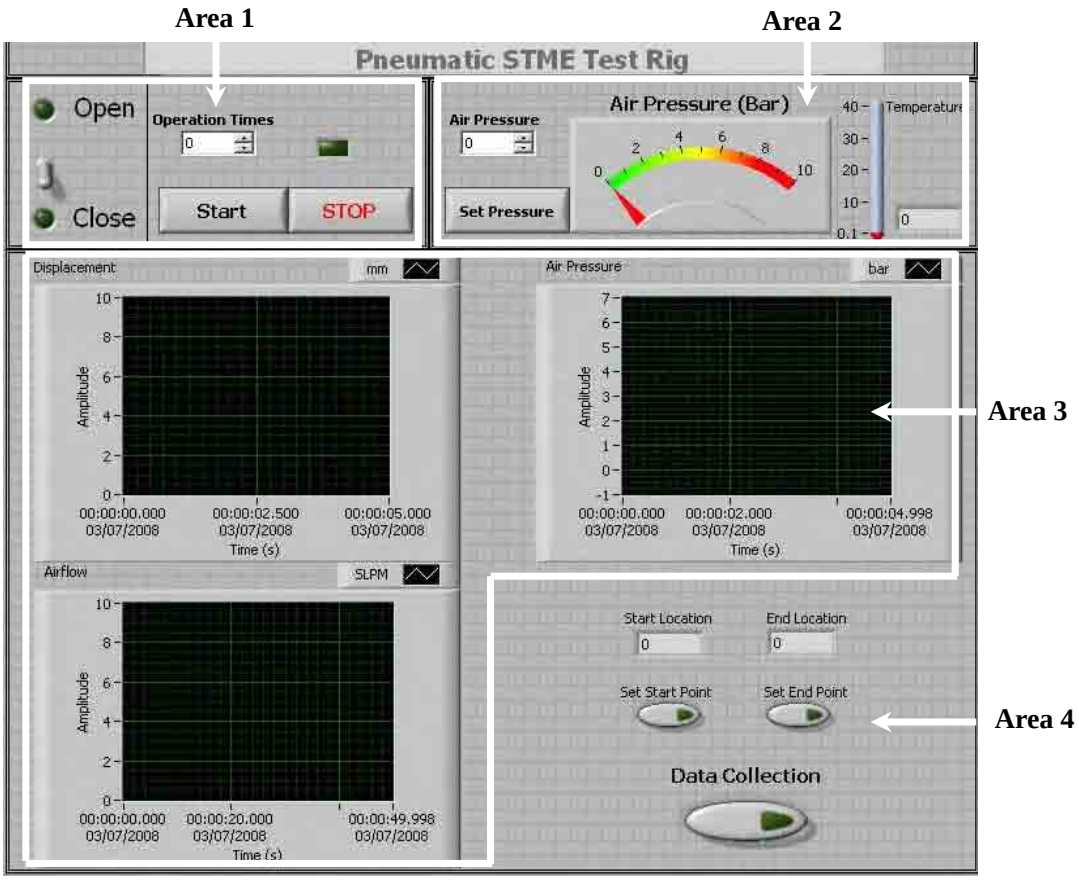


Figure B.1 LabVIEW interface for pneumatic STME test rigs.

In Figure B.1, the interface is divided into 4 areas and the functions in each of them are described as follows:

Area 1

- Indicating operation status: open (normal/forward throw) and close (reverse throw);
- Operation time counter and in-process indicator;
- Operation start/stop control buttons.

Area 2

- Air pressure setting with 0.1 bar resolution;
- Air pressure indicator up to 10 bar;
- Environmental temperature indicator.

Area 3

- Real-time visualisation of collected data including displacement, airflow and air pressure.

Area 4

- Sensor readings at the start and end of a throw for data calibration;
- Data collection button for manually selecting data for storage. The data storage function is on in default unless it is turned off manually. When data is required to be recorded, a new file is automatically named by the current time.

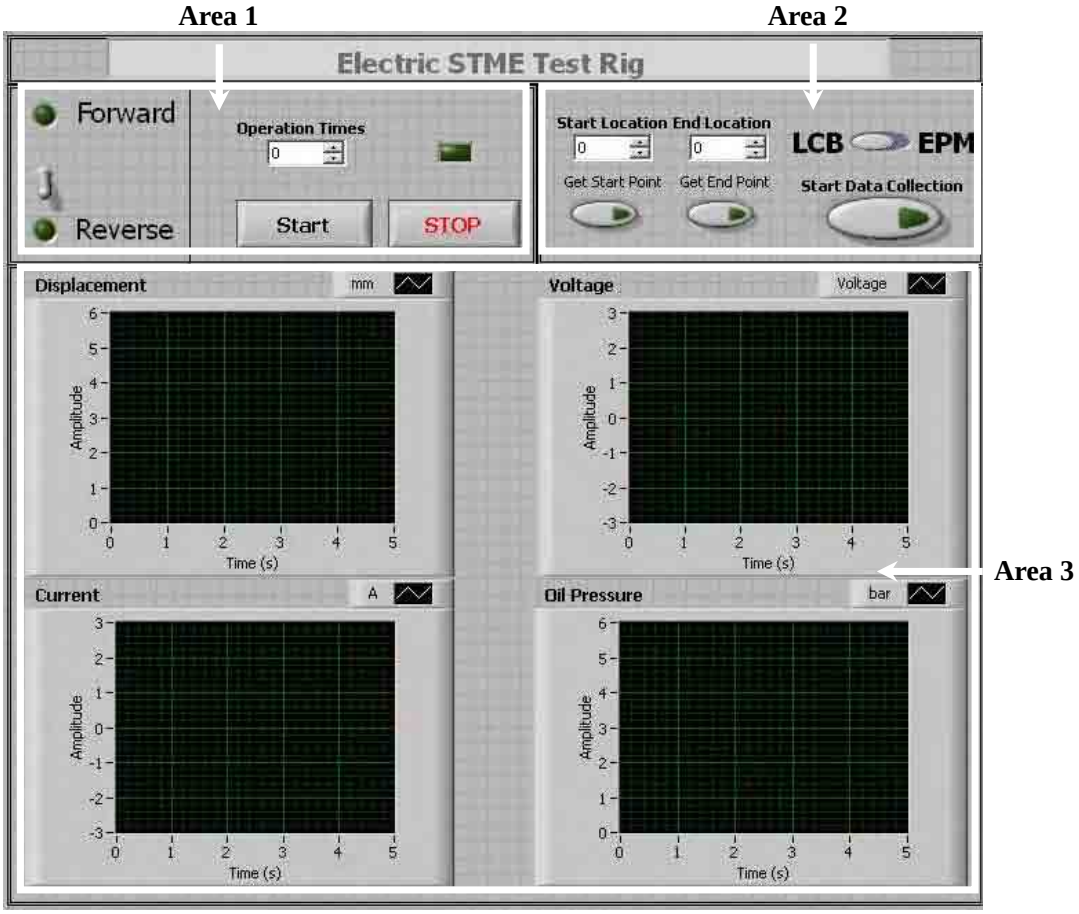


Figure B.2 LabVIEW interface for electric and electro-hydraulic STME test rigs.

In Figure B.2, the functions are distributed in 3 areas. In each area, the functions are as follows:

Area 1

- Same functions as the area 1 introduced in Figure B.1.

Area 2

- Data calibration using sensor readings at the start and end of a throw;
- A switch for selecting the two STMEs: electric point machine (EPM) and electro-hydraulic level crossing (LCB);
- Data collection start button (default status is on).

Area 3

- Area for real time data visualisation including: displacement, voltage, current, oil pressure (in the case of LCB).

Appendix C

Coefficient bank of polynomial and rational models

The coefficients of polynomial and rational models in the case of pneumatic train-stop normal throw are listed in the following table.

3-order polynomial f(x) = p1*x^3 + p2*x^2 + p3*x + p4						
0.001s	**0.01s**	**0.02s**	**0.03s**	**0.04s**	**0.05s**	**0.06s**
-0.0002419	-0.00002615	-4.837E-05	-0.000275	-0.000159	-0.0001762	-0.0003058
0.002606	-0.0002446	-0.0001269	0.002786	0.00136	0.001685	0.003177
-0.006554	0.004855	0.005726	-0.006629	-0.001419	-0.003004	-0.008284
1.02	1.008	1.004	1.021	1.016	1.018	1.023
0.07s	**0.08s**	**0.09s**	**0.1s**	**0.11s**	**0.12s**	**0.13s**
-0.0004601	-0.0002826	-0.001491	-9.647E-05	-0.0001251	0.0001615	-0.00004407
0.004907	0.002746	0.001107	0.0004272	0.001111	-0.002721	-0.0001337
-0.01405	-0.005716	0.0006457	0.003227	-0.001146	0.0154	0.005019
1.029	1.018	1.011	1.008	1.016	0.9934	1.006
0.14s	**0.15s**	**0.16s**	**0.17s**	**0.18s**	**0.19s**	**0.2s**
-7.949E-05	-0.00014	-0.000271	-0.0004145	-0.0002959	-0.0003486	-0.0002036
0.0003604	0.0008696	0.002898	0.004537	0.003009	0.00352	0.001869
0.002953	0.002051	-0.008204	-0.01382	-0.007628	-0.009171	-0.003066
1.009	1.009	1.025	1.031	1.023	1.025	1.017
0.21s	**0.22s**	**0.23s**	**0.24s**	**0.25s**	**0.26s**	**0.27s**
-0.0001223	-0.0002695	0.0000852	0.00005468	-0.0001454	-0.0006188	0.0004741
0.0008242	0.002631	-0.001787	-0.001222	0.002876	0.009462	-0.002444
0.001297	-0.005927	0.01147	0.008823	-0.01162	-0.03602	0.007992
1.011	1.021	0.9996	1.003	1.032	1.058	1.007
0.28s						
0.003708						
-0.0209						
0.0303						
1.014						
3/2 Rational f(x) = (p1*x^3 + p2*x^2 + p3*x + p4) /(x^2 + q1*x + q2)						
0.29s	**0.3s**	**0.31s**	**0.32s**	**0.33s**	**0.34s**	**0.35s**

0.0732	0.07275	0.1047	0.168	0.1602	0.1913	0.198
0.3082	0.3127	0.1053	-0.09733	0.277	0.1807	0.2772
-6.471	-6.498	-5.444	-4.305	-5.396	-4.902	-5.081
17.35	17.42	14.73	10.95	11.53	9.695	9.176
-8.577	-8.589	-7.936	-6.582	-6.053	-5.537	-5.241
19.34	19.39	16.93	12.38	11.13	9.133	8.118
0.36s	**0.37s**	**0.38s**	**0.39s**			
0.1888	0.1826	0.2075	0.2395			
0.5475	0.7318	0.4125	-0.02008			
-5.71	-5.985	-4.724	-2.911			
9.116	8.798	7.451	5.727			
-4.897	-4.621	-4.699	-4.793			
7.13	6.408	6.675	7.262			
4-order polynomial f(x) = p1*x^4 + p2*x^3 + p3*x^2 + p4*x + p5						
0.4s	**0.41s**	**0.42s**	**0.43s**	**0.44s**	**0.45s**	**0.46s**
0.003793	-0.0006964	-0.005361	-0.008356	-0.009435	-0.008585	-0.006102
-0.03562	0.03811	0.1114	0.1536	0.1641	0.143	0.09617
-0.0421	-0.4691	-0.8717	-1.067	-1.073	-0.8958	-0.5754
1.396	2.397	3.28	3.59	3.437	2.846	1.936
-1.339	-2.061	-2.637	-2.673	-2.327	-1.63	-0.6981
0.47s	**0.48s**	**0.49s**	**0.5s**	**0.51s**	**0.52s**	**0.53s**
-0.002176	0.002713	0.006601	0.006139	0.004156	0.006785	0.01125
0.02655	-0.05697	-0.1218	-0.1199	-0.09514	-0.1376	-0.2025
-0.1303	0.3807	0.7628	0.768	0.6424	0.8632	1.176
0.7582	-0.5297	-1.433	-1.415	-1.066	-1.473	-2.033
0.4141	1.567	2.317	2.252	1.891	2.126	2.452
0.54s	**0.55s**	**0.56s**	**0.57s**	**0.58s**		
0.01348	0.01303	0.01137	0.009441	0.007018		
-0.2305	-0.214	-0.1793	-0.1429	-0.1004		
1.275	1.113	0.8574	0.612	0.3453		
-2.088	-1.504	-0.7421	-0.06784	0.6173		
2.377	1.757	1.039	0.4568	-0.09874		
3/3 Rational f(x) = (p1*x^3 + p2*x^2 + p3*x + p4) / (x^3 + q1*x^2 + q2*x + q3)						
0.59s	**0.60s**	**0.61s**	**0.62s**	**0.63s**	**0.64s**	**0.65s**
4.349	4.095	3.888	3.986	3.894	3.84	3.789
-46.16	-40.99	-38.38	-37.99	-35.49	-33.56	-32
149.9	126.6	118.8	112.5	101.1	92.06	85.3
-114.8	-97.89	-100.1	-85.57	-76.18	-68.07	-62.8
-9.325	-9.028	-9.195	-8.685	-8.396	-8.096	-7.882
22.5	22.27	24.7	21.05	20.09	18.89	18.19
2.649	-4.153	-12.71	-5.756	-7.272	-7.397	-8.059
0.66s	**0.67s**	**0.68s**	**0.69s**	**0.70s**	**0.71s**	**0.72s**

3.749	3.727	3.704	3.679	3.663	3.653	3.633
-30.6	-29.63	-28.86	-28.02	-27.44	-26.94	-26.53
79.24	74.94	71.87	68.61	66.37	64.42	63.08
-57.91	-54.46	-52.33	-49.71	-48.11	-46.57	-45.95
-7.663	-7.494	-7.382	-7.241	-7.147	-7.051	-7.012
17.42	16.81	16.5	16.04	15.77	15.45	15.44
-8.357	-8.512	-8.872	-8.877	-9.033	-9.007	-9.408
0.73s	**0.74s**	**0.75s**	**0.76s**	**0.77s**	**0.78s**	**0.79s**
3.624	3.613	3.608	3.59	3.583	3.577	3.572
-25.87	-25.58	-25.11	-24.92	-24.66	-24.44	-24.31
60.42	59.44	57.6	57.12	56.17	55.42	54.97
-43.31	-42.7	-40.87	-40.9	-40.09	-39.4	-39.06
-6.854	-6.818	-6.7	-6.715	-6.666	-6.626	-6.609
14.82	14.75	14.3	14.48	14.32	14.19	14.15
-8.856	-9.018	-8.603	-9.061	-9.002	-8.921	-8.958
0.80s	**0.81s**	**0.82s**	**0.83s**	**0.84s**	**0.85s**	**0.86s**
3.559	3.548	3.537	3.531	3.526	3.522	3.522
-24.4	-24.5	-24.61	-24.52	-24.48	-24.18	-24.06
55.52	56.09	56.71	56.32	56.21	54.9	54.3
-39.97	-40.83	-41.77	-41.44	-41.46	-40.13	-39.5
-6.682	-6.754	-6.829	-6.824	-6.832	-6.761	-6.729
14.52	14.87	15.23	15.21	15.26	14.95	14.8
-9.514	-10.01	-10.53	-10.57	-10.7	-10.42	-10.28
0.87s	**0.88s**	**0.89s**	**0.90s**	**0.91s**	**0.92s**	**0.93s**
3.517	3.51	3.509	3.506	3.501	3.501	3.498
-23.91	-23.57	-23.36	-22.97	-22.76	-22.58	-22.19
53.73	52.39	51.49	49.88	49.09	48.35	46.75
-39.07	-37.81	-36.88	-35.28	-34.58	-33.83	-32.18
-6.706	-6.631	-6.574	-6.474	-6.431	-6.382	-6.276
14.73	14.43	14.19	13.78	13.62	13.42	12.98
-10.28	-10.04	-9.796	-9.408	-9.295	-9.101	-8.649
0.94s	**0.95s**	**0.96s**	**0.97s**	**0.98s**	**0.99s**	**1s**
3.495	3.494	3.49	3.486	3.485	3.483	3.482
-22.03	-21.85	-22.04	-22.12	-21.87	-22.18	-22.29
46.15	45.42	46.29	46.69	45.66	46.99	47.48
-31.65	-30.92	-31.95	-32.46	-31.39	-32.82	-33.37
-6.239	-6.193	-6.262	-6.298	-6.229	-6.325	-6.364
12.85	12.66	12.97	13.14	12.85	13.26	13.43
-8.547	-8.366	-8.746	-8.949	-8.655	-9.103	-9.292
1.01s	**1.02s**	**1.03s**	**1.04s**	**1.05s**	**1.06s**	**1.07s**
3.477	3.478	3.476	3.475	3.474	3.476	3.473
-22.39	-22.46	-22.55	-22.48	-22.52	-22.49	-22.3

47.97	48.25	48.64	48.36	48.56	48.37	47.62
-33.96	-34.25	-34.69	-34.4	-34.64	-34.41	-33.67
-6.407	-6.426	-6.457	-6.438	-6.457	-6.44	-6.393
13.62	13.7	13.83	13.76	13.84	13.76	13.57
-9.521	-9.605	-9.757	-9.679	-9.777	-9.689	-9.501
1.08s	**1.09s**	**1.10s**				
3.474	3.473	3.472				
-22.29	-22.19	-22.08				
47.57	47.15	46.7				
-33.6	-33.18	-32.73				
-6.389	-6.363	-6.334				
13.55	13.44	13.32				
-9.477	-9.372	-9.253				

Table C.1 Coefficient bank of polynomial model of the train-stop normal throw (4.1 bar).

Appendix D

Programme flowcharts for pneumatic STMEs

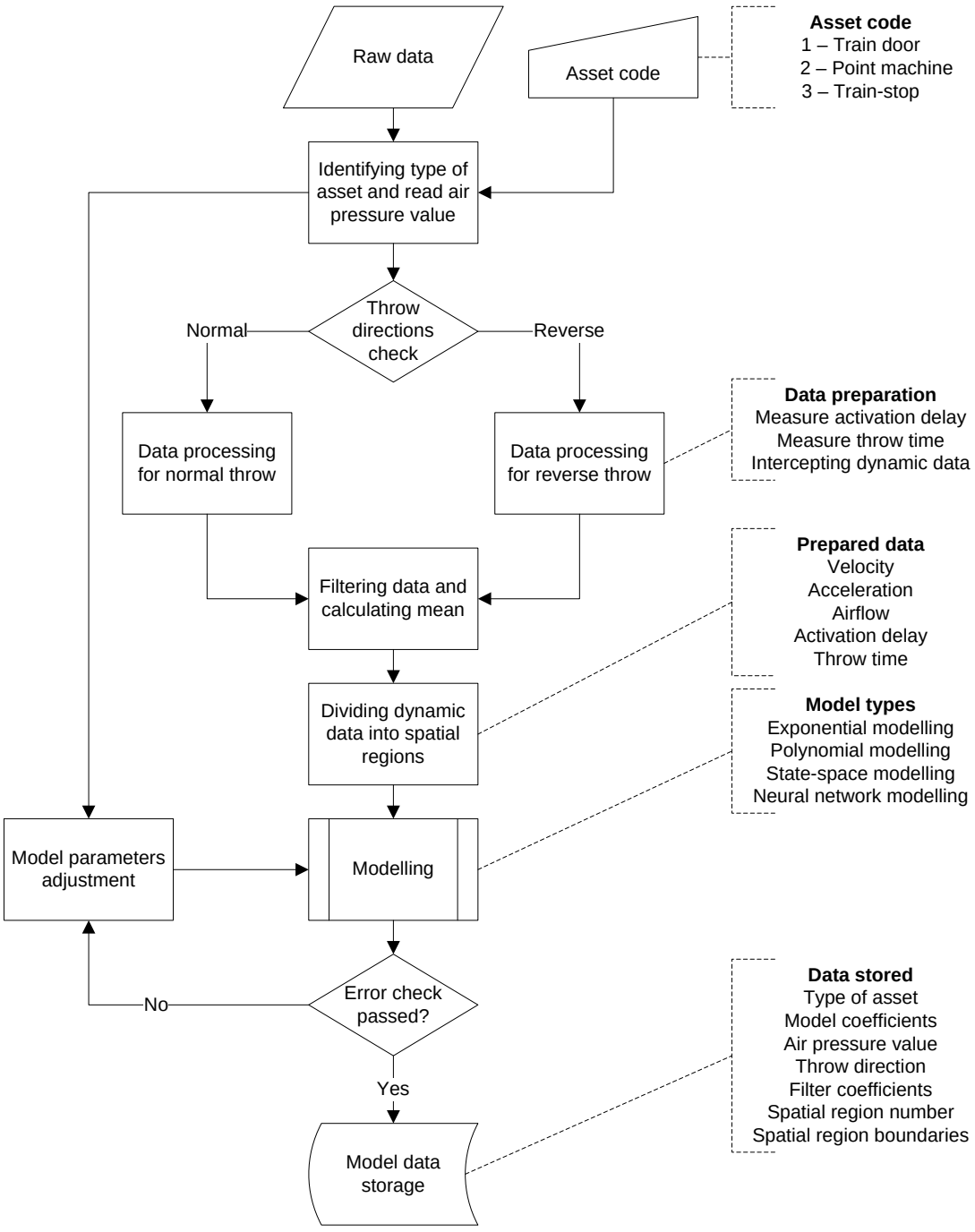


Figure D.1 Flowchart of modeling process for pneumatic STMEs.

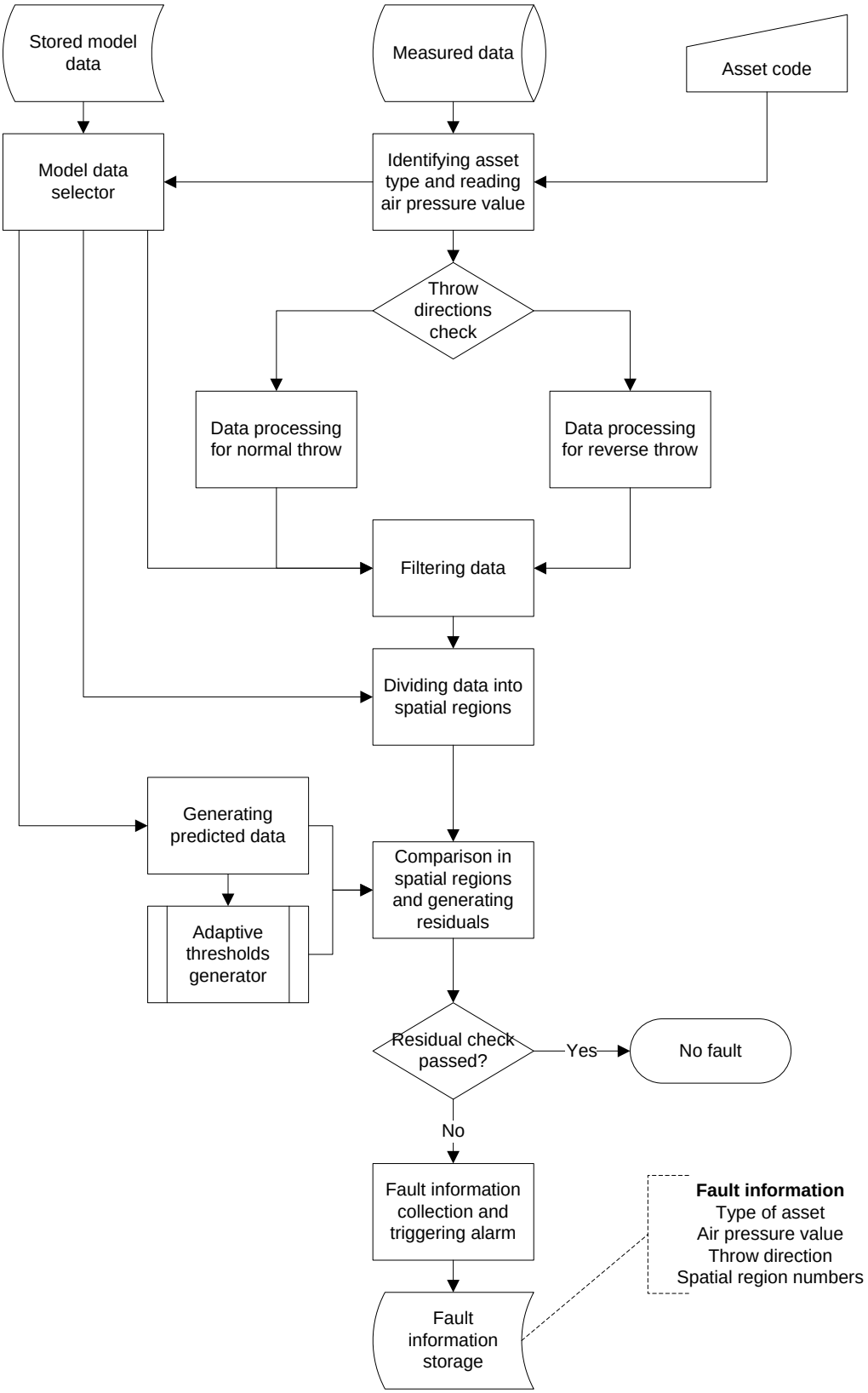


Figure D.2 Flowchart of fault detection process for pneumatic STMEs.

Appendix E

Publication

Bai, H., Roberts, C. and Goodman, C., (2008). A generic fault detection and diagnosis approach for railway assets. *IET International Conference on Railway Engineering 2008, ICRE 2008, IET Seminar Digest,* vol. 2008 (1), pp. 164-171.

References

[1] Alexandru, M., (2004). Neuro-fuzzy diagnosis in final control elements of AC motors. *Proceedings of the 2004 American Control Conference*, vol. 4, pp. 3759-63.

[2] Allan, J., (1993). An Overview of Railway Signalling and Control Systems, *IEE Vacation School on Electric Traction.*

[3] Ayoubi, M. and Isermann, R., (1997). Neuro-fuzzy systems for diagnosis, *Fuzzy Sets and Systems*, vol. 89, pp. 289-307.

[4] Ballè, P., and Fuessel, D., (2000). Closed-loop fault diagnosis based on a non-linear process model and automatic fuzzy rule generation. *Engineering application of artificial intelligence*, vol. 13, pp. 695-707.

[5] Barber, M. J. (1986). *Handbook of power cylinders, valves and controls*. Morden: Trade and Technical. ISBN: 0854611002

[6] Basila, M. R., Stefanek, Jr. G. and Cinar, A., (1990). A model – object based supervisory expert system for fault tolerant chemical reactor control. *Computers & Chemical Engineering*, vol. 14 (4-5), pp. 551-560.

[7] Becker, E. and Poste, P., (2006). Keeping the blades turning: Condition monitoring of wind turbine gears. *Refocus*, vol. 7(2), pp. 26-32.

[8] Bishop, C. M. (2003). *Pattern Recognition and Machine Learning*. Springer. ISBN 0-38-731073-8.

[9] Brinkmann, U. and Spalmann, F.H., (1996). The 'Level Crossing Online' PC-based remote diagnosis system for Pintsch type level crossing signal systems. *Signal und Draht*, vol. 88 (3), pp. 20-4.

[10] Bunday, B. D., (1984). *Basic optimization methods*. Edward Arnold, London, ISBN: 0713135069.

[11] Chen, J. and Patton, R., (1999). *Robust model-based fault diagnosis for dynamic systems*. Kluwer Academic Publishers. ISBN: 0792382536.

[12] Chen, J. Q., Xi, Y. G., and Zhang, Z. J., (1996). A clustering algorithm for fuzzy model identification. *Fuzzy Sets and Systems*, vol. 98, pp. 319-329.

[13] Chow, E. Y., and Willsky, A. S., (1984). Analytical redundancy and the design of robust failure detection systems. *IEEE Transactions on Automatic Control*, vol. 29 (7), pp. 603-614.

[14] Collacott, R. A., (1977). *Mechanical fault diagnosis and condition monitoring*. Halsted Press, USA. ISBN: 0470990953

[15] Combastel, C., Gentil, S. and Rognon, J. P., (1998). A symbolic reasoning approach to fault detection and isolation applied to electrical machines, *Control Application*, vol. 1, pp. 475-479.

[16] Crowder, M. J. and Hand, D. J., (1990). *Analysis of Repeated Measures*. Chapman and Hall, London. ISBN: 041231830X.

[17] Cybenko, G., (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303-314.

[18] Dash, S. and Venkatasubramanian, V., (2000). Challenges in the industrial application of fault diagnostic systems. *Computers and Chemical Engineering*, vol. 24(2-7), pp. 785-791.

[19] Dassanayake, H., (2001). *Fault diagnosis for a new generation of intelligent train door systems*. PhD thesis, University of Birmingham, UK.

[20] Dassanayake, H., Roberts, C. and Goodman, C., (2001). An architecture for system wide fault detection and isolation. *Proceedings of the Institution of Mechanical Engineers, Part I*: *Journal of Systems and Control*, vol. 215 (1), pp. 37-46.

[21] Daugherty, R. L., Franzini, J. B. and Finnemore, E. J., (1985). *Fluid Mechanics with Engineering Applications*. New York, McGraw-Hill.

[22] Ding, X. and Frank, P. M., (1991). Frequency domain approach and threshold selector for robust model-based fault detection and isolation. *Proceedings of the IFAC symposium SAFEPROCESS'91*, Baden-Baden, Germany, pp. 307-312.

[23] Ding, X., Guo, L. and Jeinsch, T., (1999). A characterization of parity space and

its application to robust fault detection, *IEEE Transactions on Automatic Control*, vol. 44 (2), pp. 337-343.

[24] Dougherty, E. R. (1990). *Probability and statistics for the engineering, computing and physical sciences*. Englewood Cliff, N.J.: Prentice Hall. ISBN: 013711995X

[25] Emami-Naeini, A., Akhter, M. M. and Rock, S. M., (1988). Effect of model uncertainty on failure detection: the threshold selector. *IEEE Transactions on Automatic Control*, vol. 33(12), pp. 1106-1115.

[26] Everitt, B. S., (1998). *The Cambridge dictionary of statistics*. Cambridge University Press, 1st edition. ISBN: 0521593468.

[27] Evsukoff, A., Gentil, S. and Montmain, J., (2000). Fuzzy reasoning in co-operation supervision systems. *Control Engineering Practice*, vol. 8, pp. 389-407.

[28] Förstner, D. and Lunze, J., (2000). Qualitative model-based fault detection of a fuel injection system. *Proceedings of the 4th IFAC symposium on Fault Detection, Supervision and Safety for Technical Processes – SAFPROCESS 2000*, Budapest, Hungary, pp. 28-39.

[29] Frank, P. M. and Ding, X., (1996). Survey of robust residual generation and evaluation methods in observer-based fault detection systems. J*ournal of Process Control*, vol. 7 (6), pp. 403-424.

[30] Frank, P. M., (1996). Analytical and qualitative model-based fault diagnosis – A survey and new results. *European Journal of Control*, vol. 2 (1), pp. 6-28.

[31] Fullér, R., (2000). *Introduction to Neuro-Fuzzy Systems*. Physica-Verlag Heidelberg, New York. ISBN: 3790812560.

[32] Garcia Marquez F. P., Lewis W. R., Tobias M. A. and Roberts C., (2008). Life cycle cost for railway condition monitoring. *Transportation Research Part E*, vol. 44 (6), pp. 1175–1187.

[33] Garcia Marquez F. P., Weston P. and Roberts C., (2007). Failure analysis and diagnostics for railway trackside equipment. *Engineering Failure Analysis*, vol. 14(8), pp. 1411-1426.

[34] Garcia Marquez, F. P. and Pedregal, D. J., (2007). Applied RCM^2 algorithms based on statistical methods. *International Journal of Automation and Computing*, vol. 4, pp. 109-116.

[35] Garcia Marquez, F. P., (2006). An approach to remote condition monitoring systems management. *The IET International Conference on Railway Condition Monitoring*, pp. 156-160.

[36] Garcia Marquez, F. P., Pedregal, D. and Roberts, C., (2007). New methods for the condition monitoring of level crossings. *Proceedings of the 5th International Conference on Quality and Reliability*, Chiang Mai, Thailand.

[37] Garcia Marquez, F. P., Pedregal, D. J. and Schmid, F., (2007). Unobserved component models applied to the assessment of ware in railway points: a case study. *European Journal of Operational Research*, vol. 176, pp. 1702-1703.

[38] Gertler, J. J., (1998). *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker Inc., New York. ISBN: 0-8247-9427-3.

[39] Gomm, J. B., Weerasinghe, M. and Williams, D., (2000). Diagnosis of process faults with neural networks and principal component analysis. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical*, vol. 214 (2), pp.131-143.

[40] Gonzalez, T., Sahni, S. and Franta, W. R., (1977). An efficient algorithm for the Kolmogorov-Smirnov and Lilliefors tests. *ACM Transactions on Mathematical Software*, vol. 3 (1), pp. 60-4.

[41] Gurney, K., (1997). *An Introduction to Neural Networks*. University College London (Ucl) Press, London. ISBN 185728673.

[42] Himmelblau, D. M., (1978). Fault detection and Diagnosis in Chemical and Petrochemical Processes. *Chemical Engineering Monograph*, 8, Elsevier Science.

[43] Hunt, K. J., (1993). *Polynomial methods in optimal control and filtering*, IEE control engineering series, London. ISBN: 0863412955.

[44] International Electrotechnical Commission (1993). IEC 1158 - *Fieldbus standard for use in industrial control*. International Electrotechnical Commission (IEC).

[45] Isermann, R., (1984). Process fault detection based on modelling and estimation methods – A survey. *Automatica*, vol. 20 (4), pp. 387-404.

[46] Isermann, R., (1994). On the applicability of model-based fault detection for technical processes. *Control Engineering Practice*, vol. 2 (3), pp. 439-450.

[47] Isermann, R., (1997). Supervision, fault detection and fault-diagnosis methods-an introduction. *Control Engineering Practice*, vol. 5 (5), pp. 639-652.

[48] Isermann, R., (2005). Model-based fault-detection and diagnosis – status and applications. *Annual Reviews in Control*, vol. 29 (1), pp. 71-85.

[49] Ishak, S. Z., Yue, W. L. and Somenahalli, S. V. C., (2008). The methodology development of railway level crossing safety systems – South Australia case study. *WIT Transactions on the Built Environment*, vol. 103, pp. 629-637.

[50] Jang, J. S. R., Sun, C. T. and Mizutani, E., (1997). *Neuro-fuzzy and soft computing – a computational approach to learning and machine intelligence*, Prentice Hall. ISBN: 0132610663.

[51] Klir, G. J. and Yuan, B., (1996). *Fuzzy sets, fuzzy logic, and fuzzy systems: Selected papers by Lotfi A. Zadeh*, Advances in Fuzzy Systems – Applications and Theory, vol. 6, World Scientific. ISBN: 9810224214.

[52] Kolla S., Border, D. and Mayer E., (2003). Fieldbus networks for control system implementations. *Proc. Electrical Insulation Conf. & Electrical Manufacturing & Coil Winding Technology Conference*, pp. 493-498.

[53] Korbicz J., Koscielny J. M., Kowalczuk Z. and Cholewa W., (2004). *Fault Diagnosis: Models, Artificial Intelligence, Allocations*. Published by Springer. ISBN: 3540407677.

[54] Kosebalaban, F. and Cinar, A., (2001). Integration of multivariate SPM and FDD by parity space technique for a food pasteurization process, *Computer and Chemical Engineering*, vol. 25, pp. 473-491.

[55] Lee, J. H., Lee, M. R., Kim, J. T., Luk, V. and Jung Y-H., (2006). A study of the characteristics of the acoustic emission signals for condition monitoring of check

valves in nuclear power plants. *Nuclear Engineering and Design*, vol. 236(13), pp. 1411- 1421.

[56] Lehrasab, N. and Fararooy, S., (1998). Formal definition of single throw mechanical equipment for fault diagnosis. *IEE Electronics Lett.*, vol. 34(23), pp. 2231-2232.

[57] Lehrasab, N., (1999). *A generic fault detection and isolation approach for single throw mechanical equipment.* PhD thesis, University of Birmingham, UK.

[58] Lehrasab, N., Dassanayake, H., Roberts, C., Fararooy, S. and Goodman, C., (2002). Industrial fault diagnosis: pneumatic train door case study. *Proceedings of the Institution of Mechanical Engineers*, *Part F*: *Journal of Rail and Rapid Transit*, vol. 216(3), pp. 175-183.

[59] Li, P. and Googall, R.M., (2003). Model Based Approach to Railway Vehicle Fault Detection and Isolation, *Electronic Systems and Control Division Research*, pp. 7-10.

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

[61] Lilliefors, H., (1967). On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, vol. 62, pp. 399–402.

[62] Ljung, L., (1987). *System identification: Theory for the user*, P T R Prentice Hall. ISBN: 0138816409.

[63] Mamdani, E. H. and Assilian, S., (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man Machine Studies*, vol. 7, pp. 1-13.

[64] Mandel, J., (1964). *The Statistical Analysis of Experimental Data*. Interscience, New York. ISBN: 0486646661.

[65] Marcos, A., Ganguli, S. and Balas, G. J., (2004). Application of ∞ *H* fault diagnosis and isolation to a transport aircraft. *Control Engineering Practice*, vol. 13, pp. 105-119.

[66] Miller R. W. (1996). *Flow Measurement Engineering Handbook (Third Edition ed.)*. McGraw Hill. ISBN 0070423660.

[67] Narendra, K. S. and Parthasarathy, K., (1990). Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, vol. 1 (1), 4-27.

[68] Narendra, K. S., (1996). Neural networks for control: Theory and practice. *Proceedings of the IEEE*, vol. 84 (10), pp. 1385-1406.

[69] Nash, M. and Roberts, C., (1999). Condition monitoring of level crossings – event recorders to $(RCM)^2$. *In Proceeding of the IRSE ASPECT '99 Conference*, London, UK.

[70] Network Rail (2008). *Delivering for you*. Network Rail Infrastructure Limited Annual Report and Accounts 2008.

[71] Noel, J., (2002). Digital bus selection – It's not your typical quick pick at 7-11. *Proceedings of 2002 ISA EXPO*, Chicage, IL.

[72] Oyebande, B. O. and Renfrew, A. C., (2002). Condition monitoring of railway electric point machines. *Electr Power Appl, IEE Proc*, vol. 149(6), pp. 465–73.

[73] Patton R. J., Lopez-toribio C. J. and Uppal F. J., (1999). Artificial intelligence approaches to fault diagnosis for dynamic systems. *Int. J. Appl. Math. and Comp. Sci.*, vol.9 (3), pp. 471-518.

[74] Patton, R. J. and Chen, J., (1994). A review of parity space approaches to fault diagnosis for aerospace systems. *Journal of Guidance Control and Dynamics*, vol. 17, pp. 278-285.

[75] Patton, R. J. and Chen, J., (1997). Observer-based fault detection and isolation: Robustness and application. *Control Engineering Practice*, vol. 5, pp. 671-682.

[76] Patton, R. J., (2000). A case for quantitative modelling for fault diagnosis. *Vacation school on qualitative methods for fault diagnosis*, Automatic Control

Department, Technical University of Catalonia, Terrassa (Spain).

[77] Patton, R. J., Frank, P. and Clark, R., (1989). *Fault diagnosis in dynamic systems: Theory and applications*. Hall International, UK. ISBN: 0133082636.

[78] Patzke, R., (1998). Fieldbus basics. *Computer Standards & Interfaces*, vol. 19, pp. 275-293.

[79] Pratt, W., (2003). Evaluating fieldbus networks: Choose the right tool for the job. Retrieved from http://www.hartcomm.org.

[80] Redeker, R., (2006). Condition monitoring for railway infrastructure objects. *In proceedings of 19th International Congress on Condition Monitoring and References* 18 6 *Diagnostic Engineering Management* (*COMADEM 2006*), Luleå, Sweden, pp. 689- 699.

[81] Richer E. and Urmuzlu Y., (2000). A high performance pneumatic force actuator system part 1-nonlinear mathematical model. *ASME J. Dyn. Syst. Meas. Control*, vol. 122 (3), pp. 416-425.

[82] Roberts C., Dassanayake H. P. and Goodman C. J., (2001). Diagnosis of low cost industrial assets using distributed models and data networks. In *Proceedings of the 10th International Conference on System Modelling and Control*, Lodz, Poland.

[83] Roberts, C. and Fararooy, S., (1998). Remote condition monitoring into the next millennium. *In Proceedings of the 6th International Conference on Computers Aided Design, Manufacture and Operations in the Railway and other Mass Transit Systems* (COMPRAIL '98), Lisbon, Portugal, pp. 715-722.

[84] Roberts, C., (2007). *Methods for fault detection and diagnosis of railway actuators*. PhD thesis, University of Birmingham, UK.

[85] Roberts, C., Dassanayake, H., Lehrasab, N. and Goodman, C., (2002). Distributed quantitative and qualitative fault diagnosis: railway junction case study. *Control Engineering Practice*, vol. 10, pp. 419-429.

[86] Roberts, C., Yazdi, H. R. and Fararooy, S., (1999). An intelligent fault detection system using fieldbus technology. *Condition Monitor*, vol. 155, pp. 5-8.

[87] Ruiz, D., Nougués, J. M. and Puigjaner, L., (2001). Fault diagnosis support system for complex chemical plants. *Computers & Chemical Engineering*, vol. 25 (1), pp. 151-160.

[88] Schneider, H. and Fank, P. M., (1994). Fuzzy logic based threshold adaption for fault detection in robots. *Proc. 3rd IEEE Conf. Control Applications*, Glasgow, Scotland, pp. 1127-1132.

[89] Shi, Z., Gu, F., Lennox, B. and Ball, A. D., (2005). The development of an adaptive threshold for model-based fault detection of a nonlinear electro-hydraulic system. *Control Engineering Practice*, vol. 13, pp. 1357-1367.

[90] Silmon J. and Roberts C., (2006). A systems approach to fault detection and diagnosis for condition-based maintenance. *1st IET International Conference on Railway Condition Monitoring*, Birmingham, UK.

[91] Suthasinekul, S., Chang, J. K., Dwyer, S. J., Fu, Y. S. and Feik, R. L., (1976). Level crossing pattern of time series: applications to signal processing. *3rd international Joint Conference on Pattern Recognition*, pp. 524-8, USA.

[92] Tanaka, K., (1996). *An Introduction to Fuzzy Logic for Practical Applications*. Springer. ISBN: 0387948074.

[93] Terano, T., Asai, K. and Sugeno M., (1991). *Fuzzy Systems Theory and Its Applications*. Academic Press, INC. ISBN: 0126852456.

[94] Thompson, L. M., (1997). *Industrial data communications, 4th edition*. *ISA*. USA. ISBN: 978-1-934394-24-3

[95] Tian, G., Zhao, Z. and Baines, R., (2000). A Fieldbus-based intelligent sensor. *Mechatronics*, vol. 10 (8), pp. 835-849.

[96] Vaclavek, V., (1974). Gross systematic errors or biases in the balance calculations. *Papers of the Prague Institute of Technology*, Prague, Czechoslovakia.

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

[98] Venkatasubramanian, V., Rengaswamy, R., Kavuri S. and Yin K., (2003b). A review of process fault detection and diagnosis Part II: Qualitative models and search strategies. *Computers and Chemical Engineering*, vol. 27 (3), pp. 313-326.

[99] Venkatasubramanian, V., Rengaswamy, R., Kavuri S. and Yin K., (2003c). A review of process fault detection and diagnosis Part III: Process history based models. *Computers and Chemical Engineering*, vol. 27(3), pp. 327-346.

[100] Warring, R. H., (1969). *Pneumatic engineering calculations*. Morden, Surrey: Trade and Technical Press. ISBN: 0854610383

[101] Weerasinghe, M., Gomm, J. and Williams, D., (1998). Neural networks for fault diagnosis of a nuclear fuel processing plant at different operating points. *Control Engineering Practice*, vol. 6 (2), pp. 281-289.

[102] Winston, P. H., (1992). *Artificial intelligence*. Addison-Wesley, Reading, MA, 3rd edition. ISBN: 0-201-53377-4.

[103] Wolberg, J. R., (1967). *Prediction analysis*, D. Van Nostrand Company, Ltd.

[104] Yazdi, H., Roberts C. and Fararooy S., (1998). Intelligent condition monitoring of railway signalling equipment using simulation. *IEE Seminar - Condition Monitoring for Railway Transport*, Birmingham.

[105] Yu, D. and Shields, D. N., (1996). A fault isolation method based on parity equations with application to a lathe-spindle system. *International Conference on Control*, vol. 427, pp. 317-322.

[106] Zadeh, L. A., (1965). Fuzzy Sets. *Information and Control*, vol. 8, pp. 338-353.

[107] Zhou, F., Archer, N., Bowles, J., Duta, M., Henry, M. P., Tombs, M. S., Zamora, M. E., Baker, S. and Burton, C. (2002). Remote condition monitoring and validation of railway points. *IEE Computing & Control Engineering Journal*, pp. 221-230.