Fault detection and diagnosis in HVAC systems using analytical models

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: https://dspace.lboro.ac.uk/2134/7337

Publisher: © Timothy Ian Salsbury

Please cite the published version.
This item is held in Loughborough University’s Institutional Repository (https://dspace.lboro.ac.uk/) and was harvested from the British Library’s EThOS service (http://www.ethos.bl.uk/). It is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
Fault detection and diagnosis in HVAC systems using analytical models

Timothy Ian Salsbury
Department of Civil and Building Engineering
Loughborough University

A Doctoral Thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University

August 1996

© Timothy Ian Salsbury 1996
Abstract

Faults that develop in the heat exchanger subsystems in air-conditioning installations can lead to increased energy costs and jeopardise thermal comfort. The sensor and control signals associated with these systems contain potentially valuable information about the condition of the system, and energy management and control systems are able to monitor and store these signals. In practice, the only checks made are to verify set-points are being maintained and that certain critical variables remain within predetermined limits. This approach may allow the detection of certain abrupt or catastrophic faults, but degradation faults often remain undetected until their effects become quite severe.

This thesis investigates the appropriateness of using mathematical models to track the development of degradation faults. An approach is developed, which is based on the use of analytical models in conjunction with a recursive parameter estimation algorithm. A subset of the parameters of the models, which are closely related to faults, is estimated recursively. Significant deviations in the values of the estimated parameters from nominal values, which represent 'correct operation', are used as an indication that the system has developed a fault. The extent of the deviation from the nominal values is used as an estimate of the degree of fault. This thesis develops the theory and examines the robustness of the parameter estimator using simulation-based testing. Results are also presented from testing the fault detection and diagnosis scheme with data obtained from a simulated air-conditioning system and from a full size test installation.

Keywords: air-conditioning, heat exchanger, physical modelling, fault detection and diagnosis, condition monitoring, non-linear recursive parameter estimation, analytical redundancy.
Acknowledgements

I would like to thank my supervisor Dr P. Haves for his enthusiasm, guidance, and critical discussion during the work leading to this thesis. I would also like to thank my second supervisor Dr J. A. Wright for his help and for providing practical insight into certain aspects of the work. I gratefully acknowledge the financial support of the UK Building Research Establishment for making this work possible. In addition, I acknowledge Professor V. I. Hanby who provided support during the writing period of this thesis and who also broadened my research interests into the area of pulse combustion.

I would like to thank Dr C. F. Scott for his help with mathematical matters, and my other colleagues in the Department of Civil and Building Engineering. Thanks are also due to members, past and present, of the Engineering Science Department at the University of Oxford, in particular, Dr A. L. Dexter and Dr S. J. Hepworth. The work reported in this thesis was conducted in parallel with the International Energy Agency Annex 25 project, and I wish to acknowledge the role played by members of this project in providing intellectual stimulation, which proved useful for my work. Credit is also due to members of the Building Technology Group at MIT, in particular Dr L. Norford who helped make my stay at MIT possible. Thanks also to Norman Hickman for proof-reading the thesis. Finally, I am indebted to my family for providing support and encouragement throughout all my student years.
## Contents

1 Introduction 1
   1.1 FDD in air-conditioning systems ............................. 1
      1.1.1 Characteristics of the problem ............................. 4
   1.2 Objectives of the thesis ..................................... 7

2 Literature review 8
   2.1 Background .............................................. 9
      2.1.1 Preprocessor and classifier components .................... 13
   2.2 Model-based methods ....................................... 15
      2.2.1 The process model ....................................... 16
      2.2.2 Output-innovations methods ................................ 24
      2.2.3 Parameter and state innovations methods .................... 26
   2.3 Other software-oriented FDD methods ......................... 32
      2.3.1 Signal analysis ......................................... 32
      2.3.2 Characteristic quantities ................................ 34
      2.3.3 Neural networks ......................................... 36
      2.3.4 Knowledge-based systems .................................. 40
   2.4 Conclusions ................................................ 42

3 Design of the FDD scheme 44
   3.1 Overview of air-conditioning systems .......................... 44
      3.1.1 Heat exchanger subsystems ................................ 48
   3.2 Categories of faults in HVAC systems .......................... 49
      3.2.1 Faults in heat exchanger subsystems ......................... 51
   3.3 Design criteria ............................................. 54
      3.3.1 Types of faults ......................................... 54
      3.3.2 Performance indices ...................................... 55
      3.3.3 Detection speed versus robustness .......................... 56
      3.3.4 Detail of diagnoses ..................................... 56
   3.4 Design solution .............................................. 58
      3.4.1 Reference of correct operation ............................. 58
      3.4.2 Modelling detail ......................................... 60
      3.4.3 Static versus dynamic ..................................... 61
      3.4.4 Fault detection method ................................... 63
      3.4.5 Fault diagnosis method .................................... 67
3.5 Proposed FDD scheme design .......................... 69

4 Subsystem models ........................................ 71
4.1 Components of heat exchanger subsystems ............. 71
  4.1.1 Designing the models for fault diagnosis ......... 73
4.2 Component models ...................................... 76
  4.2.1 Actuator model .................................... 76
  4.2.2 Valve model ..................................... 78
  4.2.3 Heat exchanger model ............................ 84
4.3 Selection of ‘fault parameters’ ......................... 91
  4.3.1 Valve model .................................... 91
  4.3.2 Heat exchangers .................................. 92
  4.3.3 Temperature sensors ............................. 93
4.4 Results of testing the models with experimental data ... 93
  4.4.1 Valve model .................................... 93
  4.4.2 Complete subsystem ............................. 95
4.5 Summary of the chapter ................................ 96

5 Signal processing and parameter estimation ............... 98
5.1 Development of a steady-state detector ................ 99
  5.1.1 Treatment of measurement noise .................. 102
  5.1.2 Accounting for activity in the input signals .... 105
5.2 Development of a non-linear recursive estimator ....... 107
  5.2.1 Preliminaries: Derivation of a batch estimator ... 108
  5.2.2 Recursive derivation ................................ 111
  5.2.3 Incorporation of forgetting ......................... 115
  5.2.4 Size of data record (memory vector) ............. 116
  5.2.5 Robustness implications of forgetting ............. 117
  5.2.6 Initialisation of P matrix.......................... 120
  5.2.7 Calculation of derivatives ......................... 121
  5.2.8 Calculation of confidence intervals ................ 122
5.3 Summary of the chapter ................................ 123

6 Analysis of estimator using simulation ................. 125
6.1 Test conditions ......................................... 126
6.2 Analytical techniques .................................. 126
6.3 Robustness to different fault types .................... 130
  6.3.1 Heating coil subsystem .......................... 133
  6.3.2 Cooling coil subsystem .......................... 137
6.4 Variability of input signals ........................... 139
  6.4.1 Sinusoidal variations in valve control signal ..... 141
6.5 Measurement noise ..................................... 145
  6.5.1 Quantification of noise effects in real systems .. 149
6.6 Specification errors .................................... 152
  6.6.1 Unmeasured disturbances ......................... 152
6.6.2 Modelling errors ........................................... 155
6.7 Conclusions .................................................... 156

7 Evaluation of FDD scheme 158
7.1 Practical issues .................................................. 158
7.2 Model calibration method ....................................... 161
  7.2.1 Calibration parameters .................................... 162
7.3 Evaluation using system simulation .......................... 163
  7.3.1 Incorporation of faults ................................... 164
  7.3.2 Generation of fault data .................................. 164
  7.3.3 Excitation conditions .................................. 165
  7.3.4 Model calibration data ................................... 166
  7.3.5 Results of model calibration ............................ 167
  7.3.6 Results of FDD tests ................................... 169
  7.3.7 Conclusions ............................................. 176
7.4 Evaluation using a real system .............................. 179
  7.4.1 Description of the ACE Facility ......................... 179
  7.4.2 Incorporation of faults .................................. 181
  7.4.3 Excitation conditions .................................. 182
  7.4.4 Model calibration data ................................... 183
  7.4.5 Problems associated with the BRE system ............... 183
  7.4.6 Results of model calibration ............................ 186
  7.4.7 Results of FDD tests ................................... 188
  7.4.8 Conclusions ............................................. 196

8 Conclusions and further work 198
8.1 Conclusions .................................................... 198
8.2 Suggestions for further work ................................ 202

Nomenclature 205

References 207

A The Complex method 218
List of Figures

1.1 An example of an FDD scheme ........................................... 3
2.1 Preprocessor and classifier components of an FDD scheme .......... 14
2.2 State observer ............................................................... 22
2.3 Topology of a MIMO multi-layered perceptron neural network ... 38
2.4 Topology of a MIMO radial-basis function neural network ......... 39
3.1 Main air-side subsystems in a typical air-handling unit .......... 45
3.2 Primary and secondary circuits in air-conditioning systems ...... 47
3.3 Measured inputs and outputs to heat exchanger subsystems ....... 49
3.4 Model complexity versus accuracy .................................... 61
3.5 Output innovations-based approach .................................. 63
3.6 Parameter innovations-based approach .............................. 63
3.7 Two-tier model ............................................................... 64
3.8 FDD scheme based on indirect parameter estimation ............ 66
3.9 The output innovations FDD scheme with diagnostics ........... 69
3.10 Schematic of the proposed FDD scheme ............................. 70
4.1 Water to air heat exchanger ............................................. 72
4.2 Hysteresis model ............................................................. 77
4.3 Standardised valve characteristics ................................... 79
4.4 Close-off characteristic of a modern valve ........................ 80
4.5 Exponential valve characteristics (standard and refined) ......... 81
4.6 Three-port valve circuit .................................................... 82
4.7 Comparison of standard and proposed valve models .......... 94
4.8 Result of heating and cooling coil model validation ............ 96
5.1 The signal processing and parameter estimation scheme ........ 99
5.2 Hammerstein representation ............................................. 100
5.3 Frequency response of system ........................................ 104
5.4 Frequency response of filter ......................................... 104
5.5 First-order approximation to second order system ............... 106
6.1 Non-linearity of the heat exchanger model in the input space .... 129
6.2 Surface of the criterion function in $R_w, l$ parameter space for varying amounts of data from the input space ................. 130
6.3 Surface of the criterion function in $b, R_u$ parameter space for varying amounts of data from the input space ......................................................... 131
6.4 Surface of the criterion function in $b, l$ parameter space for varying amounts of data from the input space ......................................................... 132
6.5 Heating coil step test .......................................................... 134
6.6 Heating coil ramp test ......................................................... 135
6.7 Cooling coil step test .......................................................... 138
6.8 Cooling coil ramp test .......................................................... 139
6.9 High frequency ($\frac{1}{100}$ cycles/sample) variation in the valve control signal .......................................................... 141
6.10 Medium frequency ($\frac{1}{250}$ cycles/sample) variation in the valve control signal .......................................................... 142
6.11 Low frequency ($\frac{1}{1000}$ cycles/sample) variation in the valve control signal .......................................................... 142
6.12 $\lambda$ and the trace of $P$ for low frequency variation in the valve control signal .......................................................... 143
6.13 Improvement gained in the tracking ability of the estimator by varying fractional air flow rate .......................................................... 145
6.14 The effect of noise: $\sigma^2 = 0.05K$ .............................................. 146
6.15 The effect of noise: $\sigma^2 = 0.1K$ .............................................. 147
6.16 The effect of noise: $\sigma^2 = 0.2K$ .............................................. 147
6.17 The effect of increasing $\gamma$ to compensate for noise .............................................. 148
6.18 Unmeasured disturbance to the water mass flow rate. $\gamma = 5$ .............................................. 153
6.19 Unmeasured disturbance to the water supply temperature. $\gamma = 5$ .............................................. 154

7.1 The simulated system showing the control scheme ......................... 163
7.2 The internal gains in each of the zones ........................................ 166
7.3 The envelope gains in each of the zones for the two test days .......... 166
7.4 Inputs and outputs of the simulated system on the March day ........ 167
7.5 Inputs and outputs of the simulated system on the June day .......... 168
7.6 Training data obtained from the simulated system ...................... 169
7.7 The fit of the model to the data from the simulated system .......... 171
7.8 Parameter estimates for the fouling fault data ........................... 173
7.9 Parameter estimates for the leakage fault data ........................... 175
7.10 Parameter estimates for the sensor fault data ........................... 177
7.11 The BRE system showing the control scheme ............................ 180
7.12 Incorporation of the two-port balancing valve ............................ 181
7.13 Training/validation data from BRE ACE facility ....................... 184
7.14 The fit of the model to the data from the BRE ACE facility ........ 187
7.15 Fault-free test data .......................................................... 189
7.16 Parameter estimates for the fault-free data ................................ 190
7.17 Fouling fault test data .......................................................... 191
7.18 Parameter estimates for the fouling fault data ........................... 192
7.19 Leakage fault test data .......................................................... 193
7.20 Parameter estimates for the leakage fault data ........................... 194
7.21 Sensor fault test data ........................................... 195
7.22 Parameter estimates for sensor fault data .................. 196
List of Tables

3.1 Faults in the components of heat exchanger subsystems ............ 53

4.1 Variables associated with the operation of the heat exchanger sub-
   system ........................................ 73
4.2 Parameters used by the valve model .................................. 83
4.3 Estimates of the valve model parameters ............................... 94
4.4 Parameters used in heating and cooling coil subsystem models ... 95

6.1 Fault parameters ................................................ 126
6.2 Values of the fixed input variables .................................. 132
6.3 Mean model derivative values with respect to air-flow rate .......... 151
6.4 Sensor noise levels ................................................ 151

7.1 Severity of the faults during the three sets of test data ............ 165
7.2 The parameters used in the FDD scheme when tested with the
   simulated system ........................................ 170
7.3 Parameter estimates at the end of each day: fouling fault data .... 174
7.4 Parameter estimates at the end of each day: leakage fault data .... 176
7.5 Parameter estimates at the end of each day: sensor fault data .... 178
7.6 Test data from the BRE system .................................... 182
7.7 The parameters used in the FDD scheme when tested with the BRE
   ACE facility ........................................... 186
Chapter 1

Introduction

Faults that develop in air-conditioning systems are conventionally detected by checking that set-points are being maintained and that certain measured variables remain within predetermined limits. In practice, these techniques only allow the detection of faults that cause the behaviour of a system to change significantly, such as the case of a constituent subsystem ceasing to operate. Degradations and other types of faults, which induce small changes in the behaviour of a system, are usually not detected because their effects are masked by the action of feedback control. Moreover, conventional techniques are unable to provide detailed fault diagnoses. The techniques that are described in this thesis are designed to track the development of degradation faults, thereby being complementary to the existing technology.

1.1 FDD in air-conditioning systems

Real-time fault detection and diagnosis involve continuously inferring the condition of a system from measurements of certain properties associated with its operation. The complexity of the task is dependent on how closely related the measured properties are to the faults that are considered. Installation of appropriate instrumentation and measuring equipment can therefore simplify the FDD task. In safety-critical applications, the cost of dedicated fault-monitoring hardware is justified, and fault detection and diagnosis are mostly reduced to checking
that the monitored variables stay within their safe limits.

The continuing technological evolution and accompanying fall in the prices of digital processing hardware has meant that instrumentation previously dedicated to the fault detection task can be replaced by software processing techniques. These techniques reduce the reliance on hardware by allowing properties that previously had to be measured to be predicted instead. The predictions are made by making use of information redundancy between the remaining measurements (Chow and Willsky, 1984). There are two types of redundancy that can exist:

1. Direct redundancy - where sensor signals can be compared directly.

2. Analytical redundancy - where sensors that measure inputs to a process are related to those that measure outputs.

The latter approach reduces the reliance on instrumentation and thus the overall cost of FDD schemes. Because of this reduced reliance on additional instrumentation, it has now become viable to apply FDD techniques to non-critical systems, such as air-conditioning systems, which have limited instrumentation. The general form of an FDD scheme based on analytical redundancy is shown in Figure 1.1, which illustrates the generation of fault diagnoses from measurements of the inputs and outputs to a system.

The advancements in control system technology, and the concerns with the energy and thermal comfort implications of faulty operation have stimulated fault detection and diagnosis research in the air-conditioning field. Buildings account for approximately 50% of the total energy used in developed countries\(^1\), and recent initiatives to address world-wide generation of CO\(_2\) have helped focus international attention on this area. A collaborative research project was established in 1991 sponsored by the International Energy Agency (IEA) to address fault detection and diagnosis in heating, ventilating and air-conditioning (HVAC) systems. The work described in this thesis has been conducted in parallel with the IEA project and has formed part of the United Kingdom contribution to the project\(^2\).

\(^1\)Figures supplied by the UK Building Research Establishment.

\(^2\)The UK contribution to the IEA Annex 25 project was made by the Building Research Establishment, and the universities of Loughborough and Oxford.
1.1. *FDD in air-conditioning systems*

Figure 1.1: An example of an FDD scheme

A complete air-conditioning system can be considered to consist of smaller sub-systems, which, in turn, consist of components; a fault can occur at any of these levels. BEMS\(^3\), which are distributed microprocessor-based control systems, are commonly fitted to large air-conditioning installations and are responsible for local-loop and supervisory control. A BEMS will usually be configured to generate alarms when a set-point is not being met, or a measured variable is outside an expected range. Hence, the criteria for generating an alarm are mostly related to the ability of the system to maintain comfort conditions within the building. Faults that lead to an increase in energy costs, such as a leakage through a valve, do not always impair the ability of the system to maintain the set-points, due to the operation of feedback control. These types of faults are therefore not usually detected using the conventional approaches (Usoro *et al.*. 1985).

An alternative to the conventional approach, described above, is to use the sensor and control signal data available from the BEMS to monitor the condition of the system and infer the nature and extent of any faults (Howell and Madison. 1995). Fault *detection* involves the determination that observed behaviour of the considered system is unacceptably different from expected behaviour. Unacceptable behaviour may occur over the whole operating range or be confined to a limited region. Fault *diagnosis* involves determining which of the possible causes is consistent with observed behaviour. In some cases, it may be possible to identify the

\(^3\)Building Energy Management System.
nature of the fault unambiguously, but often it is only possible to eliminate some of the possible causes. The process of diagnosis requires that the most important possible causes of faulty operation have been identified in advance, and that these different causes lead to behaviours that can be distinguished with the available instrumentation.

Faults in air-conditioning systems can be divided into two classes: abrupt faults, e.g. a broken fan belt, and degradation faults, e.g. fouling on a heat exchanger. Abrupt faults are easier to detect, since they generally result in a sudden failure of some part of the system, although they are not necessarily easier to diagnose. In the case of degradation faults, it is necessary to define a threshold, below which the fault is considered insignificant and above which it is considered desirable to detect the fault. The difficulty of detecting and diagnosing degradation faults depends on the threshold adopted (Pape et al., 1991). In principle, this threshold must be determined by some kind of cost-benefit analysis. The possible benefits of detecting a particular fault include energy savings and improved control (Fasolo and Seborg, 1995), avoidance of occupant discomfort or illness, and avoidance of damage to the building fabric and contents or other components of the air-conditioning system. The costs of detecting and remedying a particular fault include the cost of any additional instrumentation, computer hardware and software, and any human intervention. Both the costs and the benefits will depend on the particular building and application and must be determined on a case-by-case basis.

### 1.1.1 Characteristics of the problem

The FDD methodology that is developed in this thesis is applied to heat exchanger subsystems of the type found in the air-handling units of air-conditioning installations. These subsystems have been chosen because of the important role they play in providing thermal comfort, and the energy implications that faulty operation can have. The subsystems consist of a heat exchanger (coil), valve, and an actuator. The measurements of the inputs and outputs associated with the subsystem are in the form of sensor and control signals, and are available in digital form from the BEMS. These subsystems represent a challenging application for FDD technology for a number of reasons:
1.1. **FDD in air-conditioning systems**

- low levels of instrumentation;
- high levels of uncertainty in the measurements;
- unpredictable, random disturbances;
- high levels of non-linearity.

In practice, the level of instrumentation that is fitted to heat exchanger subsystems is the minimum required to carry out the basic control functions. The accuracy of the measurements that are made is frequently poor due to cost constraints limiting the quality of the sensors. For example, it is common in the industry for single point sensors to be used to measure the air temperature in ducts that have a cross-sectional area of more than 1 m². In this case, stratification effects and poor mixing of air streams up-stream of the sensor will result in the sensor reading being a poor representation of the true average temperature. However, reductions in the price of instrumentation and the demand for greater control performance have meant that there is a current trend for improved instrumentation. For example, some companies now install averaging sensors⁴ as standard.

In contrast to some of the systems where FDD technology has been applied, heat exchangers tend to experience large changes in their operating point. These changes are mostly unpredictable and can be caused by variations in weather and disturbances within the building, such as a group of people entering a room. In systems where deviations from a particular operating point are small, the local behaviour can be approximated as linear. However, in the case of heat exchangers, the global characteristic can be quite non-linear and large variations in operating point thus render the use of linear approximations inappropriate. Account therefore has to be taken of the non-linearities, which makes the FDD task more complex.

Abrupt faults such as a broken fan belt can usually be detected by the BEMS, since they lead to large changes in the behaviour of the system that impede its ability to maintain set-points. Degradation faults, such as fouling on the surface of a heat exchanger, do not always affect the ability of the system to maintain the set-points, unless the fault is severe. These types of faults are therefore not normally detected by the BEMS. To complement the existing technology, the FDD

---

⁴These function by averaging the readings from a bank of sensors distributed within the duct.
methodology that is investigated in this thesis is designed with the objective of being able to detect and diagnose slowly developing degradation-type faults. The severity of these faults may be considered to be slowly varying with time, usually in a monotonic fashion.

Three of the most important degradation faults in heat exchanger subsystems are selected for FDD; these are: valve leakage, coil fouling, and sensor drift. The first two of these faults are manifested such that the effects of the faults are only apparent in certain regions of the operating range. The implication of this is that these faults may exist in the system for a long time before they are able to be detected. Detection can only occur when the system is exercised in the operating regions where the effects of the faults are apparent.

Precise information about the plant and its operation is required to detect the small changes that degradation faults induce from limited sensor information. Quantitative mathematical models represent one way of describing the behaviour of a system. The information in these models is contained in the parameters and in the structure, and a variety of techniques can be used to exploit the information for fault detection and diagnosis purposes. The techniques investigated in this thesis are based on using the information that the parameters contain for both fault detection and fault diagnosis. This is made possible by using models with parameters that are physically meaningful and are directly related to the faults of interest. The values of the appropriate model parameters are estimated recursively to minimise the computational requirements so that the procedures could be implemented in the current technology of BEMS thereby allowing the development of faults to be tracked in real-time.

The approach of using models with physically meaningful parameters allows previous work in the area of air-conditioning plant modelling to be exploited. The use of these types of models in parameter estimation schemes is, however, complicated by the fact that the models are usually non-linear with respect to their parameters. A major part of the thesis is therefore dedicated toward the development and evaluation of an estimator suitable for these non-linear models.

---

5 Two previous IEA projects (Annex 10 and Annex 17) were concerned with plant modelling: Annex 25 represents a natural progression of the work carried out in these projects.
1.2 Objectives of the thesis

The overall objective of this work is to develop and evaluate a fault detection and diagnosis scheme that is capable of tracking the development of degradation faults in the class of non-linear heat exchanger subsystems used in HVAC plant. The main aims of the work, in relation to the structure of the thesis, are as follows:

- to describe the fundamental concepts of fault detection and diagnosis systems within a generic framework, and critically review the different FDD techniques that have been proposed in the literature (Chapter 2);

- to describe the system to which the FDD techniques are applied, discuss the types of faults in these systems, and develop an FDD methodology that is tailored to the considered application (Chapter 3);

- to develop models of the considered subsystems and to extend these models so they are capable of approximating the changes in behaviour resulting from the development of the selected faults (Chapter 4);

- to develop computationally efficient techniques for estimating the parameters of the models in order to allow slowly developing faults to be tracked in real-time (Chapter 5);

- to use empirical methods to analyse the behaviour of the parameter estimation techniques in relation to type of parameter change, excitation level of input signals, noise, unmeasured disturbances, and modelling errors (Chapter 6);

- to evaluate the performance of the FDD scheme using data from two air-conditioning installations: a detailed dynamic simulation of a 3-zone VAV system, and a full size experimental air-conditioning test facility (Chapter 7);

- to draw conclusions and suggest possible areas where there is the potential to conduct further research (Chapter 8).
Chapter 2

Literature review

Introduction

Fault detection and diagnosis are made possible by being able to identify changes in the behaviour of a system. One way to achieve this, which is mostly applicable to instrumentation faults, is to install duplications of the components that are prone to faults. The faults can then be detected and diagnosed by comparing the measurements from the duplicated components. This approach creates redundant information by using additional hardware. Alternatively, the redundant information can be created by using models of the components. Recently, the widespread availability of digital control systems with sufficient processing capabilities for non-control functions has helped trigger the development of FDD schemes based on models. These schemes alleviate the need for hardware duplication, and hence reduce the overall cost of implementing FDD schemes. A description of model-based approaches to FDD forms a large part of this chapter. In addition, other ‘software-oriented’ FDD methods that use partial models or other simpler representations of a priori knowledge of a system are described.
2.1 Background

Research in the area of fault detection and diagnosis has been most intensive in the safety-critical industries, such as chemical processing and power generation. Much of the research in these fields was initiated in response to disasters such as those at the Three Mile Island nuclear power station and the Bhopal chemical plant. More recent disasters, such as the Chernobyl nuclear power station, and the Challenger Space Shuttle, have seen a resurgence of activity in the field. The advances that have been made in these industries are now being exploited in other application areas that are not safety-critical, but where costly consequences may ensue when faults develop. Examples include air-conditioning systems, paper making, brewing, etc.

In the context of physical systems, a fault can be described as any kind of phenomenon that leads to an unacceptable anomaly in the overall system performance. This may be considered an arbitrary description since what constitutes ‘unacceptable’ performance is not easy to define. Nevertheless, it relates the concept of ‘fault’ to performance. Performance may be considered as a measure of how well a system performs the tasks it was originally designed for, and a fault inhibits the ability of a system to perform its design tasks. A fault should not be confused with a ‘failure’, since the latter infers a complete breakdown while the former may mean only a small change in performance. Some of the terminology that is commonly used to described different types of faults is listed below:

- abrupt, failure, catastrophic - these terms usually relate to sudden faults, i.e. step-like changes;

- incipient, degradation - these terms relate to faults whose severity increases slowly with time, i.e. ramp-like changes.

Abrupt faults can be caused by the failure, or the cessation of operation, of the monitored process or a component of it. These types of faults can have severe consequences, particularly in the safety-critical applications. They should therefore be detected early so that remedial action can be taken. Incipient faults lead to a slow reduction in performance. They can be caused by such phenomena as a gradual build-up of dirt on the fins of a heat exchanger, or a sensor that drifts slowly from its calibration point. Incipient faults are not easily detected in their
early stages and their effects can be concealed by the operation of feedback control (Watanabe et al., 1989). If these faults are allowed to increase in severity, failure can be the eventual result.

The fundamental operation performed by an automated FDD scheme is to check whether particular measured or unmeasured variables are within a certain tolerance of their normal values. If a tolerance limit is transgressed, a fault message is generated. The sequence of tasks performed by an FDD scheme can be defined as follows (Isermann, 1984):

1. Monitoring.
2. Fault detection.
3. Fault diagnosis (isolation).
4. Fault evaluation.

In the simplest case, monitoring is the act of acquiring the measurements from the sensors. In more processor-intensive schemes, the monitoring task can also involve transforming the measured properties into a form suitable for the fault detection task. Fault detection is the process of generating an alarm when an index has transgressed a limit. Fault diagnosis entails analysing information (usually both instantaneous and historical) to attribute a fault to a particular cause. The evaluation process is concerned with ascertaining the extent of the fault in quantitative terms. The extent to which each of these tasks needs to be performed depends on the application. If a human operator is always present, fault detection may be sufficient, since further tests could be carried out manually. Isolation involves identifying the faulty item of plant and is important when back-up systems are available. In the absence of back-up systems, evaluation of the extent of the fault and its quantitative effect on performance is useful to determine whether remedial action is necessary.

The main design goals for an FDD scheme are (Willsky, 1976):

- to achieve rapid response to failures and to degradation that lead to unacceptable performance:
2.1. Background

- to be insensitive to non-fault changes, such as noise, changes in operating point, normal signal variations, modelling errors, etc.

To achieve these goals, a number of trade-offs have to be considered and tailored to each particular application. According to Isermann (1984), these are:

- degree of fault versus detection time;
- speed of fault appearance versus detection time;
- detection time versus false alarm rate.

A fundamental requirement is for fault detection methods themselves to be reliable and not prone to breakdown. An unreliable FDD scheme that generates false alarms, or fails to detect obvious malfunctions, would soon be switched off by an operator. The additional cost of an FDD scheme must also be justified by the realisable economic benefits. Moreover, the methods should not be too complex; they should be understandable and tunable by engineers or operators, otherwise acceptance problems will arise.

‘Voting schemes’ (Willsky, 1976) can be used to detect faults when physical systems possess a high degree of parallel hardware redundancy. These schemes are particularly useful for detecting sensor faults. Usually, (at least) three identical sensors are employed and simple logic is used to detect failures and eliminate faulty sensors. For example, if one of the three sensor signals differs markedly from the other two, the differing signal is identified as being faulty and eliminated from any control functions. This method of fault detection can be extremely reliable, but the cost of incorporating additional hardware, if not already available, can be prohibitive in many applications.

One problem with the sensor redundancy approach is that it may not be possible to achieve exact duplication even with when using the same type of sensors. For example, it is not possible to position sensors in exactly the same place: e.g. fluid temperature measurements may differ due to stratification effects. Faults that affect all sensors in the same way will also be undetectable; e.g. common power supply failures. The standard number of sensor measurements associated with control tasks usually contain information that hardware-based methods do not
2.1. Background

utilise. In practice, many digital control systems have sufficient processing capabilities to be able to cope with additional non-control functions. Software-based methods that take advantage of the information from the controlled plant are therefore now being used as alternative to hardware redundancy. It is nevertheless common, particularly in some safety-critical applications, for some hardware duplication to be installed, to comply with regulations. Clearly, it is useful to exploit hardware duplication for fault detection purposes when it is available. The recent trend, however, is to supplement it with software-based methods to provide even greater reliability.

Because of the technological advances in digital processing hardware, and the relative reductions in cost, software-oriented methods have now become a viable alternative to hardware-based methods, for non-critical applications (Frank, 1990). The information available in the inputs and outputs of a monitored process is exploited using analysis, thus reducing the need for additional instrumentation. The method of analysis can vary significantly in complexity but usually incorporates some knowledge of the process. The knowledge may be in the form of mathematical equations describing the relationship between inputs and outputs, or be simply threshold values used to check whether a measured variable is within a tolerance band.

If there is limited hardware duplication, it is possible to combine hardware and software redundancy techniques. One example of this is where two sets of identical instruments are used; each set being supervised by a software-based FDD method. The software-based FDD is only initiated when a discrepancy arises between the two sensors outside the bounds of noise. Once the fault has been confirmed by the software scheme, the system is reconfigured to use only the healthy sensor, which is automatically identified by the software. This approach saves computational burden, and improves reliability (Deckart et al., 1977).

Fault detection and diagnosis technology can also offer benefits in less direct ways. For example, to detect and diagnose faults, 'performance indices' that relate to the condition of the system have to be calculated. These indices can be exploited to improve the efficiency and cost-effectiveness of maintenance programmes (Rossi and Braun, 1994). Traditional maintenance programmes involve performing repair, replacement, and cleaning work on a fixed cycle basis. Fault detection and diagnosis methods can be used to implement predictive maintenance schemes whereby work is only carried out when the condition of the system has deteriorated to a level at
2.1. Background

which it is warranted.

There are also many similarities between the techniques used in fault detection and diagnosis, and those used in adaptive control. Local-loop control can be directly affected by a change in the characteristic of the controlled process caused by a fault. A fixed controller that is tuned to provide a tight, but a stable response on the initial system may become oscillatory or sluggish if a fault occurs (Hepworth, 1994). Adaptive control attempts to address the problem of a changing plant characteristic by updating the parameters of the controller (Åström and Wittenmark, 1989). Hence, faults could be detected by monitoring the values of the controller parameters, thus allowing the control and FDD tasks to be carried out simultaneously.

2.1.1 Preprocessor and classifier components

The overall function performed by an FDD scheme is the transformation of raw measurements from the monitored process into decisions. The decisions can be made at two levels:

1. Has a fault occurred? (Binary decision)
2. What type of fault has occurred? (Multiple decision)

The decisions are made on the basis of quantitative values, which may be the measured inputs/outputs to the monitored process or be derived from them. Besides determining the type of fault responsible for a detected malfunction, the degree of fault is often required as part of the evaluation task. An FDD scheme can be characterised by:

- the type of transformation that the raw measurements undergo before being used in the decision process;
- the method of arriving at a decision on the basis of the (transformed) variables.
From the description given it may be argued that there are two distinct functions carried out by FDD schemes. Basseville (1988) recognised this distinction and Rossi and Braun (1993) formalised a method of classification based on two basic components: a preprocessor, and a classifier (Figure 2.1). The preprocessor is used to extract diagnostically relevant information from the raw measurements by transforming them into new variables (performance indices). The classifier then returns decisions appropriate for the performance indices.

![Figure 2.1: Preprocessor and classifier components of an FDD scheme](image)

The complexities of the preprocessor and classifier components can vary significantly. In the simplest example of a fault detector, *viz.* the limit checking of a measured variable, the preprocessor performs no function at all, and may be described simply as a unity transformation. The classifier in this case is also very simple, and may be constructed using a single (IF THEN ELSE) conditional instruction. The preprocessor and classifier classification can also be extended to an FDD method based on hardware-redundancy. In this case, the hardware duplications can be thought of as the main components of the preprocessor. The performance indices would be the measured differences from the duplicates, and the classifier would comprise simple rules.

It may be noted that the preprocessor makes transformations in the quantitative domain, while the classifier has to transform quantitative variables into qualitative decisions. Even a binary decision made by a fault detector is qualitative since the 'fault' and 'no fault' conditions are abstract qualities based on human perceptions. A quantitative estimation of the degree of a fault may be output as part of the diagnoses, but this would be generated by the preprocessor component and constitute a performance index. The degree of fault would thus be the basis upon which the decision of whether a fault exists is made.
2.2 Model-based methods

The preprocessor and classifier provide a means by which to contrast the differences between different methods that may appear from a high level to have a similar overall functionality. This may be exploited by classifying systems based on how the computational complexity is distributed between the two components. Performance indices may be generated that indicate directly different fault types if detailed mathematical models are used in the preprocessor. The classification task is then trivial. Alternatively, if no explicit preprocessing is carried out and detailed diagnostics are nevertheless required, the computational effort has to be concentrated in the classifier instead. An example of the latter case is an expert-system fault diagnosis scheme.

2.2 Model-based methods

Model-based approaches to FDD work by exploiting the fact that the inputs and outputs to a system are correlated, and that the nature of this correlation is affected by the occurrence of a fault. In general, changes can only be detected if a reference is available to make them apparent. Hence, a reference model of 'correct' behaviour is required for model-based schemes. There are essentially two ways in which the reference model can be used to detect changes in the characteristic of the system:

1. By using the reference model to predict the current outputs of the system and by comparing these with measured outputs.

2. By comparing all, or some of the internal attributes of the reference model with those of another, identically structured, model representing the current observed behaviour of the real system.

Each of these approaches is based on detecting changes (often called innovations). The first approach looks for changes in measured variables, while the second approach looks for changes in variables that are not directly measurable. In terms of the preprocessor and classifier description, the model would constitute the preprocessor. The routines that make the decisions using indices generated by the model would then constitute the classifier. The way in which the model of the process is formulated affects the choice of FDD method and the realisable robustness and
reliability. Some of the typical model formulations that have been used for FDD purposes are described in the following section.

### 2.2.1 The process model

A model, as its linguistic definition states, is a simplified representation of a real system. In the context of this work, the term process model refers to a mathematical description of a real system comprising equations that relate variables in a similar way to the real system. Mathematical models of all sorts have two defining attributes: structure and parameters. The structure is the mathematical function relating the inputs and outputs, e.g. linear, quadratic, exponential, etc. The parameters then tailor the function to a particular system.

A model is called 'analytical' (or 'physical') if its structure is derived by considering the mechanisms that generate the signals and variables within the real system. Analytical models are constructed using mathematical equations based on physical laws and established relationships that are known to govern the behaviour of the system. The parameters of analytical models have physical meaning providing all basic physical laws are adhered to during the derivation. The parameters may relate to physical properties that are directly measurable (e.g. length, mass, force, pressure, etc), or they may have to be determined from tests (e.g. heat transfer resistances).

It may not be always possible to construct an analytical model due to incomplete knowledge of a system. Furthermore, analytical modelling can be time consuming, particularly for complex systems. The alternative approach, which is often referred to as empirical modelling, involves inferring a model from recorded observations of system inputs and outputs. In this case, several candidate model structures are selected based on 'shallow' knowledge of the correlation between the inputs and outputs (e.g. linear, non-linear: high or low order, etc.). The parameters of these models do not have obvious physical meaning; their 'correct' values are the ones that allow the model to replicate best the behaviour of the real system\(^1\). The candidate model with the 'correct' parameter values that yields the closest match to the observed behaviour of the system is then selected.

\(^1\)As is apparent from the record of observations.
A model does not have to be exclusively empirical (sometimes referred to as black-box) or analytical (white-box). In practice, most models lie somewhere on the continuum that separates these two classes. Intermediate models (grey-box) have some parameters that relate to physical properties and some that have to be inferred from observed data. A model of this type enables partial analytical knowledge to be used by allowing the ‘missing’ information to be determined empirically. The ability of a model to approximate a system will be related to the amount of (correct) information (i.e. analytical and empirical) embedded within it. Hence, a prudent approach to modelling is to use as much of the information that is directly available as possible.

**Input-output models**

Input-output models have been used extensively for FDD, e.g. (Gertler, 1988; Patton et al., 1995). The way in which the outputs of the model are related to the inputs and the parameters is of most relevance for FDD applications. Three possible formulations of a MISO\(^2\) model are:

1. Linear in the parameters and the inputs:
   \[ y = \theta \phi. \]  
   (2.1)

2. Linear in the parameters, non-linear in the inputs:
   \[ y = \theta \psi. \]  
   (2.2)

3. Non-linear in the inputs and the parameters:
   \[ y = f(\theta, \phi). \]  
   (2.3)

The vectors \( \phi \) and \( \psi \) represent basis vectors and are of compatible dimensions with the parameter vector \( \theta \). The relationship between \( \phi \) and \( \psi \) is given by:

\[ \psi = [f_1(\phi) \ f_2(\phi) \ \ldots \ f_m(\phi)]. \]  
(2.4)

\(^2\)Multiple Input Single Output.
where \( f_i(.) \), for \( i = 1, 2, \ldots, m \) are arbitrary non-linear functions. The definition of the basis vector \( \phi \) depends on whether the model is static or dynamic (continuous or discrete):

**static:**

\[
\phi = [u_1 \ u_2 \ \ldots \ u_r]. \tag{2.5}
\]

**dynamic (continuous):**

\[
\phi = [u_1 \ \frac{du_1}{dt} \ \frac{d^2 u_1}{dt^2} \ \ldots \ \frac{d^p u_1}{dt^p} \ u_2 \ \frac{du_2}{dt} \ \frac{d^2 u_2}{dt^2} \ \ldots \ \frac{d^p u_2}{dt^p} \ \ldots \ u_r \ \frac{du_r}{dt} \ \frac{d^2 u_r}{dt^2} \ \ldots \ \frac{d^p u_r}{dt^p} \ y \ \frac{dy}{dt} \ \frac{d^2 y}{dt^2} \ \ldots \ \frac{d^m y}{dt^m}]. \tag{2.6}
\]

**dynamic (discrete):**

\[
\phi = [u_1(t) \ u_1(t-1) \ \ldots \ u_1(t-p) \ u_2(t) \ u_2(t-1) \ \ldots \ u_2(t-p) \ \ldots \ u_r(t) \ u_r(t-1) \ \ldots \ u_r(t-p) \ y(t-1) \ y(t-2) \ \ldots \ y(t-m)]. \tag{2.7}
\]

Note that \( m \) is the order of the system, and \( p \leq m \), reflecting the cause-effect relationship between the inputs and outputs. It is not physically possible for \( p \) to be greater than \( m \), because this would imply the ability to predict the future of the system input. Dynamic behaviour within physical systems stems from the ability of the system or components within to store energy, or information. In practice, all physical systems have this capacity to a certain extent. Static models do not take account of the dependence that the variables have on time and therefore represent a simplification of real behaviour. These models can, however, provide acceptable approximations in applications where the rate of variation in the inputs to a system is slow compared to the dominant time constant.

The dynamics of the inputs are often much slower than the system response time in the subsystems used in air-conditioning. These systems frequently have long periods of steady-state behaviour, interspersed with transient activity, usually due
to sudden load changes (e.g. a group of people entering a room, a step change in a set-point, etc). Static models can be used to represent these systems during the steady-state periods but they would become inaccurate during the periods of transient activity. Hence, the transient periods would have to be disregarded to ensure reliability. The FDD scheme that is reported in this thesis is based on the use of static models and a steady-state detector is developed in order to recognise transient periods of system operation.

In Equation 2.1, the parameter vector linearly maps the inputs to the outputs. The parameters of a model formulated in this way can be identified uniquely for a given set of input-output observations\(^3\). This type of model is often applied to systems having many inputs where there is little understanding of the underlying physical processes, e.g. social science applications (Montgomery and Peck, 1982). In this case, the parameters are not usually physically meaningful.

In its discrete dynamic form, Equation 2.1 is known as an ARX\(^4\) model and is often written in the following way:

\[
A(q^{-1})y_k = B(q^{-1})u_k, \tag{2.8}
\]

where:

\[
A(q^{-1}) = 1 + a_1q^{-1} + \ldots + a_pq^{-p}, \text{ and }
B(q^{-1}) = b_0 + b_1q^{-1} + \ldots + b_mq^{-m},
\]

where \(q\) is the backward shift operator. Note: if the model is deterministic, the outputs may be predicted using the measured inputs and the previous output predictions. If the model is partially stochastic, previous output measurements would be used in Equation 2.8.

It may be noted that the ARMAX\(^5\) model is an ARX model with the inclusion of a noise modelling term:

\[
A(q^{-1})\hat{y}_k = B(q^{-1})u_k + C(q^{-1})e_k, \tag{2.9}
\]

where:

\(^3\)Providing the number of unique observations is equal to, or greater than, the number of parameters.

\(^4\)AutoRegressive with eXogenous variables.

\(^5\)AutoRegressive Moving Average with eXogenous variables.
2.2. Model-based methods

\[ C(q^{-1}) = c_0 + c_1q^{-1} + \ldots + c_sq^{-s}. \]

Since the difference equations are approximations to the differential equations, the model predictions may diverge from the true values if predictions are fed back into the model. This divergence may increase during prolonged periods of transient activity.

Equation 2.2 allows a non-linear relationship between inputs and outputs to be modelled while maintaining linearity in the parameters. This is achieved by using non-linear functions to transform the inputs. The model is then parameterised so the outputs are linear combinations of the transformations. This type of model still allows the parameters to be determined uniquely from observations, and therefore represents a useful means by which to describe non-linear behaviour. There are many well known model types that are in the form of Equation 2.2, examples are polynomial curves, radial-basis function neural networks, etc.

Models that are derived from physical theory can not always be formulated with the outputs as a linear function of the parameters. Characterisation of fluid flow, heat and mass transfers, etc, in physical terms frequently results in the outputs of the model being non-linear functions of both the parameters and the inputs (Equation 2.3). Empirical determination of parameters is then more difficult since a unique solution cannot be guaranteed for a set of observations, and non-linear optimisation methods usually have to be used. In these circumstances it is often better to take advantage of the physical significance of the parameters and estimate the parameters from knowledge of the modelled system, i.e. using information contained in design or manufacturers’ data. Parameter estimates obtained in this way can then be refined empirically using non-linear optimisation methods.

Sometimes a model may be non-linear in its parameters, but have physically meaningless parameters. The parameters of these types of models have to be estimated empirically using non-linear optimisation methods. One of the main factors affecting the reliability of non-linear optimisation procedures is the initial parameter estimates. If these are not within the vicinity of the true optimum, the method may converge on a local minimum. When the values of the parameters do not have any meaning, it is difficult to select appropriate initial estimates. These types of models are therefore mostly avoided. One exception to this is the type of neural network known as the multi-layered perceptron (MLP). These models have a very general structure and because of this are often able to approximate
2.2. Model-based methods

arbitrary functions satisfactorily, even when the parameter estimates collapse on a local minimum. MLPs are highly parallel in nature and the training process is effective at automatically distinguishing the important inputs from those that have little correlation with the outputs. The behaviour of this type of model outside the regions where the training data were obtained is not usually predictable and can be highly inaccurate.

Models of the type given in Equation 2.3, which have physically meaningful parameters, can sometimes be transformed into the form given in Equation 2.2, by lumping together some of the parameters (and functions of them). It is then possible to estimate the physical parameters by inverting the known relationship between the model parameters and the physical parameters, $p$. That is:

$$\theta = f(p), \quad (2.10)$$

thus:

$$p = f^{-1}(\theta). \quad (2.11)$$

This type of formulation can be particularly useful for fault detection and diagnosis, since linear parameters can be estimated fairly easily using linear estimation techniques. These model parameters can then be related to the physically meaningful parameters to provide a deeper insight into the condition of the system (Isermann, 1985). It may be noted, however, that this approach relies on the function being invertible. It is therefore only applicable to simple model formulations, where the parameters of interest can be factored out of the equations.

**State-space models**

This type of model formulation is applicable to dynamic models and is particularly useful for describing MIMO\(^6\) systems. State variables are used to describe the dynamic behaviour, the number of which is equal to the order of the system. The state variables are time dependent and can be either well defined physical variables, such as position ($x_1$), speed ($x_2 = \dot{x}_1$), and acceleration ($x_3 = \ddot{x}_2$), or functions of the measured input and output signals. The general state-space form for a linear deterministic system is given by:

$$\dot{x}_k = Ax_k + Bu_k \quad (2.12)$$

\(^6\)Multiple Input Multiple Output.
2.2. Model-based methods

\[ y_k = Cx_k + Du_k. \]  \hspace{2cm} (2.13)

where \( k \) is the sample number and \( x \) is the vector of state variables, given by:

\[ x^T = [x_1 \ x_2 \ \ldots \ x_s], \ \text{i.e.} \ x \in R^s. \]  \hspace{2cm} (2.14)

\( A, B, C, \) and \( D \) are parameter matrices of appropriate dimensions. Solving Equation 2.12 is tantamount to solving \( s \) simultaneous first-order differential equations. The equations can be solved in a way analogous to numerical integration by developing a recursive form of Equation 2.12, such that:

\[ x_{k+1} = A_1x_k + B_1u_k, \]  \hspace{2cm} (2.15)

where \( A_1 \) and \( B_1 \) are functions of the original parameter matrices. The accuracy of the output predictions made by estimating the state vector recursively is highly dependent on the accuracy of the initial state estimates. In addition, the difference form of the equation represents an approximation, which can lead to divergence. In FDD applications, the states are usually estimated by using state estimators, such as observers or filters, rather than relying on the propagation of initial state estimates. State estimators compensate for the inaccuracy of initial state estimates in the model, prevent divergence, and allow the effects of faults to be more easily distinguished from the effects of unknown inputs (Frank and Wünnenberg, 1989). The basic function performed by state estimators is the reconstruction of the states of the system from the measured inputs and outputs, as shown in Figure 2.2.

![Figure 2.2: State observer](image)

Observers are applied to deterministic systems (Equation 2.12), while filters are applied to systems having noise components; e.g.

\[ \dot{x}_k = Ax_k + Bu_k + w_k \]  \hspace{2cm} (2.16)

\[ y_k = Cx_k + Du_k + e_k. \]  \hspace{2cm} (2.17)
2.2. Model-based methods

where \( w \) and \( e \) are noise vectors. In both types of state estimator the state vector is refined at each time step by adjusting it in proportion to the prediction error of the model. A state estimator is thus given by:

\[
\hat{x}_{k+1} = A_1 \hat{x}_k + B_1 u_k + K(y_k - C\hat{x}_k - D u_k),
\]  

(2.18)

where \( y \) in the above expression represents the measured output vector. In the case of observers, it is not usually necessary to estimate all the states in the same way, as \( n \) states can be inferred from the measurable outputs (where \( y \in \mathbb{R}^n \)). A less detailed observer, known as a reduced order observer, can then be constructed.

The effectiveness of state estimation clearly depends on the selection of an appropriate gain matrix, \( K \), which is a non-trivial task. Details and derivations of error stabilising matrices for observers can be found in (O’Reilly, 1984). Further details of the gains appropriate for the Kalman filter can be found in (Grewal and Andrews, 1993). It may be noted that the outputs of a system can be non-linear functions of the states, and the states and outputs may be non-linear functions of the parameters. In this situation, the linear state-space equations are obtained by linearising the equations about a datum point.

Linearisation

If a model is derived and cannot be easily transformed to a linear form, the model can be linearised about a particular operating point. The assumption has to be made that the process is locally linear and the degree of deviation from the datum point is small. To illustrate the linearisation process, consider a general non-linear relationship between multiple inputs and a single output:

\[
y = f(x_1, x_2, \ldots).
\]  

(2.19)

This is linearised by using gradient information as follows:

\[
\delta y = c_1 \delta x_1 + c_2 \delta x_2 + \ldots,
\]  

(2.20)

where

\[
c_1 = \frac{\partial y}{\partial x_1 |_{x_1}},
\]

\[
c_2 = \frac{\partial y}{\partial x_2 |_{x_2}}, \text{ etc.}
\]
2.2. Model-based methods

where $d_{xi}$ is the datum point for $x_i$ about which the linearisation takes place; $\delta y$ and $\delta x_1 \delta x_2 \ldots$ represent the change in the variables relative to the datum point. Linearisation can be used to linearise the model with respect to either the inputs or the parameters. Although, the process introduces further inaccuracies into a model, the resulting linear model form can simplify other procedures that may be necessary in an FDD scheme, such as parameter or state estimation.

2.2.2 Output-innovations methods

One way in which faults can be detected is to use a reference model of the correctly operating system to predict the system outputs from the measured inputs. In this approach, the predictions are compared with the measured outputs and any discrepancies are regarded as indications that the system has deviated from its correctly operating condition. An advantage of the approach is that the structure of the model is not important; it is treated as a black-box where only the inputs and outputs are relevant. Once a model has been obtained that completely describes the dynamic and static effects of the correctly operating system, an innovation vector, $i$ can be defined, such that:

$$i(t) = y(t) - \hat{y}(t),$$

(2.21)

where $i \in \mathbb{R}^n$. The equation, from which the innovation vector is calculated, is sometimes known as a parity equation, or consistency relation (Gertler, 1995). The components of the $i$ vector may be used directly in a limit checking scheme to detect faults, i.e. if $i_{1,\text{min}} \ i_{2,\text{min}} \ldots \ i_{n,\text{min}}$ are the lower limits and $i_{1,\text{max}} \ i_{2,\text{max}} \ldots \ i_{n,\text{max}}$ are the upper limits, the system would be deemed to be operating correctly providing:

$$i_{\text{min}} < i_k < i_{\text{max}}, \forall i.$$  

(2.22)

If the reference model describes the behaviour of the correctly operating system accurately, the innovations should only be due to the effects of noise when there are no faults. If the distribution function of the noise is known, the occurrence of a fault can be detected by performing statistical tests on the innovations (Frank, 1990), e.g. ‘whiteness’ tests. Should any of the innovations in the $i$ vector transgress a limit, prior knowledge of the system and its outputs can be utilised to localise the fault to a limited extent. In many circumstances a fault will lead to innovations in a number of the outputs. Consideration of the overall direction of
the innovation vector when one (or more) limits are transgressed can thus provide a greater insight into the nature of the fault (Dalton et al., 1995). For example, a unit vector, \( \mathbf{d} \), in the direction of the innovation vector can be defined:

\[
\mathbf{d}(t) = \frac{\mathbf{i}(t)}{||\mathbf{i}(t)||},
\]

where \( ||.|| \) denotes a vector norm. The simplest way of analysing this vector is to consider the sign of its components and compare these with predetermined sign patterns that relate to different fault types. This idea has been applied to heat exchangers by the author and co-workers (Benouarets et al., 1994), where innovation patterns gathered at different operating points were used to generate simple diagnostics. Rossi and Braun (1994) also used this method to diagnose faults in vapour compression chillers. The approach is capable of detecting both abrupt and incipient faults but its effectiveness relies heavily on the appropriate selection of thresholds, which is non-trivial in this case. The relationship between the magnitude of the innovations and the magnitude of a fault can only be established from a knowledge of how the physical properties of the system relate to the innovations. This knowledge may be available in the form of an analytical model, in which case the properties could be calculated from the innovations. However, if this procedure were followed, it would become a parameter estimation method, and these methods are discussed separately in Section 2.2.3.

Innovations can be caused by phenomena other than faults, such as unmeasured disturbances, or modelling errors. The selection of thresholds must therefore be based on an assessment of the extent of these other effects. The trade-off between fault detection sensitivity and false alarm frequency also has to be carefully considered when determining thresholds. Ideally, if the outputs of a model are to be used for FDD, they should be predictable from the measured inputs and the previous output predictions only. If past outputs measurements are used by the model to obtain the predictions, deviations of the system from its ‘correctly operation’ condition will not be as apparent. This is due to the model essentially only having to make a one-step ahead extrapolation using the measured data. In ARX models, this can be avoided by using previous output predictions, but this can lead to divergence problems if the coefficients are not accurate. The state-space model does not suffer these types of problems providing the states are measurable from the plant, but this is frequently not the case.

When the states cannot be measured directly they have to estimated from measured inputs and outputs using observers or filters. The estimated states are then
used to generate the model predictions of the system outputs. The method of reconstruc-
ting the state variables is based on minimising the prediction error. Thus, in the event of a fault, the states will be adjusted in an attempt to reduce the resulting innovations to zero. A persisting innovation can only be assured, therefore, if the model with its empirically estimated state vector is structurally incapable of describing fully the faulty system. This can often be the case, but in circumstances where a fault only leads to minor changes in the characteristic of the system, an adjustment of the states may be sufficient to 'correct' the innovations. To detect faults of this nature, the rate at which the fault causes the process to change has to be greater than the rate at which the state estimator reduces the error to zero. Innovation-based methods that use state-space models are therefore best suited to the detection of abrupt faults. Incipient faults that cause ramp-like changes in the system are more difficult to detect (Da and Lin, 1995).

Despite the problems associated with using state-space models to generate fault-
related output innovations, the approach has been popular for detecting abrupt faults. Most of the methods are geared towards fault detection rather than diagnosis and work by applying statistical tests to a time window of innovations, e.g. (Mehra and Peschon, 1971; Fathi et al., 1993; Usoro et al., 1985; Yu et al., 1995). One way to obtain diagnostic insight when using state estimation methods is to analyse the gain matrix associated with the estimation. This gain matrix can reveal fault 'signatures' and if these fault signatures are known \textit{a priori}, different fault types can be differentiated between (Willsky, 1976).

2.2.3 Parameter and state innovations methods

The fundamental reason for using models within FDD schemes is to generate variables that, under normal circumstances (i.e. when the system is operating correctly), remain constant, but change if there is a fault. The output innovations have been shown to have such a quality providing the reference model is known. However, the output innovations themselves do not offer much insight into the fault type and the magnitude of a particular fault cannot be assessed very easily.

One way to gain more insight into the nature of a fault is to consider other properties of the system that cannot be measured directly but are closely related to the potential faults (e.g. the states or the parameters). In this approach, a model,
which is usually of identical structure to the reference model, but extended to treat
faults, is employed on-line with the real system. The states and/or parameters
of the on-line model are updated so that the on-line model represents the current
observed behaviour. A fault is detected if the estimated states or parameters differ
from those of the reference model. The extra meaningfulness that the states and
parameters offer is then exploited for fault diagnosis.

Changes in state variables

Some faults will result in permanent changes to the state variables, from their
'correct' values, for example: a mass flow leakage or an electrical current leakage.
The change in a state variable is given by:

$$\Delta \hat{x}(t) = \hat{x}(t) - \hat{x}_0(t), \quad (2.24)$$

where $\hat{x}$ is the state vector constructed from measured inputs and outputs using
an observer or filter and $\hat{x}_0$ is the state vector of the correctly operating system.
Since a fault will cause a change in the relationship between measured inputs and
outputs these cannot be used to construct the correct state vector ($\hat{x}_0$). Instead,
state propagation techniques (Equation 2.15) have to be used to estimate future
states of the reference model, e.g. (Brumback and Srinath, 1987). This approach
can be problematic with the potential existing for divergence of the state estimates,
in particular:

- the initial conditions have to be known exactly;
- the approximations due to differencing that are needed for digital imple-
  mentation can lead to divergence during prolonged transient periods (Da
  and Lin, 1995).

Accurate estimates of the correct (reference) states are essential if small faults are
to be detected. The potential for error in the process of estimating the correct
states is such that reliable detection of small faults cannot be guaranteed. The
idea of looking for changes in the states of the system is more suited to detecting
larger faults such as those caused by failures. However, these faults could be
detected by simply observing the transience of the estimated states; without the
2.2. Model-based methods

need to maintain estimates of the correct states, e.g. by using methods suggested by Basseville (1988).

Another approach that has been employed is to use observers or filters in parallel with hardware redundancy. This involves reconstructing two (or more) state vectors from different (possibly overlapping) input and output measurements and then comparing them to check for consistency. If the models used in the observer schemes are structurally identical, modelling errors and noise will have equal effects, thus increasing robustness (Gertler, 1988).

Changes in parameters

Many faults cause process coefficients to change, such as resistances, capacitances, inductances, mass, stiffness, etc. Examples of faults that cause these types of properties to change are: an increased resistance to heat transfer due to the fouling of a heat exchanger; increased damping effect in a piston chamber due to an increase in friction. If the parameters of an on-line model are continually updated so the model represents the observed behaviour of the real system, changes in process coefficients, $\Delta p$, will lead to changes in the model parameters, i.e. $\Delta \theta$. If the parameters of the model relate directly to the process coefficients, i.e. $\theta \equiv p$, the physical consequences of a fault can be assessed easily thus simplifying the fault diagnosis task. The parameters of analytical models generally have a close relationship to process faults. Estimation of the parameter values thus enables faults to be detected earlier and localised more precisely than is possible with other approaches (Patton et al., 1989).

The parameters of empirical models do not have physical meaning and are not suited to the parameter estimation approach. Although a change in their estimated values may relate closely to a particular fault, the nature of this relationship cannot be determined easily. Instead, the way in which different faults cause changes in the parameters (parameter-based fault signatures) has to be determined a priori by empirical means (Dalton et al., 1995). This means that training data has to be obtained from the system in different faulty conditions. Hence, there are no obvious advantages to be gained by using empirical models within a fault diagnosis scheme that is based on parameter estimation.

The way in which the process model is formulated affects the choice of estimation
2.2. Model-based methods

algorithm and the resulting robustness of the FDD procedure. Models that are linear-in-the-parameters are the most useful. The ARX model, for example, is of this form, and methods such as recursive least-squares, which have well-understood asymptotic properties, can be used with these models. In some instances, it is useful to be able to isolate, for estimation, the parameters that affect the dynamic behaviour of the model from those that affect the static behaviour. or vice versa.

In the case of the ARX model the parameters attributed to all variables before the current sample affect the dynamics, while the variables at time $t$ affect the static behaviour. This approach can be useful when certain faults are known to affect only the static or dynamic characteristic in isolation (Isermann, 1993). Online parameter estimation for fault detection has also been applied successfully to models that are non-linear in their parameters (Letty, 1995). However, the robustness of the estimation is sensitive to the initial parameter estimates and the rate of change of the system parameters. A study of the robustness of a non-linear parameter estimator, for different types of parameter variations, forms a major part of the work reported in this thesis.

An equation that expresses the difference between the parameters of a fixed, 'correct operation', model and an adaptive model representing the current, observed behaviour, is known as a parity equation. Parity equations can be written in matrix form and the properties of the matrices can be analysed to check for faults. Different types of faults can manifest themselves in different ways within the matrices, and this property can be exploited for fault diagnosis. This approach, and the associated terminology, has been adopted by a number of researchers, e.g. (Chow and Willsky, 1984; Gertler, 1995; Isermann, 1993).

The objective of most parameter estimation algorithms is to reduce the model prediction error to zero. Subsequently, dynamic parameters may be adjusted to compensate for changes in the static behaviour that occur and vice versa. When states are not measured, this is an unavoidable problem in the state-space approach, since both states and parameters have to be constructed from measured inputs and outputs. An additional problem with the state-space representation is that the estimation of both the parameters and states is a non-linear problem. Uniqueness of the solution cannot be guaranteed therefore, and robustness problems can ensue (Ljung and Söderström, 1983). This is illustrated by considering a deterministic state-space model. A new vector, $\psi$, is defined to represent the augmented state and parameter vectors, i.e. $\psi^T = [x^T \theta]$. The resulting problem
2.2. Model-based methods

formulation is then non-linear; i.e.

$$\psi_{k+1} = f(\psi_k, u_k) + \omega_k \quad (2.25)$$

$$y_k = g(\psi_k) + e_k. \quad (2.26)$$

where:

$$f(\psi_k, u_k) = \begin{bmatrix} A(\theta)x_k + B(\theta)u_k \\ \theta_k \end{bmatrix}$$

$$\omega_k = \begin{bmatrix} w_k \\ 0 \end{bmatrix}$$

$$g(\psi_k) = C(\theta_k)x_k.$$

The non-linearity can be tackled by linearising the functions, $f(.)$ and $g(.)$ about a datum point (the current estimates) and applying a Kalman filter to the linearised state equations. The complete procedure is then known as the extended Kalman filter (EKF). The extended Kalman filter has been used for FDD by (Fathi et al., 1993; Yoshida et al., 1995). Other researchers have addressed the combined state and parameter estimation problem by attempting to apply different estimation techniques to the same model simultaneously, for example: by using recursive least squares to estimate parameters and observers to estimate states. Venkateswarlu et al. (1992) provides a review and analysis of some of these approaches; see also (DalleMolle and Himmelblau, 1987; Dalton et al., 1995).

Parameter estimation approaches to FDD based on fuzzy models have also been proposed for FDD applications (Benouarets and Dexter, 1995). Fuzzy models are used to take account of the uncertainties and imprecision associated with modelling complex, and ill-defined, systems. The models are designed to capture the principal features of the behaviour of a considered system by using fuzzy sets to represent the information. One particular advantage of the approach is that the models can easily incorporate whatever expert knowledge is available. The approach investigated by Dexter and co-workers involved identifying a fuzzy model of the considered system from observations of the inputs and outputs.

---

\(^7\)In Dexter's work, a partial model is identified, which contains information from a localised region of the operating space of the system.
obtained on-line. The on-line model is then compared with a number of reference models representing correct and faulty operation. The models are compared using a process known as fuzzy matching which provides an estimate of the degree of similarity between the reference models and the on-line model. Dexter has applied the techniques to air-conditioning system applications. One of the main aspects of the work has been the production of generic reference models suitable for application to classes of similar systems (Dexter and Benouarets, 1995).

It may be noted that there are a number of general prerequisites that should be satisfied to ensure satisfactory performance from a parameter estimation approach to FDD, these are (Isermann, 1989):

- the process models have to describe the process behaviour precisely enough so that the changes in the characteristic of the system caused by the faults of interest are outside the bounds of the model uncertainty;
- the parameter estimation methods have to be robust to noise and unmodelled disturbances;
- the process has to be excited sufficiently so that information is obtained from the system at the operating points where faults are apparent.

There is usually a trade-off between the complexity of the parameter estimation procedure and complexity of the diagnosis task. Analytical models with diagnostically useful parameters are often non-linear in their parameters, while empirical models with meaningless parameters are linear. Non-linear parameter estimation methods are usually only capable of reliably tracking slowly varying parameters, while linear methods are more robust to large, abrupt changes (Ljung and Söderström, 1983). The designer of an FDD scheme based on model parameter estimation therefore has to consider a number of points to arrive at a suitable design solution. A typical design process may be as follows:

1. Determine the types of faults targeted for detection and diagnosis.
2. Decide upon the level of diagnosis that is required (if any).
3. Evaluate the type of information available to construct model (i.e. extent of empirical data, do analytical models exist? Is sufficient time and expertise available to derive analytical models?).
4. Consider the parameter estimation algorithm required for the different types of feasible models: does the algorithm facilitate the detection and diagnosis of the target faults? Does the algorithm need to be supplemented with additional diagnostic procedures to satisfy 2?

5. Select the model and estimation algorithm that is best able to satisfy all the points listed above.

2.3 Other software-oriented FDD methods

Model-based methods represent one possible way of detecting and diagnosing faults. There are various other methods that have been proposed where the relationship between inputs and outputs to the system is not directly characterised. This section presents a review of the most popular of these methods.

2.3.1 Signal analysis

Signal analysis methods can range in complexity from simple limit checking schemes to more complex methods based on the use of signal models. These types of methods are generally used when the causality relationships between system variables are not known, or when signals are expected to remain static or vary in a way that is independent of other variables (i.e. the signal is only correlated with itself, e.g. ambient air temperature).

Limit checking of absolute values

The simplest form of signal analysis involves using the measurable signals to the monitored physical system to detect faults directly. A system can be monitored by employing simple limit checking procedures. Usually, limit checking operates solely on the outputs; the system would be deemed to be operating correctly if:

\[ y_{min} < y_k < y_{max}, \forall y. \]  

(2.27)

If the lower or upper limits are transgressed, an alarm is generated. Limit checking
of absolute values in this way can work especially well if the signals remain in approximately steady-state. If the operating point of the process varies significantly the approach is less viable. In the case of closed loops, changes due to faults in the process will be compensated for by the control action. If the controller is able to maintain the controlled variables at their set points, faults will not be evident if only the absolute values of controlled variables are considered (Isermann, 1995).

Limit checking of absolute values has the advantage of simplicity and the technique has been applied to numerous applications. Tolerance ranges are invariably over-estimated to avoid false alarms, hence the technique is usually only sensitive to a large abrupt fault or degradation that has led to a significant degree of deterioration. In addition, a detailed diagnosis is not usually possible with this technique. Limit checking can also be applied to simple signal-derived variables, such as the trend, \( \dot{y}(t) \). The trend represents a prediction of the progression of the absolute value and, if limits are selected appropriately, consideration of the trend can allow faults to be detected more quickly.

**Signal models**

The *behaviour* of a signal can be modelled by using stochastic models. These models do not have inputs in the deterministic sense, but correlate the current signal value with its previous values, or with its derivatives (or states). Auto-regressive models (AR, or ARMA if the noise is modelled) are examples of the former while state-space models are an example of the latter. An ARMA model or order \( m \) is given by:

\[
\dot{y}_k = A(q^{-1})y_k + B(q^{-1})e(t),
\]  

where:

\[
A(q^{-1}) = a_1q^{-1} + \ldots + a_mq^{-m}, \text{ and}
\]

\[
B(q^{-1}) = 1 + b_1q^{-1} + \ldots + b_rq^{-r}.
\]

A stochastic state-space model can be given by:

\[
\begin{align*}
x_k &= Fx_k + w_k, \\
y_k &= Hx_k + e_k.
\end{align*}
\]
where \( \mathbf{w} \in \mathbb{R}^n \) and \( \mathbf{e} \in \mathbb{R}^n \) are noise vectors. Estimation of the states in stochastic state-space models is achieved using a filter, such as a Kalman filter, instead of an observer, which is used with deterministic models. Providing the parameter matrices, \( \mathbf{F} \) and \( \mathbf{H} \), are known and the model is structurally correct, the state estimates of a Kalman filter are asymptotically ideal because of the linear form of the model. It may be noted that the Kalman filter algorithm can also be used to estimate recursively the parameters of linear regression-type models (e.g. ARMA or deterministic ARMAX), by expressing the equations in the following way:

\[
\begin{align*}
\theta_{k+1} &= \theta_k \\
y_k &= \phi_k \theta_k + e_k,
\end{align*}
\]  

(2.31)  

(2.32)

where \( \phi \) is the regressor vector. AR(MA) and state-space models can both be used in FDD schemes in the same ways that were outlined in Section 2.2. It should be noted that the accuracy of stochastic models, and hence the robustness of an FDD scheme based on these models, depends on causal inputs being periodic, static or non-existent.

Another method of signal analysis used successfully with a number of applications involves analysing the frequency spectrum of signals. In these methods, it is assumed the signals have a distinct spectral form and that this is altered by the occurrence of a fault. The approach has been used successfully for applications such as: transmission systems, internal combustion engines and jet engines (Isermann, 1984).

### 2.3.2 Characteristic quantities

Characteristic quantities \( (\eta) \) are unmeasurable variables that describe the performance of a considered system. They are calculated from the measured inputs and outputs to the system, such that:

\[
\eta = f(\mathbf{u}, \mathbf{y}).
\]  

(2.33)

Some examples of characteristic quantities are:

- efficiency (e.g. engines, machines, steam generators, heat exchangers, furnaces, vehicles):
2.3. Other software-oriented FDD methods

- fuel consumption per production unit or time (e.g. cement burning, milling, drying);
- consumption of lubrication oil per production unit or time (e.g. internal combustion engines, compressors);
- tool usage per production unit or time (e.g. machine tools);
- wear per production unit or time (e.g. tools, motors, grinding devices).

Characteristic quantities are generally static functions of the inputs and outputs to a system. If a dynamic relationship exists between \( y \) and \( u \), efforts should be made to only use steady-state samples. A failure to do this will lead to false estimates of \( \eta \). A fault can be detected if the change in the characteristic quantity, \( \Delta \eta \), is large. A characteristic quantity-based approach to fault detection was proposed in Pape et al. (1991), in which the power consumption of an air-conditioning system was considered. The power consumption of a non-faulty system was predicted using model based simulation and this was compared with the actual power consumption. Characteristic quantities do not allow faults to be localised easily and the strength of the approach therefore lies in the fault detection aspect. Since characteristic quantities relate to the performance of a system that usually comprises many subsystems, the effects of subtle faults and faults that occur in a single subsystem may not be revealed.

Watton and Creber (1988) used simple quantities such as pressure differentials and more general efficiencies to detect flow leakage in hydraulic control systems. Their findings indicated that the application of more detailed functions (constituting a partial model) provided better results. This implies that more diagnostically useful characteristic quantities can be obtained by using progressively greater amounts of physical knowledge of the system.

In the context of the preprocessor/classifier framework explained in Section 2.1.1, the approaches to FDD that have been described so far have the computational complexity concentrated in the preprocessor. These types of approaches, although extremely popular in recent years, represent only one viable solution to the problem. There are other approaches where the classifier performs the majority of the processing instead. Two approaches that match this description are neural networks and rule-based expert systems and these are described in the following sections.
2.3 Other software-oriented FDD methods

2.3.3 Neural networks

A complex non-linear relationship normally exists between the raw measured inputs and outputs to a physical process, and the decisions and/or quantitative evaluations associated with fault diagnostics. The problem of reconstructing this relationship is usually addressed using expert knowledge. The expert knowledge can be in the form of mathematical equations (e.g. process models), which are used to make transformations in the quantitative domain. Alternatively, the expert knowledge can be in a qualitative form (e.g. rules, fault symptom trees etc). Frequently though, FDD procedures utilise a combination of both quantitative and qualitative knowledge representations.

If little, or no, expert knowledge is available, the alternative is to use 'example' data to extract the required information. In Section 2.2.1 it was described how empirical process models can be used within the preprocessor component of an FDD scheme; neural networks can also be used in this way (Patton et al., 1994). However, an alternative, and more popular, approach has been to use them to model the overall relationship between measurements and diagnostic decisions (Hsu and Yu, 1992; Watanabe et al., 1989; Fan et al., 1993; Sorsa and Koivo, 1991; Hoskins et al., 1991). This sort of approach to fault detection and diagnosis can be considered to be truly black box. There is no means by which to separate any distinct preprocessing and classification function carried out internally within a neural network when applied in this way. Since the outputs are diagnostic decisions the network must be categorised as a classifier.

Neural networks are empirical modelling tools that have been shown to be capable of dealing with highly non-linear problems with many inputs and outputs (Cybenko, 1989; Funahashi, 1989; Park and Sanberg, 1991). It is their ability to model arbitrary non-linear relationships and their relative robustness when tested with novel data that has made them popular for FDD applications. Neural networks became more popular following the work of Rumelhart and McClelland that was published in 1986 (Rumelhart and McClelland, 1986), which described a simple algorithm for determining the parameters (known as weights) from training data. Many refinements and improvements to the training procedures and network architecture have been suggested since Rumelhart and McClelland's work (e.g. Chen and Billings, 1992; Chester, 1993; Harvey, 1994; Fausett, 1994).
Neural networks relate inputs, \( u \), and outputs, \( y \), in a non-linear fashion. i.e:

\[
y = f(u, \theta),
\]

where \( \theta \) is a vector of all network parameters, and \( f(.) \) is a non-linear function. There are two types of network that have been popular for modelling and FDD applications: radial basis functions (RBF), e.g. (Moody and Darkin, 1989), and multi-layered perceptrons (MLP). RBFs have the advantage that the parameters before the final layer can be determined heuristically thus rendering the empirical modelling task linear-in-the-parameters. In contrast, MLPs do not facilitate the heuristic determination of any parameters, and they all have to be estimated empirically using non-linear optimisation. MLPs have remained popular despite the parameter estimation difficulty, mostly due to them being able to generalise (extrapolate) more accurately than RBFs (Fargus, 1994). Despite MLPs having the potential for limited extrapolation, the behaviour outside training data regions is unpredictable. RBFs have the advantage that extrapolation can be detected, and thus be avoided to avoid resulting errors (Leonard et al., 1992).

An MLP network having three layers of nodes is depicted in Figure 2.3. Each node performs a non-linear transformation on its inputs and the output from each node is propagated along the 'synaptic' connections to the nodes in the next layer. The outputs of the nodes in each layer are linearly scaled before being passed as inputs to the following layer. These scaling coefficients, known as weights, are the parameters of the network and they have to be estimated during the training process to allow the network to produce the desired mapping between inputs and outputs.

Figure 2.4 depicts a typical RBF network. RBFs have the same parallel nodal structure as MLPs but only have two layers of nodes. The inputs to the network are distributed to each of the nodes in the first layer. These nodes perform a non-linear transformation on the inputs, governed by a parameter vector, known as a centre, which is unique to each node. The outputs of the first nodal layer are then propagated along weighted synapses to the final layer of nodes, which sums the linearly scaled outputs from each of the first layer nodes.

For an RBF model, if \( \theta \) represents the parameter vectors used in the non-linear functions within the first layer nodes, and \( \omega \) is a matrix of weights between the first and last layer of nodes, the network function can be written as a linear regression
2.3. Other software-oriented FDD methods

Figure 2.3: Topology of a MIMO multi-layered perceptron neural network

(providing $\theta$ has been predetermined):

$$y = \omega \phi,$$

where:

$$\phi^T = [f_1(u, \theta_1) \ f_2(u, \theta_2) \ldots \ f_m(u, \theta_m)]$$

$$\omega = \begin{bmatrix} 
\omega_{11} & \omega_{12} & \cdots & \omega_{1m} \\
\omega_{21} & \omega_{22} & \cdots & \omega_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
\omega_{n1} & \omega_{n2} & \cdots & \omega_{nm} 
\end{bmatrix}$$

where $\omega \in \mathbb{R}^{n \times m}$ is the matrix of final layer weights, $\phi \in \mathbb{R}^p$ is a vector of the outputs from the middle layer nodes, $n$ is the number of outputs nodes and $m$ is the number of middle layer nodes.

The way in which neural networks are usually used for FDD is to associate the output from each output node with a different fault, i.e. $y_1 =$ Fault 1, $y_2 =$ Fault
2.3. Other software-oriented FDD methods

Figure 2.4: Topology of a MIMO radial-basis function neural network

Training data would consist of a vector of measurements from the real system and the diagnosis appropriate for the time at which the measurements were made, also in vector form. The diagnosis vector would consist of ones and zeros, where a one indicates that the fault exists; a zero that it does not. The network parameters would then be estimated to allow the network to reproduce, as accurately as possible, the mapping of measurements to faults apparent in the training data. When the trained network is used, it is improbable the new input data vectors will match exactly any that were in the training data. Since the output nodes of neural networks are capable of generating continuous variables, the outputs will not be binary, but lie somewhere between zero and one. The output (fault node) having a value nearest to one is assumed to be the likeliest diagnosis based on the evidence provide by the training data.

One drawback of using neural networks for fault diagnosis is that the network has to be trained to model the relationship between measurements and faults. Faulty training data therefore has to be obtained from the real system. In practice, it is unlikely that the artificial introduction of different faults will be permitted to

\[ u \rightarrow f_1(\theta_1, u) \rightarrow \Sigma \rightarrow y_1 \]

\[ u \rightarrow f_2(\theta_2, u) \rightarrow \Sigma \rightarrow y_2 \]

\[ u \rightarrow f_m(\theta_m, u) \rightarrow \Sigma \rightarrow y_n \]

Usually constrained to lie between zero and one.
2.3. Other software-oriented FDD methods

gather training data. Possible ways of avoiding this problem are described in (Hoskins et al., 1991; Watanabe et al., 1989); examples include:

- utilise data logs of system measurements recorded at the time of past faults for training;
- utilise a simulation of the system to generate faulty and fault free data.

If faulty and fault-free training data are available, or can be obtained easily from the real system, the neural network approach may prove effective for FDD. However, the problem of how accurately the neural network is able to generalise outside training data regions has to be assessed. If data generated from simulation is used for training, the effectiveness of the neural network classifier will be dependent on how accurately the simulation represents the real system. If the simulation is accurate, it has to be asked whether the simulation itself may not be employed to a better effect within an FDD scheme.

2.3.4 Knowledge-based systems

Knowledge-based systems (KBS) allow FDD procedures to be developed in the cases where only qualitative knowledge of the relationship between symptoms and faults is available. The two basic components required to build a KBS are:

- the knowledge base that contains the expert knowledge in a qualitative form (e.g. IF THEN rules, sign directed graphs, fault trees);
- the inference engine that processes the expert knowledge and ultimately matches symptoms to decisions (e.g. forward/backward chaining, fuzzy inferencing).

Qualitative inputs are matched with qualitative outputs by the inferencing process carried out by a KBS. For FDD applications the inputs to the KBS can be the raw measurements from the real system, which are quantitative. In this case, the measurements have to be transformed into the qualitative (discrete) domain by applying quantitative knowledge of the process. An example would be to define
three discrete ranges for a continuous variable e.g. low, medium and high. Clearly, in this case, thresholds have to be set to distinguish between the different ranges, which requires a quantitative understanding of the variable. Instead of defining hard thresholds, it is possible to cater for uncertainty by merging the ranges so that there are regions of overlap. This concept has been described by Zadeh (1965) and is known as fuzzy sets. It may be noted though, that the positioning of the fuzzy sets can require more quantitative information than is needed to set hard thresholds.

It has been described how a KBS cannot directly use the raw measurements from a real system (sensor and actuator signals) without the incorporation of some quantitative knowledge of the system and its variables. It may be argued therefore that a preprocessor component will be always required for FDD schemes based on knowledge-based systems. In the simplest form, the preprocessor will transform the raw measurements into qualitative variables. More complex preprocessors will generate more diagnostically meaningful qualitative variables that can be related more easily to decisions, thus simplifying the construction of the KBS.

Rowan (1988) describes how process engineers rely on fairly precise fundamental equations (e.g. mass and energy balances) to narrow the diagnostic solutions. After performing this data reconciliation step to focus on fault candidates the engineer applies empirical knowledge, which has been gained through years of operating experience to make a diagnosis. This empirical knowledge is heuristic in nature and is usually in the form of rules. The majority of FDD schemes that have been designed for real-time use with physical systems attempt to mirror this diagnostic process carried out by humans. The preprocessor component will contain fundamental physical transformations (deep knowledge) and the classifier will contain heuristic (shallow) knowledge to match transformed variables to decisions, e.g. (Frank, 1990; Rowan, 1987; Yu and Lee, 1991).

The diagnostics that are output to a user of an FDD scheme generally have to be in a qualitative form. A knowledge-based system of some form therefore has to be used to generate these decisions irrespective of the amount of knowledge contained in the preprocessor. The construction of a KBS can be a difficult process and the method of gathering data from human experts has to be carefully planned (Brothers, 1988). The difficulties associated with eliciting accurate information in the right format for integration within a KBS have meant that the current trend has been toward concentrating more of the system-specific knowledge into
2.4 Conclusions

The chapter has described some of the techniques that have been used in automatic fault detection and diagnosis schemes. The objective of the chapter has been to describe the FDD techniques in a way that is independent of the application. It has been shown that fault detection is made possible by being able to identify changes in the behaviour of the considered system. The main difficulty associated with the fault detection task lies in distinguishing changes that are the result of faults from those that are not, such as unmeasured disturbances, noise, etc. When the considered system is controlled to a set-point, the failure of the system to meet the set-point may be used as an indication of the existence of a fault. However, this approach is not suited to the detection of small changes, such as those resulting from degradation. In this case, a more accurate representation of the 'correctly operating' system is required to detect fault-induced changes. Mathematical models can be used for this purpose, and different model formulations suitable for fault detection were described in Section 2.2.

Fault diagnosis involves distinguishing between different types of faults. This is achieved by analysing how a change resulting from a fault influences the behaviour of the system across its operating range. Models characterise the behaviour of the system across the range of operation and thus allow the differences caused by faults to be analysed at different operating points for diagnosis purposes. It was shown in Section 2.2.2 how models can be used for fault detection and diagnosis by considering innovations of the system outputs. The structure and type of the model are not important in this approach and empirical or analytical models may be used. The evaluation of the degree of a fault is, however, more complex and some expert knowledge of the system and how the faults are manifested is required. Analytical models satisfy these requirements, and it was shown in Section 2.2.3 how parameter estimation methods can be used to determine the degree of faults using these model types.

Other methods for detecting and diagnosing faults were described in Section 2.3. Signal analysis techniques range in complexity from simple limit checking to...
stochastic model-based approaches. These methods are most suited to systems that have many unmeasured variables that influence their behaviour. It was described in Section 2.3.2 how methods based on characteristic quantities utilise partial models of the system. Results have indicated that improved diagnostic performance can usually be obtained by increasing the amount of knowledge incorporated in these models. Neural networks can be employed in output-innovations FDD schemes, or they can be used to model the overall relationship between the inputs and outputs to a system, and the fault diagnostics. Neural network approaches have been reported widely in the literature, but the issue of how to gather appropriate training data (from faulty systems) does not seem to have been sufficiently addressed. In addition, the neural networks cannot be easily applied to the fault evaluation task. Knowledge-based systems, described in Section 2.3.4, are most suited to the diagnosis task. A useful role for these techniques would therefore be to supplement the other FDD techniques where diagnosis is the weaker aspect; i.e. approaches based on output innovations.

Although many of the overall concepts employed in FDD schemes can be considered within a generic framework, certain elements of FDD schemes have to be tailored to the application. Each system will have a distinct set of faults, and knowledge of these faults and the way in which they are manifested has to be incorporated in the FDD scheme to generate diagnoses. In many cases, the type, and nature of the faults that are considered determines the design of some of the principal aspects of the FDD scheme, such as the process model. Chapter 3 now considers the aspects specific to the application of heat exchanger subsystems of the type found in air-conditioning systems in order to formulate an appropriate design.
Chapter 3

Design of the FDD scheme

Introduction

This chapter develops the design of the FDD scheme by considering the pertinent factors for the application of heat exchanger subsystems in air-conditioning plant. Development of a diagnostic capability requires a priori knowledge of the potential faults, and of the way in which they are manifested. Section 3.2 thus presents results from international surveys that were carried out to identify and assess the importance of faults in air-conditioning systems. The faults associated with the considered subsystems are described and their symptoms and causes are discussed. The application-specific aspects of a fault detection and diagnosis scheme are considered in Section 3.3 and some of the FDD techniques viable for the application are discussed in Section 3.4. Finally, an FDD scheme that satisfies the design criteria is described in Section 3.5.

3.1 Overview of air-conditioning systems

The fundamental purpose of an air-conditioning system is to control the environmental conditions within an enclosed space. According to the Chartered Institute of Building Services Engineers (CIBSE, 1986), an air-conditioning system must be able to control the following properties of the internal air within the space: for
3.1. Overview of air-conditioning systems

'Tzone'): 

- purity;
- temperature;
- humidity.

To control the above properties, air-conditioning systems comprise: fans, filters and mixing dampers to ensure air purity by way of ventilation; heat exchangers to control temperature and provide dehumidification; and humidifiers to humidify the air. This thesis will concentrate on all-air systems, as these are common in large buildings and they usually have sufficient instrumentation to make an automatic fault detection system viable. These types of systems are able to provide sensible and latent cooling, preheating, and usually humidification, of an air-stream supplied to the space (ASHRAE, 1992). Additional cooling and dehumidification are not usually required in the zones. The same air-stream is used for heating, either centrally, or at each particular zone.

![Figure 3.1: Main air-side subsystems in a typical air-handling unit](image)

The main subsystems of a typical all-air system are shown in Figure 3.1. These subsystems are often contained within 'air-handling units'. and there may be a number of different air-handling units within one building. The heat exchangers transfer thermal energy between two fluids; air being one of the fluids. Compact
3.1. Overview of air-conditioning systems

Heat exchangers are normally used and these have an arrangement of finned tubes through which a chilled or heated fluid is passed (Kays and London, 1964). The fluid can be water or steam when the heat exchanger is used for heating; water (or glycol) or refrigerant when it is used for cooling and/or dehumidification. Heat exchangers that use the refrigerant to cool the air stream are known as 'direct expansion' (DX) coils, and these are not considered in this thesis.

The temperature of the air leaving the air-handling unit\(^1\) is usually a controlled variable for the air-handling unit. The air dampers that form the mixing box can be used to vary this temperature in the range between the temperature of the outside air and the temperature of the air returning from the building. Regulation of the supply air temperature outside this range requires the heat exchangers to be operated. The heat transferred by the heat exchangers can be varied either by changing the temperature difference between the two fluids, or by changing the flow rate of one of the fluids. Variation of the water flow rate through the heat exchangers is achieved using control valves and variable-speed pumps. Variable speed pumps are used with heat exchangers that have two-port valves, while the use of three-port valves allows the flow rate to be kept relatively constant in the primary circuit (see Figure 3.2 for the distinction between primary and secondary circuits). Variable speed fans alter the air-flow rate and a system that uses this facility to meet variations in the load are known as VAV\(^2\) systems.

Both two-port and three-port control valves are moved by actuators, which can be operated by electric motors or pneumatic power. In the United Kingdom electric motors are more common, while in the United States pneumatic actuators are more prevalent. Heat exchangers and other air-side subsystems are controlled by building energy management systems (BEMS), which provide local loop and supervisory control. The level of instrumentation fitted to all air-side subsystems is usually kept to the minimum required for control to reduce costs. In spite of this, there is still some redundancy, in most systems, between temperature sensors at certain operating points that could be exploited for fault detection.

In an air-handling unit, the outside, return, and supply air temperature sensors may be compared with each other when the system is at certain operating points (see Figure 3.1, where the temperature sensors are shown as boxed T's). For example, the ambient air sensor may be compared with the supply air sensor

\(^1\)Often termed 'supply air temperature' or 'discharge air temperature'.

\(^2\)Variable Air Volume.
when the mixing box is providing full outside air, and the heat exchangers are not operating. This technique has been exploited successfully for fault detection by (Glass et al., 1994), who termed the operating points where sensor signals could be compared as ‘landmarks’.

Figure 3.2: Primary and secondary circuits in air-conditioning systems

Large air-conditioning installations can comprise many air-handling units, each providing a heating or cooling effect to the air. The equipment that circulates the fluid used in the heat exchangers and extracts or supplies heat is often known as the ‘primary plant’. The heat exchangers and other subsystems that are in the air-flow are termed the ‘secondary plant’ (Figure 3.2). Systems that provide both heating and cooling have two fluid circuits (hot and cold). Chillers are used to extract heat from the fluid in the primary circuit and boilers are used to add heat.

Primary plant generally have their own dedicated controllers, which can often incorporate simple fault detection procedures. The amount of external instrumentation fitted to these plant, and the corresponding amount of information relayed to the BEMS, is usually minimal. There is, therefore, less scope for applying FDD procedures to primary plant from the BEMS than there is with the secondary (air-side) systems.
3.1. Overview of air-conditioning systems

3.1.1 Heat exchanger subsystems

Heat exchangers in the air-handling units of air-conditioning installations (indirectly) account for a large portion of the total air-conditioning energy costs. These heat exchangers also have a significant effect on the thermal comfort within a building. Faults that affect their performance are therefore of prime concern.

The operation of the heating coil and the cooling coil is normally sequenced so that both coils do not operate at the same time. The temperature of the air in the duct, measured after both coils, is a controlled variable. A proxy of the temperature of the air before the coils can usually be obtained from the temperatures of the two air-streams passing into the mixing box. This is possible since the mixing box normally operates at one of its extreme positions (full outside air or minimum outside air) when the heat exchangers are operating, due to the sequential control. Sensors are sometimes installed between the coils\(^3\) and this provides additional hardware redundancy, which can be used for detecting sensor faults. Mostly though, analytical techniques have to be employed to exploit the redundancy that exists.

Inputs and outputs to heat exchanger subsystems

Figure 3.3 shows a heat exchanger subsystem in block form, illustrating the inputs and outputs that are typically measured and available from the control system.

Measurements of the inlet and outlet water temperatures are not always available. The internal control loops of boilers and chillers regulate the temperature of the water, which is at the inlet to the heat exchanger control valves, to a constant set-point; equal to that given in the system design specifications. This input can therefore often be treated as constant. Another unmeasured variable that has an influence on the outputs is the flow-rate of the water in the primary circuit. Systems that use three-port valves are considered in this thesis, and the flow rate in the primary circuit (at the inlet to the control valve) is kept approximately constant by the pumps in these systems. This can therefore also be treated as a constant value. Disturbances in the unmeasured inputs may lead to the generation of alarms in an FDD scheme that is based on the assumption that they remain

\(^3\)Sensors are sometimes installed between the coils for the control of dehumidification.
3.2. Categories of faults in HVAC systems

Faults can be divided into three categories:

- design faults;
- installation faults;
- operational faults.

constant. However, this can be a useful feature, since malfunctions in the chillers, boilers and pumps can be indirectly detected by monitoring the performance of the heat exchangers.

Some of the sensors associated with the operation of the heat exchanger subsystem may also play an important role in the control of other subsystems, such as the mixing box and humidifier. Application of fault detection and diagnosis procedures to heat exchangers therefore allows the operation of the entire air-handling unit to be monitored, to a limited extent.

3.2 Categories of faults in HVAC systems

Faults can be divided into three categories:
3.2. Categories of faults in HVAC systems

The objective of an on-line FDD scheme is to detect faults in the last category. However, the performance of the on-line system will be influenced by the existence of faults in the other two categories. Generally, one of two possible assumptions has to be made, depending on the nature of the FDD procedures:

1. The observed behaviour at the commencement of operation represents correct operation.

2. The information contained in the design specifications is representative of the installed system in its correctly operating condition.

In most cases, design faults will not invalidate either of these assumptions providing it can be assumed that the system is constructed as detailed in the design. These faults are therefore not detrimental to most FDD schemes. However, there are some diagnostic systems that are developed to operate with classes of correctly designed systems that exhibit similar behaviour (Dexter, 1995b). In this case, design faults may cause the installed system to be unrepresentative of the assumed class, in which case they will have an influence.

The existence of installation faults will violate the second assumption. Ideally, these faults should be detected during commissioning time and remedied before the system is put into operation. The building industry has produced a number of codes of practice aimed at improving the standard of commissioning, e.g. (Pike and Pennycook, 1992). In practice, the commissioning of air-conditioning systems is often inadequate (Tong, 1989; Dexter et al., 1993). There are a number of possible reasons for this:

- the time available for commissioning is frequently reduced by delays in construction;
- there is a shortage of skilled personnel to perform commissioning work;
- it is difficult to produce a well-defined specification of certain aspects of performance, particularly dynamic performance;
- it is impossible to test fully the performance of the HVAC systems in an unoccupied building, at any one time of year.
Software-based procedures have been proposed as a means by which to automate some of the routine tasks in commissioning and alleviate these problems (Dexter et al., 1993). These automated procedures allow work to continue in parallel on different subsystems, while allowing the commissioning engineers to deal with the problems identified by the automated commissioning system.

The commissioning task is similar in many respects to the on-line FDD task. In particular, the objective in both cases is to analyse the performance of the system and compare it with a predetermined reference (model) representing 'correct operation'. The main differences lie in the amount of freedom available to exercise the systems and in the method of obtaining the reference model. At the commissioning stage, the building served by an air-conditioning system is unoccupied and freedom exists to excite the different subsystems to assess their performance (limited by the weather conditions at the time of the tests). During normal operation, however, it is not usually possible to excite the system artificially, since this will disrupt the comfort of the occupants. Hence, on-line FDD schemes generally have to be non-intrusive and rely on normal disturbances to provide excitations.

For an on-line FDD scheme, the reference models, or performance specifications, can be obtained empirically. This is not possible in the case of a commissioning system since the system cannot be operated to generate the inputs and outputs until it has been commissioned. A commissioning procedures therefore have to be developed based on the assumption that the system conforms to the design specifications. Because of the detrimental effects of poor commissioning on the performance of air-conditioning systems, it is better to make an on-line FDD scheme sensitive to these faults so that it can assist in their detection (Yoshida, 1996).

3.2.1 Faults in heat exchanger subsystems

Fault detection involves the determination that the relationship between the inputs and outputs of a system has changed significantly. Procedures to detect faults can be developed from information about the input-output relationship for the correctly operating system. This information can be obtained empirically from input-output observations, or analytically by utilising expert knowledge of how the system operates and by using design specifications. Fault diagnosis requires the potential faults and the way in which they are manifested to be known as well.
Since systems are not designed to develop faults, it can be difficult to predict the different fault conditions from theoretical knowledge. The frequency, importance and effects of different faults are ascertained more reliably from practical experience.

The persons involved with overseeing the operation of air-conditioning systems and in dealing with faults when they occur possess knowledge that is needed in the development of fault diagnosis procedures. As part of the IEA Annex 25 project, a number of these persons were questioned about the importance and frequency of faults in heating ventilating and air-conditioning systems, and the results are presented in (Hyvarinen, 1995). These fault surveys were carried out in France, Germany, Canada, Finland and Japan. Although the results are relevant to systems in the United Kingdom, the differences in design, installation and commissioning between the countries may affect their validity. To address these concerns, two building services maintenance contractors in the United Kingdom were questioned to supplement the other surveys. The faults in heat exchanger subsystems are presented in Table 3.1, representing a collation of the results from all the surveys. The table shows the symptoms and causes of different types of faults in the four main components: heat exchanger, actuator, sensors, and hydraulic valve.

The two operational faults in electric motor actuators are of the failure type. However, the decoupled linkage fault can occur gradually; beginning with slippage. The slippage would lead to anomalies between the expected and actual valve or damper stem position. The effect would progressively become worse until the actuator decouples from the controlled element (i.e. valve or damper). In the case of a motor failure, the actuator will fail at any position. However, if it has a spring return, it will fail at one of its extreme positions (usually closed). Incipient faults are more common in pneumatic actuators due to loss of air pressure stemming from leakage in the actuator itself or in associated piping. The operation of a pneumatic actuator is dependent on the operation of the compressor driving it. Faults in the compressor can therefore affect the actuator performance and subsequently the heat exchanger.

Sensors are relatively inexpensive compared with other components (e.g. such as actuators, heat exchangers), but they play an essential role when used in control loops. Faults in control loop sensors can cause disruptions to thermal comfort and lead to a wastage of energy (Kao and Pierce, 1983). A fault exists when
Table 3.1: Faults in the components of heat exchanger subsystems

<table>
<thead>
<tr>
<th>FAULT TYPE</th>
<th>SYMPTOM</th>
<th>CAUSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ELECTRIC (E), PNEUMATIC (P) ACTUATORS</td>
<td></td>
</tr>
<tr>
<td>installation</td>
<td>system moves in wrong direction (E)</td>
<td>reverse acting</td>
</tr>
<tr>
<td>installation</td>
<td>no response from controlled system (E)</td>
<td>incorrectly wired</td>
</tr>
<tr>
<td>installation</td>
<td>low system gain (E,P)</td>
<td>range mismatch</td>
</tr>
<tr>
<td>operational</td>
<td>no response from controlled system (E)</td>
<td>motor failure</td>
</tr>
<tr>
<td>operational</td>
<td>no response from controlled system (E,P)</td>
<td>decoupled linkage</td>
</tr>
<tr>
<td>operational</td>
<td>reduced system gain (P)</td>
<td>pneumatic leakage</td>
</tr>
<tr>
<td></td>
<td>SENSORS</td>
<td></td>
</tr>
<tr>
<td>installation</td>
<td>no reading</td>
<td>not connected</td>
</tr>
<tr>
<td>installation</td>
<td>inappropriate reading</td>
<td>wrongly positioned</td>
</tr>
<tr>
<td>operational</td>
<td>bias/drift</td>
<td>dirt on sensor</td>
</tr>
<tr>
<td></td>
<td>CONTROL VALVES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>design unstable closed loop control</td>
<td>wrongly sized</td>
</tr>
<tr>
<td></td>
<td>installation control system moves in wrong direction</td>
<td>wrongly installed</td>
</tr>
<tr>
<td></td>
<td>operational no response from controlled system</td>
<td>sticking or binding</td>
</tr>
<tr>
<td></td>
<td>operational leakage at closed position</td>
<td>physical obstruction</td>
</tr>
<tr>
<td></td>
<td>operational reduced maximum flow</td>
<td>build-up of debris</td>
</tr>
<tr>
<td></td>
<td>HEAT EXCHANGERS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>design inability to meet load requirements</td>
<td>wrongly sized</td>
</tr>
<tr>
<td></td>
<td>operational reduced capacity</td>
<td>fouled tubes</td>
</tr>
<tr>
<td></td>
<td>operational reduced capacity and increased fan power</td>
<td>dirt on fins</td>
</tr>
<tr>
<td></td>
<td>operational fractured piping due to freezing of fluid</td>
<td>wrong control law</td>
</tr>
</tbody>
</table>

the reading given by a sensor is not representative of the physical property it is supposed to measure. If a sensor becomes disconnected, or is physically damaged, the reading will be outside a feasible range and hence easily detectable. It is more difficult to detect a malfunction when the reading becomes inaccurate but stays within its feasible range.

Hydraulic valves link the actuator and the heat exchanger. If valves are not very active, deposits on the internal surface can cause the friction within the valve to increase. More force is then required to overcome the increased friction and this can cause actuator failure, or a linkage decoupling. The fluid flowing through the valve often contains dirt, which may accumulate within the valve. This effect
can prevent the valve from fully opening or fully closing. In the former case the effective capacity of the heat exchanger will be reduced, which may lead to thermal comfort problems. If the valve cannot close fully, energy will be wasted since unwanted cooling or heating will be provided when the valve is supposed to be closed.

The two main operational faults in heat exchangers are due to the build-up of dirt, either on the inside of the tubes or on the external fins. Both of these faults will increase the overall thermal resistance between the two fluids and hence reduce the heating or cooling capacity. These faults also have secondary effects that increase energy consumption. Air-side fouling increases the resistance imposed on the fan while water-side fouling increases the resistance on the pump. A more catastrophic fault that was identified during the surveys is the freezing of a cooling coil, which is due to inadequate protection in the control strategy.

3.3 Design criteria

A FDD scheme has to be (partially) tailored to the application with which it is intended to be used. This section considers the main aspects of an FDD scheme that have to be application-specific. A number of objectives are defined that the FDD scheme design should satisfy to be appropriate for the considered subsystem.

3.3.1 Types of faults

Operational faults can be categorised as being either degradation- or failure-type. Conventional methods of fault detection involve checking that set-points are being maintained and that certain variables (e.g. pressures and temperatures) are within predetermined limits. These methods are sensitive to faults that lead to a significant change in the behaviour of the system. The methods can therefore be capable of detecting failure faults and long-lasting degradations. Faults that induce a small change, such as degradation faults in their early stages and failure faults in minor components, cannot generally be detected using limit-checking techniques. Fault detection is more difficult when the induced changes are small and this is where more advanced techniques are required. The primary objective
of the FDD scheme will therefore be the detection of degradation faults and other small faults that conventional methods would usually be unable to detect.

3.3.2 Performance indices

'Performance indices' are produced by the preprocessor component of an FDD scheme (Section 2.1.1). These indices form the basis of the fault decision process, and selection of the indices represents an important part of the design of an FDD scheme. The performance indices should, ideally, be selected so that they simplify the diagnosis task. The selection of indices should be based on the following criteria:

- to simplify the determination of thresholds;
- to enable faults to be easily distinguished.

These criteria can usually be satisfied by ensuring the performance indices have physical meaning. Thresholds may then be set heuristically and the type, and magnitude, of a fault can be assessed more easily.

A limit checking approach to fault detection considers only the instantaneous magnitude of change in a variable and ignores the cumulative effect that some faults can have. Many of the faults that develop in air-conditioning systems incur a cost penalty from the moment they occur and continue to do so until they are remedied. For example, a persisting fault can:

- increase the rate of energy consumption;
- reduce profits of an organisation due to decreased productivity of the workforce resulting from inadequate thermal comfort;
- increase the complexity and costs of an inevitable maintenance task.

The indices should therefore be of a form (or be capable of being easily transformed into a form) that enables their accumulated change over time to be physically meaningful. The types of indices that meet these requirements are additional,
3.3. Design criteria

unwanted, energy consumption, or the costs resulting from this. Consideration of the integrated effect of changes, along with the instantaneous magnitude of change, will also make the FDD scheme more sensitive to small changes that persist over long time periods.

3.3.3 Detection speed versus robustness

A fault is detected when a performance index transgresses a threshold. If the performance indices are not directly measurable properties, as in all but the simplest schemes, a procedure is required to calculate the indices from the measured variables. There can be uncertainties in this procedure such as noise, unmeasured disturbances, and inaccuracies in embedded models or physical equations. These uncertainties are reduced by making use of additional, corroborative, data samples (information), which, for the application of on-line FDD, are obtained sequentially in time. This approach causes the performance indices that are calculated to lag behind the ‘true’ value and thus affects the speed of detection. A threshold will also have to be selected with uncertainties in mind and a threshold that is too high (or low, depending on the index that is considered) will take longer to be transgressed if there is a fault. The potential speed of fault detection is therefore related to the uncertainty levels associated with the FDD scheme and the system to which it is applied.

The objective of the FDD scheme is to detect degradations and other small faults before the induced changes reach a level that would allow them to be detected with conventional limit-based methods. This represents a constraint on the speed of detection that should not be violated if the system is to be of use. The other objective that will influence the achievable speed of detection is the FDD scheme should generate only minimal false alarms.

3.3.4 Detail of diagnoses

By virtue of applying FDD procedures to individual subsystems, a fault detection facility can provide a degree of fault isolation, without performing any explicit fault diagnosis. A more detailed diagnosis can be obtained by:
3.3. Design criteria

- conducting a 'postmortem' analysis;
- using the information accumulated up to the time of the alarm.

The 'postmortem' approach involves investigating the cause of the fault by performing tests (e.g. by applying test signals and analysing behaviour, manual inspection, etc). The aim of this approach is to eliminate faults from a list of possibilities. This kind of analysis invariably involves disrupting the operation of the system to obtain information at different operating points. This is the only viable approach for failure faults since these occur suddenly and the resulting disruption of a postmortem analysis is normally unavoidable, and is warranted in order to eliminate the disruptive effects of the fault.

Degradation faults develop slowly and their symptoms can be apparent before their detrimental effects are sufficient to warrant generating an alarm. Diagnostic information can be accumulated before the fault reaches a critical level by relying on normal disturbances to exercise the system and provide information at different operating points. The criteria upon which a decision to carry out remedial action should be based can change from one installation to another, due to different cost constraints and different uses of the building. Assessment of these criteria, to select thresholds, is a non-trivial task. In most cases, it is better, therefore, to design the FDD scheme to make physically meaningful quantities available to the operators. The decision whether remedial action is required can then be left to the operator.

There is a need for both the postmortem and pre-alarm diagnosis procedures and each would perform a complementary function. This work will focus only on the latter, non-intrusive approach, as this is more closely linked to degradation faults.

---

4 Assuming that an alarm is only generated when remedial action is required.
3.4 Design solution

3.4.1 Reference of correct operation

The prime objective of the FDD scheme is to detect the faults that conventional procedures cannot, i.e. faults that result in small changes in the behaviour of the system. To detect small changes, an accurate reference of the 'correct' behaviour is required. For the system to be non-intrusive, normal disturbances have to be relied upon to exercise the system throughout its range of operation. A reference of correct operation must therefore be available to cover all the possible operating points so that the available information is best utilised. A quantitative mathematical model of the system satisfies these requirements, and has already been shown to be a viable tool for fault detection schemes (Isermann, 1995).

Having ascertained that a quantitative mathematical model meets the basic requirements of the FDD scheme, the most suitable type of model must now be determined. Firstly, the dichotomy relating to the type of model must be considered. Empirical models and analytical models both have been used for fault detection and each model type has particular advantages and disadvantages that need to be considered and evaluated in the context of this work.

Empirical models

Standard functional forms, such as polynomials, neural networks, etc, are used in empirical models. This therefore enables them to be applied to different systems without the need to alter the fundamental structure. The models are made to be system-specific by using training data to calibrate their parameters. Since only one aspect of empirical models, viz. the parameters, has to be tailored to each application, these models have a lower 'design cost' than do analytical models, which require both the structure and parameters to be system-specific.

The system-specific a priori information that is lacking in empirical models has to compensated for by the information inherent within the training data. The training data therefore has to contain examples of system inputs and outputs spanning the operating space. The operating space of the system is related to the
input space and thus increases in size exponentially with the number of inputs. For multiple input systems the amount of training data required to provide a dense coverage of the operating space can be very large and often practically unrealisable within realistic time frames (Fargus, 1994). In addition, some inputs may be uncontrollable, thereby preventing the acquisition of training data from certain regions of the operating space, e.g. ambient air temperature. The difficulty in obtaining training data thus stems from either the lack of time available to excite the system or the inability to exercise some of the inputs.

As a consequence of the generic structure of empirical models, they are normally unable to interpolate or extrapolate reliably. Hence, as the models are used outside the training data regions, their level of accuracy can deteriorate. The deterioration in accuracy is expected to be worse for empirical models than for analytical models, which are calibrated using training data. This is because analytical models employ physical constraints in their structure, which prevent variables from violating physical laws and thus prevent unrealistic predictions from being made outside the training data. If empirical models are used in fault detection systems, this deterioration in accuracy may be construed as a fault (Leonard et al., 1992). An additional disadvantage of using empirical models for fault detection is that the internal parameters and structure are physically meaningless and cannot be utilised to provide an insight into the nature of a fault.

**Analytical models**

Analytical models are derived by considering the physical cause and effect relationships between the inputs and outputs. Physical laws are used to derive the equations, such as energy and mass balances. These models have to be derived individually for each particular class of system or subsystem. Satisfactory performance can only be obtained from the analytical models if the parameter values are accurate. The parameters of analytical models often relate to physical properties of the real system and can therefore be estimated by direct measurement of the actual property (e.g. length), or by referring to design information. If the property cannot be measured or obtained from published material, it may be estimated empirically instead by using training data. Even when parameter values can be estimated from published information, previous work has shown that analytical models still benefit from empirical refinement of their parameter values.
3.4. Design solution

(Salsbury et al., 1994). There are two possible reasons for this:

- the installed system is different from that described in the design information;
- the model is not structurally correct, causing the parameters to be 'effective' rather than exact representations of physical properties.

As explained in the previous section, analytical models are generally better able to interpolate and extrapolate from regions of the operating space where training data are obtained due to the incorporation of physical constraints in the structure. Moreover, the models have the advantage of having physical meaningful parameters and structure, which facilitates the formulation of meaningful performance indices. Analytical models more closely match the requirements sought from the FDD scheme. These types of models have thus been selected for the FDD scheme reported in this thesis.

3.4.2 Modelling detail

Analytical models can be derived based on different levels of detail. At the most complex level, the interactions at micro-scale can be considered (e.g. computational fluid dynamic model) and the resulting model is then computationally intensive. It would not be possible to implement this level of detail in the current technology of air-conditioning controllers, and a simpler representation is therefore sought.

Although the potential to reproduce the behaviour of the real system is reduced by simplifying the mathematical form of the model, a simpler model has fewer parameters and is therefore easier to configure. Moreover, simpler models are less sensitive to 'holes' in the training data, thus allowing more reliable extrapolation and interpolation. If there is sparse training data, model simplification can yield better global accuracy, but at the expense of local accuracy. Akaike (1974) showed that the amount by which the accuracy of a model improved for a fixed change in complexity decreased in a logarithmic fashion, as shown in Figure 3.4.

---

5 For example, the mean of the absolute prediction errors.
6 Ascertained by the number of parameters.
Akaike derived an 'information criterion' to provide a basis for the selection of competing models. This criterion includes the number of parameters and the 'fit' of the model to the training data as the governing factors. The criterion was developed primarily for empirical (or statistical) models and it is less useful for analytical models, as some, or all, of the parameters of these models are meant to be estimated by non-empirical means. Nevertheless, the basic principle of Akaike's work remains applicable; that increased complexity generally has diminishing returns, in terms of accuracy. Models are therefore sought that produce the desired level of sensitivity from the FDD scheme with the minimum number of parameters.

### 3.4.3 Static versus dynamic

Another dichotomy that has to be considered is whether the model should be static or dynamic. This decision is best made by considering the operational behaviour of the system considered for fault detection and diagnosis. The disturbances on air-conditioning systems are in the form of external weather changes and variations in internal heat gains and losses within a treated zone. The ambient temperature usually follows a daily sinusoidal pattern and the rate of change is typically much slower than the dynamics of the system. Cloud movement can cause variations in the solar gains within the zones, which are faster than the system dynamics, but these are dampened by the thermal capacitance of the building materials. In
office buildings and other working environments, internal heat gains are the result of equipment and personnel and they remain fairly constant during a typical day. Buildings such as cinemas, conference centres, etc, can experience large and sudden changes in internal gains due to the movement of a large group of people. The operational pattern of most air-conditioning systems can therefore be summarised as being near to steady-state for most of the time, but having the potential for step-like transients at random intervals.

Because of the operational pattern of air-conditioning systems, valuable information for FDD purposes can be obtained by modelling the steady-state periods using static models. Static models do not have time derivative terms and can be written as algebraic expressions. This produces a major computational advantage and helps avoid some of the additional uncertainties associated with modelling the dynamic behaviour.

The transient periods in the mostly steady-state data must not be used by static models as these will lead to prediction anomalies that may be construed as faults. It is important therefore for the FDD scheme to recognise when transient periods are occurring and suspend operation until they are completed. The information available during the transient periods is sacrificed and in some applications these data can contain important information for fault diagnosis. However, all the faults that were identified in Section 3.2.1 for heat exchanger subsystems in HVAC plant will be apparent as a change in the static characteristic.

One potential problem with using only steady-state data can occur when the heat exchanger is used as part of a feedback control scheme and the controller is badly tuned. If the controller gain is too high for the system, steady-state may never be reached and the system will oscillate about the set-point. In this case, there will be a complete lack of steady-state data and all faults will be masked. This type of oscillatory behaviour is detrimental to the system, and will cause the valves and actuators to wear. The commissioning process should include the tuning of feedback controllers and poor tuning is the result of inadequate commissioning. It may be noted that a poorly tuned controller can be detected as part of an FDD scheme based on static models, by configuring the scheme to generate an alarm if there is a lack of steady-state data over a predetermined epoch.
3.4.4 Fault detection method

Faults are detected by monitoring performance indices and recognising when they transgress a threshold. The performance indices therefore have to be sensitive to changes in the behaviour of the system that are the result of a fault. Static models can be used to generate indices that are sensitive to faults based on parameter or output innovations. Figures 3.5 and 3.6 illustrate how these two types of innovations can be generated.

Figure 3.5: Output innovations-based approach

Figure 3.6: Parameter innovations-based approach

Previous work performed by the author and co-workers investigated the effective-
ness of the output-innovations approach for detecting faults in heat exchanger subsystems (Benouarets et al., 1993; Salsbury et al., 1995b). In these approaches, analytical models were used to predict the temperature of the air leaving the heat exchanger and then compared with the measured value. The magnitude of the difference between the predicted and measured temperature was compared with a threshold to determine whether to generate an alarm.

The main problem encountered was how to determine a threshold value that did not lead to the generation of false alarms, and yet maintained maximum sensitivity to fault-induced discrepancies. The problem was addressed by combining the analytical model with a linear-in-the-parameters radial-basis function model, as shown in Figure 3.7. The training data was used to optimise the fit of the combined model in a least-squares sense. This process then facilitated the derivation of a statistical threshold, which was variable across the range of operation, being a function of the density of the calibration data and the local accuracy (Salsbury, 1995).

![Figure 3.7: Two-tier model](image)

The approach was successfully applied to detect small faults, e.g. a 1% flow leakage through a heat exchanger. It was found that the sensitivity of the approach was strongly related to how well the model represented the real system. The sensitivity could be improved significantly by increasing the amount and quality of the calibration data. The approach becomes less attractive, however, as the need for data increases.

The parameter-innovation approach has also been explored by the author and co-workers (Haves et al., 1996b; Salsbury et al., 1995c). Analytical models of heat exchanger subsystems were utilised in these approaches and their parameter values
3.4. Design solution

Estimated using data from the system. Estimating the parameters of analytical models has the distinct advantage that the physical meaning of the parameters can allow detection thresholds to be determined more easily. However, analytical models are generally not linear-in-the-parameters and the parameter estimation task is not analytically soluble. Non-linear optimisation methods therefore have to be used to estimate parameter values, and these can suffer robustness problems. Two methods for estimating the parameter values have been investigated:

- indirect estimation using radial-basis functions (Haves et al., 1996b);
- recursive prediction error method (Salsbury et al., 1995c).

In the first approach, direct estimation of the analytical model parameters is avoided by estimating the parameters of an intermediate model that is linear-in-the-parameters. This intermediate model is used periodically to generate data covering the complete operating range of the system. These data are then used in the estimation of the parameters of the analytical model. Figure 3.8 is a schematic of the scheme. The RBF models the local behaviour of the target system and is updated using the recursive least-squares algorithm when the system is (approximately) in steady state. The RBF is then exercised over the operating range of the system. The data that is generated is then used as part of a non-linear batch optimisation process to estimate the parameters of the analytical model. Because the RBF is a local model, it provides an estimate of the most recently observed behaviour of the system in different parts of the operating space, responding relatively quickly to fault-induced changes.

The method was found to be reasonably robust, but the use of two models (RBF and analytical model) served to increase the uncertainty in the final parameter estimates. The inaccuracies associated with each of the models are unavoidably compounded and there is no obvious analytical means by which to estimate the resulting uncertainties. A further disadvantage with the approach was that it was computationally intensive (particularly the batch optimisation) and this raises questions regarding practical implementation in the current technology of air-conditioning control systems.

The second approach involves estimating the parameters of the analytical model directly, by using the prediction error of the model to govern the parameter adjustment. The method allows the uncertainty in the parameter estimates to be
3.4. Design solution

ascertained more easily than with the combined physical model and RBF model, and the method was successfully applied to the detection and tracking of three degradation faults in heat exchanger subsystems. The estimator became unreliable when the behaviour of the system changed in a sudden and significant manner (i.e. in the case of failure faults). Although the diagnosis of failure faults was unreliable, the method was sensitive to these faults in a way that would allow them to be detected, if not diagnosed.

For any parameter estimation method, the accuracy of the estimated parameters is dependent mostly on the structural accuracy of the model. The overall objective of a parameter estimator is normally to minimise the prediction error, and structural inadequacies in the model lead to errors. The parameters that are estimated will be sensitive to these errors and will fluctuate as the model is exercised through regions of varying accuracy. The amount by which the parameters fluctuate relates to the uncertainty in the parameters due to the inherent structural inadequacy. This band of variation has to be taken into account when determining thresholds. The structural precision of the model therefore directly affects the potential sensitivity of the fault detection system.
Another important factor affecting the performance of an FDD scheme, based on parameter estimation, is the accuracy of the reference parameters, which correspond to the correctly operating system. The innovations (indications of change) are generated by comparing the estimated parameters with the reference parameters. These reference parameters are supposed to be the parameter values that enable the model to reproduce the behaviour of the correctly operating system. A training phase may not be required if these values can be accurately obtained from design and manufacturers' data. Otherwise, the reference values can be determined by either:

- using training data from the system when subjected to artificially generated disturbances, gathered before the fault detection routines are put into operation; or
- by heuristically determining some initial values and then allowing these estimates to be refined during a designated training phase where normal disturbances are the source of system excitation.

In spite of the potential robustness problems with the direct parameter estimation method, the approach has the following specific advantages:

- computationally undemanding;
- suited to the detection of degradation faults;
- capable of detecting failure faults;
- estimation of physically meaningful parameters simplifies diagnosis task and the setting of thresholds.

Due to these advantages, the direct parameter estimation method has been selected as the basis for the FDD scheme considered in this thesis.

### 3.4.5 Fault diagnosis method

The parameter estimation approach greatly simplifies the diagnosis task and helps to avoid some of the arbitrariness associated with the generation of diagnoses in
schemes based on output innovations. The only directly available way to determine
the nature of a fault from innovations of the output is to consider:

1. The size of the innovation, which is correlated to the size of the fault.
2. The direction of the innovations.

In heat exchangers, many of the faults that occur change the characteristic of the
system in a non-linear way, and the size and direction of the innovation vector
can vary across the operating range. This greatly increases the complexity of
the diagnosis task and is responsible for most of the arbitrariness that has to be
introduced.

In the output-innovation approach developed and tested by the author and co-
workers, the detection aspect was designed to be as sensitive as possible in order
to detect a fault when any change was outside the uncertainty bounds. Different
faults were distinguished by an implicit process of elimination that relied on
natural disturbances to exercise the system and provide the innovation informa-
tion required. The variation in the innovations across the operating range was
characterised by using ‘bins’, the number of which was determined by the dimen-
sionality of the operating space and consideration of the faults that needed to
be distinguished. These bins were used to store the average innovation values in
mutually exclusive regions of the input space. The positioning and sizing of the
bins were done in a heuristic way relying heavily on expert knowledge of how the
symptoms of the faults are manifested. The approach was used successfully to
distinguish three faults: water-side coil fouling; valve leakage; sensor drift. It may
be noted, however, that a significant amount of empirical refinement of thresh-
olds was necessary to achieve the desired sensitivity. The scheme is illustrated in
Figure 3.9.

The parameter estimation approach simplifies the diagnosis task since the esti-
mated parameters can be selected to relate to physical properties of the system
that are influenced directly by the occurrence of faults. The estimated parameters
are compared with the reference values corresponding to the correctly operating
system to produce the parameter innovations. These innovations can provide an
insight into the extent of the fault in understandable physical terms. The FDD
scheme developed in this thesis is based on parameter estimation and analytical
models, and the proposed scheme is presented in the following section.
3.5 Proposed FDD scheme design

Figure 3.4 schematically depicts the proposed FDD scheme, which incorporates the techniques discussed in the previous section. The scheme is based on estimating the parameters of a static model using the measured inputs and outputs from the real system. The static model predicts the outputs of the system and these are compared with the measured values. The prediction error is then passed to a parameter estimator, which adjusts the parameters of the model. A steady-state detector is included in the scheme so that the parameters of the static model are only updated when the real system is in steady-state. The parameters that are estimated relate directly to the faults of interest, and thus provide an indication of the degree of fault for detection and diagnosis purposes.

Each of the individual components of the scheme is now derived and investigated in detail in the next two chapters. Chapter 4 derives the models of the system and Chapter 5 derives the signal processing and parameter estimation components.
Figure 3.10: Schematic of the proposed FDD scheme
Chapter 4

Subsystem models

Introduction

The FDD scheme that is described in this thesis is applied to heat exchanger subsystems of the type found in the air-handling units of air-conditioning systems. These subsystems consist of three components: actuator, hydraulic valve, and heat exchanger. Models of these components are derived in this chapter using physical theory and established correlations. The majority of the model parameters are physically meaningful and can be initially estimated from the design information that is normally available. Some empirical parameters are, however, included in the models and training data from the considered system are required to estimate their values. The models are considered in the context of the FDD task, and 'fault parameters' are selected for estimation by the FDD scheme. The selection of the parameters is based on consideration of diagnostic relevance, fault priorities, and the robustness implications to the estimation process. Finally, the results obtained from testing the models with experimental data are presented.

4.1 Components of heat exchanger subsystems

Figure 4.1 shows the components of a heat exchanger subsystem. A three-port valve is shown in the figure, which is the most common type of valve used with
heat exchangers in central air-handling plant.

Table 4.1 lists the variables associated with the operation of the subsystem. The variables labelled with an asterisk are those that are normally measured, or are able to be calculated easily from other measured properties.

The purpose of developing a model of the subsystem is to allow the relationship between the inputs and outputs to be characterised. Hence, the model should be able to predict the measured output(s) based on the measured inputs. In most air-handling units, the only measured output is the temperature of the air leaving the coil, $T_{ao}$. A static model is used and the dynamic interactions between variables are therefore not irrelevant. A static function, $f(.)$, is therefore required such that:

$$T_{ao} = f \{T_{ai}, T_{wi}, H_{ai}, \dot{m}_o, \dot{m}_w, u\}.$$  \hspace{1cm} (4.1)

A model of the heat exchanger alone cannot be used as the mass-flow rate of water into the coil is not a measured variable. This flow rate is governed by the valve.

---

**Figure 4.1:** Water to air heat exchanger

The diagram shows a simple heat exchanger system with air entering at $T_{ai}$ and $H_{ai}$, passing through the heat exchanger, and exiting at $T_{ao}$ and $H_{ao}$. Water enters at $T_{wi}$ and $H_{wi}$, flows through the heat exchanger, and exits at $T_{wo}$ and $H_{wo}$. The actuator and 3-port valve control the flow rate and temperature.
which is controlled by the actuator. The control signal to the actuator represents the measured variable that is most closely related to the flow rate. Models of the valve and actuator are therefore required to predict the water flow rate from the control signal, and facilitate the use of a heat exchanger model. Static models of the actuator, valve, and heat exchanger are derived in the Sections 4.2.1-4.2.3. When used together, the models are able to approximate the function $f(\cdot)$ given in Equation 4.1.

### 4.1.1 Designing the models for fault diagnosis

As discussed in Section 3.4.1, analytical models are used in the FDD scheme developed in this thesis. These models have physically meaningful parameters and thus allow initial estimates of the parameter values to be made from information that is normally available in design and manufacturers’ data. The objective of the FDD scheme is to track the development of degradation faults by estimating the model parameters using sampled input-output data. To achieve this objective:

- the models have to be capable of representing a correctly operating system so that changes in the characteristics due to faults can be distinguished from modelling errors;
4.1. Components of heat exchanger subsystems

- the model has to have parameters that, when adjusted, allow the model to represent the characteristic of the system when affected by the degradation faults of interest.

The component models are intended to be used together to represent a heat exchanger subsystem. Hence, the predictions (outputs) made by one particular component model have to be sufficiently accurate so that the accuracy of another component model that uses predictions as inputs is not jeopardised. The implication of this is that the level of accuracy sought from a particular component model is not solely determined from the faults that are likely to occur in that component. The effects of faults, such as a sensor off-set, are manifested across the range of operation of the heat exchanger subsystem. A global subsystem model (and hence component models) is thus required to detect and track a sensor off-set and be able to distinguish it from other faults.

The attributes that the models need to have in order to diagnose the faults of interest are considered in the sections below. Suitable models based on the specified requirements are then derived in Section 4.2.1-4.2.3.

**Actuator model**

All the faults identified in Table 3.1 for electric motor actuators are of the failure type. Discrete parameters (binary) would be required to model the changes caused by failure faults. Parameter estimation algorithms are generally not suited to models having discrete parameters and these faults will not be considered here. Specific fault parameters are therefore not incorporated in the actuator model.

**Valve model**

Some of the typical (and important) operational faults that can occur in three-port valves were discussed in Section 3.2.1; these are:

- leakage;

- exit flow constriction (fouling);
4.1. Components of heat exchanger subsystems

- sticking.

Leakage is apparent in the region where the valve is almost closed. The model of the correctly operating valve therefore has to be accurate enough in this region to allow leakage to be detected. Moreover, the model has to have a parameter that corresponds to the degree of leakage through the valve to allow the development of leakage to be tracked by the FDD scheme.

The build-up of matter at the exit port of a valve can have the effect of changing the characteristic of the valve across its range of operation. The valve model therefore has to be able to represent the correctly operating valve across its entire range of operation if this fault is to be detected and diagnosed correctly. It is therefore necessary to include parameters that relate to flow resistances so the effects of this fault can be tracked if required.

In the case of a sticking valve, the effective operational range of the heat exchanger subsystem is reduced to zero. This is a failure fault and will therefore not be considered explicitly. It may be noted, however, that any parameter estimator based (directly or indirectly) on the magnitude of the model prediction error will be sensitive to faults that result in a change to the characteristic of the system. Hence, if there is a failure fault, other parameters may be adjusted by the estimator in an attempt to make the model match the observed behaviour at the stuck operating point. This would therefore allow such faults to be detected, although not diagnosed.

Heat exchanger model

The most important degradation fault encountered in heat exchangers is fouling. Fouling is a build-up of matter on either the air- or water-side of the heat exchanger and it has been discussed in detail in (Chenoweth and Impagliazzo, 1981; Bott and Bemrose, 1983). Fouling leads to an increase in the thermal resistance of the coil wall and a reduction in the maximum duty of the heat exchanger. The fault has thermal comfort implications when the heat exchanger is operated at high duty. In practice, coils are often oversized and because of this are operated mostly at low duties. In this situation, fouling may not affect thermal comfort as set-points can still be maintained via a greater opening of the control valve.
A parameter that relates to the thermal conductance of the coil is required to allow the build-up of fouling to be tracked.

Temperature sensors

A common fault identified in Section 3.2.1 was sensor drift. This fault causes the reading obtained from the sensor to deviate from the true value of the physical property it is supposed to measure. Sensor signals are used as both inputs and outputs to the models. An error in any of these sensors changes the nature of the relationship between the inputs and outputs. If a sensor error were to occur, prediction errors would result. If sensor errors are not explicitly modelled other parameters would vary due to the parameter estimator attempting to reduce the prediction errors. Hence, these faults would be detectable even if they are not explicitly modelled. As sensor faults are relatively easy to model they can be represented by using a 'sensor model'. For example, consider the sensor that measures the temperature of air leaving a heat exchanger:

\[
\hat{T}_{ao} = T_{ao} + b, \tag{4.2}
\]

where \(\hat{T}_{ao}\) is a prediction of the sensor reading, \(T_{ao}\) is the physical property, i.e. as predicted by the heat exchanger model, and \(b\) is the sensor off-set parameter.

4.2 Component models

This section presents component models for finned-tube, water to air, heat exchanger subsystems. The models are derived using physical laws and established correlations. Correspondingly, the models have physically meaningful parameters that can be initially estimated from system design information, or from information obtained during commissioning.

4.2.1 Actuator model

Control signals issued by the control system are received by the actuator virtually instantaneously. The actuator then moves the valve to the position requested by
the controller over a period of time. This dynamic behaviour can be ignored since a static representation is sought. Ideally, the control signal to the actuator should be linearly related to the fractional position of the valve, when in steady-state. However, in practice, slack may exist in the linkage between the valve stem and the actuator. If this is the case, the valve will fail to move until the actuator has moved by an amount sufficient to overcome the slack. Slack is then re-introduced when the direction in which the valve is moving changes. This hysteresis effect jeopardises the accuracy of the predictions if it is assumed that the valve stem position \((s)\) is equivalent to the control signal \((u)\). The hysteresis model described by Clark (1985) is therefore used in the actuator model. A graphical representation of the model is shown in Figure 4.2.1.

The hysteresis is modelled by means of the following algorithm:

\[
\text{IF } (u_k - s_{k-1}) > v \text{ THEN } s_k = u_k - v\\
\text{ELSE IF } (u_k - s_{k-1}) < 0 \text{ THEN } s_k = u_k\\
\text{ELSE } s_k = s_{k-1}
\]

The valve stem position is then mapped onto the range \(0 \leq s \leq 1\):

\[
s = \frac{s_k}{1 - v}.
\]
The hysteresis model has one single parameter, \( v \). The previous valve stem position also has to be stored so the amount of slack is known. Ideally, the sampling frequency of the FDD scheme (and hence the actuator model) would be equivalent to that of the controller so any reversals are detected. In practice, the sampling frequency used by the FDD scheme is likely to be less than that of the controller. The assumption therefore has to be made that there are no reversals between the samples.

### 4.2.2 Valve model

The objective of the valve model is to predict the flow rate of water into the coil from the valve stem position, which is predicted by the actuator model. Three-port valves are favoured over two-port valves in air-conditioning systems as they allow a constant flow rate to be maintained in the primary circuit (see Figure 3.2). The way in which the flow rate through the relevant port varies with the valve stem position is known as the valve characteristic. The installed characteristic of a valve depends on the inherent characteristic and the other resistances in the flow circuit.

### Preliminaries

Any type of constriction that impedes the flow of fluid induces a pressure drop. For the high Reynold's numbers that are associated with the flows in the primary circuit of air-conditioning systems, the drop in pressure can be assumed to be proportional to the velocity head of the flow:

\[
\Delta P \propto \frac{v^2}{2g} 
\]

\[
\Delta P = \xi \rho \frac{v^2}{2g},
\]

(4.5)

where \( v \) and \( \rho \) are the velocity and density of the fluid respectively, and \( \xi \) is a dimensionless resistance coefficient. To avoid having to specify cross-sectional areas of tubes and pipes it is more useful to express the pressure drop in terms of the mass flow rate (\( \dot{m} \)):

\[
\Delta P \propto \dot{m}^2
\]

\[
\Delta P = R\dot{m}^2,
\]

(4.6) (4.7)
where $R \text{ (kg}^{-1} \cdot \text{m}^{-1})$ is a resistance coefficient.

### Inherent characteristic

The inherent characteristic describes the way in which the fluid flow rate varies with the valve position when there is a constant pressure drop across the valve. Information about this characteristic for a particular valve is usually available from the valve manufacturer. Two flow characteristics are standardised\(^1\): linear and equal percentage. These characteristics are defined by the Equations 4.8 and 4.9, and valves are usually manufactured to approximate these forms:

$$f(s) = s, \text{ linear} \quad (4.8)$$

$$f(s) = \exp(\beta(s - 1)), \text{ equal percentage,} \quad (4.9)$$

where the function, $f(s)$, is the fractional flow at the stem position, $s$; i.e.

$$f(s) \triangleq \frac{\dot{m}_{w,c}}{\dot{m}_{w,d}}, \quad (4.10)$$

$\beta$ is a parameter that determines the curvature of the exponential characteristic, and $\dot{m}_{w,d}$ is the maximum flow rate through the valve. The characteristics resulting from the use of these equations are shown in Figure 4.2.2.

![Figure 4.3: Standardised valve characteristics](image)

The linear and exponential equations have been used to model the behaviour of valves; e.g. (Clark, 1985). It may be noted (and confirmed from Figure 4.2.2) that

\(1\)According to IEC534-1.
Equation 4.9 results in a leakage through the closed valve, the extent of which is determined by curvature parameter. Modern valves are manufactured so that they have negligible leakage at the fully closed position. The characteristic of these valves therefore deviates from being purely exponential as the valve approaches its fully closed position. This effect is illustrated in Figure 4.2.2, which shows a possible close-off characteristic for a modern valve, indicated as a dashed line.

Figure 4.4: Close-off characteristic of a modern valve

The basic exponential equation would thus result in an over-estimation of the leakage at the closed valve position. Haves (1994) describes one approach to modelling an improved close-off characteristic, based on the specification of a major valve manufacturer. The approach involves linking a linear part to the exponential curve allowing a more accurate approximation of the close-off characteristic, while maintaining the exponential curvature over the majority of the range. A modification to the basic exponential equation is proposed here that requires fewer parameters than the model described by Haves. The proposed model allows a leakage to be specified (for fault diagnosis purposes) and the function has a smooth characteristic, which is beneficial for parameter estimation procedures that are based on derivative information. The model is given by:

\[
f(s) = l + (1 - l) \frac{1 - e^{\beta s}}{1 - e^{\beta}}, \quad \text{for } \beta \neq 0, \tag{4.11}
\]

\[
f(s) = l + (1 - l)s, \quad \text{for } \beta = 0. \tag{4.12}
\]

Figure 4.5 shows how the characteristic of the proposed model differs from the
standard exponential model. The leakage in the proposed model is fixed at zero, while the leakage in the standard model is determined by the curvature parameter. The characteristics are shown for a curvature parameter of three, which is most common in practice. The characteristics are the same when \( s = 1 \), but they diverge as \( s \to 0 \), with the modified model giving a more realistic close-off characteristic.

![Exponential valve characteristics (standard and refined)](image)

Figure 4.5: Exponential valve characteristics (standard and refined)

A three-port model is developed by taking account of the other resistances in the flow circuit. A suitable model for the considered application is described in the following section.

**Installed characteristic**

The inherent characteristic of the valve describes the relationship between flow and valve stem position at a constant pressure. When a valve is installed in a circuit in which the flow varies, the pressure drop across the valve also varies. Variations in the pressure drop across the valve cause the actual (installed) characteristic to differ from the inherent characteristic. A dimensionless quantity known as authority is commonly used to describe the degree by which the installed characteristic differs from the inherent characteristic. The authority, \( A \), is given by:

\[
A = \frac{\Delta P_v}{\Delta P_v + \Delta P_s} = \frac{R_v(0)}{R_v(0) + R_s}, \quad \text{where } 0 \leq A \leq 1,
\]

where \( \Delta P_v \) is the pressure drop across the valve, and \( \Delta P_s \) is the pressure drop across the other components in series with the flow port of the valve. \( R \) is the...
flow resistance or pressure loss coefficient; $R_v(0)$ for the open valve, and $R_s$ for the other components in the circuit. A valve with a unity authority has an installed characteristic that is identical to its inherent characteristic. A lower authority leads to a greater difference between the inherent and installed characteristics. Typical values of $A$ are in the range 0.5 – 0.8.

To derive an expression for the installed characteristic of the valve, a simplified schematic of a three-port valve circuit is considered, which is depicted in Figure 4.6. The resistance associated with the heat exchanger is denoted by $R_c$, the bypass valve port by $R_{b,b}$, the control port by $R_{v,c}$, and a balancing valve by $R_b$. The mass flows are denoted by $\dot{m}_{w,c}$, which is the variable flow into the coil, and $\dot{m}_{w,d}$, which is the maximum flow through the coil (usually specified in the design data).

![Figure 4.6: Three-port valve circuit](image)

A pressure balance across the flow route through the coil and the control port of the valve yields:

$$\Delta P = \dot{m}_{w,c}^2 [R_c + R_{v,c}(s)]. \quad (4.14)$$

Three-port valve circuits used with coils in air-conditioning systems are usually commissioned so that the total pressure drop across the circuit is constant. By assuming that $\Delta P$ remains constant, a pressure balance at the maximum flow rate (denoted by $\dot{m}_{w,d}$) can be equated to Equation 4.14 such that:

$$\dot{m}_{w,d}^2 [R_c + R_{v,c}(0)] = \dot{m}_{w,c}^2 [R_c + R_{v,c}(s)]. \quad (4.15)$$
where $R_{v,c}(0)$ is the resistance across the control port when it is open, and $R_{v,c}(s)$ is the resistance across the control port at the stem position $s$. The fractional flow, modified by the other resistances, will be denoted by $f'(s)$, and is thus given by:

$$f'(s) = \left( \frac{R_v + R_{v,c}(0)}{R_v + R_{v,c}(s)} \right)^{\frac{1}{2}}. \quad (4.16)$$

The authority of the control port of the valve can be defined as:

$$A = \frac{R_{v,c}(0)}{R_{v,c}(0) + R_v} \quad (4.17)$$

Hence the resistance across the coil can be given in terms of the authority:

$$R_c = R_{v,c}(0) \left( \frac{1}{A} - 1 \right). \quad (4.18)$$

An expression for the fractional flow into the coil, modified by the effect of the authority, can now be obtained by substituting Equation 4.18 into Equation 4.16:

$$f'(s) = \left( \frac{\frac{1}{A} - 1 + 1}{\frac{1}{A} - 1 + \frac{R_{v,c}(0)}{R_{v,c}(0)}} \right)^{\frac{1}{2}} \quad (4.19)$$

$$= \frac{1}{\sqrt{1 + A (f^{-2}(s) - 1)}} \quad (4.20)$$

where the function $f(.)$ is defined according to Equation 4.11. The mass flow rate of the fluid into the heat exchanger can thus be calculated using the parameters listed in Table 4.2.

Table 4.2: Parameters used by the valve model

<table>
<thead>
<tr>
<th>SYMBOL</th>
<th>MEANING</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l$</td>
<td>fractional leakage through the control port</td>
</tr>
<tr>
<td>$\beta$</td>
<td>curvature parameter of the control port</td>
</tr>
<tr>
<td>$A$</td>
<td>authority of the control port</td>
</tr>
</tbody>
</table>

It may be noted that the resistance coefficient used in the valve model equations is not the same as the $K_v$ value (or flow coefficient) that is often provided in practice. The $K_v$ value is defined by:

$$K_v = 3.6 \cdot 10^4 \frac{\dot{m} \sqrt{\rho}}{\sqrt{\Delta P}}, \quad (4.21)$$
where \( \dot{m} \) is a flow rate in m\(^3\)s\(^{-1}\), \( \Delta P \) is a pressure drop in Pascals, and \( \rho \) is the fluid density in kg/m\(^3\). The relationship between this flow coefficient and the resistance coefficient used in the model is given by:

\[
R = 1.296 \times 10^9 \frac{\rho}{K_v^2}.
\]

(4.22)

The \( K_v \) values for the relevant components in the circuit should be obtainable from design or manufacturer’s literature, thus allowing the authority of the valve to be calculated.

### 4.2.3 Heat exchanger model

Air-handling units in air-conditioning systems usually contain two finned tube heat exchangers, one for heating, and one for cooling and dehumidification. The inherent characteristics of heating and cooling coils differ due to differences in their sizing and geometrical arrangement. Moreover, the mass transfer involved in dehumidification complicates the modelling of cooling coils. A number of heating and cooling coil models have been proposed for simulation and performance assessment purposes. Contributions have been made by the IEA Annex 10 (System Simulation) and Annex 17 (Building Energy Management Systems) groups; other contributions have been made by (Elmahdy and Biggs, 1979; Holmes, 1982; Sauer and Ganesh, 1988; Ding et al., 1990; Braun, 1988). Models of heating and cooling coils suitable for fault diagnosis are presented in the sections that follow.

**Preliminaries**

A heat exchanger model can be developed around the concept of the overall conductance \((UA)\), which is defined in terms of the total thermal resistance to heat transfer, between the two fluids:

\[
\frac{1}{UA} = \frac{1}{(\eta_o h A)_a} + R_w + \frac{1}{(h A)_w},
\]

(4.23)

where the subscripts \( a \) and \( w \) denote the air- and water-sides respectively. \( h \) is convective resistance, \( R_w \) is the conductive resistance through the material separating the two fluids, and \( \eta_o \) is the overall surface efficiency, which is calculated from fin
4.2. Component models

efficiency. For standard fin configurations, the appropriate coefficients needed to calculate the $UA$ of a coil can be found in (Kays and London, 1965).

Once the $UA$ of the coil is known, the heat transferred between the fluids can be calculated using the NTU-effectiveness method (Incropera and De witt, 1990), where the rate of heat transferred between the two fluids in the heat exchanger is calculated from:

$$Q = \varepsilon C_{\text{min}}(T_{hi} - T_{ci}),$$

(4.24)

where $h$ and $c$ denote the hot and cold fluids respectively, and

$$\varepsilon = \frac{Q}{Q_{\text{max}}} = f(NTU, C_r).$$

(4.25)

$$NTU = \frac{UA}{C_{\text{min}}}$$

(4.26)

$$C_r = \frac{C_{\text{min}}}{C_{\text{max}}}.$$  

(4.27)

where $C = \dot{m}c_p$ is the heat capacity, and $C_{\text{min}}$ and $C_{\text{max}}$ are the minimum and maximum of the air and water heat capacities. These equations thus allow the outlet conditions to be calculated by using the input temperatures only, without the need for iteration. The function describing effectiveness is dependent on the flow arrangement and different functions are normally used to represent the heating and cooling coils in air-conditioning systems. The temperatures of the fluids leaving the coil can be calculated by applying an energy balance; e.g. for the air-stream:

$$C_a(T_{ao} - T_{ai}) = \varepsilon C_{\text{min}}(T_{wi} - T_{ai}),$$

(4.28)

$$T_{ao} = T_{ai} + \varepsilon \frac{C_{\text{min}}}{C_a}(T_{wi} - T_{ai}).$$

(4.29)

**Heating coil**

Heating coils can be assumed to have a cross-flow configuration in which the water-flow is mixed and the air-flow unmixed. The effectiveness function for this configuration depends on which fluid has the minimum heat capacity rate; i.e.

$$\varepsilon = \left(\frac{1}{C_r}\right)(1 - \exp\{-C_r[1 - \exp(-NTU)]\}) \quad \text{for } C_{\text{min}} = C_a$$

(4.30)

$$\varepsilon = 1 - \exp(-C_r^{-1}\{1 - \exp[-C_r(NTU)]\}) \quad \text{for } C_{\text{min}} = C_w.$$  

(4.31)
The effectiveness can then be used to calculate the exit temperature of the air by performing an energy balance (Equation 4.29).

Heating coils in air-handling units provide sensible heating only; the moisture content can therefore be assumed to remain constant across the coil. The performance of the model depends on the accuracy of the overall heat transfer coefficient. For an unfouled coil, $R_w$ will be negligible in comparison to the convection resistances. Evaluation of the $UA$ is not as straightforward however, since the convection coefficients are complex functions of the fluid properties, surface geometries and flow conditions. There is therefore a large number of independent variables affecting the convection coefficients, which ultimately results from the fact that the convective heat transfer is determined by the boundary layers that develop on the surfaces. The relatively complex geometry of heat exchanger surfaces and the resulting complex flow conditions make a theoretical derivation of the convection coefficients infeasible. This has encouraged researchers to develop empirical correlations based on experimental data as an alternative. Different correlations have been proposed of varying complexity, e.g. (Stephan and Gruschka, 1994). A simple correlation described in (Holmes, 1982; Clark, 1985) is:

\[
(\eta hA)_a = \kappa_a \dot{m}_a^{a1} \\
(hA)_w = \kappa_w \dot{m}_w^{a2},
\]

More specifically, it has been suggested by Holmes that a fixed exponent value of 0.8 for both $a1$ and $a2$ is a reasonable approximation for the range of heat exchangers found in practice. This then reduces the total number of tunable parameters in the heating coil model to just three ($R_w$, $\kappa_w$, and $\kappa_a$), thereby greatly simplifying the parameter estimation process.

### Cooling coil

Cooling coils will dehumidify the air stream if the dew point temperature of the air is greater than the temperature of the coil surface. The phase change invalidates the energy balance given in Equation 4.28 and account has to be taken of the latent as well as the sensible heat transfer.

The air-side surfaces of a cooling coil can be in one of three possible conditions: dry, wet, or partially wet. The condition of the coil can be determined by comparing
4.2. Component models

the dew-point with the surface temperature of the coil. The surface temperature of the coil will vary, and some regions may be above the dew point while other regions are below. In this situation, the coil would be partially wet and the performance of the coil can only be predicted accurately by determining the fraction of dry and wet areas. The ASHRAE 3-line method (ASHRAE, 1988) is based on this approach and involves tracking the surface temperature along the air-flow path to locate the wet/dry interface. The inherent complexity of this method makes it an unattractive option for an on-line application such as fault detection. A simpler approach known as the 'sensible heat ratio' method is adopted here, which can be used to model wet and partially wet coils by assuming a fictitious overall surface temperature; this method is explained below.

Sensible heat ratio (SHR) method

The heat transfer rate for a wet coil depends on the moisture content of the air. As with a dry coil, the heat transfer rate can be expressed in terms of the effectiveness, $\varepsilon$, and the minimum capacitance rate, $C_{\text{min}}$. There are two notable ways of expressing the total heat transfer of a wet coil in terms of effectiveness: sensible heat ratio (SHR), and wet-bulb temperature difference (WBTD) (Ding et al., 1991). The SHR method is based on using the dry-bulb temperature difference as the driving potential, and the total heat transfer under wet conditions is calculated using the SHR as follows:

$$\dot{Q}_{\text{total}} = \varepsilon \cdot \frac{C_{\text{min}}}{\text{SHR}} \cdot \Delta T_a,$$

where $\Delta T_a$ is the temperature difference of the air across the coil. The WBTD method uses the difference between the wet bulb temperatures of the air as the driving potential:

$$\dot{Q}_{\text{total}} = \varepsilon' \cdot C'_{\text{min}} \cdot \Delta T'_a,$$

where $C'_{\text{min}}$ and $\varepsilon'$ are both calculated using modified heat capacities; $\Delta T'_a$ is the wet bulb temperature difference of the air across the coil. These two methods, and the ASHRAE 3-line method, have been compared and validated with experimental data from two heat exchangers by Stephan and Gruschka (1994). Theoretically, the methods are identical if an infinitesimally small area is considered, but the methods have functional differences when the whole heat exchanger is treated. However, results presented by Stephan and Gruschka indicate that the differences...
4.2. Component models

in performance between the approaches are small. Each method was able to pre-
dict the heat transfer rate more accurately at high heat transfer rates (maximum
relative error \( \approx 10\% \)) than at low heat transfer rates (maximum relative error \( \approx 20\% \)). Overall, the SHR method was found to be the most accurate.

The wet-bulb temperature difference method has the apparent advantage of not
requiring iteration to predict the heat transfer rate, and this makes it more attrac-
tive for on-line implementation. However, it can be shown that calculation of the
wet-bulb temperatures from the available measurements also requires iteration.
This iteration is often overlooked since it is hidden from a model user by being
incorporated in library functions that are used to evaluate relevant psychrometric
properties. Overall, there is not a significant difference between the processing
requirements of the WBTD and SHR methods.

The SHR method was the predecessor to the 3-line method in the ASHRAE stan-
dards and it therefore represents a well-established approach that has been widely
used. Although it was replaced by the 3-line method, which allows wet and dry
areas to be treated separately, work by Braun (1988) and Holmes (1982), in partic-
ular, has shown that satisfactory (and comparable) performance can be obtained
from the SHR method when the coil is both fully wet and partially wet. In view of
these results, it has been decided to adopt the SHR method for the fault detection
system.

A simple test is first carried out to see whether the coil is (partially) wet. The coil
is taken to be wet (or partially wet) if the dew point of the inlet air is above the
surface temperature of the coil \( (T_s) \), where \( T_s \) is calculated using:

\[
T_s = \frac{T_{ao} - b_f \cdot T_{ai}}{1 - b_f},
\]

where \( b_f \) is the bypass factor, which is defined as:

\[
b_f = \exp(-\eta h A) a / C_a \]

\[
= \exp(-\kappa a \dot{m}_a^{0.8} / C_a). \]

The bypass factor is based on the theory that a portion of the air flow passes through
the heat exchanger with its temperature unchanged. The remaining fraction of air
is reduced to the (average) surface temperature. If \( T_s \), calculated in Equation 4.36,
happens to be above the dew point of the inlet air\(^2\), the dry coil model is used

\(^2\)The dew point of the inlet air is calculated from the dry bulb temperature and the relative
to predict $T_{ao}$. Otherwise, the SHR model is used to calculate $T_{ao}$ for the wet condition. If the coil has more than four rows of tubes, a counterflow configuration can be assumed (Clark, 1985). A dry coil model for this configuration is then given by:

$$\dot{Q}_{dry} = \varepsilon C_{min}(T_{ai} - T_{ci})$$

$$\varepsilon = \frac{1 - \exp(-NTU(1 - C_r))}{1 - C_r \exp(-NTU(1 - C_r))}$$

$$NTU = \frac{UA}{C_{min}}$$

$$UA = \frac{1}{\frac{(\eta h A)_{a}}{a} + \frac{1}{h A}_{w} + R_{w}}$$

$$\frac{(\eta h A)_{a}}{a} = \kappa_{a} \dot{m}_{a}^{0.8}$$

$$\frac{(h A)_{w}}{w} = \kappa_{w} \dot{m}_{w,c}^{0.8}$$

$$T_{ao} = T_{ai} - \frac{\dot{Q}_{dry}}{C_{a}}.$$  

The only difference between this model and the model of the heating coil is in the way the effectiveness is calculated. Cooling coils generally require a greater surface area than heating coils due to the smaller temperature difference between the fluids, and they correspondingly have to have more rows of tubes. When the number of rows increases beyond four, the performance of a cross-flow heat exchanger approaches that of a counterflow heat exchanger. The effectiveness equation suitable for a counterflow arrangement is thus used instead of cross-flow.

If the coil is determined to be wet or partially wet, the estimate of $\dot{Q}_{dry}$ can be used as an initial estimate of the total heat transfer rate in the SHR model equations. These equations are presented below in the order in which they should be evaluated.

Non-iterative preliminary equations:

$$\omega = f_1(RH_i, T_{ai})$$

$$c_{p,ai} = (1 - \omega)c_{p,a} + \omega c_{p,s}$$

$$h_{ai} = c_{p,ai} T_{ai} + \omega(c_{p,s} T_{ai} + h_{fg})$$

$$C_{a} = \dot{m}_{a} c_{p,a}$$

$$C_{w} = \dot{m}_{w,c} c_{w}$$

humidity.
4.2. Component models

\[ C_{\text{min}} = \min(C_a, C_w) \]  \hspace{2cm} (4.51)
\[ C_{\text{max}} = \max(C_a, C_w). \]  \hspace{2cm} (4.52)

Iterative equations:

\[ h_{ao} = h_{ai} - \frac{\dot{Q}_{\text{total}}}{m_a} \]  \hspace{2cm} (4.53)
\[ h_s = \frac{(h_{ao} - b_f h_{ai})}{(1 - b_f)} \]  \hspace{2cm} (4.54)
\[ T_s = f_2(h_s) \]  \hspace{2cm} (4.55)
\[ T_{ao} = b_f(T_{ai} - T_s) + T_s \]  \hspace{2cm} (4.56)
\[ SHR = \frac{C_a(T_{ai} - T_{ao})}{m_a(h_{ai} - h_{ao})} \]  \hspace{2cm} (4.57)
\[ UA' = \frac{1}{\frac{SHR}{(\eta_0 h_A)_a} + \frac{1}{(h_A)_{w}} + R_w} \]  \hspace{2cm} (4.58)
\[ \varepsilon = \frac{1 - \exp(-NTU(1 - C_r))}{1 - C_r \exp(-NTU(1 - C_r))} \]  \hspace{2cm} (4.59)
\[ NTU = \frac{UA'}{C_{\text{min}}} \]  \hspace{2cm} (4.60)
\[ \dot{Q}_{\text{total}} = \frac{\varepsilon C_{\text{min}}}{SHR}(T_{ai} - T_{wi}). \]  \hspace{2cm} (4.61)

The value for \( \dot{Q}_{\text{total}} \), determined from Equation 4.61, is used in Equation 4.53 to effect the iterative process. The process is terminated when the difference between successive estimates of \( \dot{Q}_{\text{total}} \) falls below a threshold\(^3\). The functions \( f_1(.) \) and \( f_2(.) \) are psychrometric functions: \( f_1(.) \) calculates the moisture content of the air from the relative humidity and the dry bulb temperature; and \( f_2(.) \) calculates the wet-bulb temperature from the saturation enthalpy. For simplicity, the surface efficiency \( \eta_0 \) is not corrected for the wet condition as the effect of this is small compared to the correction made in Equation 4.58.

It may be noted that the cooling coil model does not require the specification of any more parameters than the heating coil model. There are thus three parameters that have to be determined for the heating and cooling coil models. These parameters are the tube wall conductive resistance, \( R_w \), and the two convective coefficients, labelled \( \kappa_a \) and \( \kappa_w \), in Equation 4.32. The latter two parameters are

\(^3\)Empirical results have shown that, under normal conditions, convergence to machine precision can be obtained in less than 20 iterations.
4.3. Selection of 'fault parameters'

The robustness of a parameter estimation process depends on the independence of the estimated parameters (Gill et al., 1981). In addition, the parameters can only be estimated when information (in the form of data samples recording the inputs and outputs at one sample time) is obtained from \( n \) independent operating points; where \( n \) is greater than the number of parameters. In practice, data from additional points may be required to counteract the effects of measurement noise and other uncertainties associated with the system.

In order to minimise the requirement for data and to increase the potential robustness of the parameter estimation, a subset of the total number of model parameters is selected for estimation by the FDD scheme. The subset contains the 'fault parameters', which are directly related to the faults of interest. Suitable 'fault parameters' are selected from the valve and heat exchanger models in the following sections.

4.3.1 Valve model

A leakage through the control port of a three-port valve can significantly affect the observed behaviour of the heat exchanger due to the exponential characteristics of coils, where the gain is higher at low flow rates. This fault can be diagnosed by selecting the parameter \( l \) in Equation 4.11 for estimation by the FDD scheme.

A leakage that develops in the control port has energy cost implications. When the valve is supposed to be closed, heat is transferred between the water and air thus imposing an unwanted load on the boiler or chiller. The effect is not, however, limited to the plant serving the leaky valve. To meet the set-points other plant
items will have to compensate for the effects of the leakage by increasing their load. For example, imagine an air-handling unit containing a cooling and heating coil, where the cooling coil is switched off and the heating coil is operating on partial load. If the valve serving the cooling coil develops a leak, not only is an additional load imposed on the chiller, but the heating coil has to operate at a higher load point to counteract the extra cooling due to the leakage. An estimate of the leakage can be used in the coil model to predict the additional heat transferred to the air-stream when the valve is supposed to be closed. When feedback control causes other plant items in the air stream to ‘cancel’ the effects of a leakage, the total energy cost is then double the amount estimated using the coil model.

The build-up of matter at the exit port of a valve can change the installed characteristic, and this fault could lead to control problems if the characteristic becomes very non-linear. The authority of the valve could be selected for estimation by the FDD scheme in order to diagnose this fault. However, the effects of changes in the authority may be hard to distinguish from fouling throughout large regions of the operating range. Moreover, it is unlikely that the effects of this fault would be severe enough to jeopardise thermal comfort or energy consumption in practice. In view of these points, and in order to maintain orthogonality between the parameters that are estimated, this parameter will not be selected for estimation.

### 4.3.2 Heat exchangers

Fouling can cause minor variations in the operating point of the boiler or chiller plant, which can increase energy consumption. Moreover, fouling leads to an increased resistance to fluid flow and it therefore results in a greater load being imposed on the fluid pumping equipment (fans and pumps). However, these effects on energy consumption will be small in all but the severest cases. The most apparent effect of fouling is a reduction in the maximum air-side temperature difference across the coil, which can inhibit the ability of the system to meet setpoints. The effects of fouling can be modelled by increasing the magnitude of \( R_w \), which represents the thermal resistance of the coil material. This parameter is therefore selected for estimation by the FDD scheme. It should be noted that it will not be possible to distinguish between air- and water-side fouling without considering inlet and outlet fluid pressures.
By estimating the increase in the thermal resistance of the coil it is possible to estimate the reduction in the maximum heat transfer rate of the heat exchanger based on design specifications; i.e.

\[ \Delta \dot{Q}_{\text{max}} = C_{\text{min}}(T_{hi} - T_{ci})(\epsilon_0 - \epsilon_1). \]  

(4.62)

where \( T_{hi} \) and \( T_{ci} \) are the design hot and cold fluid temperatures respectively. \( \epsilon_0 \) is the effectiveness of the coil calculated using the design fluid flow rates and the correctly operating overall conductance \( UA \), and \( \epsilon_1 \) is the effectiveness of the coil calculated using the conductance appropriate for the currently observed coil.

### 4.3.3 Temperature sensors

The most important sensors in air-conditioning installations are those that measure controlled variables used in control loops. An error in these sensors will cause the controller to operate the energy consuming plant at inappropriate levels. Faults that develop in these sensors can therefore have energy and thermal comfort implications. The most important sensor used in the thermal control of air-handling units is the discharge air temperature sensor. It is therefore most useful to incorporate the modeling of an off-set in this sensor in the FDD scheme (modelled using Equation 4.2).

### 4.4 Results of testing the models with experimental data

Data obtained from a valve manufacturer was used to test the modified exponential function that was described in Section 4.2.2. In addition, the complete subsystem model was tested using data obtained from heating and cooling coil subsystems installed in an air-conditioning test facility.

#### 4.4.1 Valve model

The valve model was tested using experimental data obtained from (Palmertz, 1993), which represents the inherent characteristic (i.e. tested under constant pressure) of
an equal percentage valve. Figure 4.7 shows the accuracy of the standard exponential function (Equation 4.9) in the left graph, and the modified exponential model (Equation 4.11) in the right graph. The graphs show the fractional mass flow rates (y-axis) plotted against the valve stem positions (x-axis), where the y-axis is on a logarithmic scale. On each graph, circles indicate the experimental data points, while the solid lines are the model predictions.

Figure 4.7: Comparison of standard and proposed valve models

The curvature parameters for the two models were determined by minimising the prediction errors of the model using a gradient descent method. For the modified exponential model (right graph), the leakage was set to the fractional flow at the closed valve position recorded in the data. The parameters of each model and the mean of the absolute prediction errors (MAE) are presented in Table 4.3.

Table 4.3: Estimates of the valve model parameters

<table>
<thead>
<tr>
<th>MODEL TYPE</th>
<th>CURVATURE (( \beta ))</th>
<th>FRACTIONAL LEAKAGE (( l ))</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eqn. 4.9</td>
<td>3.10</td>
<td>0.045</td>
<td>0.0202</td>
</tr>
<tr>
<td>Eqn. 4.11</td>
<td>2.75</td>
<td>0.028</td>
<td>0.0120</td>
</tr>
</tbody>
</table>

A reduction of 40% in the MAE achieved is by using the modified exponential model. The model is better able to represent the behaviour of the valve near to the closed position, which is where the standard model is least accurate. To compensate for moving the point of intersection on the y-axis, the curvature parameter has to be increased to reproduce the behaviour of the real valve throughout the
rest of the range. The reduction in the curvature parameter, compared with the standard exponential function, is approximately 11% for this particular valve.

### 4.4.2 Complete subsystem

The heating and cooling coil subsystem models (actuator, valve, and heat exchanger combined) were tested using data obtained from the Air-conditioning Evaluation Facility, at the Building Research Establishment; this system is described in Section 7.4.1. Step tests were conducted on the heating and cooling coil subsystems at a constant air flow rate of 1.5 m³s⁻¹. Initial values for the three heat transfer parameters were set based on values recommended by Holmes (1982). These values were then refined by using the direct search method described in Appendix E to reduce the prediction errors of the model for the data. The rest of the model parameters were estimated from the design, manufacturers', and commissioning data that was available from the system. The parameter values estimated for the heating and cooling coil subsystems are presented in Table 4.4.

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>HEATING COIL</th>
<th>COOLING COIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v$ (-)</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$\beta$ (-)</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td>$l$ (-)</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>$\dot{m}_w$ (kg · s⁻¹)</td>
<td>0.53</td>
<td>2.64</td>
</tr>
<tr>
<td>$A$ (-)</td>
<td>0.55</td>
<td>0.55</td>
</tr>
<tr>
<td>$\kappa_a$ (kW · K⁻¹ · kg⁻0.8 · s⁰.8)</td>
<td>0.59</td>
<td>13.0</td>
</tr>
<tr>
<td>$\kappa_w$ (kW · K⁻¹ · kg⁻0.8 · s⁰.8)</td>
<td>26.17</td>
<td>24.1</td>
</tr>
<tr>
<td>$R_w$ (kW⁻¹ · K)</td>
<td>0.38</td>
<td>0.286</td>
</tr>
</tbody>
</table>

The inlet water temperatures to the coils were 80°C and 7°C for the heating and cooling coils respectively. The cross-sectional area of the ducting was 1.44 m² (1.2 m by 1.2 m), which was also assumed to be the face area of the coils. The diameter of the heat exchanger pipes was 13 mm. Figure 4.8 shows the accuracy of the heating and cooling coil subsystem models compared with the experimental data. The tests involved stepping the control signal to the actuator between the following values (0.0, 0.1, 0.3, 0.5, 0.7, 1.0). The x-axis on each graph represents
4.5. Summary of the chapter

the control signal and the y-axis is the air-side approach, given by:

\[ \alpha = \frac{T_{ai} - T_{ao}}{T_{ai} - T_{wi}}. \] (4.63)

The data samples from the subsystems are shown as circles, while the model predictions are shown as solid lines. It can be observed that the maximum air-side approaches for each subsystem reach the values expected for the coils usually installed in air-handling units (≈ 0.2 for heating coils; ≈ 0.6 for cooling coils).

![Graph showing heating and cooling coil model validation](image)

Figure 4.8: Result of heating and cooling coil model validation

It was assumed that the cooling coil was dry throughout the tests; the performance of the wet coil model is therefore not evaluated here. A linear thermal fan model was used to approximate the temperature rise across the supply fan, which was sited between the available sensors. The temperature rise across the fan was calculated from:

\[ \Delta T = \max(\Delta T) \frac{\dot{m}_a}{\max(\dot{m}_a)}. \] (4.64)

where \( \max(\Delta T) \) was set to 1K.

4.5 Summary of the chapter

This chapter has presented models of the constituent components of the heat exchanger subsystems that constitute the application for the FDD scheme. When used together, the models are capable of representing the global characteristic of
4.5. Summary of the chapter

The heat exchanger subsystems typically used in air-conditioning installations. The majority of the parameters of the subsystem model relate to physical properties, and are thus able to be initially estimated from design and manufacturers' data that is normally available. The models are used for fault detection and diagnosis by estimating some of the parameters of the model using sampled input-output data from the real system. A number of 'fault parameters' has been selected for estimation by the FDD scheme that relate to the faults of interest (as identified in Section 3.2.1). The fault parameters represent a subset of the total number of model parameters, and the number of fault parameters has been kept to a minimum to increase robustness of the estimation. Chapter 5 now describes the signal processing and parameter estimation aspects of the FDD scheme.
Chapter 5

Signal processing and parameter estimation

Introduction

Figure 5.1 shows the signal processing and parameter estimation aspects of the FDD scheme. Signal processing is performed to detect when the system from which data are obtained is in steady-state. This is necessary to facilitate the use of static component models, and a steady-state detector is described in Section 5.1 for this purpose. The steady-state detector requires the derivatives of the signals to be evaluated. Measurement noise affects the estimation of derivatives and a suitable digital filter is described in Section 5.1.1 to reduce the noise effects.

The subsystem models, which were described in the Chapter 4, have outputs that are non-linear functions of the parameters. Estimation of the parameters of these models is not an analytically soluble problem and non-linear optimisation methods are required. The parameter estimation is made more complex when on-line adaptation is sought, due to the storage of large amounts of data being impracticable. The storage requirements are minimised by formulating the parameter estimation in a recursive fashion, where information from previous data samples is stored implicitly in the parameter estimates. Certain approximations and assumptions have to made to derive a recursive parameter estimation algorithm for non-linear models, and awareness of these is important for ascertaining the limitations of the
5.1 Development of a steady-state detector

The signals associated with the system that is considered for FDD are monitored in real-time. Since the FDD procedures are based on the use of static models, the signals should only be used when dynamic effects are negligible. A steady-state detector is therefore required that operates like a switch, only allowing signals to be used by the FDD scheme when the system is in steady-state. Heat exchanger subsystems have an infinite impulse response (IIR) characteristic, meaning that once the inputs have been excited, the system will never exactly reach steady-state, but will asymptotically approach it. Hence, for practical implementation a steady-state detector has to be able to let signals pass through when the system is close to steady-state, according to a threshold.

Two steady-state detectors have been developed as part of IEA Annex 25. each of
5.1. Development of a steady-state detector

which is based on different criteria. Dexter and Benouarets (1995a) have developed a ‘transient detector’, which considers the gradient of all the measured signals over a moving time window. Glass (1995) has developed a similar steady-state detector that is based on evaluating the time averaged variance of all the measured signals. The main problem with each of the approaches is that thresholds have to be determined for all the signals (inputs and outputs) associated with the system. These thresholds do not relate to meaningful quantities and suitable values can be difficult to determine. Instead of considering the isolated behaviour of the signals, it will be shown that the model of the system, which is used in the FDD scheme, can be used to simplify threshold determination.

To develop the steady-state detector, it will be assumed that the system of interest can be separated into static and dynamic components. This model is illustrated in Figure 5.2 and is known as a Hammerstein representation. The inputs are passed to a static transformation that generates an output, which is the input to a dynamic component. The limitation of the representation is that the same dynamic relationship is assumed to exist between the output and each of the inputs.

\[ y(t) = h(t) - \tau \frac{dy(t)}{dt}, \]  
\[ h(t) = f[\theta, u(t)], \]

where:
\[ u \] is a vector of the inputs, \[ \theta \] is a vector of parameters, and \[ y \] is a scalar output. The function \( f(\cdot) \) represents the characteristic of the static component. It may be

![Figure 5.2: Hammerstein representation](image-url)
noted that for a system that is statically linear then \( h(t) = \theta^T u(t) \). If the system reaches steady-state, the differential term in Equation 5.1 will disappear and:

\[
y(t) = h(t).
\] (5.3)

The amount by which the output, at any instant of time, differs from its steady-state value will be called the steady-state error, \( e(t) \), and is given by:

\[
e(t) = h(t) - y(t) = \tau \frac{dy(t)}{dt}.
\] (5.5)

In steady-state, this error will equal zero, i.e \( e(t) = 0 \). However, since the first-order system described by Equation 5.1 is an IIR system it is necessary to select a threshold that defines the proximity to steady-state that is acceptable. Providing the assumption of a first-order system is reasonable, the threshold that is selected will correspond to the prediction error resulting from using a static model with the dynamic data. Hence, the test for steady state can be given by:

\[
\text{IF } \tau \frac{dy(t)}{dt} < \kappa \text{ THEN system is in steady-state}
\]

\[
\text{ELSE system is transient,}
\]

where \( \kappa \) is the threshold value. Time constants appropriate for each of the inputs can be determined by applying step changes to the real system and observing the response of the output. An approximate estimate of the time constants can be obtained by measuring the time taken for the output to reach 63% of its steady-state value following a step. The time constant with the highest estimated value is the dominant time constant of the system and \( \tau \) should be set to this value. Systems that have non-linear dynamics\(^1\) would cause an estimate of a time constant to vary, depending on the operating region over which a step test was carried out. In this case, the highest time constant value for each input should be considered.

\(^1\)Such as cooling coil subsystems.
5.1. Development of a steady-state detector

5.1.1 Treatment of measurement noise

The sensor signals are sampled by the control system and differencing methods are used to calculate the derivatives of the signals. For example, a simple backward differencing equation can be used:

$$\frac{dy}{dt} \approx \frac{y_k - y_{k-1}}{\Delta t},$$  \hspace{1cm} (5.6)

where $k$ represents sample number and $\Delta t$ is the time between samples. The accuracy of derivatives calculated in this way is sensitive to noise in the sensor measurements and the existence of noise would cause the steady-state error to be over-estimated. Signals would therefore not be deemed to be in steady-state when they were close enough to steady-state, but noisy. This effect is undesirable as measurement noise is independent of dynamic activity and not indicative of it.

One way to reduce the effects of noise on the derivative calculation is to apply a low-pass filter to the noisy signal. The derivative can then be calculated from the clean signal thereby reducing the effects of noise. Since the FDD procedures may operate on-line and in real time, a physically realisable filter is sought, i.e. one that does not require future samples. A digital filter that operates on a window of past samples meets this causality requirement and is given by:

$$A(q^{-1})\bar{y} = B(q^{-1})y,$$  \hspace{1cm} (5.7)

where $A(q^{-1})$ and $B(q^{-1})$ are polynomials in the delay operator, $y$ is the noisy signal, and $\bar{y}$ is the filtered signal. The order of the polynomials $A$ and $B$ represents the order of the filter. If there is no recursion of $\bar{y}$, i.e. $A = 1$, the filter is said to be finite impulse response (FIR), otherwise, with recursion, the filter is infinite impulse response (IIR). The latter type of filter is well-known to have a superior level of performance for an equal number of coefficients. Since it is desirable to reduce the memory requirements for the on-line FDD scheme, an IIR filter is adopted here. There are various methods available for calculating the coefficients in the polynomials $A$ and $B$ that use different frequency response criteria (e.g. Butterworth, Chebyshev, elliptic, etc). Each of the different methods produces filters that have particular frequency response characteristics, e.g. ripples, sharp cut-off, etc.

The Butterworth filter design method is able to produce a smooth attenuation to the signal as the frequency increases, without ripples in either the pass-band or the
stop band. The sharpness of the cut-off is determined primarily by the order that is selected. The method is also well-known and algorithms are widely available that calculate the coefficients to meet specified criteria, e.g. see (Ackroyd, 1973; Press et al., 1992). The Butterworth filter method was therefore selected to determine the coefficients of the filter.

The sampling frequency, cut-off frequency, and the order of filter have to be specified to obtain the coefficients. In air-conditioning monitoring applications a sampling frequency of 0.0167 Hz is fairly typical (i.e. one minute sampling intervals). Assuming that a heat exchanger subsystem can be approximated as first-order, the magnitude of the frequency response is then calculated from:

$$|G(j\omega)| = \frac{1}{\sqrt{1 + \omega^2\tau^2}},$$

where $G(.)$ is the transfer function of the system. Assuming that 5 minutes is a reasonable estimate for the maximum dominant time constant of a heat exchanger subsystem, a frequency response graph can be plotted as is shown in Figure 5.3, where the frequency is shown as a fraction of the sampling frequency. It can be observed that frequencies that are greater than 10% of the sampling frequency are attenuated quite significantly. Activity in the system at frequencies exceeding 0.00167 Hz would therefore be difficult to distinguish from noise, which will be of a relatively low amplitude compared with the amplitude of the true signal variations. Heat exchanger subsystems will be unlikely to cycle at frequencies greater than 0.00167 Hz, and the cut-off frequency for the noise filter is therefore set to this value.

The following design specifications were used to obtain the coefficients:

- sampling frequency: 0.0167 Hz;
- cut-off frequency: 0.00167 Hz (10% sampling frequency);
- filter order: 4.

The values for the coefficients were calculated using the Butterworth filter described in (Press et al., 1992):

$$A = [1.0000, -2.3695q^{-1}, 2.3140q^{-2}, -1.0547q^{-3}, 0.1874q^{-4}]$$

\(^2\)A time constant of 5 minutes represents the longest time constant likely to be experienced in practice. This magnitude of time constant may be experienced when small perturbations are applied to the valve control signal at low water flow rates.
5.1. Development of a steady-state detector

\[ B = [0.0048, 0.0193q^{-1}, 0.0289q^{-2}, 0.0193q^{-3}, 0.0048q^{-4}]. \]

The frequency response graph of the filter with these coefficients is presented in Figure 5.4. The graph shows the magnitude attenuation of the filter function against the fraction of the sampling frequency.

Figure 5.4: Frequency response of filter

The derivative of the measured output can be estimated by using the output of
5.1. Development of a steady-state detector

the noise filter as follows: \[
\frac{dy}{dt} \approx \frac{\hat{y}_k - \hat{y}_{k-1}}{\Delta t}.
\] (5.9)

Despite being able to reduce the effects of noise in the output measurements, the accuracy of the steady-state detector can be jeopardised by two other factors:

1. Inadequacy of the first-order approximation to the true dynamics of the system.

2. Activity in the inputs to the system at frequencies exceeding the cut-off frequency of the noise filter.

Each of these may lead to data being inappropriately identified as steady-state. Hence, a modification to the steady-state detector is necessary to prevent this from happening. One technique that can be employed is to take account of the activity of the inputs signals to the system within the steady-state detector. However, instead of considering all of the inputs separately as proposed by Dexter and Benouarets (1995a), and by Glass (1995), the model that is used in the FDD scheme can be employed to simplify the determination of thresholds. This is described in the following section.

5.1.2 Accounting for activity in the input signals

It was assumed during the derivation of the steady-state detector that the dynamics of the system are first-order. In practice, this is unlikely to be the case and the estimate of the steady-state error used in the steady-state detector will therefore be in error. In the time domain, an 's-shaped' response to a step change in an input is typical of heat exchangers in air-conditioning plant. For this type of response, the use of the first-order approximation will lead to an under-estimation of the derivative in the period immediately following a change. As the effects of the disturbance diminish, the steady-state error will be over-estimated. This is illustrated in Figure 5.5. The upper graph shows the output signal of a first order and a second order system in response to a unity-step input signal. The bottom graph shows the gradient of these output signals. The largest difference between gradients calculated from these signals occurs in the initial stages of the
5.1. Development of a steady-state detector

disturbance, it is therefore at this time when the steady-state detector is likely to be most unreliable.

![Figure 5.5: First-order approximation to second order system](image)

For the system to be in steady-state, both the inputs and the outputs should be static. The input signals can be evaluated for variation when the first-order approximation is inaccurate, or when the noise filter has attenuated true variations in the outputs below the threshold. Ideally, an evaluation of the activity of the input signals should be related to a steady-state error in the same way as the output. For the FDD application considered here, a model of the system is available, thus allowing the effect of a change in an input variable on the output to be evaluated. The variation of all the inputs can be evaluated by using the measured inputs with a first-order Hammerstein model:

\[
\dot{y}(t) = f(u_k, \theta) - \tau \frac{d\hat{y}(t)}{dt},
\]

where \( f(.) \) is the static model used in the FDD scheme, and \( \hat{y}(t) \) is the estimate of the dynamic output based on the measured inputs \( u(t) \) and the model parameters \( \theta \). Equation 5.10 may be solved analytically to give the following particular solution for the initial condition of \( y = y_0 \) when \( t = 0 \):

\[
\hat{y}(t) = f(u_k, \theta) - [f(u_k, \theta) - \hat{y}_0(t)] \exp\left(\frac{-t}{\tau}\right).
\]
5.2 Development of a non-linear recursive estimator

This may be formulated in a recursive fashion as follows:

\[
\hat{y}_k = \left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right] f_k(u_k, \theta) + \left[\exp\left(-\frac{\Delta t}{\tau}\right)\right] \hat{y}_{k-1}.
\]  

(5.12)

where \(k\) is the sample number, \(\Delta t\) is the sampling interval, and \(\tau\) is an estimate of the dominant time constant of the system. The condition for steady-state is then given by:

\[
\begin{align*}
\text{IF } & \tau \frac{\hat{y}_k - \hat{y}_{k-1}}{\Delta t} < \kappa \\
\text{AND } & \tau \frac{\hat{y}_k - \hat{y}_{k-1}}{\Delta t} < \kappa \\
\text{THEN } & \text{ system is in steady-state} \\
\text{ELSE } & \text{ system is transient,}
\end{align*}
\]

where \(\hat{y}\) is the filtered measured output, \(\hat{y}\) is the prediction obtained from Equation 5.12 and \(\kappa\) is the threshold, which corresponds to an estimate of an acceptable steady-state error. Consideration of the measured output signal of the system allows account to be taken of dynamic activity resulting from variations in all the inputs (measured and unmeasured). Consideration of the model output is necessary to account for inaccuracies in the first-order approximation and unwanted attenuation of true signal variations by the noise filter. Figure 5.1 may be referred to for a schematic representation of the scheme. It may be noted that if all the inputs that affected the output were measured, there would be no need to consider the measured output.

5.2 Development of a non-linear recursive estimator

The subsystem models that were described in Chapter 4 cannot be written in a linear-in-the-parameters form. Consequently, the optimum parameter values for a set of training data points cannot be calculated analytically. Similarly, an exact recursive estimation procedure cannot be derived for these types of models. Approximations have to be made, therefore, to derive a suitable recursive algorithm for the FDD scheme.

One way to derive an algorithm is to factor the parameters that are to be estimated out of the non-linear equations by means of a series expansion. A common
approach involves expanding a criterion function, which is the fit of the model to a set of data points, to quadratic precision. This approach leads to the recursive prediction error method (Moore and Weiss, 1979; Ljung, 1980). An alternative approach is adopted here, which involves making a linear approximation to the non-linear model by a first-order Taylor expansion. This approach makes the implications of the assumptions that need to be made more apparent. It may be noted, however, that both the approaches to the derivation yield identical algorithms.

The non-linear recursive algorithm that is derived makes adjustments to the parameter vector based on the prediction errors and accumulated search direction information. The basic form of the algorithm assumes time-invariant parameters and the sensitivity of the estimator to the prediction errors reduces as $t \to \infty$. Hence, modifications are necessary to make the estimator remain responsive to parameter changes. The approach that is adopted here is to incorporate 'forgetting' within the algorithm so that data samples are weighted in the calculations according to their age.

In its basic form, forgetting can lead to robustness problems when there is a lack of excitation in the plant. The main problem with having a constant rate of forgetting is that the weighting applied to the implicitly stored information is reduced as time progresses, regardless of whether the latest information is different. As a consequence, the uncertainty associated with the parameter estimates can increase when the plant is not excited, which can ultimately cause false alarms to be generated by the FDD scheme. Since air-conditioning systems have a tendency to spend long periods at one operating point, measures are required to ensure robustness. The prediction error forgetting method (PEF) (Fortescue et al., 1981) is therefore adopted here, which involves adjusting the rate of forgetting based on an assessment of the information content of the data. The PEF method is described in Section 5.2.5.

5.2.1 Preliminaries: Derivation of a batch estimator

In order to derive the recursive parameter estimator, an off-line formulation of the problem is first considered. The heat exchanger subsystem models that were presented in Chapter 4 are non-linear in both the parameters and the inputs. The
complete subsystem model has a vector of inputs, \( \mathbf{u} \), and a vector of parameters, \( \boldsymbol{\theta} \). In most situations the only measurable system output will be the dry-bulb temperature of the air leaving the coil. Hence, the output of the model is a scalar quantity, which will be denoted by \( y \). A single function, \( f(.) \), may therefore be used to represent the complete subsystem model, such that:

\[
\hat{y}_k = f_k(\mathbf{u}_k, \boldsymbol{\theta}),
\]

where \( k \) denotes the sample number. To simplify the notation, the input vector at sample \( k \) will be assumed to be implicit in the function, such that:

\[
\hat{y}_k = f_k(\boldsymbol{\theta}).
\]

The objective of parameter estimation is to locate the parameter values that minimise the prediction errors of the model for a batch of data. To achieve this, it is necessary to define a criterion function representing the performance of the model:

\[
V_N(\boldsymbol{\theta}) = \frac{1}{2} \sum_{k=1}^{N} c_k^2(\boldsymbol{\theta}),
\]

where \( V_N(\boldsymbol{\theta}) \) is a scalar quantity representing the sum of the square of the model prediction errors over a data record containing \( N \) samples, multiplied by 0.5 to simplify later calculations. The error, \( \epsilon(.) \), is defined as the difference between the predictions made by the model and the measured outputs in response to the measured inputs; i.e.

\[
\epsilon_k(\boldsymbol{\theta}) = f_k(\boldsymbol{\theta}) - y_k.
\]

The parameter values that minimise the criterion function are found by differentiating the criterion function with respect to the parameter vector and calculating the parameters that give a zero gradient. The parameter vector should thus be calculated from:

\[
\sum_{k=1}^{N} \{ f_k(\boldsymbol{\theta}) \nabla f_k(\boldsymbol{\theta}) \} - \sum_{k=1}^{N} \{ y_k \nabla f_k(\boldsymbol{\theta}) \} = 0,
\]

where:

\[
\nabla f_k(\boldsymbol{\theta}) \triangleq \begin{bmatrix}
\frac{\partial f_k(\boldsymbol{\theta})}{\partial \theta_1} \\
\frac{\partial f_k(\boldsymbol{\theta})}{\partial \theta_2} \\
\vdots \\
\frac{\partial f_k(\boldsymbol{\theta})}{\partial \theta_p}
\end{bmatrix}.
\]
5.2. Development of a non-linear recursive estimator

Equation 5.17 cannot be solved easily as the parameter vector cannot be factored out of \( f(\cdot) \). One way to address this problem is to linearise the model with respect to the parameters. If it is assumed that an initial parameter estimate is available \((\theta_0)\), and that the optimum parameters are close to these initial values, a Taylor expansion can be used to linearise the model function such that:

\[
    f_k(\theta) \approx f_k(\theta_0) + [\theta - \theta_0]^T \nabla f_k(\theta_0).
\]

This linearised model can now be substituted in the criterion function, thus giving:

\[
    V_N(\theta) = \frac{1}{2} \sum_{k=1}^{N} \left\{ f_k(\theta_0) + [\theta - \theta_0]^T \nabla f_k(\theta_0) - y_k \right\}^2.
\]

Differentiation of this expression with respect to \( \theta \) yields:

\[
    \nabla V_N(\theta) = \sum_{k=1}^{N} \left\{ \nabla f_k(\theta_0) \left( f_k(\theta_0) + [\theta - \theta_0]^T \nabla f_k(\theta_0) - y_k \right) \right\},
\]

where \( \nabla V_N(\theta) \) is the vector of partial derivatives of the criterion function with respect to the parameters. The parameters can now be factored out quite easily. The parameters corresponding to the minimum criterion function value are found when the gradient vector of the criterion function given by \( \nabla V_N(\theta) \) is zero. Hence, the estimate of the optimum parameter vector is given by:

\[
    \hat{\theta} = \theta_0 + \left\{ \sum_{k=1}^{N} \nabla f_k(\theta_0) (y_k - f_k(\theta_0)) \right\} \left\{ \sum_{k=1}^{N} \left( \nabla f_k(\theta_0)[\nabla f_k(\theta_0)^T] \right) \right\}^{-1}.
\]

Application of Equation 5.21 causes the parameter vector to move in the Newton direction. When the model is linear-in-the-parameters such that:

\[
    f(\theta) = \theta \phi^T, \quad \text{where} \ \phi \text{ is a regressor vector},
\]

an expansion is not necessary and the optimum parameters can be found from any starting point, hence \( \theta_0 \) could be set to zero. In this case, Equation 5.21 reduces to the general linear-least-squares solution given by:

\[
    \hat{\theta} = \left( \sum_{k=1}^{N} \phi_k \phi_k^T \right)^{-1} \left( \sum_{k=1}^{N} \phi_k \epsilon_k(\theta_0) \right).
\]
where \( f(0) = 0 \) in Equation 5.23. The accuracy of the parameters estimated using Equation 5.21 will depend on how well the linear expansion is able to approximate the local non-linear characteristic of the model. Since the accuracy of the linear approximation is likely to deteriorate as the (Euclidean) distance between \( \theta_0 \) and the true optimum \( \theta \) increases, only small changes from \( \theta_0 \) are likely to be estimated accurately in one step (Gill et al., 1981).

The estimate of the parameter vector obtained by using Equation 5.21 is the optimum (in a least-squares sense) for the linearised model. Providing the linear approximation is reasonable, this estimate is likely to improve the fit of the non-linear model to the data, but the estimate is unlikely to be the true optimum. Hence, to move closer to the true optimum it is necessary to repeat the calculation of the parameter estimate by linearising the model at the new parameter estimate. This process is repeated in an iterative fashion, using the entire data record at each iteration, until the optimum estimate for the non-linear model is found.

It can take a long time for air-conditioning plant items to explore their operating range sufficiently enough to allow changes due to degradation faults to be identified. Hence, large amounts of data would be needed to estimate the fault parameters reliably using the batch estimation method. This approach is therefore unsuited for an on-line application where minimisation of storage requirements is sought. The alternative to batch optimisation is to formulate the equations in a recursive form so that the information from the data samples is stored implicitly in the parameters. Not being able to iterate on the whole data record means that approximations have to be made to derive a recursive form of the parameter estimator. The approach that is adopted here is to re-linearise the model at each sample based on the previous parameter estimates, which amounts to a form of sequential iteration. The recursive equations based on this approach are derived in the following section.

### 5.2.2 Recursive derivation

A recursive set of equations can be derived for the non-linear model in a similar manner to the way in which the recursive least-squares algorithm is derived, e.g. see (Åström and Wittenmark, 1989). Before proceeding, it is useful to define the
5.2. Development of a non-linear recursive estimator

following notation:

\[ e_k = f_k(\theta_{k-1}) - y_k \] (5.25)

\[ \psi_k = \nabla f_k(\theta_{k-1}) \] (5.26)

\[ P_k = \left\{ \sum_{i=1}^{k} \psi_i \psi_i^T \right\}^{-1} \] (5.27)

\[ \Delta \theta_k = \theta_k - \theta_{k-1} \] (5.28)

Using this notation, Equation 5.22 may be re-written as:

\[ \Delta \theta_k = -P_k \sum_{i=1}^{k} \psi_i e_i \] (5.29)

\[ = -P_k \left\{ \left( \sum_{i=1}^{k-1} \psi_i e_i \right) + \psi_k e_k \right\}. \] (5.30)

At this stage it is useful to give the inverse of the \( P \) matrix in a recursive form:

\[ P_k^{-1} = P_{k-1}^{-1} + \psi_k \psi_k^T \] (5.31)

\[ P_k^{-1} = P_k^{-1} - \psi_k \psi_k^T. \] (5.32)

The summation term in Equation 5.30 can be given as follows:

\[ \sum_{i=1}^{k-1} \psi_i e_i = -P_{k-1}^{-1} \Delta \theta_{k-1} \] (5.33)

\[ = -\left( P_k^{-1} - \psi_k \psi_k^T \right) \Delta \theta_{k-1} \] (5.34)

\[ = \left( \psi_k \psi_k^T - P_k^{-1} \right) \Delta \theta_{k-1}. \] (5.35)

Substituting Equation 5.35 into Equation 5.30 gives:

\[ \Delta \theta_k = -P_k \left\{ \left( \psi_k \psi_k^T - P_k^{-1} \right) \Delta \theta_{k-1} \right\} + \psi_k e_k \] (5.36)

\[ = \Delta \theta_{k-1} - P_k e_k - e_k \] (5.37)

\[ = \Delta \theta_{k-1} - \psi_k \left[ e_k + \psi_k \Delta \theta_{k-1} \right]. \] (5.38)
The accuracy of the parameter estimates at \( k \) depends on the accuracy of the estimates at \( k - 1 \) and \( k - 2 \). The errors resulting from the linearisation are fed back into the estimator and amplified by \( P_k \psi_k \). This will lead to stability problems even when the linearisation is reasonable. A direct feedback of the parameter estimates in this way should therefore be avoided. One way to address the problem is to assume that previous changes in the parameter values were zero (or very small); i.e.

\[
\psi_k^T \Delta \theta_{k-1} \ll \epsilon_k, \text{ and } \Delta \theta_{k-1} \ll P_k \psi_k \epsilon_k.
\]

Hence, the parameters at sample \( k \) can be estimated from:

\[
\Delta \theta_k = -P_k \psi_k \epsilon_k \quad \text{ (5.39)}
\]

\[
\theta_k = \theta_{k-1} - P_k \psi_k \epsilon_k. \quad \text{ (5.40)}
\]

The parameter vector at sample \( k \) can thus be calculated from the model gradient vector, and the prediction error based on the last parameter vector. The derivation is not yet complete, however, since Equation 5.31 only allows the inverse of the \( P \) matrix to be updated. The \( P \) matrix itself therefore has to be found from a matrix inversion at each sample, which is undesirable for an on-line application. This can be avoided by applying the Sherman-Morisson formula (Press et al., 1992), which enables the \( P \) matrix to be updated directly at each sample.

Lemma:

\[
(A + BCD)^{-1} = A^{-1} - A^{-1} B (C^{-1} + DA^{-1} B)^{-1} DA^{-1}, \quad \text{ (5.41)}
\]

where \( A, B, C \) and \( D \) are matrices of compatible dimensions, so that the product \( BCD \) and the sum \( A + BCD \) exist. Application of this lemma to Equation 5.31 by setting:

\[
A = P_{k-1},
\]

\[
B = \psi_k,
\]

\[
C = I, \text{ and}
\]

\[
D = \psi_k^T,
\]

yields the following updating formula for \( P_k \):

\[
P_k = P_{k-1} - \frac{P_{k-1} \psi_k \psi_k^T P_{k-1}}{I + \psi_k^T P_{k-1} \psi_k}. \quad \text{ (5.42)}
\]
5.2. Development of a non-linear recursive estimator

Note that \( [I + \psi_k^T P_{k-1} \psi_k] \) evaluates to a scalar for the scalar output case that has been considered (i.e. \( I = 1 \)). A ‘gain’ vector \( L \) may be introduced whereby:

\[
L_k = P_k \psi_k \tag{5.43}
\]

\[
= \left\{ P_{k-1} - \frac{P_{k-1} \psi_k \psi_k^T P_{k-1}}{I + \psi_k^T P_{k-1} \psi_k} \right\} \psi_k \tag{5.44}
\]

\[
P_{k-1} \psi_k \left[ I - \frac{\psi_k^T P_{k-1} \psi_k}{I + \psi_k^T P_{k-1} \psi_k} \right] \tag{5.45}
\]

\[
P_{k-1} \psi_k \left[ \frac{I + \psi_k^T P_{k-1} \psi_k - \psi_k^T P_{k-1} \psi_k}{I + \psi_k^T P_{k-1} \psi_k} \right] \tag{5.46}
\]

\[
P_{k-1} \psi_k \left[ I + \psi_k^T P_{k-1} \psi_k \right]^{-1} \tag{5.47}
\]

The updating expression for the \( P \) matrix can now be given in terms of the \( L \) vector:

\[
P_k = P_{k-1} - L_k \psi_k^T P_{k-1} \tag{5.48}
\]

\[
= (I - L_k \psi_k^T) P_{k-1}. \tag{5.49}
\]

Hence, the complete algorithm can now be re-stated as:

\[
\theta_k = \theta_{k-1} - L_k \epsilon_k \tag{5.50}
\]

\[
L_k = P_{k-1} \psi_k (I + \psi_k^T P_{k-1} \psi_k)^{-1} \tag{5.51}
\]

\[
P_k = (I - L_k \psi_k^T) P_{k-1}. \tag{5.52}
\]

The algorithm moves the parameter estimates in an approximate Newton direction. The \( P \) matrix is the inverse of the Hessian\(^3\) and the \(-\psi\) vector is the (current) steepest descent direction. The \( P \) matrix contains curvature information that modifies the steepest descent direction and provides the step size, based on locating the minimum of an approximate quadratic model of the criterion function.

It can be verified from Equation 5.31 that \( ||P|| \rightarrow \infty \) as \( k \rightarrow \infty \). Hence, \( ||P|| \rightarrow 0 \) and the sensitivity of the algorithm diminishes as \( k \rightarrow \infty \). For applications with time-varying parameters, incorporation of ‘forgetting’ is therefore necessary to keep the gain \( ||P|| \) and hence \( ||L|| \) high so that the estimator has tracking ability. In addition, curvature information is parameter-dependent as the model.

\(^3\) The Hessian is sometimes termed the ‘information matrix’.
is re-linearised at each sample. Thus, as the parameters change, old curvature information will become less appropriate, which may cause the search direction to stop being a descent direction. The incorporation of forgetting involves placing a greater weighting on new curvature information in the calculations. This is therefore beneficial for maintaining an appropriate search direction.

### 5.2.3 Incorporation of forgetting

One way to make the estimator maintain its sensitivity to the most recent data samples is to weight the data according to its age. This can be achieved by first expressing the criterion function as follows:

\[ V_N(\theta) = \frac{1}{2} \sum_{k=1}^{N} \lambda^{N-k} \epsilon_k^2(\theta), \]  

where \( \lambda \) is the 'forgetting factor', or discounting factor. The forgetting gives the information an exponential profile with respect to sample time. Based on this interpretation, \( \lambda \) is included in all the summation terms used to derive the estimator. Thus, using the notation defined in Section 5.2.2, Equation 5.29 can be re-written as:

\[ \Delta \theta_k = -P_k \sum_{i=1}^{k} \lambda^{k-i} \psi_i \epsilon_i. \]  

Forgetting is incorporated into the \( P \) matrix by amending the recursive updating expression to:

\[ P_k^{-1} = \lambda P_{k-1}^{-1} + \psi_k \psi_k^T. \]  

Since the recursive derivation (Equations 5.29-5.38) expresses \( \sum_{i=1}^{k} \psi_i \epsilon_i \) in terms of the \( P \) matrix, the simple forgetting applied to the updating of the \( P \) matrix is equivalent to weighting the criterion function directly.

The matrix lemma given in Equation 5.41 can be used as before to derive the updating formula for the \( P \) matrix by this time including the scalar \( \lambda \) in \( A \), such that \( A = \lambda P_k^{-1} \):

\[ P_k = \frac{P_{k-1}}{\lambda} - \left[ \frac{P_{k-1} \psi_k \psi_k^T P_{k-1}}{1 + \frac{\psi_k^T P_{k-1} \psi_k}{\lambda \lambda}} \right] \]  

\[ = \frac{1}{\lambda} \left[ P_{k-1} - \frac{P_{k-1} \psi_k \psi_k^T P_{k-1}}{\lambda + \psi_k^T P_{k-1} \psi_k} \right]. \]
5.2. Development of a non-linear recursive estimator

The complete algorithm with forgetting is therefore given by:

\[ \theta_k = \theta_{k-1} + L_k \epsilon_k \]  
(5.58)

\[ L_k = P_{k-1} \psi_k \left( \lambda + \psi_k^T P_{k-1} \psi_k \right)^{-1} \]  
(5.59)

\[ P_k = \frac{1}{\lambda} (I - L_k \psi_k^T) P_{k-1} \]  
(5.60)

If the linear approximation to the model function is exact, the curvature information is independent of the parameter estimates (the case of models that are linear-in-the-parameters). The algorithm would then be equivalent to the recursive least-squares algorithm and forgetting would cause the estimator to behave similarly to a first-order filter (under the condition of persistent excitation). However, this approximation will break down as the curvature of the surface of the criterion function deviates from being quadratic.

5.2.4 Size of data record (memory vector)

The assumption that is made implicitly by assuming that \( \Delta \theta_{k-1} \) in Equation 5.38 is zero is that:

\[ \sum_{i=1}^{k-1} \psi_i \epsilon_i = 0. \]  
(5.61)

Hence, it is assumed that the parameter estimate at sample \( k - 1 \) is the optimum estimate yielding a zero prediction error. No previous first-derivative (steepest descent) direction information is therefore used in the parameter adjustment at each sample. The steepest descent direction is calculated from the current sample and is modified using (inverse) Hessian information to provide the Newton direction. It may be recalled that, without forgetting, the \( P \) matrix is updated at each sample using:

\[ P_k^{-1} = P_{k-1}^{-1} + \psi_k \psi_k^T. \]  
(5.62)

Each unique update resulting from a new \( \psi_k \) vector effectively represents a new equation. The evaluation of the \( P \) matrix (inverse of \( P_k^{-1} \)) is tantamount to solving \( p \) simultaneous equations, where \( p \) is the number of parameters. Hence, \( p \) unique \( \psi_k \) vectors are required to make \( P_k^{-1} \) invertible. The forgetting factor determines the amount of information stored in the \( P \) matrix and its magnitude should therefore be related to the number of model parameters. For a fixed forgetting factor,
5.2. Development of a non-linear recursive estimator

The updating of the $P$ matrix is first-order in nature. A single time constant (in samples), $\tau$, can therefore be used to describe the dynamic characteristic of the update process; i.e.

$$\lambda = \exp\left(-\frac{1}{\tau}\right). \quad (5.63)$$

Ljung (1983) describes $\tau$ as a 'memory' time constant since it governs the extent in which past information is 'recalled' in the calculations. Matrix updates that are equal in sample age to the time constant are thus weighted by approximately 0.37. The magnitude of the time constant (and hence $\lambda$) determines the effective number of samples that are used in the $P$ matrix. This number of samples may be calculated by summing the weightings of the past samples, given by the following infinite series:

$$n = 1 + \lambda^1 + \lambda^2 + \lambda^3 + \ldots \quad (5.64)$$

$$= \frac{1}{1 - \lambda}. \quad (5.65)$$

It should be noted that this series is only convergent when $\lambda < 1$. $n$ must therefore be greater than $p$ if $P^{-1}$ is to be invertible. If it could be assumed that the system generates unique data points ad infinitum (i.e. assuming persistent excitation) the following condition defining the minimum forgetting factor can be given:

$$\lambda > 1 - \frac{1}{p}. \quad (5.66)$$

5.2.5 Robustness implications of forgetting

The incorporation of forgetting in the algorithm means that old data is discounted in favour of new data irrespective of whether any new information exists. Hence, in the likely scenario of the system remaining at one operating point, the Hessian matrix ($P^{-1}$) will approach singularity. Assuming the matrix is initialised to be positive-definite, singularity is theoretically not possible since the forgetting does not truly disregard old information, but asymptotically diminishes its weighting in the calculations toward zero. In practice, limitations in the precision of the processor do eventually lead to singularity when the inputs remain static. As the Eigen values of the Hessian matrix tend toward zero, the norm of the $P$ matrix tends towards infinity. Any other numerical errors in the estimator are thus amplified, as the $P$ matrix determines the gain vector $L$. This phenomenon is known as
5.2. Development of a non-linear recursive estimator

'blow-up' (or estimator wind-up) and various measures have been proposed in the literature to treat this problem, and these are known as 'regularisation' measures. A review of some of the more common methods is provided by Ljung (1983).

For the non-linear estimation problem considered here, the matrix singularity problem may be actually be avoided when the parameters are changing by virtue of the $P$ matrix being a function of the parameter estimates, viz. $\psi = f(u, \theta)$. Hence, changes in the parameter estimates themselves can lead to diversity in $\psi$ over a period of time. However, under stable conditions, this effect is expected to be small compared with the changes due to variations in the inputs to the model. However, it is conceivable that parameter variations induced by incipient singularity may in fact prevent eventual numerical break down in the estimation process (this effect is confirmed in tests carried out in Section 6.4.1).

The regularisation approach adopted here involves varying the magnitude of the forgetting factor. This involves varying $\lambda$ according to a criterion related to the information content of the data. Fortescue et al. (1981) proposed using a criterion based on the information content, which includes the magnitude of current prediction errors and variations in the gradient vector. The convergence properties of the prediction error forgetting (PEF) method, which is described below, have been investigated in (Cordero and Mayne, 1981). Although the method was designed to operate with linearly parameterised models it can be applied to the non-linear problem considered here since the estimation algorithm derived in Section 5.2.2 is analogous to the linear form.

Another regularisation technique considered that can be combined with PEF is known as 'directional forgetting' (DF), which was proposed to address the potential lack of diversity in the gradient vector (i.e. $\psi$). The objective behind the DF method is to compensate for the possible non-uniform distribution of information conveyed in the operating space. The method operates by allowing the $P$ matrix to be incremented or decremented by deriving an updating expression from:

$$P_{k}^{-1} = P_{k-1}^{-1} + \beta_k \psi_k \psi_k^T,$$  \hspace{1cm} (5.67)

$\beta$ can be positive or negative and is varied according to how orthogonal $\psi$ is to the information already stored in the $P^{-1}$ matrix. The DF method was originally proposed by Kulhavy and Karny (1984) and is described by Hägglund (1985). The convergence properties of the method have been investigated by Campi, 1991. When the PEF and DF methods are combined, the result is the PEDF method.
5.2. Development of a non-linear recursive estimator

(Bertin et al., 1985). The DF aspect tends to dominate the PEDF algorithm and in the event of a lack of excitation and yet changing parameters, the DF lacks sensitivity. Hence, it would be difficult to detect faults by observing the parameters while the system stays at one operating point. The PEF method does not suffer this problem since the parameter estimates remain sensitive to any error-inducing phenomena. A comparison of the three approaches is given by Bertin et al. (1986).

The PEF algorithm is favoured for the FDD application considered here since it remains sensitive to errors. The algorithm is given by:

\[ \epsilon_k = f_k(\theta_{k-1}) - y_k \]  
\[ r_k = \psi_k^T P_{k-1} \psi_k \]  
\[ L_k = \frac{P_{k-1} \psi_k}{1 + r_k} \]  
\[ \theta_k = \theta_{k-1} - L_k \epsilon_k \]  
\[ \lambda_k = max \left\{ \lambda_0, 1 - \frac{\epsilon_k^2}{(1 + r_k)\gamma} \right\}, \quad \lambda_0 \in (0, 1) \]  
\[ P_k = \frac{1}{\lambda_k} (I - L_k \psi_k^T) P_{k-1}. \]  

Equation 5.72 thus modifies the forgetting factor. An approximation is made in Equation 5.70 by substituting unity for the forgetting factor. This is done since the inclusion of \( \lambda_k \) in this expression would mean that \( \lambda \) would have to be updated before \( L_k \). This would result in a more complex algorithm where a quadratic would need to be solved at each sample. The use of unity instead of \( \lambda_k \) in Equation 5.70 therefore reduces the complexity involved in estimating \( \lambda \) and the practical differences resulting from this simplification are small in most cases (Fortescue et al., 1981). The constraint in Equation 5.72 is nevertheless necessary to protect against \( \lambda \) being too low. An appropriate lower bound is obtained from \( 1 - \frac{1}{p} \), as described in Section 5.2.4.

The parameter, \( \gamma \), determines how the size of the prediction error affects the forgetting factor. Fortescue et al. (1981) discuss the significance of this parameter on the estimator behaviour and provide the following guideline to its estimation:

\[ \gamma = \tau_0 \sigma^2. \]  

where \( \tau_0 \) is a nominal time constant that \( \lambda \) will be asymptotically derived from, \( \sigma^2 \) is an estimate of the variance of the output signal in the absence of parameter induced prediction errors, i.e. the variance of the measurement noise. When the
5.2. Development of a non-linear recursive estimator

model is perfect. The PEF algorithm is simply convergent, in contrast to the exponential convergency of the basic forgetting algorithm (Bittanti et al., 1989). This slower convergence rate has implications for the tracking ability of the estimator, as outlined by Anderson and Johnstone (1983). The main problem is that when $\lambda$ is stuck on its upper bound of 1, the $P$ matrix will tend toward zero. Eventually the precision of the machine that is running the algorithm will truncate the elements of the $P$ matrix to zero and the estimator will lose its tracking ability. This can be avoided by incorporating an upper bound on $\lambda$, which is less than unity, thereby guaranteeing alertness; for example $1 - \delta$, where $\delta$ is a small number close to the machine precision.

Implications of structural inadequacies in the model

Another problem affecting on-line parameter estimation methods that incorporate forgetting is related to the nature of the inherent structural inaccuracy of the model. In practice, the model will have a varying level of accuracy across the range of operation due to structural deficiencies in the non-linear equations. Since the parameter estimator will always endeavour to reduce the prediction error to zero, the estimator may try and improve the fit of the model in a region of structurally poor fit, at the greater overall expense of another better approximated region. This effect has been investigated by Hepworth (1994) and was termed 'over-training'. This can be avoided by reducing the sensitivity of the estimator (i.e. by increasing $\gamma$). The effect of this is to increase the size of the implicit data record, which is used in the estimation of the parameters. Hence, by making the implicit data record long enough so that it contains samples from a large portion of the operating range, the sensitivity to local structural inaccuracies will be reduced. This technique is only appropriate, however, if excitation in the input signals can be guaranteed over the period of the implicit data record.

5.2.6 Initialisation of P matrix

For the recursive algorithm described, the initial values: $\theta_0$ and $P_0$ are required. The accuracy of the initial parameter vector in relation to the real system parameters represents the amount of a priori information embedded in the model\(^4\).

\(^4\)The model structure also constitutes a priori knowledge of the system.
5.2. Development of a non-linear recursive estimator

For a truly linear-in-the-parameters model the $P$ matrix (the inverse Hessian) is directly related to the parameter covariance matrix. When the model is close to being linear the relationship becomes approximate; thus:

$$\text{cov}(\hat{\theta}_0) \approx \sigma^2 P_0.$$  \hspace{1cm} (5.75)

This relationship holds for the scalar output model case. $\sigma$ represents the variance of the output residuals and $\text{cov}(\hat{\theta}(0))$ denotes the covariance matrix of the initial parameter estimates. The usual approach is to set $P$ to be some large multiple of the identity matrix. However, this results in initial large steps in the steepest descent direction, which can result in initial instability in the face of noise, unmeasured disturbances, etc. For the application considered here, the intention is to obtain initial estimates of the parameter values before the estimator is put into operation. In this case where the parameters are known a priori, and the prediction errors are likely to be small. Hence, the initial matrix can be set to the identity matrix to ensure initial stability; thus:

$$P_0 = I.$$  \hspace{1cm} (5.76)

5.2.7 Calculation of derivatives

Calculation of the derivatives for the heat exchanger subsystem models described in Chapter 4 is a non-trivial task. Analytical evaluation of the derivatives for these models increases the design 'cost' of the FDD scheme. If the estimator is made to be reliant on analytical derivatives, the advantage of being able to utilise models developed for other purposes by simply interchanging a 'model module' is lost. The derivatives are therefore calculated numerically here, thus allowing a generic procedure to be developed that is applicable to different models.

Derivatives are approximated numerically using finite differencing, which is an application of the Taylor series expansion. The minimum requirement for the numerical evaluation of the first derivative is that the function be continuous in the region of interest. The success of the recursive prediction error method is critically dependent on the accuracy of the derivatives. One source of errors in the derivative approximation stems from neglecting higher order terms in the Taylor expansion. The effects of this 'truncation error' can be reduced, however, by employing a central differencing approach, which eliminates alternating terms in
5.2. **Development of a non-linear recursive estimator**

the error component; i.e.

\[ f'(x) = \frac{f(x + h) - f(x - h)}{2h} + \phi_T. \]  

(5.77)

where:

\[ \phi_T = \frac{h^2 d^3 f(x)}{3!} + \frac{h^4 d^5 f(x)}{5!} + \ldots \]  

(5.78)

Hence, the truncation error tends to zero as the differencing interval tends to zero. In practice, account has to be taken of the precision of the machine. An interval that is too small may cause the differences between \( f(x - h) \) and \( f(x + h) \) to be less than the machine precision, thus causing the numerator and hence the gradient to be always zero. An approximate formula for estimating the differencing interval that produces the minimum combined error due to truncation error and machine precision constraints is given by Gill et al. (1981):

\[ h = \varepsilon_R^{1/2}, \]  

(5.79)

where \( \varepsilon_R \) is the minimum value able to satisfy \( 1 + \varepsilon_R > 1 \) on the machine being used.

### 5.2.8 Calculation of confidence intervals

Confidence limits for the estimated fault parameters can be calculated quite easily by exploiting the analogy between the estimation method and multiple linear regression. As mentioned previously, the \( P \) matrix represents the covariance matrix of the parameters in linear regression. The variance of each parameter estimate is therefore given by:

\[ \text{var}(\hat{\theta}_i) = s_e^2 P_{i,i}, \]  

(5.80)

where \( \text{var}(\hat{\theta}_i) \) is the variance of the \( i \)-th parameter, \( s_e^2 \) is the sample variance of the prediction errors and \( P_{i,i} \) is the \( i \)-th diagonal element of the \( P \) matrix. The student-\( t \) statistic can be used to calculate a confidence interval on \( \hat{\theta}_i \) such that the true parameter value is attributed a 100(1 - \( \alpha \)) percentage chance of being in the range:

\[ \hat{\theta}_i \pm t_{\alpha/2, n-p} \left( \frac{\text{var}(\hat{\theta}_i)}{n - p} \right), \]  

(5.81)

where \( n \) is the effective number of data samples, \( p \) is the number of parameters, and \( t \) is the student-\( t \) distribution with \( n - p \) degrees of freedom. Since the forgetting
factor is variable, $s_{ek}^2$ and $n$ have to be calculated recursively thus taking into account the variations in $\lambda_k$, such that:

\[ n_k = \lambda_k n_{k-1} + 1 \quad (5.82) \]
\[ s_{ek}^2 = \lambda s_{ek-1}^2 + (1 - \lambda_k)\epsilon_k^2. \quad (5.83) \]

where $\epsilon_k^2$ is the square of the model prediction error. Since $n$ will be a continuous real value, $t$ should also be continuous. This continuity can be achieved using a polynomial function to enable the discrete values of the $t$ statistic to be interpolated between at a specified confidence (Leonard et al., 1992). It should be noted that the accuracy of the confidence limits calculated using these techniques is dependent on how well the quadratic approximation describes the local characteristic of the criterion function.

### 5.3 Summary of the chapter

This chapter has described a recursive parameter estimator suitable for application to models whose outputs are non-linear functions of the parameters, such as those described in Chapter 4. The algorithm is suited to the tracking of parameters that change slowly in relation to the rate at which steady-state samples are obtained from the real system. The estimator is able to remain alert to recent parameter changes as the information used in the estimation is weighted according to its age (in samples) by a process of ‘forgetting’. The problem of estimator wind-up has been considered and a modification to the basic algorithm has been described. The modification involves varying the forgetting factor according to whether the most recent data contains information that is new, compared with the data that is already implicitly stored in the parameters. When the change in the parameters between samples is small, the linear approximation to the model function will be most accurate. In this situation the $P$ matrix is analogous to the covariance matrix for the parameters. This fact is exploited by using the $P$ matrix to calculate confidence intervals for the parameter estimates.

The chapter has also described a steady-state detector used to extract steady-state data samples from the raw data obtained from the control system. The model of the subsystem is utilised so that only a single threshold is required by the steady-state detector. Noise in the measurements is treated by using a Butterworth filter.
The behaviour of the estimator is now evaluated in Chapter 6, using simulation-based techniques.
Chapter 6

Analysis of estimator using simulation

Introduction

One of the main factors influencing the robustness of the parameter estimator is the accuracy of the linear approximation to the model function. Since the validity of the approximation will depend on the functional nature of the model, the robustness of the estimator cannot be assessed independently of the model. This is in contrast to estimators that are designed to work with models that are linear-in-the-parameters.

There are a number of other factors, besides the functional form of the model, that influence the robustness of the estimator and the resulting accuracy of the estimated parameters. The most important of these are considered separately of this chapter. Throughout this chapter, a model, which is equivalent in form to the model used in the estimator, is used to generate the data. This removes uncertainty from the assessment that would be unavoidable if data from a real system were used. Moreover, by using a model to generate the data, effects such as noise, model mismatch, and unmeasured disturbances, which are present in real data, can be investigated independently.

1 The model used to generated the data is termed the 'system'.

125
6.1 Test conditions

The analysis is carried out using the heating and cooling coil subsystem models described in Chapter 4. The initial parameter values selected for the models determine the local non-linearity about which the linear approximation is made during the estimation process. The initial values therefore have the potential to affect the robustness of the estimator. Since there is a fairly high number of parameters, it is not feasible to conduct the robustness tests at every possible combination of initial values. Consequently, the initial values will remain constant throughout the tests at the values given in Table 4.4, in Chapter 4.

The 'system' used to generate the data is the same static model that is used in the estimator. All the data are in steady-state and the steady-state detector is therefore not used. In each of the tests, the $P$ matrix is initialised to the identity matrix. The estimator is used to estimate the three fault parameters selected in Section 4.3, and these are listed in Table 6.1. All other parameters of the model remain fixed at their initial values.

<table>
<thead>
<tr>
<th>FAULT PARAMETER</th>
<th>MEANING</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l$ (-)</td>
<td>fractional flow leakage through valve control port</td>
</tr>
<tr>
<td>$R_w$ (kW$^{-1}$·K)</td>
<td>tube wall resistance of heat exchanger</td>
</tr>
<tr>
<td>$b$ (K)</td>
<td>off-set of supply air temperature sensor</td>
</tr>
</tbody>
</table>

During the estimation process, each of the fault parameters was constrained to be within the following feasibility limits: $0.0001 \leq l \leq 0.7$, $0 \leq R_w \leq 15$, $-10 \leq b \leq 10$. The limits for the forgetting factor used in the recursive algorithm (Equations 5.68-5.73) were: $0.66 \leq \lambda \leq 1 - \delta$; where $\delta$ is a small number.

6.2 Analytical techniques

Most of the analytical techniques that have been proposed for analysing the behaviour of recursive estimators only consider the asymptotic behaviour, such as the values to which $\hat{\theta}$ converge to as $t \to \infty$. These techniques assume the gain of
6.2. Analytical techniques

the estimator\(^2\) tends to zero as \(t \to \infty\), and are therefore most suited to analysing the asymptotic behaviour of the estimator to step changes in the system parameters. The most common method of analysis involves representing the recursive equations as a set of deterministic differential equations. Stability theory can then be used to evaluate the properties of the differential equations, and the findings can be related to the original estimator equations. Asymptotic analyses of the PEF recursive equations applied to linear models can be found in (Cordero and Mayne, 1981).

It is the behaviour of the estimator in response to time-varying system parameters that is of most interest for the application considered in this thesis. An analytical analysis is significantly more complex for the time-varying parameter case and simulation-based techniques are therefore usually preferred (Ljung and Söderström, 1983). For the non-linear model that is considered, there is the additional problem of the parameters having the potential for collapsing on a local minimum (or a saddle point).

The 'surface' of the criterion function essentially determines the performance of the estimator. It may be recalled that the criterion function is defined by:

\[
V_N(\hat{\theta}) = \frac{1}{2} \sum_{k=1}^{N} \left[ f_k(\hat{\theta}, u) - y_k \right]^2,
\]

where \(y_k\) is the measured output from the system, which is a function of the inputs \((u)\) and the system parameters \((\theta)\), i.e. \(y_k = f(u, \theta)\). The shape of the criterion function surface is therefore characterised by the values of the vectors \(u, \hat{\theta}, \) and \(\theta\) for the data record (from \(k = 1\) to \(k = N\)). The incorporation of forgetting simply makes \(V\) more sensitive to recent values of \(u, \hat{\theta}, \) and \(\theta\). Two factors that will affect the shape of the criterion function and hence the performance of the estimator are listed below.

1. Accuracy of the linear approximation to the model function over the region of discrepancy between the system parameters \((\theta)\) and the model parameters \((\hat{\theta})\).

2. Coverage of the input space \(u_k\), for \(k = 1\) to \(k = N\) and how well this accentuates the differences between the parameters.

\(^2\)In the estimator considered in this thesis, the gain is ultimately determined by the time-varying forgetting factor \(\lambda_k\).
As the accuracy of the linear approximation to the model function deteriorates, the surface of the criterion function becomes less quadratic, and a potential for local minima arises. For the problem of estimating the parameters for a batch of data (Section 5.2.1), the number of iterations (parameter movements in the Newton direction) required to converge on the solution increases as the surface of the criterion function deviates from being quadratic. The recursive implementation of the estimator does not iterate on the whole data record, instead an additional iteration can only take place when a new data sample is obtained. The existence of a non-quadratic minima will therefore impede the ability of the estimator to track changes in the parameters. Moreover, the linear approximation may be inaccurate enough to induce local minima in the criterion function and the parameters may subsequently collapse on a false set of values.

In practice, the coverage of the model input space will be non-uniform due to stochastic disturbances and the effect of the non-linearity in the input space. An illustration of this non-linearity, for the heating coil subsystem model, is presented in Figure 6.1. This figure shows how the normalised air-side approach (defined in Equation 4.63) of the model varies in the input space of the fractional air flow rate and the valve control signal.

A sparse coverage of the input space has the effect of flattening the minimum. Consider the two dimensional problem of estimating the degree of thermal resistance in the heat exchanger ($R_w$) and the degree of leakage ($l$). This problem is illustrated in Figure 6.2, which shows the surface of the criterion function in the parameter space of $R_w$ and $l$, in the form of a contour plot. Each of the four surfaces in the figure was generated by defining the optimum (system) parameter vector as $\theta = [R_w, l, b] = [0.5, 0.05, 0.0]$. Different numbers of input-output vectors were then used to generate each of the surfaces. All the inputs to the model were constant except the valve control signal, which was varied between the ranges specified at the top of each graph. It can be observed that when only one data vector was used (denoted by $u = 0.5 - 0.5$ on the graph), a trough exists on the surface. Any combination of $R_w$ and $l$ in the trough could therefore minimise the criterion function, and the problem does not have a unique solution. As the number of data vectors is increased, a minimum can be seen to develop and become more pronounced.

When the minimum is not well defined and the curvature is small, there is increased uncertainty involved in the estimation of the parameters. The method for
6.2. Analytical techniques

Figure 6.1: Non-linearity of the heat exchanger model in the input space

calculating the confidence intervals, described in Section 5.2.8, exploits this fact by
narrowing the interval in proportion to the curvature of the minimum. Surfaces
of \( V \) have been generated for the other two combinations of fault parameters \( (b, R_w, \text{ and } b, l) \) and these are presented in Figures 6.3 and 6.4 respectively.

The surfaces in Figures 6.3 and 6.4 undergo similar topological changes to the
surface in Figure 6.2 as the number of data vectors is increased. The surfaces
are examples that were contrived to demonstrate the variation in the curvature
of the minimum by using increasing numbers of (distinct) input-output vectors
from regions of the operating range that were known to allow the parameters to
be distinguished. Hence, leakage and fouling can be distinguished when the valve
control signal is either high or low; sensor off-set and fouling when the control
signal is low; and sensor off-set and leakage when the control signal is high.

In the case where the surface of the criterion function is globally quadratic, the pa-

\(^3\)The confidence interval is related to the magnitude of the prediction error in addition to the
number of unique input-output data vectors.

\(^4\)As is the case for a linear regression model: \( y(t) = \phi^T \theta \).
6.3 Robustness to different fault types

Parameter estimates will always converge on the best estimates (in the least-squares sense) for the particular input-output data. The nature of the changes in the system parameters does not therefore alter the asymptotic convergency of the parameters. In the case of the non-linear models considered here, there is the possibility that there may be local minima, or saddle points on the surface of the criterion function. Hence, the estimator may converge on parameter values that are not the best estimates for the data. The nature of the change in the parameters is therefore of particular importance for the non-linear estimator and this is investigated in the next Section.

6.3 Robustness to different fault types

One of the main factors affecting the performance of the estimator is the nature of the parameter changes. Two types of faults have been mentioned previously: abrupt and incipient. Abrupt faults cause a step change in the parameters, while
6.3. Robustness to different fault types

Figure 6.3: Surface of the criterion function in $b, R_w$ parameter space for varying amounts of data from the input space

Incipient faults develop slowly causing a change in parameters that is a slow, continuous function of time. In the case of incipient faults, the rate of change of the parameters may vary. For simplicity, it will be assumed that this type of fault causes the parameters to change at a constant rate, thereby allowing the behaviour to be modelled as a ramp change. By assuming that abrupt and incipient faults can be modelled as step and ramp changes respectively, two factors need to be investigated: the step size and the gradient of the ramp.

The effects of varying the step size and the gradient of the ramp change are analysed empirically in Sections 6.3.1 and 6.3.2 for the heating and cooling coil subsystem models respectively. The two input signals that are most likely to vary rapidly during normal operation of these subsystems are the flow rate of the inlet air ($\dot{m}_a$) and the valve control signal ($u_v$). Data is generated from the simulated system by varying these two inputs in a random way to ensure data diversity and hence the identifiability of the model:

$$\dot{m}_a(t) = r(t)\dot{m}_{a,max}$$  \hspace{1cm} (6.2)

$$u_v(t) = r(t),$$  \hspace{1cm} (6.3)
6.3. Robustness to different fault types

Figure 6.4: Surface of the criterion function in $b, l$ parameter space for varying amounts of data from the input space.

where $0 \leq r(.) \leq 1$ is a random number, $m_{a,\text{max}}$ is the maximum air-flow rate deliverable by the supply fan. All other inputs are kept constant during the analysis at the values shown in Table 6.2.

Table 6.2: Values of the fixed input variables

<table>
<thead>
<tr>
<th>INPUT VARIABLE</th>
<th>VALUE</th>
<th>Heating Coil Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>inlet air temperature</td>
<td>30 °C</td>
<td>10 °C</td>
</tr>
<tr>
<td>inlet water temperature</td>
<td>7 °C</td>
<td>80 °C</td>
</tr>
<tr>
<td>inlet air relative humidity</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

The inlet air is set to be dry so that the cooling coil is tested in its dry mode only to avoid the possibility of evaluating the derivatives around the point where the coil model is switched between its wet and dry modes.

Sequences of steps tests and ramp tests of varying magnitudes and gradients were performed on the data gathered for the tests in the previous section.
performed on the fault parameters, for the heating and cooling coil models. The sequence of step sizes was as follows: \( \Delta \theta, 2\Delta \theta, 3\Delta \theta, 4\Delta \theta \); where \( \Delta \theta \) is given by:

\[
\Delta \theta^T = \begin{bmatrix} \Delta R_w & \Delta l & \Delta b \end{bmatrix}
\]

\[= \begin{bmatrix} 1.0 & 0.05 & 0.5 \end{bmatrix},\]

\( \Delta \theta \) was selected to correspond approximately to the degree of fault for which detection would be expected to be possible, in view of the noise and uncertainties associated with real systems.

The parameters \( R_w \) and \( l \) were stepped in the following directions: +, +, −, +, −. The sensor off-set parameter, \( b \), was stepped in alternate positive and negative directions. Each of the parameter changes was held for 300 samples. The ramp tests involved ramping the parameters at a rate that enabled them to reach the same magnitude of change as each of the steps over the 300 sample period. The sequence of (unsigned) gradient changes (per sample) is thus given by: \( \frac{1}{300} \Delta \theta, \frac{1}{150} \Delta \theta, \frac{1}{75} \Delta \theta, \frac{1}{37.5} \Delta \theta, \frac{1}{37.5} \Delta \theta \). In addition, the fault parameters were fixed at their initial values for the first 100 samples of each test sequence to let the \( P \) matrix stabilise. The parameter \( \gamma \) in the estimator (Equation 5.72) was set to 0.1. Such a low value is reasonable since the model and ‘system’ are identical and there was no uncertainty associated with the signals. All fault parameters are varied during the tests so that the ability of the estimator to track changes in each of the parameters can be compared simultaneously.

### 6.3.1 Heating coil subsystem

The results of the step and ramp tests on the heating coil subsystem are presented in Figures 6.5 and 6.6 respectively. Each figure contains five graphs. The top three graphs show the variation in the fault parameters with sample time, the fourth graph from the top shows the trace of the \( P \) matrix, and the bottom graph shows the magnitude of the forgetting factor. The dash-dot line on each of the parameter graphs is the true system parameter values, while the solid line represents the model estimates. In addition, each of the parameter estimates is surrounded by two solid lines, symmetrical about the estimates, and these lines represent a 99.9% confidence interval (estimated using the method described in Section 5.2.8).
6.3. Robustness to different fault types

![Figure 6.5: Heating coil step test](image)

The behaviour of the estimator is similar to a first order filter for the smallest step changes, but oscillatory behaviour ('hunting') can be observed during the period immediately following the largest parameter changes (1000 samples, and 1300 samples). The direction in which the parameter vector is updated is determined by the current gradient information and the accumulated curvature information in the $P$ matrix. At the instant of change this information is based on the parameter values of the system prior to the step. The appropriateness of the search direction calculated using this information is therefore dependent on the validity of the linear approximation, over the region of parameter change.
6.3. Robustness to different fault types

As the instantaneous error determines the size of the parameter change (Equation 5.58) a large error caused by a step change will result in a large movement in the search direction calculated using old parameter values. Providing successive parameter adjustments manage to move the parameters closer to their target values, the directional information will become increasingly appropriate and the estimates will converge. Hence a period of initial oscillation is to be expected for the non-linear estimator. The estimates of the confidence interval reflect the initial uncertainty immediately following a parameter change. The intervals widen in response to the parameter changes, and narrow as more information at the new
6.3. **Robustness to different fault types**

parameter position is obtained. The estimator was able to track the parameter changes closely during the ramp tests. The only significant inaccuracies in the estimates occurred during the final ramp, which caused the estimates of \( l \) and \( R_w \) to become erroneous. The trace of the \( P \) matrix can be seen to increase in magnitude during the period that the accuracy of estimates deteriorates. The \( P \) matrix affects the gain in the estimator and may thus be responsible for moving the estimates too far in an inappropriate direction during the period of the final ramp change.

It may be noted that the model function is linear in the direction of the \( b \) parameter. The gradient of the model output with respect to \( b \) is constant \((b = 1)\) over the range of operation and is independent of the inputs and other parameters. Hence, the linearisation in the direction of this parameter carried out at each sample is exact. Estimation of this single parameter in isolation would thus be a linear least-squares problem. However, as the estimation task is multi-dimensional, the search direction can be affected by the approximations made in the other non-linear dimensions. Since the derivative of the model function with respect to the off-set parameter is constant across the range, the estimator always remains active in the direction of this parameter. Hence, the inclusion of this parameter in the estimation process will contribute to the uncertainties associated with the estimation of the other parameters.

During the step test sequence, the forgetting factor remains on its upper bound for most of the time but is reduced in response to the initial, large prediction errors resulting from the steps. During the ramps, the forgetting factor is oscillatory within its upper and lower bounds. As the model is non-linear, the magnitude of the prediction error will vary with the inputs. Hence, the random nature of the input excitations is responsible for the noisy behaviour of the forgetting factor. The trace of \( P \) is the sum of the Eigen values and can thus be used as a measure whether the matrix is approaching singularity (when \( \text{trace}(P) \to \infty \)). Moreover, the trace may be used as an indication of the condition of the solution. From this interpretation it appears from the step test sequence as though the condition\(^5\) deteriorates as the magnitude of change in the parameters is increased. This effect can also be observed during the ramp sequence, where the trace decreases when the parameters approach their original values.

\(^5\)For the non-linear estimator, a poorly conditioned solution may indicate a break-down in the quadratic approximation.
6.3. Robustness to different fault types

The results of the step tests and ramp tests on the heating coil subsystem show that the tracking ability of the estimator does, as anticipated during the derivation, deteriorate with increasing step size and ramp gradient. The estimator loses most of its tracking ability during the initial response to the step sizes that were greater than or equal to $2\Delta \theta$. For the ramp changes, the parameter estimation errors are relatively small for all the gradients below $\frac{1}{75} \Delta \theta$. It may be noted that the confidence intervals manage to predict the level of inaccuracy in the parameter estimates quite well.

6.3.2 Cooling coil subsystem

Figures 6.7 and 6.8 show the performance of the estimator when the dry cooling coil subsystem is subjected to the step and ramp test sequences.

The response of the estimator is dependent on the local non-linearity of the model function. As the characteristics of the heating and cooling coil models are different and have different initial parameter values, the performance of the estimator is expected to differ between the two models. This is evident in both the step tests and ramp tests. The estimator appears to be relatively stable for the smaller steps but the larger steps induce severe instability. A large oscillation in the parameter estimates can be observed for the third step in the sequence, however the estimates do converge to the correct values toward the end of the step. The fourth step causes the most problems, with the estimates failing to converge on the correct values. This could be due to the estimator converging on a local minimum, or saddle point. The forgetting factor is kept low (and hence the gain is kept high) during the period following the fourth step due to the magnitude of the prediction errors resulting from an inadequate model 'fit' at the position of the false solution. Changes in the system parameters cause the characteristic of the criterion function to change. The final step causes a significant enough change to enable the parameter estimates to leave the false solution and converge on the correct values. As with the step sequence, the accuracy of the parameter estimates in response to the different ramp changes is inferior to that experienced for the heating coil subsystem.

The performance obtained from the estimator for the heating and cooling coil models was similar for the smaller changes in parameters. However, the estimator
was less able to track the large parameter changes in the cooling coil model than in the heating coil. One conclusion that could be made from this is that the cooling coil is more non-linear than the heating coil in the local region about which the linearisations are made. Since each model uses identical valve models, the only differences in model structure lie in the NTU-effectiveness expressions. The differences in the non-linearities are therefore likely to be mostly due to differences between the initial estimates of all the other parameters of the models, e.g. valve authority, heat transfer resistances, etc.

Figure 6.7: Cooling coil step test
6.4 Variability of input signals

Changes in the parameters of the system cannot be identified unless information is obtained from regions of the operating range where the effects of the changes are apparent. For non-intrusive FDD schemes this means that the detection of certain faults may be delayed until they affect the operation. For example, fouling may accumulate in a heat exchanger while the coil is operating at low duties, but the effect of the fouling (in terms of air temperature variations) will only become apparent when the coil is operated at high duties. There are three operating
6.4. Variability of input signals

regions where data should be obtained to enable the fault parameters of the heat exchanger models to be made easily identifiable:

- low duty - where leakage is most apparent (i.e. closed valve);
- mid duty - where the effects of leakage and the fouling are small compared to the effects of sensor off-set;
- high duty - where fouling is most apparent.

As discussed in Section 6.2, the performance of the estimator is affected by the variation in the input variables. For example, if there is a change in the leakage parameter it will be apparent at low duty. However, any observed change at a single point in the low duty region could also be due to a sensor off-set. The variation in the inputs must therefore be such that information is obtained from the mid duty region. This then eliminates the possibility of a change in the off-set parameter being the cause of the anomaly. For the three fault parameters considered, the reliable estimation of parameter changes requires information from all regions. The rate by which the inputs are required to vary in order to identify the three parameters depends on how quickly the parameters are changing, i.e. as the parameters vary more quickly, the inputs also have to vary more quickly to ensure identifiability.

The problem of obtaining sufficient input excitation is investigated in this section under the circumstance of all parameters varying slowly with time. Information from low, medium, and high duties is required to ensure identifiability of all the parameters. The differences between the effects of changes in the three fault parameters can be made to be more apparent by changing the control signal to the coil valve rather than the air-mass flow rate. Hence, it is the nature of the variations in this signal that mainly determine the tracking ability of the estimator. The performance of the estimator is therefore examined in the next section by varying the frequency of this signal. The heating coil model is used during the tests and the initial parameter and input values are set equal to the those used in the previous section. The fault parameters of the system are all ramped slowly at the following constant rates (per sample): \( \Delta l = \frac{1}{15000} \), \( \Delta R_w = \frac{1}{1500} \), \( \Delta b = \frac{1}{3000} \).
6.4. Variability of input signals

6.4.1 Sinusoidal variations in valve control signal

Figures 6.9, 6.10, and 6.11 show the results of varying the control signal to the valve actuator sinusoidally with maximum amplitude, at different frequencies. It can be observed that as the frequency decreases, the ability of the estimator to track the parameter changes deteriorates. The parameter estimates oscillate about the true parameter values with the same frequency as the input signal for the two higher frequencies, but at the lower frequency the estimates become unstable. The breakdown of the estimator at low frequencies is due to the lack of information. The estimator is unable to reduce the prediction error at each sample since not enough directional information has been accumulated. The parameters are thus moved in an inappropriate direction and subsequent prediction errors, which determine the size of the parameter step, increase.

The variable forgetting procedure responds to an increase in prediction errors by reducing the magnitude of the forgetting factor. This occurs at approximately 350 samples in Figure 6.11. The variation in the forgetting factor and the trace of $P$ for this low frequency case is shown in Figure 6.12. It can be observed that while the forgetting factor is low, the trace increases rapidly\(^6\). When the

---

\(^6\)The range of the trace is so large that a log scale is used for the y axis.
Figure 6.10: Medium frequency ($\frac{1}{250}$ cycles/sample) variation in the valve control signal

Figure 6.11: Low frequency ($\frac{1}{1000}$ cycles/sample) variation in the valve control signal

trace is high the gain of the estimator is also high, which results in the large
6.4. Variability of input signals

Parameter changes. It may be observed that the period of instability around 400 samples is soon followed by re-stabilisation of the parameter estimates. This is an interesting phenomenon, which is due to the fact that $\psi$, and hence $P$, is a function of the parameters as well as the inputs. Thus, due to the non-linear nature of the estimation problem, instability that is induced due to the $P$ matrix tending toward infinity is self-correcting. The effect can be observed in Figure 6.12 where the trace of $P$ peaks at around $10^6$ but is reduced to around $10^2$ when the parameter estimates oscillate\(^7\).

![Graph showing trace of P and forgetting factor over sample time](image)

Figure 6.12: $\lambda$ and the trace of $P$ for low frequency variation in the valve control signal

Prediction error forgetting is intended to avoid the Hessian matrix tending toward singularity when the parameters are approximately constant, and while the system remains at one operating point. It also improves the ability of the estimator to track parameters that vary at a non-constant rate. However, the procedure does not explicitly protect against a lack of input excitation, since the information content of the data is determined jointly by the variation in $\psi$ and the magnitude of the error. Hence, instability in the estimator will result if the inputs are not varying at a rapid enough rate to keep up with parameter variations. This instability is, however, an indication that the parameters are changing. The estimator

\(^7\)The oscillations are constrained within the feasibility limits that are imposed.
therefore retains its fault detection capabilities despite being presented with insufficient data. The potential for instability is not a desirable attribute, however, and techniques that utilise expert knowledge to constrain the estimation process may be required in practice; this approach is sometimes known as ‘jacketing’ (Dexter and Haves, 1989). One possible technique would be to limit the rate of change of the parameters in accordance with expectations of the likely rate of change of the system parameters.

The results presented in Figures 6.9, 6.10, and 6.11 illustrate the effects of variations in the valve control signal while maintaining the other inputs at a constant level. Since the models are designed to be used with VAV systems it is unlikely that the air-mass flow rate will be constant. Hence, the results represent a worst-case scenario. Variations in the mass flow rate of the air also serve to accentuate differences between parameters, although not to the same extent as variations in control signal. The robustness and tracking ability of the estimator is therefore improved when the air flow varies. This is illustrated in Figure 6.13, which shows the improvement to the tracking at the cycling frequency of 1/250 samples when the air flow is also varied at a high frequency, with a small amplitude about its mean value. The improved tracking ability is accompanied by narrower confidence intervals, thus confirming the improved level of certainty in the estimates.

Implications of variability in inputs for real systems

The results of the tests with different frequency variations in the input signals show that instability is induced in the estimator at a frequency of $\frac{1}{1000}$ cycles per sample. The change in the system parameters for one cycle at this frequency is:

$$
\Delta \theta^T = [\Delta R, \Delta l, \Delta b]
$$

$$
= [0.67, 0.067, 0.33]
$$

In practice, diurnal variations in inputs to the system represent the dominant form of cyclic excitation. Hence, based on the empirical analysis, the estimator would be susceptible to instability if the degree of fault increased by more than the $\Delta \theta$ given above, each day.

The estimator is able to track the system parameters reasonably well at the frequency of $\frac{1}{250}$ cycles/sample. The change in parameters per cycle at this frequency
6.5. Measurement noise

This section investigates the effect that noise has on the performance of the estimator. In practice, noise is present in the signals obtained from sensors that are used to monitor physical properties, i.e. air-flow, temperature, pressure, etc. Although there is no noise directly associated with control signals these signals can contain high frequency components due to the operation of feed-back control based on noisy sensors.

Noise effects can help alleviate some of the numerical problems that can occur in the estimator. In particular, the singularity problem associated with the Hessian
6.5. **Measurement noise**

Matrix is avoided if the noise level is sufficient to induce significant variation in $\psi$ between samples. Noise also affects the estimates of the parameters, since the size of the parameter adjustment at each sample is governed by the prediction error. Variations in the prediction error, due to noise, therefore pass directly through to the parameter estimates. In addition, the problem is accentuated since the estimator was derived based on a least-squares criterion function. The effect of noise on the parameter estimates will therefore increase quadratically with magnitude.

The effect of noise on the parameter estimates is demonstrated under the condition of maximum input signal variability and slowly ramped parameters. Gaussian noise having zero mean and different variances is applied as an additive component to the temperature of the air leaving the coil.

![Figure 6.14: The effect of noise: $\sigma^2 = 0.05K$](image)

Figures 6.14, 6.15, and 6.16 show the results obtained from the estimator with $\gamma$ fixed at 0.1. It can be observed that the parameter estimator acts as a filter to the noise (in the prediction errors) and the parameter estimates are thus fairly insensitive to the noise effects. In contrast, the confidence interval is related directly to the variance of the noise in the output as the prediction error is used in its calculation (Equation 5.80). Although the mean of the parameter estimates can be observed to correspond quite well with the true parameters, the non-linearity
of the criterion function means that the optimum solution may be pushed outside the region where a local quadratic model yields the correct parameters. This may be the cause of the deterioration in the accuracy of the mean, after 1000 samples.
6.5. Measurement noise

for the 0.1 K variance level.

Noise in the output signal can be counteracted by increasing the magnitude of \( \gamma \), which ultimately reduces the sensitivity of the estimator to real prediction errors. Reductions in the sensitivity of the estimator are unavoidable if the effects of noise are to be sufficiently diminished. Hence, the sensitivity and reactiveness of the parameter estimator in response to real parameter changes will be proportional to the level of noise in the measurements. Figure 6.17 shows the improvement to the estimator stability at the noise variance level of 0.2K that is achieved by increasing \( \gamma \) to twenty\(^8\). The parameter estimates are more stable but the lag increases between the estimates and the real parameters. It can also be observed that the confidence interval is narrower during the initial part of the response to the ramp. This is due to the initial values in the \( P \) matrix\(^9\) taking longer to diminish due to the decreased sensitivity.

![Figure 6.17: The effect of increasing \( \gamma \) to compensate for noise](image)

\(^8\)Using the Equation 5.74, where \( \gamma = \sigma^2 \tau_0 \), \( \tau_0 \) is thus calculated as 100 samples.

\(^9\)The \( P \) matrix is initialised to the identity matrix.
6.5. Measurement noise

6.5.1 Quantification of noise effects in real systems

Noise in the model inputs can be treated as additional noise in the output and hence also counteracted by adjusting $\gamma$. Consider a SISO system, which is non-linear in the inputs:

\[ y = f(x + \eta) + \mu, \quad (6.8) \]

where $f(.)$ is the model function, $\eta$ is the noise in the input signal and $\mu$ is the noise in the output signal. Equation 6.8 can be expanded as follows:

\[ y = f(x) + \left[ \eta \frac{df(x)}{dx} + \mu \right]. \quad (6.9) \]

Hence, the total noise present in the prediction error\(^{10}\) may be calculated by summing the output noise and the input noise multiplied by the derivative. For the MISO case the derivatives become partial and the total additive noise is given by:

\[ v = \varphi^T \eta, \quad (6.10) \]

where:

\[ \varphi^T = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \frac{\partial f(x)}{\partial x_2} & \cdots & \frac{\partial f(x)}{\partial x_i} \end{bmatrix}, \quad (6.11) \]

\[ \eta^T = [\mu \eta(x_1) \eta(x_2) \cdots \eta(x_i)]. \quad (6.12) \]

The total noise therefore depends on the partial derivatives of the model function, with respect to each of the inputs. These derivatives vary with operating point and current parameter vector, since the model output is a non-linear function of the inputs and the parameters. The main inputs for the dry coil models are considered below.

**Input temperature sensors**: The input temperatures to the heat exchanger models are the inlet air temperature $T_{ai}$ and the inlet water temperature $T_{wi}$. The partial derivative (gain) with respect to the model output ($T_{ao}$) at a specific operating point is easily calculated from the following:

\(^{10}\)The prediction error directly affects the updating of the parameters, and any noise present comes from the measured output and the model output; the noise in the model output originating in the measured inputs.
6.5. **Measurement noise**

**Cooling Coil:**

\[
T_{ao} = T_{ai} - \frac{\epsilon C_{\min}}{C_a} (T_{ai} - T_{wi}) \quad (6.13)
\]

\[
\frac{\partial T_{ao}}{\partial T_{ai}} = 1 - \frac{\epsilon C_{\min}}{C_a} \quad (6.14)
\]

\[
= 1 - \alpha \quad (6.15)
\]

\[
\frac{\partial T_{ao}}{\partial T_{wi}} = \frac{\epsilon C_{\min}}{C_a} \quad (6.16)
\]

\[
= \alpha \quad (6.17)
\]

where \(\alpha\) is the air-side approach of the coil (Equation 4.63). Note: the partial derivatives for the heating coil and cooling coil are equivalent. If necessary, these derivatives can be calculated on-line since the required variables are readily available in the model.

**Valve control signal:** As discussed previously, the noise manifested in the valve control signal originates in the sensor used in the control loop as the controlled variable. In the heat exchanger subsystems considered here, the control variable is the temperature of the air leaving the coil. The noise in the control signal may thus be estimated from the noise in this sensor, providing the control law is known. Most of the controllers used with heat exchanger subsystems use a PI\(^1\) control law. If it can be assumed that the noise in the output sensor has zero mean, the integral term does not contribute to the transmission of noise to the control signal, when in steady-state. The transmitted noise is then simply determined from the proportional control action:

\[
u(t) = k [w(t) - (y(t) + \mu(t))]\]

where \(w(t)\) is the set-point, \(u(t)\) the control signal, \(y(t)\) the controlled variable and \(\mu(t)\) the noise component. The magnitude of the noise in the control signal is therefore simply \(k\mu(t)\).

**Air-flow sensors:** The derivative of the model with respect to the air-flow rate is more difficult to evaluate. Analytical derivatives of the model are not easily evaluated and numerical estimation would impose additional computational load in the on-line calculations, which is undesirable. An estimate of the mean partial derivative \(\frac{\partial T_{ao}}{\partial m_a}\) was obtained for the heating and cooling coil subsystem models\(^2\) by testing the models across the range of operation.

\(^1\)Proportional plus integral.

\(^2\)Parameterised for the BRE ACE facility.
calculating the derivatives numerically. The values that were estimated are shown in Table 6.3.

Table 6.3: Mean model derivative values with respect to air-flow rate

<table>
<thead>
<tr>
<th>HEATING COIL</th>
<th>COOLING COIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\partial T_{ae}}{\partial m_a}$</td>
<td>2.4 (K·kg(^{-1})·s)</td>
</tr>
</tbody>
</table>

The noise levels for temperature and air-flow sensors were established empirically using data obtained from a full size air-conditioning test rig, built to typical industry standards\(^\text{13}\). The noise variance levels are presented in Table 6.4.

Table 6.4: Sensor noise levels

<table>
<thead>
<tr>
<th>SENSOR</th>
<th>NOISE VARIANCE ($\sigma^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>temperature</td>
<td>$2.05 \cdot 10^{-2}$ K(^2)</td>
</tr>
<tr>
<td>air-flow</td>
<td>$7.63 \cdot 10^{-4}$ Kg(^2)·s(^{-2})</td>
</tr>
</tbody>
</table>

The total noise variance $\sigma_T^2$ for a cooling coil subsystem model, based on these measurements, is therefore calculated as follows:

$$
\sigma_T^2 = \text{var}(m_a) \cdot 2.3 + \text{var}(T_{ui}) \cdot \alpha + \text{var}(T_{ai})(1 - \alpha)
\text{ + var}(T_{ao}) k_p + \text{var}(T_{ao}).
$$

Hence, using typical maximum values:

$$
\sigma_T^2 = 7.63 \cdot 10^{-4} \cdot 2.3 + 2.05 \cdot 10^{-2}(0.2 + 0.8 + 0.1 + 1)
= 4.48 \cdot 10^{-2}.
$$

Assuming the model describes the system perfectly, $\gamma$ should therefore be set to a minimum value of approximately 0.05$\tau_0$, where $\tau_0$ is the asymptotic time constant (in samples), the significance of which was discussed in Section 5.2.5.

\(^{13}\)The test system was the BRE ACE facility. The temperature sensors were single point thermistors (Vaisala) and the air-flow sensors were velocity pressure manometers (Staefa fka-v2).
6.6 Specification errors

A specification error can develop between the model and the system when any attribute of the system, other than the estimated parameters, changes. Modelling errors and unmeasured disturbances therefore both induce specification errors, and these are considered in this section.

6.6.1 Unmeasured disturbances

The only disturbances that are of relevance to the parameter estimator are those that affect the output signal. If all the inputs to the heat exchanger subsystem are measured, there is very little potential for unaccounted phenomena to disrupt the model output. There are some very minor effects that are not accounted for such as heat transfer through the ductwork, but these are negligible in comparison to the modelled phenomena. The main potential for error is incurred when it is assumed that certain inputs remain constant due to them being unmeasurable for a particular plant. The two inputs to the model for which this is most likely to be the case are:

- supply water temperature;
- maximum water flow rate though the coil (i.e. delivered by the pump).

The supply water temperature may be measured in some installations, but it is common for the measurement to be available only to isolated controllers of the boiler and chiller plant. The water flow rate delivered by the pump is very rarely measured. These two properties therefore often have to be assumed to be constant at their design values. Deviations in the actual values from the design values, during the operation of the parameter estimator, will induce prediction errors, which will cause the estimator to make adjustments to the parameters. Hence, the variation in these inputs is relevant to the operation of the FDD scheme.

It should be noted that by assuming that $T_{wi}$ and $\dot{n}_{\text{max}}$ remain constant, it is implicitly assumed that the primary plant circuit operates without faults. If a fault does develop in this circuit the assumption will be violated and variations
in the parameter estimates will occur. Although the resulting parameter changes will not be indicative of the true nature of the fault, the sensitivity that is offered facilitates the detection of such faults.

Figures 6.18 and 6.19 respectively show the effect that unmeasured disturbances in $\dot{m}_{\text{max}}$ and $T_{\text{wi}}$ have on the parameter estimates for the heating coil subsystem. In each of the tests the air-mass flow rate and the valve control signal were varied randomly within their ranges. In Figure 6.18 $\dot{m}_{\text{max}}$ is ramped slowly from its design value, which is used in the estimator, to twice this value over a period of 1500 samples. The parameter estimates respond to the unmeasured change with the most noticeable adjustment being made to the sensor off-set parameter\textsuperscript{14}. Smaller adjustments are also made to the other two parameters. A similar bias toward the sensor off-set parameter can be observed in Figure 6.19, when the hot water supply temperature is ramped down from its design value to half that. Note that the change in the off-set parameter is negative in this test.

Figure 6.18: Unmeasured disturbance to the water mass flow rate. $\gamma = 5$

\textsuperscript{14}$R_w$ is unresponsive since it is constrained to be greater than zero.
6.6. Specification errors

The effect that unmeasured disturbances and faults have on the estimator are very similar. Both phenomena cause the model to be incapable of predicting the system output based on the previously applicable parameter values. The estimator responds to this situation by adjusting the parameters to minimise the prediction errors over the (weighted) number of samples, determined by the forgetting factor. The difference between faults and unmeasured disturbances lies in the ability of the model structure to describe the system behaviour following the change. If a fault occurs in the system that corresponds to a change in one of the estimated fault parameters, the estimator will be able to adapt the relevant parameter to maintain structural adequacy. Conversely, if a fault or an unmeasured disturbance occurs in the system that cannot be represented by a change in a fault parameter, the achievable ‘best fit’ of the model to the data samples, and hence structural adequacy, will be reduced.

The estimator will always attempt to minimise the prediction errors by adjusting the parameters. However, when the change is not explicitly catered for within the model structure, a mismatch develops between the model and the system. This leads to a reduction in the overall attainable accuracy of the model and serves to keep the magnitude of the prediction greater than zero, even when the parameters have converged on their ‘best fit’ values. Since the magnitude of error determines

Figure 6.19: Unmeasured disturbance to the water supply temperature. $\gamma = 5$
6.6. Specification errors

the size of the parameter adjustment the parameters will remain active while the prediction error cannot be reduced to zero. The sensitivity of the estimator will also be increased by the variable forgetting procedure since the amount by which the forgetting factor is reduced is governed by the error. Further implications of mismatches between the system and the model are now discussed in the next section.

6.6.2 Modelling errors

A model is supposed to be a simplified representation of the real system and there will always be, in practice, at least a small underlying structural mismatch. The model parameters that are estimated only correspond to the true system parameters when the structures of the model and system are equivalent. Since a perfect model-system match is not realisable in practice it is usually accepted that the model parameters are a ‘good approximation’ to the system parameters providing the model is designed well. However, this assumption is usually based on obtaining the best global fit to the system.

When the structures of the model and system are different, the best global fit is likely to be inferior to the best local fit\textsuperscript{15}. It may be recalled that the parameter estimator always endeavours to reduce the prediction errors toward zero. To achieve this, the estimator adjusts the parameters based on an (implicit) data record, the size and diversity of which is determined by the the forgetting factor and the variation of the system inputs. If the data record is biased toward a small local region of the operating space, the estimator will endeavour to locate the parameter estimates that best fit the local region. These estimates will then be different to the global best estimates. The parameter estimates of a structurally incorrect model will thus fluctuate as data is obtained non-uniformly from the system.

\textsuperscript{15}This is not true, however, when the structures are equivalent, since the best fit would yield zero prediction errors.
6.7 Conclusions

The performance of a parameter estimator has been investigated empirically in this chapter. The results of the investigations revealed characteristics of the estimator that have implications for FDD in real systems. The conclusions drawn from the results are summarised below:

Fault types: It was found that the tracking ability of the estimator deteriorated as rate of change in the parameters was increased. The estimator was tested with the heating and cooling coil subsystem models for ramp changes and step changes in the parameters. For small steps and ramps, the behaviour of the estimator was similar to a first-order filter. The larger changes were characterised by more oscillatory behaviour ('hunting') as the linear approximations to the model function became less appropriate. The confidence intervals were able to indicate the level of uncertainty in the estimates quite well by widening as the size of parameter change was increased. Overall, the performances obtained from the estimator were similar for the heating and cooling coil models.

Excitation level of inputs: The estimator was tested under the condition of slowly ramping parameters and varying levels of excitation, controlled by varying the frequency of sinusoidal input signals. In the case of very low frequencies the variable forgetting failed to protect against singularity developing in the Hessian matrix, and the norm of the $P$ matrix became very large, inducing instability. The variable forgetting is based on information content, as jointly determined by variations in the gradient vector, $\psi$, and the prediction errors. Hence, the slowly ramped parameters induce prediction errors that serve to keep the forgetting factor less than unity. It was found that the ability of the estimator to track parameter variations depends on the rate of change and absolute variation of $\psi$ relative to the rate of change of $\theta$.

Noise: Noise in the measurements directly affects the parameter estimates. If the noise level is high, there is a risk that the parameters may be pushed too far from the optimum values thus causing a deterioration in the accuracy of the quadratic model of the criterion function. The effects of noise can be reduced by making the estimator less sensitive to the prediction errors. However, the
estimator is then less able to track fault-induced changes. Reduced tracking ability is therefore a consequence of having to use noisy signals.

**Modelling errors and unmeasured disturbances:** These two phenomena lead to specification error; i.e. the model becomes unable to approximate the behaviour of the system across its operating range. Severe specification errors will cause the parameter estimates to fluctuate as data is obtained non-uniformly from the operating range.

Chapter 7 will now evaluate the performance of the FDD scheme using data from air-conditioning test systems.
Chapter 7

Evaluation of FDD scheme

Introduction

This chapter presents the results of tests carried out to evaluate the FDD scheme with data from cooling coil subsystems that are part of complete air-conditioning installations. The subsystems are thus subject to the disturbances and load changes resulting from the interactions between the different plant items. Two systems are considered: a dynamic simulation of a VAV system; and a large experimental VAV test facility. Both systems use schedules of internal load changes and real weather data. Three faults are introduced separately in the systems (coil fouling, leakage, and sensor off-set), and the ability of the FDD scheme to estimate the degree of these faults is evaluated. In addition, data obtained from the correctly operating system is used to calibrate the model parameters and to assess the ability of the FDD scheme to avoid false alarms.

7.1 Practical issues

This section discusses the practical issues associated with the testing and evaluation of FDD schemes with real systems.

Although data can be obtained relatively easily from a real installation when it is operating correctly, it is more difficult to obtain data from a faulty system to test
7.1. Practical issues

an FDD scheme. There are two approaches that can be adopted:

1. Wait for a real fault to occur and develop.
2. Introduce the faults of interest artificially.

The first approach is not viable as a test regime due to the lack of controllability and the potential time required for the faults of interest to occur being excessive. The second approach is clearly more useful, but in practice the incorporation of any unnecessary faults in an operating system is undesirable and hence difficult to negotiate with the building owner.

To obtain realistic variations in the input signals to the considered subsystem, it is necessary to obtain data from a system that is subject to the disturbances associated with normal operation. Effects such as weather variations and changes in loads caused by fluctuations in internal gains should therefore be considered. These combined effects are only experienced in systems that serve occupied buildings. However, the implementation of certain faults in these systems can disrupt the comfort of the occupants. One possible way to avoid this is to apply simulated disturbances that are representative of internal load variations to the test subsystem when the building is unoccupied. These disturbances can be generated by varying the operating point of other heat exchangers and plant items upstream of the subsystem of interest. Although, the behaviour of the overall system in these circumstances is not completely realistic, that of the test subsystem, if considered in isolation, can be made to be representative of occupancy periods. This approach was adopted to obtain realistic data from a heat exchanger subsystem that was part of a full-size test system. The results from these experiments are presented in Section 7.4.

One of the main reasons why certain faults in air-conditioning systems are considered to be important is that they cause damage to the plant that cannot be easily rectified without incurring large costs. Consequently, it is also a costly exercise to implement faults of these types to test FDD procedures. To avoid incurring these costs it is therefore often necessary to simulate some of the faults. One example where a simulated approach is desirable is in the case of water-side coil fouling. This fault can cause permanent damage to a heat exchanger if it is severe, and treatment of the less severe cases can take long periods. Section 7.4.2 describes
how three faults (leakage, fouling, and sensor off-sets) were simulated in a real test system.

Another problem associated with the testing of FDD procedures with real systems is the difficulty involved in implementing the faults on realistic time scales. The degradation faults that have been described can all take long periods to develop to significant levels in practice. Air-handling unit plant items may typically take several years to degrade to levels that would warrant remedial action. Operational information from the different plant items will only be obtainable during certain seasons during the year, e.g. winter for heating plant, summer for cooling plant. Moreover, the variation in the operating point of each of the subsystems will be on a diurnal basis, within the relevant seasons. For testing FDD schemes, it is therefore often desirable to consider ‘accelerated’ time scales and introduce the artificial faults at a more rapid rate than would occur in practice. In addition, ramp-like changes in fault severity can be difficult to implement artificially and often only step changes are possible.

Due to the problems that have been described above associated with implementing faults in real systems, simulation represents a valuable tool for the evaluation of FDD procedures. Simulations of complete air-conditioning installations can be used to generate realistic data by using real weather data, and faults can be incorporated relatively easily by changing the values of certain simulation model parameters. When the faults in real systems have to be simulated or introduced in an unrealistic manner, the versatility and repeatability that simulation offers makes it potentially more useful for evaluating FDD schemes than real systems.

The testing of FDD methods with real data does, however, allow certain aspects of the FDD methods to be tested more thoroughly than does simulation. In particular, evaluation of the ability of the models to approximate a real system is made possible, as is the evaluation of the robustness of the parameter estimation to the uncertainties associated with real systems. Although the ability of the subsystem models to approximate real systems may be evaluated independently of FDD tests, the effect that modelling errors have on the parameter estimation can only be assessed by using real data. Simulated systems and real systems thus provide complementary features for the testing of FDD schemes. Simulation offers versatility and repeatability, while real systems allow the robustness of the FDD scheme to modelling errors, unmeasured disturbances, etc. to be ascertained. This chapter presents results from tests carried out using both simulated and real
7.2 Model calibration method

The majority of the parameters of the subsystem model used in the FDD scheme that is considered here relate to physical properties of the real system. The values of these parameters can normally be obtained from design or manufacturers' information. However, some of the parameters, such as the heat transfer coefficients of the coil, are empirical in nature and their values can only be estimated reliably for a particular system by using training data. Training data comprise measurements of inputs and outputs to the considered system at different operating points. These data can be used in batch to estimate the unknown parameter values by minimising the differences between the model predictions and the measured outputs for the recorded inputs.

Training data has to be obtained when the system is operating correctly. In practice, the system can only be assumed to be in a fault-free condition when it has just undergone testing, such as following commissioning or extensive maintenance. In both of these circumstances there is likely to be short time available before the system has to be put into operation. The amount of training data is therefore likely to be limited. For the optimisation to be well conditioned, the training data should be obtained the regions of the operating range where changes in the estimated parameters influence the model output. Generally, the amount of training data required to ensure a stable solution is related to the number of parameters included in the optimisation. Hence, it is beneficial to minimise the number of estimated parameters to reduce the reliance on training data.

The batch version of the non-linear estimator, described in Section 5.2.1, could be used to estimate the parameters from training data. This method is only robust, however, when the Hessian matrix is well conditioned. This cannot be guaranteed in practice, due to the possibility of only being able to obtain small amounts of training data. Direct search methods are less sensitive to sparse data and can be applied to non-linear problems having local minima. The Complex direct search method (Box, 1965) has thus been selected since it is able to deal with constraints on the parameters. The algorithm is described in Appendix A.
If the problem is underdetermined, valleys will exist in the criterion function surface and the optimisation may progress along these valleys, causing the parameter estimates to enter unfeasible regions (Rao, 1987). The Complex method allows upper and lower bounds to be applied to each of the parameters preventing this from happening. It may be noted though, that a parameter collapsing on a feasibility bound is a good indication of insufficient data for the number of parameters. This can be addressed by either reducing the number of parameters in the estimation or by obtaining more training data. As with all non-linear optimisation problems, the initial estimate of the parameters is of paramount importance in obtaining a reliable solution. For the heat exchanger models considered here, an initial guess can be made using design information or expert knowledge.

### 7.2.1 Calibration parameters

The parameters from the heat exchanger subsystem models, described in Section 4.2, that are selected for calibration using training data are listed below.

- hysteresis \((v)\);
- curvature of the valve control port \((\beta)\);
- leakage through the valve control port \((l)\);
- heat transfer coefficient on the air-side of the heat exchanger \((\kappa_a)\);
- heat transfer coefficient on the water-side of the heat exchanger \((\kappa_w)\);
- coil wall resistance of the heat exchanger \((R_w)\);

All other parameters of the subsystem models are estimated from design and manufacturers' information. The initial estimates and feasibility regions are determined from typical values expected for the type of coil. The criterion function used in the optimisation is the mean absolute prediction error for the steady-state samples. The mean of the absolute errors is chosen as the criterion function instead of the mean of the squared errors since it gives less emphasis to any outlier points that may exist, which can skew the parameter estimates.
7.3. Evaluation using system simulation

The component-based simulation program HVACSIM+ (Park et al., 1985) was used to simulate a building and plant whose behaviour is representative of an air-conditioned office building\(^1\). The simplified building has a variable-air-volume air-conditioning system with a single air-handling unit and three zones. The air-handling unit consists of: fresh, return, and exhaust air dampers; a pre-heating coil; a cooling coil; and a fan. Each zone has a reheat coil and a variable-air-volume (VAV) box to vary the temperature and flow rate of the air supplied to that zone. The chiller and boiler plant are assumed to operate ideally and the temperatures and flows of the chilled and hot water supplied to the heat exchangers therefore remain constant at their design values.

![Diagram of the simulated system showing the control scheme]

Figure 7.1: The simulated system showing the control scheme

The simulated system is depicted in Figure 7.1, which shows the main plant items and the control strategy. The circles in the figure represent temperature sensors and the squares labeled 'A' represent the actuators. The control scheme, which is typical of those used in commercial office buildings, has two main control-loops:

1. Control of the air temperature supplied by the air-handling unit.

---

\(^1\) The simulation was developed as part of IEA Annex 25 project and was used as a means by which to evaluate and compare different FDD methods.
7.3. Evaluation using system simulation

2. Control of the air temperature in the zones.

The zone models use two capacitances, which represent the thermal storage capabilities of the building materials. One capacitance represents the elements that have a thermal time constants that exceed one hour and the other capacitance represents the air and the light-weight thermal mass. The simulation is similar to that described in (Dexter and Haves, 1994) but the sizing of the equipment is taken from the detailed designs of a recently completed office building in London.

7.3.1 Incorporation of faults

Faults are incorporated into the simulation models by changing the values of certain parameters. Fouling is represented in the heat exchanger model by changing the tube wall material to calcium carbonate (CaCO$_3$). The extent of the fouling is then varied by altering the tube wall thickness. The valve models utilise a leakage parameter, which represents the fractional flow through the valve when the stem position is zero (at constant pressure). This parameter can therefore be varied to model the effects of leakage. For more severe leakages, the 'rangeability' of the valve has to be adjusted to preserve the monotonicity of the inherent characteristic of the valve$^2$. Sensor faults are incorporated by adding an off-set to the sensor values reported by the simulation. The erroneous sensor signals are then used by the control system, which changes the operating points over which the plant is exercised.

7.3.2 Generation of fault data

The degradation faults that have been considered may take very long periods to reach the level of severity that would warrant remedial action (e.g. years). In order to provide the realistic disturbances for the simulation, weather data for a complete year would have to be obtained. The data would have to be sampled at a frequency that is high enough to allow the variations in operating point within the diurnal cycles to be captured. However, consideration of a whole year of weather data at the appropriate sampling frequency would be computationally

$^2$Details of the valve model used in the simulation are provided by Haves (1994).
intensive. Moreover, in real systems, many of the days in a year would not cause the subsystem of interest to be operated in the regions where the effects of the considered faults are exhibited. It was therefore decided to introduce the faults over five days of operation using two days of recorded weather data (a March day and a June day) in the manner shown in Table 7.1.

Table 7.1: Severity of the faults during the three sets of test data

<table>
<thead>
<tr>
<th>DAY</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>WEATHER DATA USED</td>
<td>June</td>
<td>March</td>
<td>June</td>
<td>March</td>
<td>June</td>
</tr>
<tr>
<td>Fouling fault data</td>
<td>0.5mm</td>
<td>1.0mm</td>
<td>1.5mm</td>
<td>2.0mm</td>
<td>2.5mm</td>
</tr>
<tr>
<td>Leakage fault data</td>
<td>1%</td>
<td>2%</td>
<td>3%</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>Sensor fault data</td>
<td>-0.5 K</td>
<td>-1.0 K</td>
<td>-1.5 K</td>
<td>-2.0 K</td>
<td>-2.5 K</td>
</tr>
</tbody>
</table>

The two days of weather data, in combination, cause the cooling coil subsystem to be operated across a large portion of its range, thus making all three faults apparent during the five days. Although the faults are introduced at a more rapid rate than is likely to occur in practice, there is also a greater level of excitation induced over the five-days than would be expected. Hence, the rate of fault development relative to the rate in which the operating range is explored may be considered as being reasonably typical.

### 7.3.3 Excitation conditions

Realistic behaviour is obtained by subjecting the simulated building to disturbances typical of those found in actual buildings. The weather data used for the tests were minute by minute solar radiation and hourly temperature measurements made near London. Two days were used for the tests: a March day that causes the cooling coil to be exercised in its low duty region, and a June day during which the cooling coil is operated throughout most of its range. Both of the days were selected for their rapid fluctuations in solar radiation, which provide disturbances for the control system over a wide range of frequencies. The heat gains in the building envelope induced by the fluctuations in solar radiation are shown in Figure 7.3.

Schedules of the internal gains, with step changes at various times of day, are
7.3. Evaluation using system simulation

Figure 7.2: The internal gains in each of the zones

Figure 7.3: The envelope gains in each of the zones for the two test days

used to represent internal gains due to lights, equipment and people in each of the three zones. The internal gains applied to each of the three simulated zones are presented graphically in Figure 7.2. The activity of the signals associated with the cooling coil subsystem in response to the disturbances is shown in Figures 7.4 and 7.5, for the March day and the June day respectively. The upper graphs in each figure show the supply air temperature and the outside air temperature and the lower graphs show the fractional air flow rate and the mixing box and cooling coil control signals.

7.3.4 Model calibration data

Training data were generated by performing open and closed loop tests on the simulated plant, as shown in Figure 7.6.
Tests of this sort could be carried out as part of the commissioning process in a real building (Haves et al., 1996). Constant loads and inlet conditions were maintained during the tests. In the first phase, all closed loop control was disabled and the control signal to the cooling coil valve actuator was varied in a series of steps at four different air flow rates. In the second phase, local loop control was restored, and the set-point for the supply air temperature varied in a series of steps. During the each of the test phases, the VAV boxes were stepped between fully open and fully closed positions to provide different air-flow rates. It may be noted that the supply temperature exceeds the outside air temperature when the valve is closed due to the heat produced by the supply fan.

7.3.5 Results of model calibration

Table 7.2 shows the parameter values of the FDD model. The parameters that were listed Section 7.2.1 were estimated using the data presented in Figure 7.6. Ordinarily, the other parameters would be estimated from design and manufacturers' information. However, since the test system was simulated, the other parameters
7.3. Evaluation using system simulation

Figure 7.5: Inputs and outputs of the simulated system on the June day were obtained directly from the models used in the simulated system. The result of the optimisation was a mean absolute prediction error (MAE) of 0.34 K. The fit of the model to the data using the parameter values listed in Table 7.2 is shown in Figure 7.7.

The solid line in the upper graph in Figure 7.7 is the measured supply air temperature recorded in the calibration data. The model predictions are shown as crosses at the points when the system was determined to be in steady-state. It can be observed that the model provides a good level of accuracy for most of the training data samples. The accuracy varies across the range of operation due to structural differences between the FDD model and the system model.

The steady-state detector allows the majority of the samples collected during transient periods to be eliminated. However, the detector appears to be too insensitive when large step changes are applied; e.g. at $t \approx 7$ hours. This is likely to be due to the action of the noise filter, which is providing too much smoothing in the initial stages of a large change. These points do not, however, affect the estimates of the parameters, and the model provides a good approximation to the system.
7.3. Evaluation using system simulation

7.3.6 Results of FDD tests

The results of the testing the FDD scheme with the data are presented in this section. During each of the tests, the parameter estimation algorithm was only activated when the following conditions were satisfied:

- steady-state was deemed to exist;
- air-flow existed in the duct (i.e. the fans were on);
- mixing box dampers were set to deliver 100% outside air.

The latter condition is specified so that the measurements from the ambient air temperature sensor can be used as the proxy for the inlet air temperature to the coil. This is preferred over using the mixed air sensor since this is not always fitted in practice. The overall sensitivity of the estimator is determined by $\gamma$ (defined in Section 5.2.5), which, in turn, determines the magnitude of the forgetting factor (note: $0.66 \leq \lambda \leq 1 - \delta$; where $\delta$ is a small number). A value for $\gamma$ was determined...
Table 7.2: The parameters used in the FDD scheme when tested with the simulated system

<table>
<thead>
<tr>
<th>SYMBOL</th>
<th>VALUE (UNITS)</th>
<th>MEANING</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>cooling coil subsystem</td>
</tr>
<tr>
<td>$v$</td>
<td>0.0 (-)</td>
<td>actuator hysteresis</td>
</tr>
<tr>
<td>$\beta$</td>
<td>2.65 (-)</td>
<td>control port curvature</td>
</tr>
<tr>
<td>$l$</td>
<td>0.001 (-)</td>
<td>leakage through control port</td>
</tr>
<tr>
<td>$\dot{m}_w$</td>
<td>6.3 (kg · s$^{-1}$)</td>
<td>chilled water flow rate</td>
</tr>
<tr>
<td>$A$</td>
<td>0.55 (-)</td>
<td>authority (control port)</td>
</tr>
<tr>
<td>$\kappa_a$</td>
<td>8.6247 (kW · K$^{-1}$ · kg$^{-0.8}$ · s$^{0.8}$)</td>
<td>air-side convective coefficient</td>
</tr>
<tr>
<td>$\kappa_w$</td>
<td>16.2461 (kW · K$^{-1}$ · kg$^{-0.8}$ · s$^{0.8}$)</td>
<td>water-side convective coefficient</td>
</tr>
<tr>
<td>$R_w$</td>
<td>0.0828 (kW$^{-1}$ · K)</td>
<td>tube wall resistance</td>
</tr>
<tr>
<td>$b$</td>
<td>0.0 (K)</td>
<td>sensor offset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>supply fan</td>
</tr>
<tr>
<td>$\Delta T$</td>
<td>1.5 (K)</td>
<td>fan temperature rise</td>
</tr>
<tr>
<td></td>
<td></td>
<td>steady-state detector</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.1 (K)</td>
<td>threshold</td>
</tr>
<tr>
<td>$\tau$</td>
<td>500.0 (s)</td>
<td>overall time constant</td>
</tr>
</tbody>
</table>

empirically by progressively increasing it from a small value up to the point where the parameters began to respond when the estimator was tested with the training data. Using this approach, $\gamma$ was determined to be 5 for the training data obtained from the simulated system.

**Fault-free**

The estimator was tested firstly using data generated from the fault-free simulation using the March and June days (Figures 7.4 and 7.5 respectively). The estimator did not adjust the parameters throughout each of the data sets, and the confidence intervals were narrow, indicating a high level of confidence in the values. The estimator was therefore unaffected by variations in the operating point of the system and the modelling errors that were apparent in the training data (Figure 7.7).
7.3. Evaluation using system simulation

Figure 7.7: The fit of the model to the data from the simulated system

Fouling fault

Figure 7.8 shows the parameter estimates obtained using the fouling fault data. The parameter estimates are shown as dash-dot lines and the estimates are surrounded by 99% confidence intervals, which are shown as solid lines. The parameters that correspond to the correctly operating system are shown as dotted horizontal lines. When a correct operation parameter is not contained within the confidence interval this implies that there is a 99% confidence that the parameter of the real system has deviated from its correct value; i.e. a fault has occurred. Sometimes, the confidence interval becomes very wide due to a lack of information or large prediction errors and the interval is outside the scale of the graph at certain times. It may be noted that the parameter estimates are shown in continuous time despite the estimator only being activated when the conditions listed in Section 7.3.6 are satisfied. A graph showing whether these conditions are satisfied is therefore included at the bottom of the figure; a one indicates that the estimator was active, while a zero indicates that it was inactive.

Table 7.3 shows the numerical values for the parameters at the end of each day.
with the actual parameter values of the system shown in brackets. The symbol \( t \) ('leakage parameter') is the fractional flow leakage through the closed valve, and \( b \) ('off-set parameter') is the error in the temperature sensor measuring the air leaving the heat exchanger. The two heat transfer parameters that relate to the convective heat transfer of the coil \((\kappa_a, \kappa_w)\) are empirical in nature and these parameters combine with the conduction parameter \((R_w)\) to determine the overall heat transfer behaviour of the coil model. Hence, inaccuracies in the way the convective heat transfer is modelled will be compensated for by variations in the conduction parameter. The \( R_w \) parameter may therefore not retain its physical meaning. Moreover, the area of the coil has to be known to calculate a useful measure of the thermal conductivity of the coil. The estimate of the \( R_w \) parameter has therefore been used to calculate the overall conductance of the coil \( U_A \) at the maximum flow rates, using Equation 4.23.

It can be observed from Figure 7.8 that the parameter estimates oscillate in response to the step change that is introduced at the beginning of each day. The parameters do, however, converge on appropriate values toward the end of each day. The initial variations are accompanied by a widening of the confidence interval. After each step change, the estimates become increasingly stable, and the confidence interval narrows to reflect the increase in the amount of information available at the new parameter values. The fouling parameter increases in a ramp-like fashion to reflect the build-up of fouling in the simulated system and the other parameters can be seen to deviate marginally from their correct values. The fluctuations in the estimated amount of leakage are negligible, but the fluctuations in the sensor off-set parameter are more significant. However, the 99% confidence intervals associated with this parameter is wide enough during the fluctuations so that the zero off-set (corresponding to 'correct operation') does not fall outside the interval.

The 'correct operation' fouling parameter falls outside the confidence interval, toward the end of days 1-4. This interval would thus constitute a reliable fault detection threshold in this instance. The trend of the estimates also indicates that the better convergence could be obtained if there was a longer period between steps. It can be observed from Table 7.3 that the \( U_A \) calculated from the initial estimates of the heat transfer parameters in the FDD model (estimated from the training data) is lower than the \( U_A \) of the coil in the simulated system. Structural differences between the FDD model and the simulation model may account for the erroneous initial estimate of the \( U_A \). The uncertainty associated with the
initial estimate cannot be assessed, however, since confidence intervals were not calculated as part of the training process. In spite of the initial error, the \( UA \) that is calculated using the fouling parameter \((R_w)\) estimated at the end of each day is close to the true \( UA \). The estimator thus provides a reliable estimate of the \( UA \) during the estimation process.

The momentary impulse-like increase in the confidence interval at approximately one hour in Figure 7.8 is caused by the effective number of data points, \( n \) (in Equation 5.81) being very close to the number of parameters. When this happens, the \( t \) statistic tends toward infinity. The interval can be observed to narrow quickly, when additional samples are obtained that meet the conditions for the estimator to be activated.
Table 7.3: Parameter estimates at the end of each day: fouling fault data

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>( U_A (\text{kW} \cdot \text{K}^{-1}) )</th>
<th>( l_c (-) )</th>
<th>( b (\text{K}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>8.26 (9.66)</td>
<td>0.001 (0.001)</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>Day 1</td>
<td>6.99 (7.59)</td>
<td>0.000 (0.001)</td>
<td>-0.002 (0.0)</td>
</tr>
<tr>
<td>Day 2</td>
<td>5.68 (6.08)</td>
<td>0.016 (0.001)</td>
<td>-0.468 (0.0)</td>
</tr>
<tr>
<td>Day 3</td>
<td>5.10 (4.90)</td>
<td>0.011 (0.001)</td>
<td>-0.614 (0.0)</td>
</tr>
<tr>
<td>Day 4</td>
<td>4.07 (3.95)</td>
<td>0.014 (0.001)</td>
<td>-0.481 (0.0)</td>
</tr>
<tr>
<td>Day 5</td>
<td>3.02 (3.17)</td>
<td>0.009 (0.001)</td>
<td>-0.168 (0.0)</td>
</tr>
</tbody>
</table>

Leakage fault

Figure 7.9 and Table 7.4 show the results obtained using the data obtained from the system when the leakage fault was incorporated. The estimates of the leakage parameter at the end of each day closely mirror the actual increments in leakage implemented in the simulated system. The estimate of the fouling parameter fluctuates during the data, with the most noticeable deviation from the correct operation value occurring at the end of day 2. For the particular combination of heat transfer parameters listed in Table 7.2, the \( U_A \) is quite sensitive to small changes in the fouling parameter and relatively large variations in the \( U_A \) are thus evident in Table 7.4.

There is a more significant variation in the sensor off-set parameter, which follows a general upward trend indicating a positive off-set. A positive sensor off-set has an opposite effect to leakage (i.e. residuals resulting from the faults are of opposite sign). The identification of a positive off-set in the model therefore has a compensatory effect to the identified leakage. This is due to the structural differences between the way leakage is modelled in the FDD model and the simulation model. The leakage in the simulation model is more localised in the close-off region of the valve, due to the use of a piecewise model (Haves, 1994). In contrast, the effect of leakage in the FDD model is apparent over a greater region of valve operation. The leakage that is estimated as being appropriate for the closed (or very nearly closed) valve position thus causes too much of an effect as the valve in the simulated system is opened. The sensor off-set parameter is therefore being varied to compensate.
7.3. Evaluation using system simulation

The 'correct operation' leakage parameter falls outside the confidence interval, toward the end days 2-5. The interval is wide enough for the majority of the test to contain the fluctuations in the estimated off-set and fouling parameters. However, marginal transgressions of the interval occur at the end of day 2 and 3 for the fouling parameter, and at the end of day 3 for the off-set parameter. These transgressions are nevertheless small in comparison to those experienced for the leakage parameter. Moreover, a decrease in the degree of fouling is implied at the end of day 3, which would not constitute a fault. A more significant transgression of the confidence interval occurs at the end of day 5 for the sensor off-set parameter, but the leakage fault would have already been diagnosed by this point.
Table 7.4: Parameter estimates at the end of each day: leakage fault data

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>UA (kW·K⁻¹)</th>
<th>l_c (-)</th>
<th>b (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>8.26 (9.66)</td>
<td>0.001 (0.001)</td>
<td>0.000 (0.0)</td>
</tr>
<tr>
<td>Day 1</td>
<td>8.33 (9.66)</td>
<td>0.001 (0.011)</td>
<td>0.000 (0.0)</td>
</tr>
<tr>
<td>Day 2</td>
<td>5.88 (9.66)</td>
<td>0.016 (0.021)</td>
<td>-0.012 (0.0)</td>
</tr>
<tr>
<td>Day 3</td>
<td>11.76 (9.66)</td>
<td>0.034 (0.031)</td>
<td>-1.586 (0.0)</td>
</tr>
<tr>
<td>Day 4</td>
<td>9.09 (9.66)</td>
<td>0.044 (0.041)</td>
<td>-1.103 (0.0)</td>
</tr>
<tr>
<td>Day 5</td>
<td>13.33 (9.66)</td>
<td>0.053 (0.051)</td>
<td>-2.201 (0.0)</td>
</tr>
</tbody>
</table>

Sensor fault

Figure 7.10 and Table 7.5 show the results obtained using the data with the negative sensor off-set fault. The off-set parameter can be seen to be the most sensitive to the test data. The estimated degree of off-set increases during each day, but the estimated value is lower than the true value at the end of each day. It can be observed from Figure 7.10 that this fault induces variation in both the other parameters, although the estimates tend toward their nominal values by the end of each day. The leakage parameter varies most on days 2 and 4, which are the March days when the coil is operated in its low duty region. Conversely, the fouling parameter varies most on days 1, 3, and 5, which are the June days when the coil is operated in its high duty region. The increase in the estimated leakage that occurs during the test may be responsible for the underestimation of the off-set, since leakage and negative sensor offset produce residuals of the same sign. The variations in the fouling parameter are compensatory to the negative sensor off-set and when these occur the degree of off-set is overestimated (e.g. at 22 hours).

7.3.7 Conclusions

The tests using the data from the simulated system have shown that the estimator is capable of tracking the development of three degradation faults. The confidence intervals, derived in Section 5.2.8, have been shown to have the potential to act as thresholds for fault detection. A fault would be deemed to exist when any of the
nominal ('correct operation') parameters fall outside the confidence intervals that surround the estimated fault parameters. The fluctuations in the parameters that were observed during the tests are caused by phenomena such as modelling errors and lack of variability in the input signals, which were investigated in Chapter 6. The fluctuations in the parameters are an unavoidable consequence of the overall uncertainty associated with the parameter estimation. Although the confidence intervals contain the fluctuations that were apparent in the tests, large fluctuations are undesirable and some form of ‘relaxation’ may be necessary to limit the magnitude of the parameter changes in practice.

Figure 7.10: Parameter estimates for the sensor fault data

Another important characteristic of the estimation process, which was revealed during the tests, was the sensitivity to structural errors in the model and how these
7.3. Evaluation using system simulation

Table 7.5: Parameter estimates at the end of each day: sensor fault data

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>( UA \ (\text{kJ} \cdot \text{K}^{-1}) )</th>
<th>( l_c \ (-) )</th>
<th>( b \ (\text{K}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>8.26 (9.66)</td>
<td>0.001 (0.001)</td>
<td>0.000 (0.0)</td>
</tr>
<tr>
<td>Day 1</td>
<td>9.35 (9.66)</td>
<td>0.001 (0.001)</td>
<td>-0.002 (-0.5)</td>
</tr>
<tr>
<td>Day 2</td>
<td>8.77 (9.66)</td>
<td>0.021 (0.001)</td>
<td>-0.147 (-1.0)</td>
</tr>
<tr>
<td>Day 3</td>
<td>10.20 (9.66)</td>
<td>0.012 (0.001)</td>
<td>-0.611 (-1.5)</td>
</tr>
<tr>
<td>Day 4</td>
<td>7.87 (9.66)</td>
<td>0.016 (0.001)</td>
<td>-1.429 (-2.0)</td>
</tr>
<tr>
<td>Day 5</td>
<td>10.0 (9.66)</td>
<td>0.009 (0.001)</td>
<td>-1.696 (-2.5)</td>
</tr>
</tbody>
</table>

affect the parameter estimates. This was particularly noticeable when testing the estimator with the data from the system with the leakage fault. The differences in the way that leakage was modelled between the simulation model and the FDD model appeared to cause compensatory variations in the off-set parameter. However, it would be expected that the extent of the compensatory variation could be reduced significantly by obtaining more data from the regions where the coil was operated in its high duty region. This thus re-affirms the need, identified in Chapter 6, for the ‘window’ of data samples implicitly used in the estimation to contain (well distributed) data across the operating range.

The results from the tests show that the FDD scheme is capable of tracking the development of the degradation faults within the accelerated time scale that was considered. The sensitivity of the estimator was determined by \( \gamma \), which was set to five for the tests. Fortescue et al. (1981) interpret \( \gamma \) in the following way:

\[
\gamma = \tau_0 \sigma^2,
\]

(7.1)

where \( \tau_0 \) is an asymptotic time constant, which determines the size of the ‘memory vector’, discussed in Section 5.2.4; \( \sigma^2 \) is the variance of the model prediction errors when tested with the data. For the simulation, the variance of the prediction errors determined from the training data was 0.32 K\(^2\), hence \( \tau_0 = 15.63 \). The effective time constant of the memory vector is thus only approximately 16 samples. This is a small value relative to the rate at which faults would be expected to develop in practice. Because of this high level of sensitivity, the parameter estimates were quite sensitive to modelling errors and other anomalies in the data. In practice, the sensitivity could be significantly reduced, and this would allow the samples from a greater portion of the operating range to be used in the estimation. This
would then serve to dampen the parameter fluctuations and provide more stable estimates.

7.4 Evaluation using a real system

During the IEA Annex 25 project, data were collected from the Air-Conditioning Evaluation (ACE) Facility at the UK Building Research Establishment. These data were collected by BRE personnel for the original purpose of evaluating two FDD schemes developed by the UK participants\(^3\) in the Annex 25 project. The scheme developed at Loughborough University is described by Salsbury \textit{et al.} (1996), and the scheme developed at Oxford University is described by Dexter and Benouarets (1996).

The BRE test system was operated for periods of about 60 hours with different faults introduced in the cooling coil subsystem, independently at the beginning of each test period. The relevant inputs and outputs to the subsystem were sampled at one minute intervals and the data were provided in batch form for the evaluation of the FDD schemes. The faults remained unidentified until the FDD schemes had been tested with the data. The two schemes were able to detect most of the faults, however, diagnosis proved more difficult and both schemes failed to diagnose any of the faults unambiguously. The data collected from the BRE system has also been used to evaluate the FDD scheme developed in this thesis, and the experimental procedure and results of these tests are presented in the sections that follow.

7.4.1 Description of the ACE Facility

The ACE facility is a full size installation at the Building Research Establishment. The system is capable of providing a maximum air flow of 3.4 m\(^3\)·s\(^{-1}\) at a static pressure of 1000 Pa, with a 65 kW cooling load and a 25 kW heating load. The design and level of instrumentation are typical of that found in practice in the United Kingdom. A diagram of the installation is presented in Figure 7.11, which shows the relevant subsystems and the control strategy.

\(^3\)Loughborough University and Oxford University.
The inlet air stream, which can comprise a mixture of fresh air and air drawn from the hall where the system is housed, is mixed with return air according to the position of the recirculation damper. The inlet air is blown by the supply fan onto the cooling coil and the heating coil before passing through nine VAV boxes. Most of the air exhausted from the VAV boxes passes straight back into the return air stream. However, some of the supply air is discharged into the hall: the hall air is then extracted into the extract air stream in order to introduce an additional thermal capacitance between the supply and the return flows. The return fan maintains air flow in the return duct and variable proportions of the return air can be recirculated or expelled through the exhaust duct, depending on the position of the mixing box dampers.

The purpose of the controller C1 is to maintain the temperature of the air discharged from the air-handling unit at a fixed set-point by operating the cooling coil, mixing box dampers, and heating coil in sequence, according to demand. Controller C2 varies the speed of the supply fan in order to maintain a constant static pressure in the supply duct. The return fan speed is varied according to the supply fan speed so that a small, relatively constant, positive pressure is maintained in the zone. The system was able to supply air to a real zone but this facility was not used as the conditions in the zone could not be varied easily. A simulated zone was therefore used instead based on a simple building model (Crabb et al., 1987).
Internal gains were then instigated using predetermined schedules. The controller C4 operates a heater battery sited in the return duct so that the return temperature is representative of the simulated zone temperature. Controller C3 adjusts the position of the VAV box damper serving the simulated zone, in response to the simulated zone temperature. The other eight VAV boxes are also controlled by C3 in order to produce the effect of there being nine similar zones.

7.4.2 Incorporation of faults

To prevent permanent damage to the plant, faults were simulated using the techniques outlined below:

**Fouling:** Fouling was simulated by increasing the hydraulic resistance in the primary circuit by incorporating a two-port balancing valve in the return leg of the heat exchanger (Figure 7.12). The effect of this is to reduce the flow delivered by the pump to the heat exchanger circuit. The stem position of the two-port valve was adjusted to reduce the temperature difference across the coil by 2 K when at maximum air and water flow rates.

![Figure 7.12: Incorporation of the two-port balancing valve](image)

**Leakage:** This was implemented by preventing the valve control signal from falling below a minimum value. The minimum value was selected so that the
7.4. Evaluation using a real system

Temperature of the air leaving the coil was altered by 2 K when the control signal to the valve was zero.

**Sensor off-set:** This fault was simulated by adding a 2 K off-set to the sensor measuring the temperature of the air leaving the air-handling unit. The erroneous sensor reading was used in the control system as part of the feedback control strategy. The fault thus affected the operating point of the controlled plant.

Limited time was available to use the test facility. It was therefore decided that introducing the faults as step changes was the best way to enable a number of different faults to be considered in the time available. Table 7.6 lists the faults that were implemented.

<table>
<thead>
<tr>
<th>FAULT IMPLEMENTED</th>
<th>DEGREE OF FAULT</th>
</tr>
</thead>
<tbody>
<tr>
<td>simulated fouling</td>
<td>2 K at maximum air flow and full cooling</td>
</tr>
<tr>
<td>simulated leakage</td>
<td>2 K at maximum air flow</td>
</tr>
<tr>
<td>sensor error</td>
<td>2 K positive offset</td>
</tr>
<tr>
<td>fault free</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 7.6: Test data from the BRE system

7.4.3 Excitation conditions

The two main sources of excitation for the cooling coil subsystem were the simulated zone and the ambient air. Weather data was supplied by the BRE, which contained one minute samples from near London on a cloudy September day and a cloudless June day. This weather data together with scheduled internal gains similar to those used with the simulated system were used as inputs to the simulated zone, which generated the internal dry-bulb temperature of the zone as its output. This temperature was then used as the controlled variable in the VAV box control loop. Variation in the position of the VAV box dampers causes the pressure drop in the supply air circuit to vary. The purpose of the fan control loop is to maintain a constant supply air static pressure, by varying the rotational speed of the supply fan. The set-point for the zone was 21°C, and the set-point of the air discharged from the air-handling unit was 16°C.
The mixing dampers normally provide full outside air when the system is in cooling mode. However, a dry-bulb economiser is installed that reverses the direction of the dampers when the return air temperature falls below that of the ambient. The return air is heated so that its temperature is equal to that of the simulated zone, but this does not affect the inlet temperature of the air to the cooling coil for the majority of time. The simulated zone therefore causes the variations in the flow rate of the supply air, while the ambient air is responsible for variations in temperature and humidity of the inlet air to the coil. The set-point for the temperature leaving the coil is constant and the controller varies the control signal to maintain this set-point when subjected to variations in temperature, humidity and flow rate of the inlet air. The plant is operated through both the day and night time to obtain data for warm and cool conditions.

### 7.4.4 Model calibration data

The system was operated in its correctly operating condition for two and a half days to test the operation of system with the artificially generated loads. The data obtained from this test period were used to calibrate the models, and these data are shown in Figure 7.13. The top graph shows the temperatures associated with the air-handling unit. The middle graph shows the control signals to the control valve and the mixing box. The bottom graph shows the humidities and fractional air-flow rate. The first data were collected at a start time of mid-day, which can be seen to correspond approximately to the peak in the ambient air temperature cycle. The return air temperature and the air-flow rate vary in response to the schedule of gains used with the zone model. The control signal to the valve responds to variations in supply air-flow rate and ambient air temperature.

### 7.4.5 Problems associated with the BRE system

A number of the sensors used with the BRE system suffered from problems. These are outlined below:

---

4The fractional air-flow rate is calculated by assuming 2.5kg·s⁻¹ to be the maximum value.
7.4. Evaluation using a real system

Figure 7.13: Training/validation data from BRE ACE facility

**Velocity sensor:** The sensor reading, although correlated with the true velocity, was not directly indicative of the true value. The maximum air-flow deliverable by the fan was supposed to be $3.4 \, \text{m}^3 \cdot \text{s}^{-1}$, but the maximum in the data was only $2.3 \, \text{m}^3 \cdot \text{s}^{-1}$. It was discovered after the experiments was that the velocity readings may have been saturating, and this may have been responsible for the low flows.

**Chilled water temperature sensor:** The chilled water temperature was measured at a point distant from the heat exchanger. It was suspected that the chilled water was subject to heat gain between the coil and the point at which the temperature measurement was made.

**Ambient air temperature sensor:** The ambient air sensor was sited away from the fresh air entry point to the duct system. This sensor may therefore not always have been representative of the true air temperature entering the duct.

**Humidity sensors:** There is evidence of erroneous humidity sensor readings in Figure 7.13. When the valve is closed, at $t \approx 58$ hours, the dry-bulb tem-
7.4. Evaluation using a real system

Temperatures of the supply and outside air are 16 °C and 15 °C respectively. The relative humidities at this time are 35% for the supply and 85% for the outside air. Based on these figures, the change in enthalpy across the cooling coil is approximately 13 kJ·kg⁻¹. According to the sensor measurements there is therefore ~ 16 kW of cooling duty when the valve is closed.

Analytical models rely on using the measured variables according to their physical meaning. It is therefore particularly important for sensor readings to be representative of the physical properties they are supposed to measure for these types of models. There are two ways that sensor readings may be affected:

1. The readings represent some unknown transformation of the true property. i.e. \( y = f(x) \). (where \( y \) is the measurement and \( x \) is the true property).

2. The readings are corrupted by some other unmeasured and unaccounted for effect, i.e. \( y = x + v \). (where \( v \) is the unmeasured effect).

In the first example listed above, a method based on physical equations will be at a disadvantage compared to a method that uses more general 'black box' models. The parameters that are estimated during the training process may be able to compensate for a transformed property measurement, but the parameters will lose their physical meaning. If the sensor transformation is not able to be sufficiently approximated by the model by adjusting the parameters, the model will have an inherent specification error. This will cause the accuracy of the model to vary over the range of system operation leading to fluctuations in the fault parameters that are estimated on-line.

In the second example numbered above, the component \( v \) represents an unmeasured disturbance. In an ideal situation, \( v \) has a zero mean over a small period, as in the case of measurement noise. Any effect that causes \( v \) to have a non-zero mean, over a period that exceeds that over which averaging (via forgetting) is carried out, will lead to variations in the estimated fault parameters. Any method that exploits the redundancy in the information between the sensors will be affected by this problem, and variations in \( v \) will, in most cases, be indistinguishable from a fault. For a parameter estimation approach, the result will be that the parameters will be varied in a continual attempt to obtain the best fit of the model to the observed data. The problems with the sensors in the BRE system are therefore likely to make the FDD scheme susceptible to false alarms.
7.4. Evaluation using a real system

7.4.6 Results of model calibration

Table 7.7 shows the parameter values estimated for the BRE ACE facility. The upper graph in Figure 7.14 shows the model predictions of the air temperature leaving the coil indicated by crosses at the times when the system was deemed to be in steady-state. The measurements of the same air temperature are shown as a continuous solid line. The lower graph in Figure 7.14 shows the apparatus dew point, calculated using Equation 4.36, and the dew point of the inlet air. The coil is deemed to be wet (or partially wet) when the dew point of the inlet air is below the apparatus dew point.

Table 7.7: The parameters used in the FDD scheme when tested with the BRE ACE facility

<table>
<thead>
<tr>
<th>SYMBOL</th>
<th>VALUE (UNITS)</th>
<th>MEANING</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
<td>0.2982 (-)</td>
<td>actuator hysteresis</td>
</tr>
<tr>
<td>β</td>
<td>2.0228 (-)</td>
<td>control port curvature</td>
</tr>
<tr>
<td>l</td>
<td>0.0203 (-)</td>
<td>leakage through control port</td>
</tr>
<tr>
<td>( \dot{m}_w )</td>
<td>2.64 (kg \cdot s(^{-1}))</td>
<td>chilled water flow rate</td>
</tr>
<tr>
<td>A</td>
<td>0.55 (-)</td>
<td>authority (control port)</td>
</tr>
<tr>
<td>( \kappa_a )</td>
<td>1.5334 (kW \cdot K(^{-1}) \cdot kg(^{-0.8}) \cdot s(^{0.8}))</td>
<td>air-side convective coefficient</td>
</tr>
<tr>
<td>( \kappa_w )</td>
<td>6.4036 (kW \cdot K(^{-1}) \cdot kg(^{-0.8}) \cdot s(^{0.8}))</td>
<td>water-side convective coefficient</td>
</tr>
<tr>
<td>( R_w )</td>
<td>0.0861 (kW(^{-1}) \cdot K)</td>
<td>tube wall resistance</td>
</tr>
<tr>
<td>b</td>
<td>0.0 (K)</td>
<td>sensor offset</td>
</tr>
<tr>
<td>( \Delta T )</td>
<td>1.0 (K)</td>
<td>fan temperature rise</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>0.1 (K)</td>
<td>threshold</td>
</tr>
<tr>
<td>( \tau )</td>
<td>500.0 (s)</td>
<td>overall time constant</td>
</tr>
</tbody>
</table>

The parameter calibration process was unable to reduce the prediction errors to the level that was achieved for the simulated system; the mean of the absolute prediction errors being 0.64 K. The inferior fit of the model to the data was most likely due to the problems associated with the sensors, which were discussed in

\(^5\)Fictitious effective surface temperature of the coil.
Section 7.4.5. Also, in contrast to the simulation, the wet coil model is used for most of the time. The wet coil model has the potential for greater inaccuracy than the dry coil model due to the readings from the humidity sensors having to be used to calculate the temperature of the air leaving the coil.

Figure 7.14: The fit of the model to the data from the BRE ACE facility

The parameter calibration process involves minimising the prediction errors for the training data. If unmodelled phenomena exist in the data, such as the faulty sensor readings discussed in Section 7.4.5, the parameters that are estimated during the model calibration will become erroneous in order to compensate. Unusual parameter estimates may therefore be used as an indication of initial faults in
7.4. Evaluation using a real system

the system. Evidence of compensation for faulty sensor readings is apparent in the estimated parameter values shown in Table 7.7. The convective heat transfer coefficient for the water-side of the coil is unusually low, and a 2% leakage was identified through the control port of the valve. An additional effect that is present in the data is hysteresis, which is estimated to be 30% of the range.

It can be observed from Figure 7.14 that the times when the model predictions deviate most markedly from the measured values correspond approximately with the times when the dew point of the inlet air and the apparatus dew point cross. Overall, the model appears to be least accurate when the equations for the dry are used (30-40 hours; 50-60 hours). The humidity sensors are known to be unreliable and errors in the sensor reading the humidity of the inlet air to the coil may be causing the condition of the coil (wet or dry) to be wrongly predicted. Hence, the dry coil may have been used when the real coil was still wet or partially wet.

In spite of the problems with the data, the mean absolute error of 0.64 K is below the magnitude of the residuals caused by the implemented faults (2 K). The FDD scheme is therefore expected to show some evidence for the faults. However, it can be observed from Figure 7.14 that there are regions of the operating range where the prediction errors exceed 2 K; the scheme is therefore susceptible to false alarms.

7.4.7 Results of FDD tests

As with the tests carried out using the simulated system, the parameter update algorithm was only activated when the conditions listed in Section 7.3.6 were satisfied. The minimum value of \( \gamma \) that caused the estimated parameters to respond to the training data was 1200; \( \gamma \) was thus set to this value for the tests. The effect of setting \( \gamma \) to such a high value is that the forgetting factor is less sensitive to the size of the prediction error. The forgetting factor is therefore kept high except if there are very large errors. When the forgetting factor is high, information from previous samples has a greater weight in the estimation process. The parameters are then effectively estimated for data accumulated over a longer time. The estimates will therefore be less sensitive to the local structural inaccuracies in the model, providing the system is reasonably active within its operating range.
7.4. Evaluation using a real system

Fault-free

The estimator was tested first with data gathered from the correctly operating system, which allows the false alarm rate to be assessed. The data are presented in Figure 7.15, which shows the relevant inputs and outputs to the subsystem. It can be observed from the figure that the mixing box is not set to give full outside air during certain periods in the data. The parameters are not adjusted unless the mixing box is at one of its extreme positions due to the lack of reliability associated with estimating the mixed air temperature.

![Temperatures](image)

![Control Signals](image)

![Fractional air flow + relative humidities](image)

Figure 7.15: Fault-free test data

Figure 7.16 shows the estimated parameters for the fault-free data. Before $t \approx 48$ hours, the parameter estimates do not vary significantly and the confidence intervals contain the 'correct operation' parameters for the majority of the time. However, there is a significant change in the fouling parameter at $t \approx 48$ hours. A change of this sort indicates that the model has predicted more cooling than is apparent from the data. The change in the fouling parameter appears to coincide with the valve control signal changing direction. Inconsistencies between the way hysteresis was modelled in the FDD model and way it was manifested in the real system is one possible reason for this parameter change. In spite of this effect
the confidence interval associated with the fouling parameter is wide enough to contain the 'correct operation' value.

Figure 7.16: Parameter estimates for the fault-free data

Fouling fault

The data gathered from the system when the fouling fault was incorporated is presented in Figure 7.17. It may be observed from the graph that the control signal to the valve never exceeds 50% of its maximum range. Since fouling is most apparent at high duties the effects of this fault will be small for these data.

In spite of the coil being exercised in its low duty region, the estimator does manage to produce estimates that indicate the type of fault. The estimated magnitude of
7.4. Evaluation using a real system

the fouling parameter increases at approximately 20 hours, which corresponds to the time when the control signal is highest. For most of the time between 20 hours and 50 hours, the correct operation fouling parameter is outside the confidence interval; constituting strong evidence for a change in this parameter. A small positive sensor off-set is also estimated during the data, due to it being difficult to distinguish between the symptoms of fouling and positive sensor off-set.

There is a significant increase in the estimated leakage parameter at the start of the data. This may be due to the initial estimate of the valve stem position, which is required for the hysteresis model, being inaccurate. The initial estimate of valve stem position determines the amount of slack to be taken up, and at least one reversal is required to eliminate any initial errors. If data from before the first reversal is used to estimate the parameters, the initial error may take several steady-state samples before any changes that are induced in the parameters are eliminated. This effect can be observed in the test, which shows that the estimate of the leakage parameter does return to its correct value when steady-state data is obtained after the first reversal.
7.4. Evaluation using a real system

Figure 7.18: Parameter estimates for the fouling fault data

Leakage fault

The data gathered from the system with the leakage fault incorporated is presented in Figure 7.19. It can be observed that the mixing box was being operated during certain periods in the data. Leakage is most apparent when the control valve is supposed to be closed. However, the control strategy that was implemented in the BRE system caused the mixing box dampers to be modulated when the control signal to the valve was driven to zero. Since the estimator is de-activated when the control signal to the mixing box dampers is at any mid value, there is little potential to obtain very much information at the operating point where leakage is most apparent. This problem could be overcome by incorporating a dead-band
7.4. Evaluation using a real system

between the sequenced operation of the mixing box and the heat exchangers. This approach is sometimes adopted in order to reduce the control activity and hence the wear on the controlled elements. Alternatively, sensors suitable for poorly mixed air-streams, such as averaging sensors, could be installed to measure the mixed air temperature more reliably.

Figure 7.19: Leakage fault test data

Figure 7.20 shows that the estimates of the parameters for the leakage data. Leakage is identified initially (≈ 1 hour), but this may be due to the initial estimate of the slack in the hysteresis model being inaccurate. A significant leakage (8%) is estimated at approximately 25 hours, which causes the ‘correct operation’ leakage to be outside the confidence interval. The estimated leakage fluctuates when t ≈ 50 hours, and a compensatory effect is provided via the estimation of a positive sensor off-set. This compensatory effect may have been due to the leakage in the model not being sufficiently localised at the closed valve position, as was experienced for the tests with the simulated system. Toward the end of the data, the estimate of the leakage is reduced. This is most likely to be caused by modelling errors or inconsistencies in the data caused by sensor inaccuracies.
7.4. Evaluation using a real system

Figure 7.20: Parameter estimates for the leakage fault data

Sensor fault

The data incorporating the sensor fault are presented in Figure 7.21.

Figure 7.22 shows the parameter estimates for the sensor fault data. It can be observed that all the parameter estimates vary during the test. The sensor off-set parameter varies the most, indicating a positive off-set in the initial stages of the test, with narrow confidence intervals. The fouling parameter is also adjusted in the early stages of the test with the 'correct operation' value falling outside the confidence intervals during the first ten hours. Fouling and positive sensor off-set faults have similar symptoms at high and mid duty regions of the operating range, and this is why an increase in the fouling parameter is estimated. The
7.4. Evaluation using a real system

The accuracy of the parameter estimates deteriorates after approximately 42 hours, with a negative sensor offset being estimated. It may be observed from Figure 7.21 that at this time the chilled water temperature begins to rise, which was caused by a chiller fault. The control signal to the valve rises in response to the change in chilled water temperature, and saturates at its maximum value. The estimate of a negative sensor offset indicates that the model was not predicting enough cooling based on the sensor readings provided. It may be recalled from Figure 7.13 that the model was initially calibrated using data where the control signal never exceeded 70%. When the control signal saturates at its maximum value the FDD model is therefore being tested outside the region of the original training data. Inaccuracies in the model outside the region of the training data may therefore be one possible cause of the apparent deterioration in the accuracy of the parameter estimates. Although the chilled water temperature returns to its

Figure 7.21: Sensor fault test data

fouling parameter does, however, return to its nominal value when the system is exercised across a greater part of its operating range (≈ 12 hours). The estimate of the sensor offset at 15 hours is close to the true offset and the confidence intervals are narrow enough at this point for the estimate to be attributed credence.
nominal value, the parameter estimates do not recover sufficiently in the remaining time to re-diagnose the positive offset.

7.4.8 Conclusions

The FDD scheme has been tested using data from a full size air-conditioning system. Although the tests were relatively unrealistic due to the faults being introduced as step changes, it has been demonstrated that the FDD scheme is capable of detecting the three degradation faults that have been considered. In each of the tests, the parameter corresponding to the fault type changed most
significantly, thus demonstrating the ability of the FDD scheme to distinguish between the different faults. The estimation of the degree of fault was less accurate and the parameters tended to fluctuate during the tests. This fluctuation was due, in part, due to unreliable sensors.

In the review of the literature that was presented in Chapter 2, it was described how there is a trade-off between the false alarm rate and the sensitivity to faults. This has been demonstrated in the tests using the data from the BRE system. The sensitivity was made such that changes in the parameter estimates were (only just) avoided for the training data. Hence any minor deviation from the relationship between the measured inputs and outputs that was implicit in the training data would thus be expected to induce a change in the parameters. The estimator was therefore made to be as sensitive as possible for the data that was available. The consequence of this was that the scheme was over-sensitive to the test data, which contained samples from new operating points.

It was necessary to make the scheme sensitive due to the time-scale for each test being short. In practice, the sensitivity could be reduced quite significantly. This would then dampen the fluctuations in the parameter estimates and ultimately lead to a more robust estimate of the parameter values.
Chapter 8

Conclusions and further work

8.1 Conclusions

This thesis has described the development of an FDD scheme based on analytical models and recursive parameter estimation. The scheme has been applied to heat exchanger subsystems of the type used in air-conditioning systems, and the ability of the scheme to estimate the degree of three degradation faults (coil fouling, sensor drift, valve leakage) has been investigated.

Models of the constituent components of finned-tube, water to air heat exchanger subsystems were described in Chapter 4. Most of the parameters in the models relate to physical properties of the real system and these can thus be estimated initially from design or manufacturers' information. A small number of the parameters are empirical, and these are estimated from training data. The combination of training data, and design and manufacturers' information allows the models to be calibrated to represent the 'correctly operating' system.

The models used in the FDD scheme are extended to enable the behaviours resulting from the development of the considered faults to be modelled. Certain model parameters relate directly to the considered degradation faults, and these parameters are estimated recursively from the data that is available from the system at each sample. Detection and diagnosis of faults are facilitated by monitoring the changes in the estimated parameters from their nominal ('correct operation') values. The magnitudes of the parameter changes represent a direct estimate of
8.1. Conclusions

Robust estimation of the model parameters in a recursive fashion is one of the main challenges of the work. The non-linear relationship between the outputs and the parameters in the model that was considered was one significant cause of difficulty. This non-linearity was addressed by linearising the model in the parameter space at each sample in order to calculate the direction and magnitude of the parameter updates. The resulting algorithm is analogous to the recursive least-squares method, to which it reduces for a linear-in-the-parameters model, and the analogy has been exploited to allow confidence intervals to be calculated for the parameters that are estimated.

Forgetting was used to make the estimator continually sensitive to the most recent data. The effect of forgetting is to weight the data exponentially in the time domain. The sensitivity of the estimator is then governed by the forgetting factor. In air-conditioning systems, the inputs can remain at one operating point for long periods and estimator wind-up can occur when forgetting is employed. This leads to numerical problems and instability in the parameter estimates. This was addressed by using the prediction error forgetting (PEF) algorithm (Fortescue et al., 1981) to adjust the forgetting factor according to an assessment of the information content of the data, which is jointly determined by the variations in the inputs and the prediction error.

It was demonstrated in Section 6.4.1 that the wind-up problem can still occur when the inputs are varied very slowly while the parameters are changing. An interesting phenomenon that was observed in Section 6.4.1 was that instability in the parameter estimates, induced by estimator wind-up, was self-correcting due to the non-linear form of the model. This is because the gradient vector of the model function, and hence the $P$ matrix, which determines the gain of the estimator, is a function of the parameters as well as the inputs.

The derivation of the estimator, presented in Chapter 5, showed that it is equivalent to the recursive least-squares algorithm when the model function is linear with respect to the parameters. This approximation deteriorates as the model deviates from being linear over the interval of a parameter change. The similarity of the algorithm to recursive least-squares thus depends on the size of the change in the parameters and the degree of non-linearity in the model function. The behaviour of the estimator is thus specific to the model with which it is used. The behaviour
of the estimator was examined in Chapter 6 for different magnitudes of parameter change using the models described in Chapter 4.

In Section 6.3.2, the parameters converged on a wrong set of values for the largest step changes that were applied to the parameters of the cooling coil model. This therefore appears to indicate the existence of local minima on the surface of the criterion function caused by a deterioration in the linear approximation to the model function. The tests provided a quantitative indication of the largest step change over which the linear approximation was acceptable. However, since the local non-linearity of the model depends on the estimates of the other (non fault) parameters, the results are not generic. It may be noted that the sizing and design of typical heat exchanger subsystems in air-conditioning systems are not expected to differ significantly from the type studied in the thesis. Hence, the results do provide an approximate estimate of the magnitude and rate of parameter change that is likely to lead to inaccuracies in the estimation.

Chapter 6 also investigated the effect of variability in the inputs, noise, model mismatch, and unmeasured disturbances on the performance of the estimator. It was demonstrated that the performance was sensitive to the rate of change in the inputs relative to the rate of change in the parameters. An attempt was made to quantify this relationship in Section 6.4.1, and an estimate of the maximum rate of change in the parameters (when changing simultaneously) for diurnal cycles was provided (Equation 6.6). Modelling errors and unmeasured disturbances were discussed in Chapter 6 and it was described how these two phenomena both induce specification error in the model. Specification errors lead to changes in the estimated parameters and these can be difficult to distinguish from faults.

The performance of the estimator was evaluated in Chapter 7 using data from two complete air-conditioning systems: a simulated and a real system. These systems were subject to realistic disturbances and structural differences existed between the FDD model and the system. The results of the tests demonstrated that the proposed FDD scheme is capable of distinguishing the three degradation faults that were considered. In the simulation tests, the scheme was able to provide an accurate estimate of the degree of the faults, and 99% confidence intervals provided a useful fault detection threshold.

In the tests using the data from the real system, the sensor information was unreliable and there were inconsistencies in the data. The sensitivity parameter,


8.1. Conclusions

\( \gamma \), was thus set to a high value, which corresponds to low sensitivity to prediction errors. In spite of this, the estimator was able to detect the faults and distinguish between them. However, the estimates were less stable than they were for the simulated system, largely because of the inaccurate sensor readings. The tests thus demonstrated the importance of obtaining reliable sensor information in order to estimate the degree of fault accurately. In addition, the importance of being able to model the effect of faults accurately was demonstrated. This was particularly apparent in the results obtained when trying to detect the leakage fault. The way that leakage was modelled in the FDD model was not consistent with the way it was modelled in the test systems and this caused other parameters to be varied to provide a compensatory effect.

Chapter 7 demonstrated that in the face of inaccurate sensors and specification errors little credence can be attributed to the estimated values for the system parameters. Low cost sensors are generally used in air-conditioning systems and these sensors are likely to be unreliable in practice. The fault estimation aspect of the FDD scheme may thus prove unreliable for the application of air-conditioning systems. However, as demonstrated in Chapter 7, the detection and isolation aspects are more robust. The parameters could therefore be used as indications of symptoms associated with certain classes of faults, rather than accurate estimates of the absolute degree of fault.

In summary, there a number of specific conclusions that can be drawn:

- obtaining an accurate model of the correctly operating system is of prime importance for the subsequent on-line FDD task. Although the analytical models that have been presented allow most of the parameters to be estimated from design and manufacturers’ data, it is necessary to estimate the more empirical parameters using training data. This serves two purposes:

  1. It improves the fit of the model to the system, ultimately allowing the detection of smaller faults.

  2. It confirms that the system is operating correctly, according to the design specifications;

- the estimator is only capable of tracking slowly varying parameters. Large step changes and fast ramps lead to instability, due to a breakdown in the
8.2 Suggestions for further work

quadratic approximation to the criterion function or the rate of change in
the inputs being too slow relative to the rate of change in the parameters:

- the PEF algorithm does not protect against estimator wind-up when the
  inputs are constant but the parameters are changing. However, during low
  levels of input excitation, the potential for wind-up was reduced by selecting
  parameters whose effects were orthogonal;

- modelling errors or inconsistencies in the data obtained from the plant af-
  fect the performance of the estimator by causing the parameter estimates to
  fluctuate as the errors in the model vary across the operating range. These
  fluctuations can be attenuated by reducing the sensitivity of the estimator:
  i.e. by increasing the magnitude of $\gamma$. This approach leads to a larger ‘win-
  dow’ of samples being used to estimate the parameters. The sensitivity to
  local variations in model accuracy will be reduced if the system explores its
  complete operating range during the window time period;

- unmeasured disturbances can be treated in the same way as modelling errors
  by reducing the sensitivity so that the average effect of the disturbances over
  the length of the window is reduced. However, if the disturbances have a
  significant non-zero mean the estimates will be affected;

- for analytical model-based schemes, it is important that the sensor readings
  are accurate representations of the physical properties they are supposed to
  measure. If this is not so the parameters will be biased and will thus lose
  their physical meaning.

8.2 Suggestions for further work

The effect that non-linearities in the parameter space of the model have on the
estimation process needs to be investigated further. A number of different models
could be contrived that produce particular classes of non-linearity in the criterion
function, e.g. varying levels of deviation from quadratic, saddle regions, local min-
ima. Software that enabled multi-dimensional problems to be visualised would
be useful for this work, and visualisation of the way in which the parameters are
moved by the estimator for non-quadratic surfaces would help increase the under-
standing of the process. In addition, the parameters that were included in the
models to represent the three considered degradation faults had relatively orthogonal effects on the model output. Further work needs to be carried to examine how the robustness of the estimator is by varying the degree of orthogonality between the parameters that are estimated.

It was demonstrated in Chapter 6, that the shape of the criterion function surface is affected by the excitation of the inputs. The degree of curvature at the minimum determines the confidence that can be attributed to the parameter estimates. This confidence can be increased by obtaining a greater number of independent input-output samples from the system. Since the models considered in this thesis are non-linear in the input space, the curvature of the minimum will be more sensitive to data from certain regions of the input space. Further work therefore needs to be carried out to ascertain the regions of the operating range where data should be obtained to allow the parameters to be estimated reliably. The results from this work would be particularly useful for designing tests for collecting training data from the system.

The tests carried out using the data from the BRE system demonstrated the sensitivity of the estimator to anomalies in the data. The estimator is quadratically sensitive to prediction errors since the criterion function is defined as the sum of the squares of these errors. The parameters that are estimated can therefore be changed significantly when a large error occurs. If this error is due to a momentary problem with a sensor and is not sustained, many samples may then be required to reverse any parameter changes that are made. This problem could be alleviated by using square root algorithms (Åström and Wittenmark, 1989). Further work could involve applying these algorithms and investigating their effect on the overall sensitivity of the estimator.

The FDD scheme described in this thesis involved estimating the values of parameters that related directly to particular faults. If a fault were to occur that was not explicitly modelled, the parameters that are estimated would still be altered. Although the parameters could not be used directly to diagnose the fault, the overall direction of the change in the parameter vector would be related to the type of fault. Hence, there is the potential to supplement the parameter estimation scheme with a classifier (e.g. based on expert rules) so that faults not directly represented by the estimated parameters may be diagnosed. The parameters may be used as indicators of particular symptoms, which could be used to distinguish between different classes of faults. As explained in the Section 8.1, this approach
may prove useful when the sensor measurements are unreliable. This idea needs further investigation.

This thesis has concentrated on one class of subsystem in HVAC plant: other air-side subsystems, such as VAV boxes, fans, dampers, etc., also warrant attention. Analytical models of these items are available, e.g. (Clark, 1985), and the potential exists for applying the parameter estimation techniques that have been described to the models of these subsystems. The techniques may also be applied to primary plant, such as boilers and chillers. In contrast to the bespoke air-side plant, these systems are generally mass-produced. Models of these systems may therefore be identified by the manufacturer rather than on site. In this case it should only be necessary to identify the model once, rather than for each item.

The work has reinforced the need for adequate commissioning of air-conditioning systems. One aspect of commissioning is the detection, diagnosis, and elimination of faults. It seems logical therefore to address on-line FDD and commissioning using the same technology. The type of faults at commissioning time will clearly not be the result of degradations, but will most likely have been caused by poor installation work or even poor design. Their effects will be more obvious than degradation faults and do not therefore require the use of highly accurate models to detect them. One idea, therefore, would be to utilise the models intended for real-time FDD by first configuring them using design data only. The predictions of these models could then be compared with the outputs that are measured during tests on the system in order to detect any faults. Data obtained from commissioning tests could be used to calibrate the models for the on-line FDD task. The idea of combining commissioning and on-line fault detection and diagnosis is currently being investigated by the author and co-workers as part of a collaborative research project.
Nomenclature

\( \alpha \) Air-side approach of the heat exchanger
\( \beta \) Curvature parameter for inherent valve characteristic
\( \gamma \) Sensitivity parameter used in PEF algorithm
\( \delta \) A small number
\( \epsilon \) Model prediction error
\( \theta \) Model parameter vector
\( \kappa \) Threshold in steady-state detector
\( \kappa_a \) Air-side convective heat transfer parameter
\( \kappa_w \) Water-side convective heat transfer parameter
\( \lambda \) Forgetting factor
\( \tau \) Time constant
\( \psi \) Vector of model first derivatives (Jacobian)
\( \omega \) Angular frequency
\( \varepsilon \) Effectiveness of the coil

\( A \) Authority of the valve
\( C_a \) Capacity rate of air
\( C_w \) Capacity rate of water
\( C_r \) Ratio of fluid capacities
\( I \) Identity matrix
\( L \) Gain vector in estimator
\( NTU \) Number of transfer units for the coil
\( P \) Inverse of the Hessian matrix used in the parameter estimation
\( Q_{total} \) Heat transfer rate of wet coil
\( Q_{dry} \) Heat transfer rate of dry coil
\( R_w \) Conductive resistance of tube wall in heat exchanger
\( SHR \) Sensible heat ratio
\( T_{ai} \) Air temperature entering coil
\( T_{ao} \) Air temperature leaving coil
\( T_{wi} \) Water temperature entering coil
\( T_{wo} \) Water temperature leaving coil
\( T_s \) Average surface temperature of coil
UA Overall heat transfer conductance
\(V(.)\) Criterion function used in parameter estimation

\(b\) Temperature sensor off-set parameter
\(b_f\) Bypass factor
\(f\) Frequency
\(h_{ai}\) Humidity of air entering coil
\(h_{ao}\) Humidity of air leaving the coil
\((\eta_A h_A)\_a\) Air-side convective heat transfer coefficient
\((h_A)\_w\) Water-side convective heat transfer coefficient
\(k\) Sample number
\(l\) Valve model leakage parameter
\(m_a\) Mass flow rate of air entering coil
\(\dot{m}_w\) Mass flow rate of water coil entering coil
\(\dot{m}_{w,d}\) Maximum mass flow rate of water through coil (i.e. supplied by pump)
\(q^{-1}\) Backward shift operator
\(s\) Valve stem position
\(t\) Time
\(u_v\) Control signal to valve actuator
\(\mathbf{u}\) Vector of model inputs
\(v\) Hysteresis parameter
References


Appendix A

The Complex method

The complex method is an adaptation of the simplex direct search method and was proposed by Box (1965) to allow the incorporation of constraints. The method works by defining a geometric shape, known as a ‘complex’ in the search space. The objective function is evaluated at each vertex, and the shape reflected about its centroid in a direction towards the optimum.

The complex rolls over and over, normally expanding, and contracts and flattens itself if a constraint boundary is encountered. It is able to roll along the constraint boundary or leave if the direction of the optimum is away from the constraint. The complex is also able to accommodate more than one boundary and turn corners. If the complex straddles a minimum, with no better solutions possible through reflection, it will collapse onto its centroid and converge on the local minimum.

The method is often able to avoid converging on a local minimum due to the complex initially straddling a large portion of the search area. It is then able to reflect itself over the local solutions towards the global minimum. The method can be effective for problems having a small number of parameters, but the performance deteriorates as the number of parameters becomes very large (e.g. >> 20).

The Complex method has been used to estimate the parameter values for models using data that records the inputs and outputs to the modelled system. The objective function ($V$) for the parameter estimation was defined as the mean of the absolute prediction errors. For a scalar output model the objective function is defined by:
\[ V(\theta) = \frac{1}{N} \sum_{k=1}^{N} |f(\theta_k, x_k) - y_k| \] (A.1)

where \( \theta \) is the vector of parameters to be estimated, \( x \) is the vector of measured inputs, \( y_k \) is the measured output, \( f(.) \) is the model function, and \( N \) is the number of data vectors. The absolute prediction error is used instead of the squared prediction error in order to reduce the effect that anomalous data points have on the estimation. The complex method assumes that an initial feasible point \( \theta_1 \) is available; where \( \theta \in R^p \). A sequence of geometric figures, each having \( K \geq p + 1 \) vertices is formed to find the constrained minimum point of the function \( V(.) \).

The search procedure, according to (Rao, 1987), is described below:

1. Construct the initial complex.

   Find \( K \geq n + 1 \) points, each of which satisfy all the constraints. With the exception of \( \theta_1 \), the remaining \( K - 1 \) points are generated randomly. Each component, \( \theta_{i,j} \), of the vectors, \( \theta_i \) is selected as a random point within predetermined limits:

   \[ \theta_{i,j} = \theta_{i,\text{low}} + r_{i,j}(\theta_{i,\text{high}} - \theta_{i,\text{low}}), \quad i = 1, 2, \ldots p; \quad j = 1, 2, \ldots K. \] (A.2)

   where \( r_{i,j} \) is a random number, and \( 0 \leq r_{i,j} \leq 1 \). If a newly generated point, \( \theta_j \), does not satisfy the constraints, it is moved half way towards the centroid of the remaining, already accepted points:

   \[ \theta_j = \frac{1}{2}(\theta_j + \theta_o) \] (A.3)

   where \( \theta_o \) is the centroid of the already accepted points.

   \[ \theta_o = \frac{1}{j-1} \sum_{i=1}^{j-1} \theta_i \] (A.4)

   This process is continued until all \( K \) points have been found.

2. Evaluate the objective function at each of the \( K \) points (vertices).

   The parameter vector that produces the largest objective function value is denoted by \( \theta_k \). A new point, \( \theta_r \), is found by reflection:
\[ \theta_r = (1 + \alpha) \theta_o - \alpha \theta_h \]  

(A.5)

where \( \alpha \geq 1 \), and \( \theta_o \) is the centroid of all vertices not including \( \theta_h \):

\[ \theta_o = \frac{1}{K-1} \sum_{l=1,l \neq h}^{K} \theta_l \]  

(A.6)

3. Test the feasibility of the new point \( \theta_r \).

If \( V(\theta_r) < V(\theta_h) \), \( \theta_h \) is replaced by \( \theta_r \). A new geometric figure (complex) is then constructed, and the procedure returns to the second step. If \( V(\theta_r) \geq V(\theta_h) \), \( \theta_r \) is not the direction towards minimum and a new \( \theta_r \) is calculated by reducing \( \alpha \) by a factor of 2 until a satisfactory \( \theta_r \) is found. If \( \alpha \) is smaller than a prescribed small quantity \( \epsilon \), and an appropriate \( \theta_r \) has not been found, the process is terminated. A new search is then started using the parameter vector that produces the second largest objective function value, instead of \( \theta_h \).

4. If \( \theta_r \) is infeasible at any stage in the optimisation, it is moved halfway towards the centroid until it is feasible:

\[ (\theta_r)_{\text{new}} = \frac{1}{2} (\theta_o + \theta_r) \]  

(A.7)

This step is continued until the complex collapses onto its centroid.

5. Generate the new complex.

The new complex is generated by replacing \( \theta_h \) with \( \theta_r \).

The optimisation is terminated when either of the following conditions are satisfied:

(a) the complex shrinks to a specified small size \( \epsilon_1 \).

(b) the standard deviation of the function value becomes sufficiently small, i.e.

\[ \frac{1}{K} \sum_{j=1}^{K} [V(\theta_o') - V(\theta_j)]^2 < \epsilon_2 \]

where \( \theta_o' \) is the centroid of all the \( K \) vertices of the current complex, and \( \epsilon_2 \) is a specified small number.