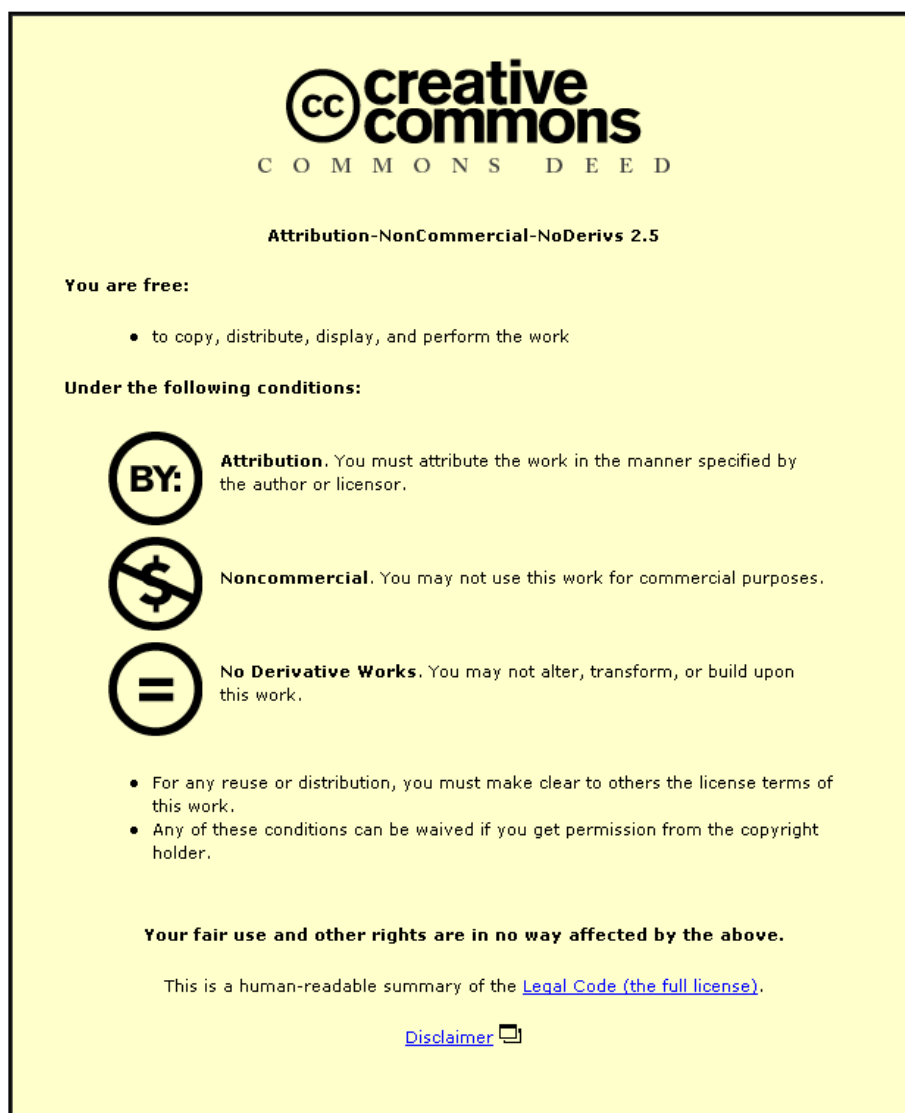




This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Model-Based Condition Monitoring of a HVAC Cooling Coil Sub-System in a Real Building

R. A. Buswell¹, P. Haves² and J. A. Wright¹

Abstract

A comparison of the performance of two fault detection and diagnosis methods applied to a cooling coil subsystem in an air-handling unit installed in a real building is presented. Both methods employ a first principles based reference model of the target system. One scheme carries out diagnosis using expert rules and the other recursively re-estimates selected parameters of the system model that correspond to particular faults. The procedures and information required to configure the schemes for condition monitoring are discussed. The results of testing the methods on an HVAC cooling coil subsystem in a commercial office building in the UK over an entire cooling season are reported. Both methods were able to both detect and diagnose faults; however, the expert rule method appears to be more robust. Issues associated with the configuration and implementation of both methods are discussed in terms of performance and cost.

1. Introduction

The increasing complexity of building control and management systems heightens the need for the development of tools to assist in the monitoring of these systems. The application of these tools is expected to lead to improved comfort, energy performance and reduced maintenance costs^[1]. The realisation of such tools is made possible by the increasing computational capacity of contemporary computer-based building management systems (BMS).

¹ Loughborough University, U.K.

Currently, BMSs are able to detect abrupt changes in the condition of heating, ventilating and air conditioning (HVAC) systems, usually through the violation of simple thresholds applied to measurements or control set points. They do not, however, offer significant fault isolation and diagnostic information and are unable to detect gradual degradations in system performance, which can result in poor system operation and unnecessarily high energy consumption. Fault Detection and Diagnosis (FDD) techniques attempt to identify faulty system performance, alerting building operators and offering diagnostic information.

FDD has been the subject of research in the process industry for some years and has more recently been expanded into the area of HVAC equipment. In the past eight years, the International Energy Agency (IEA) Annex 25 has investigated real time simulation of HVAC systems for building optimisation, fault detection and diagnosis^[2]. Most of the FDD methods developed were evaluated using simulations of building plant and controls systems. A recently completed IEA Annex (Annex 34) addressed the issues surrounding the practical application and demonstration of FDD tools in HVAC plant installed in real buildings. This research formed part of the U.K. contribution to Annex 34^[16].

This paper describes the implementation of two methods for the automatic detection and diagnosis of degradation faults in HVAC cooling coil subsystems. Both methods use a first principles model of the system to detect and diagnose changes in the system condition. The first method uses a reference model to

² Lawrence Berkeley National Laboratory, USA

determine change in the performance of the system and uses expert rules to determine the possible cause. The second method recursively re-estimates the parameters of the model that correspond to possible faults; a significant change in the value of a particular parameter indicates the presence of the corresponding fault. The paper then describes the application of these FDD techniques to a cooling coil in an air-handling unit (AHU) in a real building and reports their performance.

2. Model-based Approaches to FDD

In model-based FDD, a quantitative model is used to represent the correctly operating system process and is termed a ‘reference model’. Measured data from the real system are used as inputs to the model and the model predictions are compared to the measured output from the system, as illustrated in Figure 1. The difference between the prediction and the observed value is termed the ‘residual’. A significant residual, i.e. greater than a threshold value that is determined by the uncertainty in the model predictions and the measurements, is termed an innovation and is evidence that the system is not operating as expected.

Figure 1: A model-based fault detection scheme.

Generally, reference models can be split in to two categories ^[3]. *Black box* models are empirical models that embody no prior knowledge of the system (apart from that which is implicit in the choice of the inputs and outputs). Physical models are largely based on a first principles analysis of the system, which may include

established empirical relationships (e.g. to predict heat transfer coefficients). Black box models do not require any prior knowledge of the process. They are, however, poor at extrapolation and so must be configured using training data generated by the target system (or possibly a simulation of that system) that cover the whole region of operation. The main advantage of black box models is that they can be chosen to be linear in the parameters, making the process of estimating the parameters of the model both less demanding computationally and more robust. One advantage of physical models is that the prior knowledge that they embody improves their ability to extrapolate to regions of the operating space for which no training data are available. They also require fewer parameters for a given degree of model accuracy. A further feature of physical models is that they may be designed so that the parameters represent physically meaningful quantities, allowing abnormal values of particular parameters to be associated with the presence of particular faults.

Fault diagnosis requires knowledge of how the system behaves when faults are present. The two approaches investigated in the work reported here are:

- analysis of how the innovations vary with operating point;
- estimation of the parameters of an *on-line* model that has been extended to treat particular faults.

The first approach has been implemented using a rule-based classifier, in which the rules are obtained from experts^[3]. The current system condition is compared to a model of correct operation in which the parameters are fixed at the values estimated during model calibration. Any innovations are evidence that the system

operation has changed; a diagnosis is generated using expert rules that describe the variation of the residual with operating point expected for different faults. The second approach has been implemented by extending the system model to represent both faulty and correct operation using physical principles and engineering knowledge. The current values of model parameters that correspond to particular faults are estimated from the measured performance of the system using the recursive parameter estimation (RPE) method^[4,17]. A significant change in the value of a particular fault parameter provides an indication of the presence of the corresponding fault. The estimated values for the parameters relating to the faulty behaviour are then used for fault diagnosis^[4, 5]. In each case, first-principles steady-state models of the system are used to represent the performance of the system. The predictions of the models are only valid when the system is in steady-state and so a steady-state detector is employed to filter out unwanted transient data^[4,17].

2.1. First Principles Subsystem Models

The basis of the models used here is described by Buswell et al.^[17]. Certain criteria have become evident in the course of developing models for on-line use; the models should:

- be designed to operate within the limitations imposed by the monitoring capabilities of the HVAC control system;
- require no more input data (sensor measurements and control signals) than are available from a typical HVAC control system (minimising implementation cost);

- be able to reproduce the correct operation of the system to the accuracy required to detect the desired degree of fault, though there is little point in exceeding the accuracy of the sensors used.

The cooling coil subsystem model whose use is described below consists of separate models of the actuator, the supply air temperature sensor, the three port control valve and the cooling coil. The actuator model treats both hysteresis and any mismatch between the travel of the actuator and the valve. Leakage in the control valve is modelled by assuming there is an additional flow path through the control valve whose resistance is independent of the stem position. The installed characteristic is modelled by correcting the inherent characteristic of the valve for valve authority. The cooling coil operation is modelled using the NTU-effectiveness method, with wet coil operation modelled using the SHR method^[4, 8]. The under-capacity fault is modelled by scaling the resistance of the coil to heat transfer, thus reducing the effectiveness. The coil outlet air temperature sensor fault is modelled as an offset that is constant over the range of operation. The models and the fault parameters are described in more detail by Salisbury^[4] and Buswell and Wright^[17].

2.2. Diagnosis using Expert Rules

Figure 2 illustrates the principles of the rule-based fault diagnosis scheme. The operating range is divided into a number of different regions such that each of the faults of interest produces a unique combination of effects in the different regions. The choice of the positions of the boundaries between these regions is part of the process of defining the expert rules and is based on engineering judgement, once the characteristics of the subsystem have been established from the calibration

tests. An accumulator ‘bin’ is set up for each region of the operating range and the value of the residual for each steady-state sample is accumulated in the ‘bin’ that corresponds to the region that includes the current operating point. The current bin value, B_n , is an exponentially-weighted moving average of the prediction errors in the corresponding part of the operating range:

$$B_n = \lambda \varepsilon_n + (1 - \lambda) B_p \quad (1)$$

where λ is the weighting, or forgetting, factor ε_n is the prediction error for the current sample and B_p is the bin value at the previous sample. The bin value is only updated if the operating point is in the region corresponding to the bin in question and the system is deemed to be in steady-state. At each steady-state sample, the magnitude of the accumulated residuals in each bin is compared to an ‘error significance threshold’, α . The innovation at the current sample, i_n , is then calculated as:

$$\begin{aligned} i_n &= B_n - \alpha \quad \text{for } B_n > \alpha \\ &= B_n + \alpha \quad \text{for } B_n < -\alpha \\ &= 0 \quad \text{for } |B_n| < \alpha \end{aligned} \quad (2)$$

The significance threshold reflects both modelling and measurement errors and is determined from the goodness of fit of the model to the calibration data. In the general case, it may vary over the operating range.

Expert rules are used to analyse the innovation values at each sample. The rule base consists of a set of crisp ‘IF-THEN’ rules that relate the innovations to the set of possible faults, e.g.,

IF the High Duty innovation is significant and negative AND the Medium Duty innovation is insignificant AND the Low Duty innovation is insignificant THEN the coil is fouled

Table 1 lists the signs of the innovations in different operating regions expected for different faults in a cooling coil.

Table 1: Variation of innovation with operating point for different faults.

Figure 2: The expert rule fault diagnosis scheme.

Figure 3: The recursive parameter estimation scheme.

2.3. Diagnosis using Recursive Parameter Estimation

The recursive parameter estimation method is illustrated in Figure 3. The steady-state prediction errors are used by the parameter estimator to re-estimate the values of the fault parameters recursively^[4,17]. A change in the value of a parameter from the value estimated when calibrating the model indicates a change in the condition of the system. For instance, the three-port valve model includes a parameter that represents the leakage of the valve; assuming that the model structure allows a good representation of valve leakage, then an increase in the leakage parameter indicates directly that the leakage has increased. As with the expert rule approach, the effectiveness of the parameter estimation method is dependent on the system covering enough of the operating range to differentiate between the faults of interest. For example, if the system has not been operating in the low duty region, then there can be no prediction errors resulting from valve leakage and the valve leakage parameter should not change. The parameter estimation method estimates

confidence intervals (at the 99.9% level) for the parameter values, c_{ϕ}^{est} , where ϕ is the fault parameter. These limits give an indication of how well the model fits the data. Under normal operating conditions, the parameter estimates vary to a small degree due to inevitable imperfections in the model system and errors in measurement. The magnitude of the parameter variation under normal operation implicitly describes the uncertainty in the model representation of the process characteristics. This uncertainty should be accounted for to prevent false alarms. A second measure of uncertainty, c_{ϕ}^{cal} , is therefore introduced to minimise the false alarm rate. c_{ϕ}^{cal} is determined for each parameter by observing the parameter variation in a set of normal operating data, preferably covering the expected range of operation. In the current implementation, the value of c_{ϕ}^{cal} is set equal to the maximum variation observed in the corresponding parameter. The threshold for a significant change in the value of a parameter is therefore given by:

$$C_{\phi} = \sqrt{(c_{\phi}^{cal})^2 + (c_{\phi}^{est})^2}, \quad (3)$$

A significant change in the parameter value is deemed to have occurred when the following rule is satisfied;

$$IF (| \phi_{est} - \phi_{cal} | > C_{\phi}) THEN change is significant \quad (4)$$

where ϕ_{cal} is the value of the parameter after the model calibration process. If the value of one fault parameter becomes significantly different to that established during calibration, then the corresponding fault is considered to have been diagnosed. If more than one parameter value changes significantly then the

condition is diagnosed as ‘other’. A lack of significant changes in the parameter values indicates correct operation.

3. The Test Building, Plant and Control System

The fault detection and diagnosis methods were tested on a cooling coil in a constant air volume HVAC system in a UK commercial office. The coil is in one of twelve air-handling units located in a single plant room. As shown in Figure 4, the air handling units do not have individual mixing boxes; instead, return air from the space is mixed with outside air in a common mixing chamber formed by the plant room itself. The AHU monitored in this work stands vertically and is approximately 4m high by 0.8m square. Air entering the unit passes through a filter, the cooling coil, the heating coil and the supply fan. The cooling coil is inclined at approximately 45° to the vertical.

Figure 4: Section through the test AHU.

3.1. Faults to be Identified

Three faults were selected for detection and diagnosis in this study^[2, 4, 9]:

- leakage through the control port of the three port control valve for the cooling coil;
- fouling of the cooling coil (or reduction in capacity);
- fixed offset in the coil outlet air temperature measurement.

This selection of faults provides a good test of diagnostic methods since each fault will affect performance of the coil in a different part of the operating range and

should therefore have the potential to be identified uniquely^[10]. The nature of any faults that may have occurred in the test period was not known in advance. Artificial faults were not introduced into the system; the approach adopted in this investigation was to monitor the system and wait for any naturally occurring changes in the behaviour of the system.

4. Implementation of the FDD Scheme

The following procedure, evolved during the field-testing, was used to implement the FDD methods:

- establish the target system and faults to be monitored;
- check that the necessary measurements exist to support the FDD method;
- establish the FDD regime, i.e. *on-line* or batch processing, and the frequency of data sampling;
- establish the mode and format of data collection (remote access, text file, DDE);
- establish communications requirements (how the FDD tool accesses the data);
- define the model calibration tests required and the means by which these tests may be performed;
- gather the required configuration data from design information and/or site inspection;
- test the installation to ensure continuity in data handling and correct operation of the FDD tool;
- carry out the calibration testing;

- if the plant operation is acceptable, calibrate the system models and initiate condition monitoring.

4.1. Sensors Used

Table 2 shows the measurements that were used by the FDD procedures. The sensors are commercial grade temperature sensors commonly used for HVAC control applications.

Table 2: Measurements used by both condition monitoring schemes.

The air temperature leaving the coil predicted by the model is compared with the measured value in order to generate the value of the prediction error. The air inlet temperature and the control signal are inputs to a first-principles model of the coil and control valve, as described above. The chilled water temperature, chilled water supply pressure and air-flow rate are assumed to be constant and are established from site inspection or design information. During the calibration testing, the coil was inspected at full duty and found to be virtually dry. In the absence of the necessary humidity measurements, it was assumed that any latent duty was insignificant.

4.2. Model Calibration and Training Data

The HVAC subsystem models developed for this application are generic and are configured for use with a specific system by identifying the model parameters from measured system input/output and design data. For the semi-empirical model used here, geometric and configuration parameters, such as the number of parallel coil

circuits; the number and internal diameter of the coil tubes; and the coil face area are required to calculate the respective fluid velocities across/through the coil. These were found by inspection or from the manufacturer's design data. Empirical parameters, such as actuator hysteresis and valve curvature, are estimated from performance measurements. Non-linear optimisation^[11] is used to fit the model to the training data. Further details of this process and the models used can be found in [5, 7, 11].

Steady-state calibration data were collected from the cooling subsystem using a systematic testing procedure:

1. Set to open loop control.
2. Divide the control signal range into discrete intervals.
3. Starting at one end of the range, measurements are recorded once the system is in steady state. The control signal is then stepped through each interval, pausing after each step until steady-state is attained.
4. The same sequence is repeated in reverse, although limitations on test time may necessitate a reduced number of test points. At least one point in the mid-range is required in order to check for hysteresis.

4.3. Thresholds and Operational Parameters

The FDD methods used in the work reported here use steady state models and so it is necessary to determine when the system being monitored is close to steady-state. The system is deemed to be in steady state, and the model predictions valid, when the following conditions are satisfied:

$$\tau \frac{\bar{y}_k - \bar{y}_{k-1}}{\Delta t} < \kappa \quad (5a)$$

$$\tau \frac{\hat{y}_k - \hat{y}_{k-1}}{\Delta t} < \kappa \quad (5b)$$

where \bar{y}_k is the current value of the system output, filtered to reduce noise, \bar{y}_{k-1} is the (filtered) value of the system output at the previous sample, \hat{y}_k and \hat{y}_{k-1} are the corresponding values of the model prediction and Δt is the sampling interval. In this case, the system output is the coil outlet temperature. The value of the threshold, κ , was 0.25 K and the value of the time constant, τ , was 300 seconds. The operational parameter that controls the sensitivity of the recursive parameter estimation method is the forgetting constant, γ , which is discussed by Fortescue et al.^[13] and Salsbury^[4]. The value of γ needs to be selected so that the forgetting rate is such that information from the whole operating range is retained. The value of γ depends on the excitation rate of the input variables relative to the data sampling rate. A value of $\gamma = 70$ was found to produce a stable response that was sensitive enough to identify faults within a day, which was the period of time for which the faults were introduced in the tests. In real systems, degradation faults are likely to develop over a period of months rather than hours, so rather larger, more conservative, values of γ could be used. Further work is required to establish more a formal process for determining suitable values of γ .

There are a number of operational parameters associated with the rule-based diagnosis scheme. The operating range is divided into three equally sized

regions, so the boundaries between the bins are at control signal values of 0.33 and 0.67. The value of the error significance threshold, α , was set to 0.75K. The forgetting factor for the innovation values stored in the bins was 0.994.

5. Model Calibration

The model calibration data were obtained by stepping the control signal to the actuator, as shown in the bottom graph of Figure 5. After stepping in 10% increments from fully closed to fully open, the valve was stepped back to two partly closed positions in order to test for hysteresis. Figure 5 compares the calibration test data and the model predictions after fitting the model parameters to the data. The continuous lines represent the measured data and the 'dots' represent the predictions of the reference model when the measured data are deemed to be in steady-state. The model predictions are within 0.75K of the measured data across the range of cooling coil operation, with the largest errors occurring at a control signal value of ~ 0.4 . The valve does not begin to open until the control signal exceeds 0.3, which indicates a significant mismatch between the range of movement of the valve stem and the control signal. In fact, inspection of the control strategy seemed to suggest that this feature was implemented deliberately in order to generate a control 'dead-band' between heating and cooling operation. The heating coil control valve in the same AHU exhibited similar characteristics.

Figure 5: Model calibration tests.

Figure 6 shows plots the normalised approach, β , of the measured data and the model predictions calculated by,

$$\beta = \frac{T_{ao} - T_{ai}}{T_{wi} - T_{ai}}, \quad (6)$$

where T_{ao} is the measured or predicted air outlet temperature, T_{ai} and T_{wi} are the air and water inlet temperatures respectively. The poor valve rangeability can be seen. The spread of the model prediction in the U_{cc} direction is due to the non-zero value of hysteresis in the model.

Figure 7 shows the prediction errors generated from all the normal operation data used during the test period. It may be noted that: the prediction errors are generally less than 0.75K except in the region $0.33 < U_{cc} < 0.47$. The large prediction errors appear to be due to the valve opening at different values of control signal. A possible reason is a variable offset between the ranges of motion of the valve and the actuator, exacerbated by a combination of poor rangability, resulting in rapid opening, combined with significant hysteresis. The true cause, however, could not be determined and since the latter effect cannot be predicted from the available measurements, the data in the range $0.33 < U_{cc} < 0.47$ have been excluded from further consideration. The rejected data are plotted lightly in Figure 6.

5.1. System and Data Characteristics

The test period lasted approximately six months, from 10th June, the date of the model calibration testing, until 31st December. The cooling coil covered its whole range of operation during this period. The measured inputs were sampled at one

minute intervals. For ~30% of these samples, the plant is either shut down or there are start up transients. Inspection of the data revealed that the system operation characteristics between start up and 18:00 hours (normal working hours) and those between 18:00 hours and shut-down were different. The difference appeared when $U_{cc} = 0.0$ and, with reference to Figure 6, gave an approach of approximately zero, slightly higher than shown. The reason for this could not be confirmed, but it occurred persistently throughout the test period. Weekend days were affected in a similar way. Because the step test was carried out from 18:00 hours onwards, the normal working day and weekend data was rejected as part of the steady-state filtering procedure. This left ~20% of the data, of which ~11% were in steady state, and hence suitable for FDD.

Figure 6: The airside approach characteristics for all the test data.

Figure 7: Prediction error for the ‘normal operation’ test data.

For the purposes of the rule-based diagnosis method, the operating range was divided into three bands to represent low medium and high operation. These bands corresponded to control signal ranges of $0.0 \rightarrow 0.33$, $0.34 \rightarrow 0.67$ and $0.68 \rightarrow 1.0$. Figure 8 shows the cumulative total number of steady-state points of each of these ranges over the test period. The trends in information availability in each of the bands of operation can be observed for the whole test period. The implications for condition based monitoring are:

- the coil is active from June to the beginning of October;

- only from mid June to mid July does the operation of the system span the full range of operating conditions and hence provide the richest information for diagnosis of all of the faults considered;
- leakage faults could be detected in June and from September onwards;
- capacity faults could only be detected from July to August, possibly into September/October depending on the severity;
- faults that occur over the whole range of operation can probably always be detected.

The ability to diagnose a fault unambiguously depends on the similarity of its symptoms to the symptoms of the other faults being considered. This similarity tends to increase as the range of operating conditions is reduced. In the case reported here, the least ambiguous diagnosis can be expected to occur over the June-July period. If more system information is required to increase confidence in the diagnosis at any given time, the plant will require artificial excitation to generate the missing data.

5.2. System Faults During the Test Period

Figure 9 shows the prediction errors for all the steady-state test data over the test period. The two dark regions indicate the days where the prediction errors were generated when there was a fault in the system.

The first fault on day 60 occurred naturally and indicates under-capacity in the cooling coil subsystem. Examination of the corresponding data for the other air-handling units in the building showed that all the air-handling units were affected in a similar way, which suggests that there was actually a fault in the chiller or the chilled water circulation system. The chilled water temperature and flow rate were

not measured variables and so, in terms of the FDD methods, these chilled water problems would be seen as under-capacity in relation to the cooling coil subsystem model. The similar prediction errors magnitudes on days ~25 and ~68 did not generate alarms because the fault manifested itself for a shorter duration than on day 60 and so did not exceed the threshold value after being smoothed by a low pass filter. The second ‘fault’ occurred on day 175. In fact this was one day’s data where the normal working hours data (before 18:00 hours) were used. Because of the unresolved discrepancy between the evening and daily operation characteristics, discussed earlier, the data *appear* to indicate a leakage fault. This day was added to the test data set in order to test the FDD methods on leakage detection.

Figure 8: The cumulative number of steady-state data for the test period.

Figure 9: Prediction errors for the test period.

5.3. Performance of the Bin Method

Figures 9 and 10 illustrate the performance of the Bin Method rule-based diagnosis scheme. For clarity, only the “end of day values” are plotted. The contents of each bin are shown in Figure 10 (solid line) compared to the significance threshold (dotted line) and zero (dashed line). The bottom plot indicates the age of the data. The younger the data, the higher the potential reliability of the diagnosis. After day 120, the operating point is never in the part of the range corresponding to the middle and high bins and so the quality of the diagnosis becomes a function of how

representative the old information is with regard to the current condition of the system.

Figure 10 shows the transgression of the significance threshold by the prediction errors generated by the two faults at day 60, in the high bin and at day 175 in the low bin. Figure 11 displays the output of the expert rules analysis of the bin values. Lack of evidence of a fault in any of the four categories indicates normal operation. If the prediction errors exceed the significance threshold but the bin values do not correspond to any one of the three fault signatures, leakage, under-capacity or sensor offset, then the fault is classified as ‘other’. Using this scheme, both faults are diagnosed as expected. Although the under-capacity fault (day 60) only lasted for one day (Figure 9), the diagnosis indicates a longer period. This is because there were no suitable data samples during the following days, and so the high bin innovation remained constant until normal operating data in the high duty range were available on day 65.

5.4. Performance of the Recursive Parameter Estimation Method

Figures 11 and 12 illustrate the performance of the parameter estimation diagnosis method. As for Figures 9 and 10, the “end of day” values only are plotted for clarity. The current values of the fault parameter estimates are shown in Figure 12 (dot-dashed line). These values are bounded by the confidence limits described by Equation 3 (solid lines). The initial parameter estimates (dashed line) are the reference against which the confidence intervals on the current parameter estimates can be compared, determining normal or faulty operation. So long as the

confidence limits associated with the current parameter estimates span the initial parameter estimates, the system is deemed to be operating normally. If the parameter estimate changes so that the confidence interval no longer includes the initial estimate, an alarm for the corresponding fault is generated. If an alarm would be generated for more than one fault, each of the specific fault indicators is set to show 'no fault' and the 'other' classification is indicated.

Figure 10: End of day bin values for the test period.

Figure 11: Fault evidence output for the bin method scheme.

Figure 13 demonstrates that the under-capacity fault (day 60) is correctly identified, and the tool continues to indicate the presence of the fault for the same reasons as for the bin method. The 'leakage fault' (day 175) is not diagnosed as expected; the expert rules applied to the parameter estimates generate an 'other' diagnosis. The size of the prediction error generated for the leakage day is so large and abrupt that it causes the parameter estimator to become unstable. This is demonstrated by the erratic estimates for all three parameters on day 175, in particular the UA scaling factor. This highlights one of the limitations of this approach: It is designed to track slow changes in system characteristics and not abrupt changes. The magnitude of the prediction error generated by the under-capacity fault on day 60 increased gradually over a few hours and so could be handled by the scheme. However, the correction of the parameters after the fault, around day 70, caused a false alarm. One solution is to decrease the sensitivity of

the parameter estimator; however, in this case, this would have resulted in the under-capacity fault not being detected. The trade-off between sensitivity and robustness is more critical when dealing with quickly developing or intermittent faults.

Another issue is the coupling between the parameter estimates. This is demonstrated in Figure 12 when all the parameters change at once (day 60 and 175 in particular). The UA scaling factor only has an influence on the output of the model when there is heat transfer taking place. When the control signal, U_{cc} , is zero and the leakage parameter, l , is zero, the UA scaling factor has no influence. If however, the parameter estimator changes the value of the leakage parameter such that $l > 0$, the model predicts that there is chilled water flowing through the coil even when $U_{cc} = 0$ and, hence, that the UA scaling factor has an influence on the model output. As a result, the estimator can reduce the prediction error by changing the estimated value of the UA scaling parameter even if $U_{cc} = 0.0$. This can lead to unexpected behaviour by the estimator in which the model prediction error is minimised through the compensatory behaviour of the parameters. This contributes to the parameters not returning to their original values when normal operation resumes after the leakage fault.

Figure 12: End of day parameter estimates for the test period.

Figure 13: Fault evidence output from the RPE scheme.

5.5. Cost of Implementation

It was found in this investigation that the majority of the cost associated with condition monitoring is in the engineering time required for configuration. The implementation of the software was carried out by the authors and others associated with the project. A subjective estimate of the effort that each of the various tasks requires is listed in Table 3. These are estimates of the effort that would be required for the authors to implement the same FDD methods on a similar subsystem in one of the other air handling units in the test building. It is expected that the effort required would decrease as the methods and their application mature.

Table 3: Estimated time required to implement the FDD software.

The times listed in Table 3 assume complete familiarity with the target system. The time required to install and configure the diagnostic procedures for the first time in an unfamiliar building is considerably longer because of the need for:

- understanding unexpected process control strategies;
- familiarisation with the plant, control system and associated data;
- assessing the communications capacity of the building management system;
- establishing how the plant can be made to perform the calibration tests;
- arranging access to the building plant, and agreeing suitable times for testing;
- fault finding on failed network communications;
- establishing how the control system handles data logging faults.

6. Conclusions

Two methods for the condition based monitoring of HVAC systems have been implemented in this work. Both methods employ identical first principle models of the sub-system. In the first method, detection is based on the transgression of a significance threshold by the filtered prediction errors. Diagnosis is based on a crisp rule set that describes the relationship between the prediction errors in three intervals of the operating range and three fault conditions. The second method used the prediction errors to recursively re-estimate the values of three model parameters that represent the three possible faults. A fault is detected when the confidence limits associated with the current parameter estimate no longer include the initial parameter estimate. The diagnosis comes directly from the parameter values. Rules are employed only to identify ambiguous diagnoses. Both methods were applied to data gathered from a cooling coil sub-system installed in a real building. The test period lasted for six months and hence, the operating conditions covered the entire coil duty.

Both methods are suitable for *on-line* application. Faults can be detected and some level of diagnosis is possible using both approaches using both approaches. The Bin method scheme is simpler algorithmically and in its implementation. The RPE method performs detection and diagnosis simultaneously but is inherently more sensitive to unmeasured disturbances. More robust fault detection could be obtained by also running a model with fixed parameters that represent correct operation. Both methods are unable to diagnose the physical cause of the fault,

given the constraints of the available measurements and the simplicity of the approaches. Both approaches do, however, offer significant fault isolation and fault symptom information useful to the maintenance engineer.

References

- [1] ASHRAE 1996. "Fault Detection and Diagnosis for HVAC Systems".
ASHRAE Technical Data Bulletin. Volume 12, Number 12.
- [2] Hyvarinen, J. 1996. "Building Optimisation and Fault Diagnosis Source Book". Published by Technical Research Centre of Finland.
- [3] Benouarets, M., A.L. Dexter, R.S. Fargus, P. Haves, T.I. Salisbury and J.A. Wright. 1994. "Model-Based Approaches to Fault Detection and Diagnosis in Air-Conditioning Systems,". Proceedings of System Simulation in Buildings '94, Liege, Belgium.
- [4] Salisbury, T. I. 1996. "Fault Detection and Diagnosis in HVAC Systems using Analytical Models". Loughborough University. PhD Thesis.
- [5] Buswell, R.A., P. Haves and T.I. Salisbury. 1997. "A Model-Based Approach to the Commissioning of HVAC Systems,". Proceedings of CLIMA 2000, Brussels.
- [6] Rossi, T., M. and Braun, J., E. 1996. "Minimizing Operating Costs of Vapour Compression Equipment with Optimal Service Scheduling". Technical papers of IEA Annex 25.
- [7] Buswell, R. A and Wright, J. A. 1998. "Final Report on Methods for Condition Monitoring". Partners in Technology: The Practical Application of

Fault Detection and Diagnosis Techniques, Loughborough University final report contribution.

- [8] Holmes, M. J. 1982. "The Simulation of Heating and Cooling Coils for Performance Analysis". Proceedings of the first international conference on System Simulation in Buildings, Liege, Belgium.
- [9] Salsbury, T. I., Haves P., Wright J. A. 1995. "A Fault Detection and Diagnosis Method based on First Principles Models and Expert Rules". Proceeding of Tsinghua HVAC-95, Beijing, PRC.
- [10] Buswell, R., A., Haves, P., Dexter, A., L., Ngo, D. and Fargus, R. 1997. "Introducing faults into HVAC plant installed in real buildings". Presented at the Boulder, Colorado meeting of IEA Annex 34.
- [11] Box, M. J. 1965. "A New Method of Constrained Optimisation and a Comparison with Other Methods". The Computer Journal. Volume 8.
- [12] Visier, J., V., Vaesi-Nejad, H. and Corrales, P. 1999. "A Fault Detection Tool for School Buildings". ASHRAE Transactions. Volume 105, Part 1.
- [13] Fortescue, T., R., Kershenbaum L., S. and Ydstie, B., E. 1981. "Implementation of Self-Tuning Regulators with Variable Forgetting". Automatica. Volume 17. Number 6. Page 831.
- [14] Petitjean, R. 1994. "Total Hydronic Balancing". Tour & Anderson. ISBN 91-630-2626-0.
- [15] Buswell, R., A. 2001. "Uncertainty in the First Principle Model Based Condition Monitoring of HVAC Systems". Loughborough University. PhD Thesis.

- [16] Dexter, A and Pakenen, J (ed). “Demonstrating Automated Fault Detection and Diagnosis in Real Buildings”. Espoo. VTT Building Technology, 2001. VTT Symposium 217.
- [17] Buswell, R, A, Wright, J, A, Salsbury, T, I and Haves, P. “Fault Detection and Diagnosis Using Recursive Parameter Estimation in Non-Linear Systems”. Proceedings of System Simulation in Buildings 2002, Lège, Belgium, 2002.

Acknowledgement

This work was funded under the Partners in Technology scheme of the UK Department of Environment and was conducted in collaboration with the University of Oxford and the Building Research Establishment. T I Salsbury contributed to the development of the theoretical basis for this research and developed the initial version of the software tool.

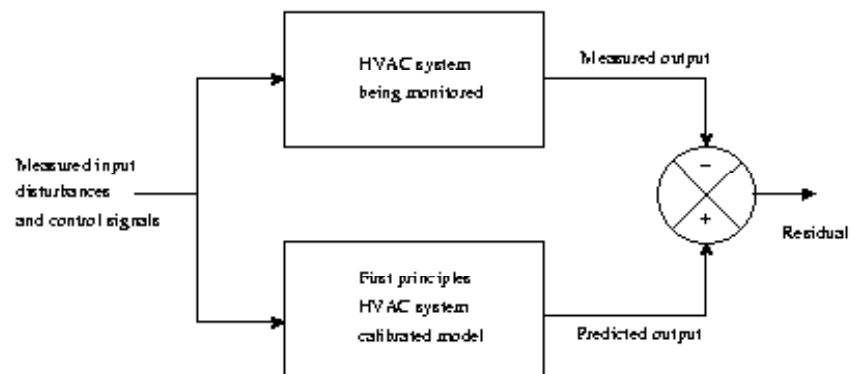


Figure 1: A model-based fault detection scheme.

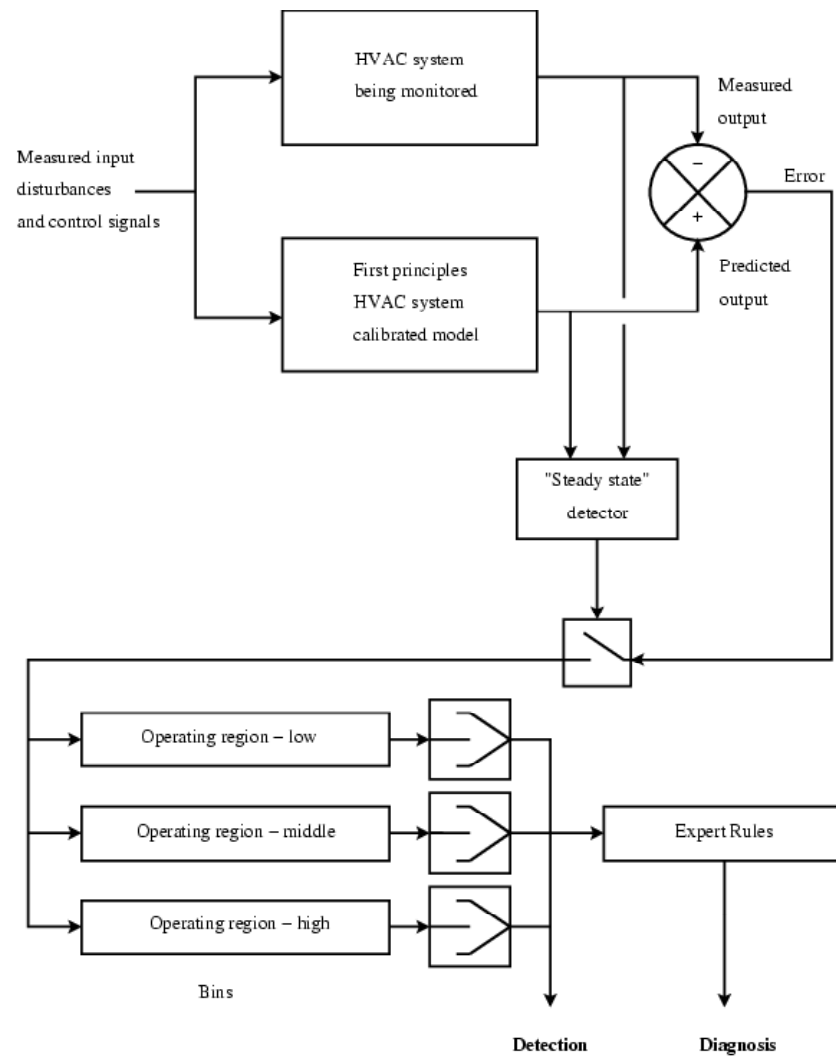


Figure 2: The expert rule fault diagnosis scheme.

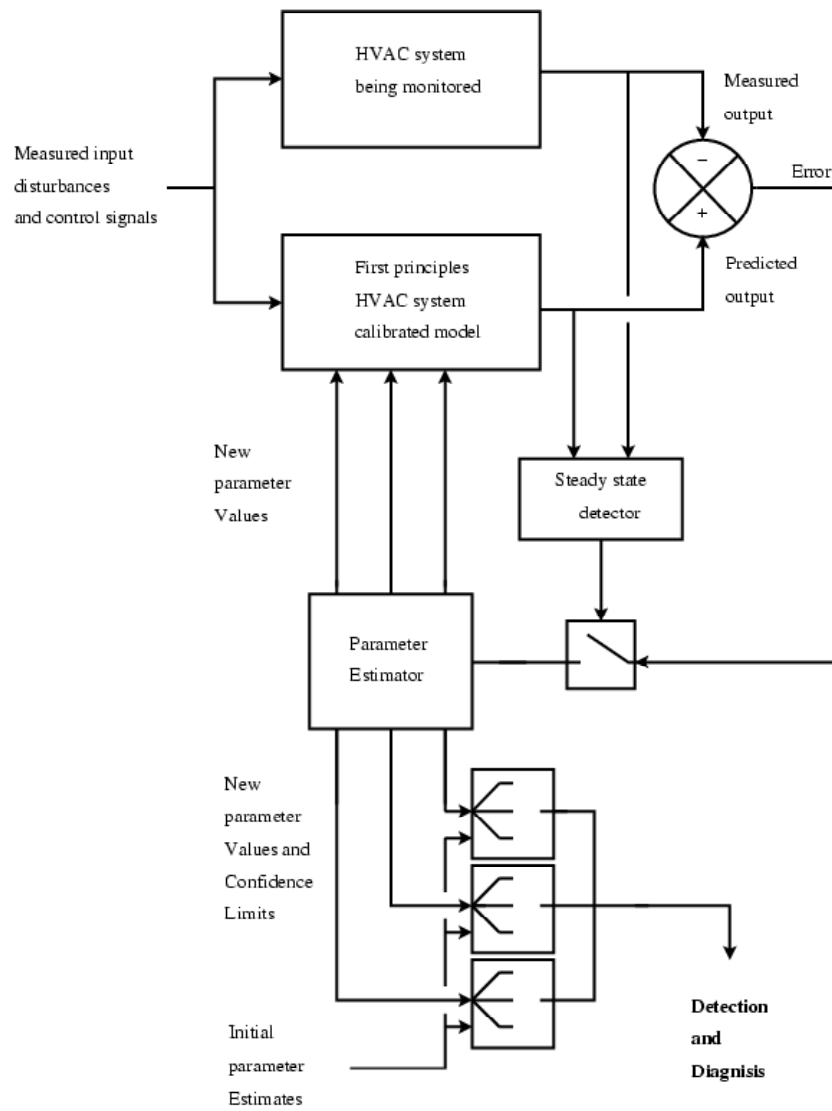


Figure 3: The recursive parameter estimation scheme.

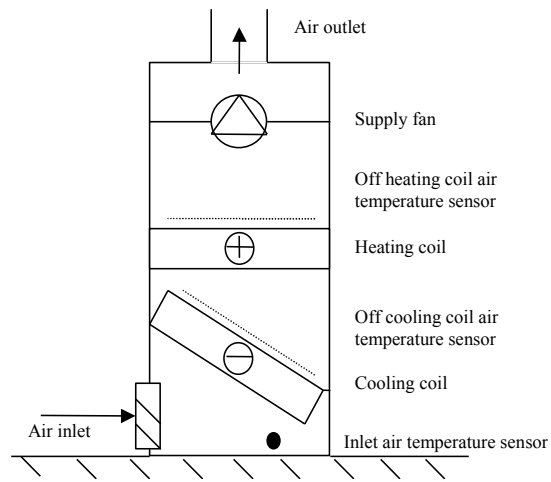


Figure 4: Section through the test AHU.

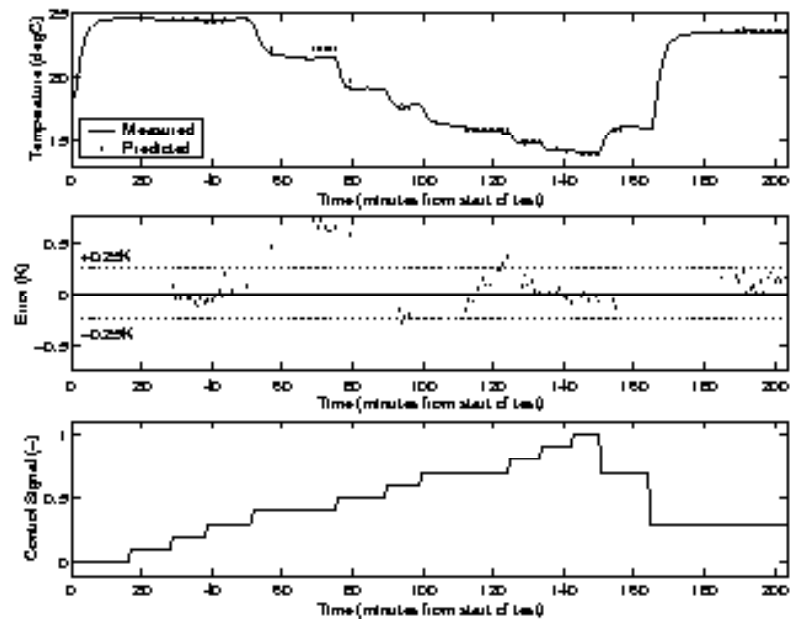


Figure 5: Model calibration tests.

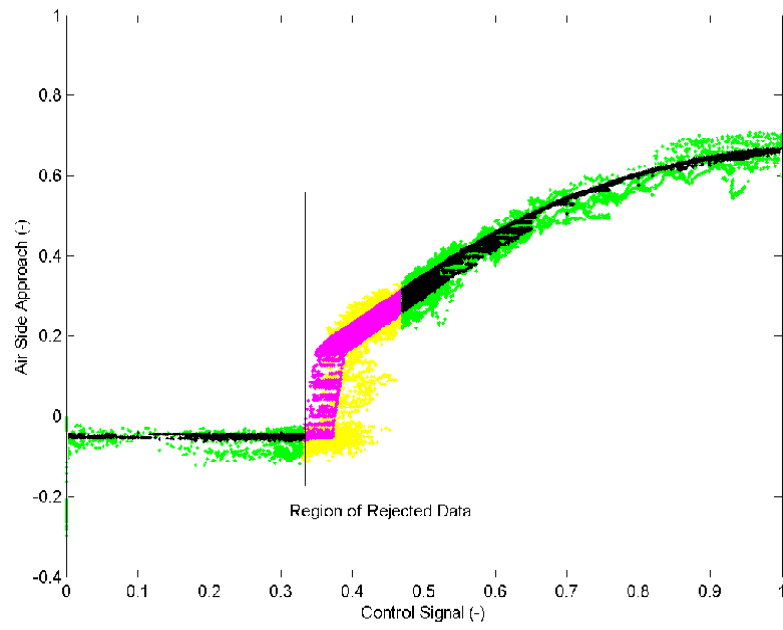


Figure 6: The airside approach characteristics for all the test data.

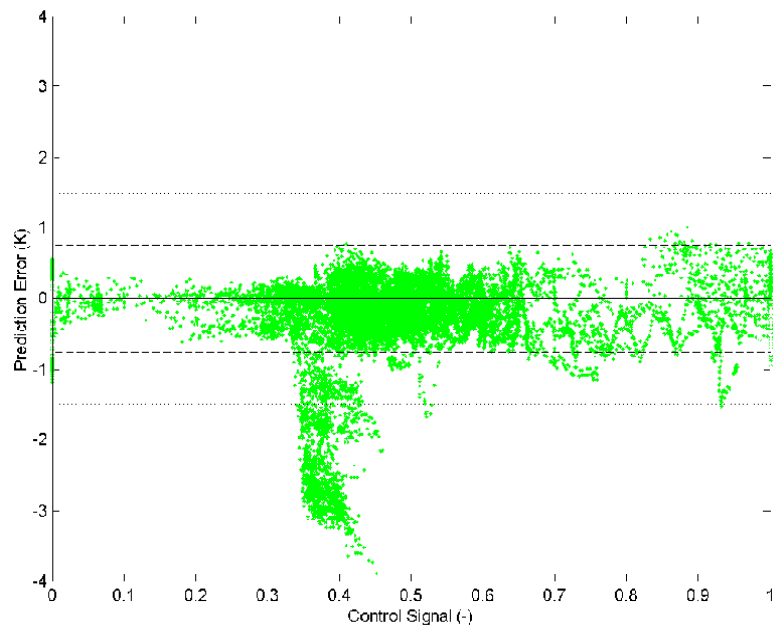


Figure 7: Prediction error for the ‘normal operation’ test data.

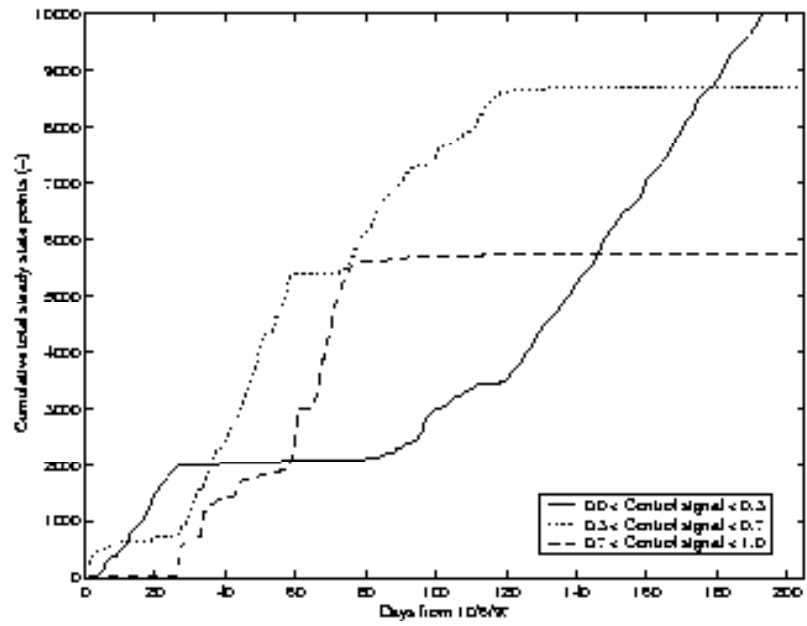


Figure 8: The cumulative number of steady-state data for the test period.

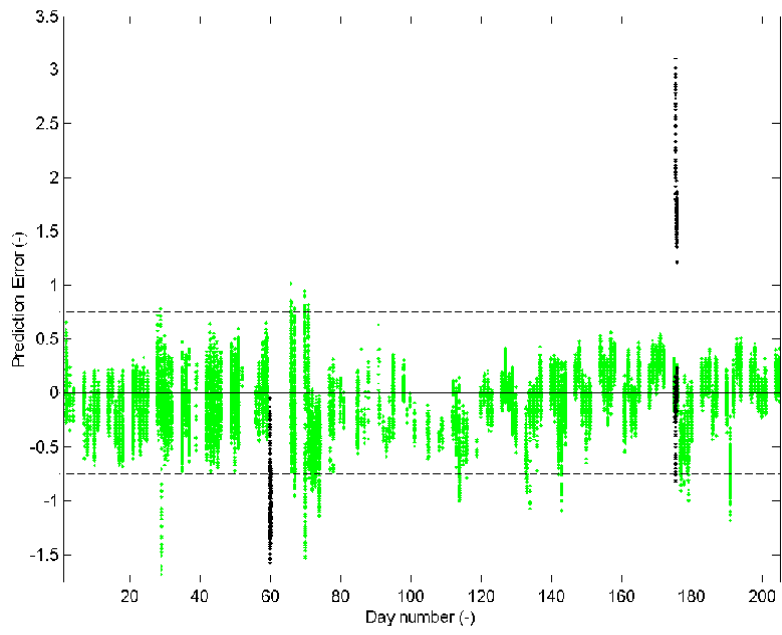


Figure 9: Prediction errors for the test period.

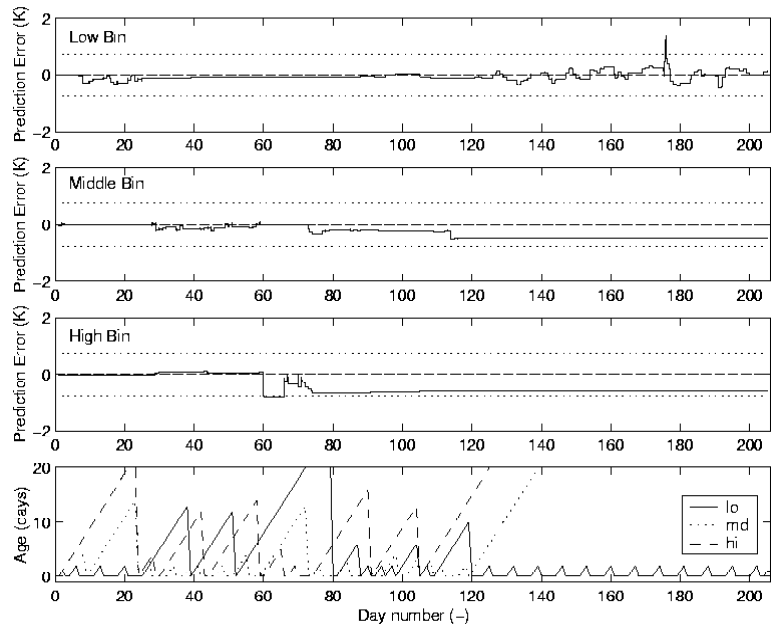


Figure 10: End of day bin values for the test period.

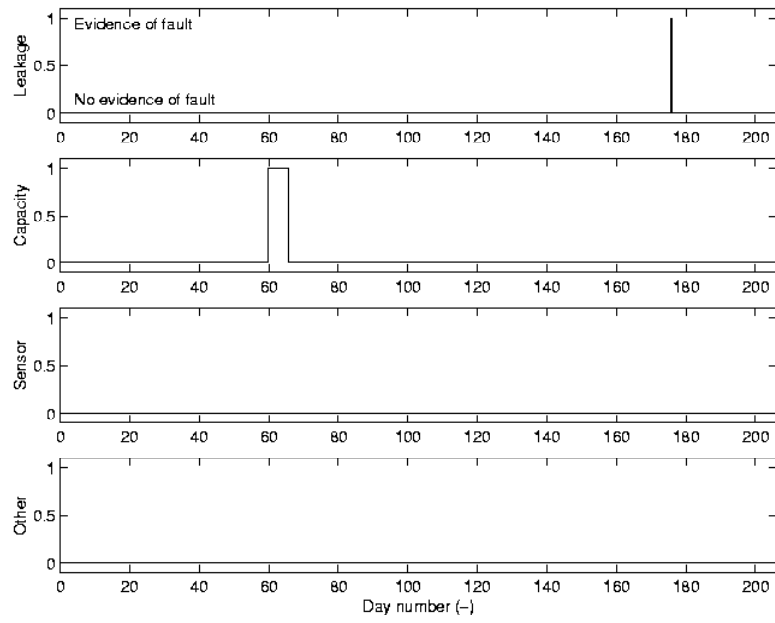


Figure 11: Fault evidence output for the bin method scheme.

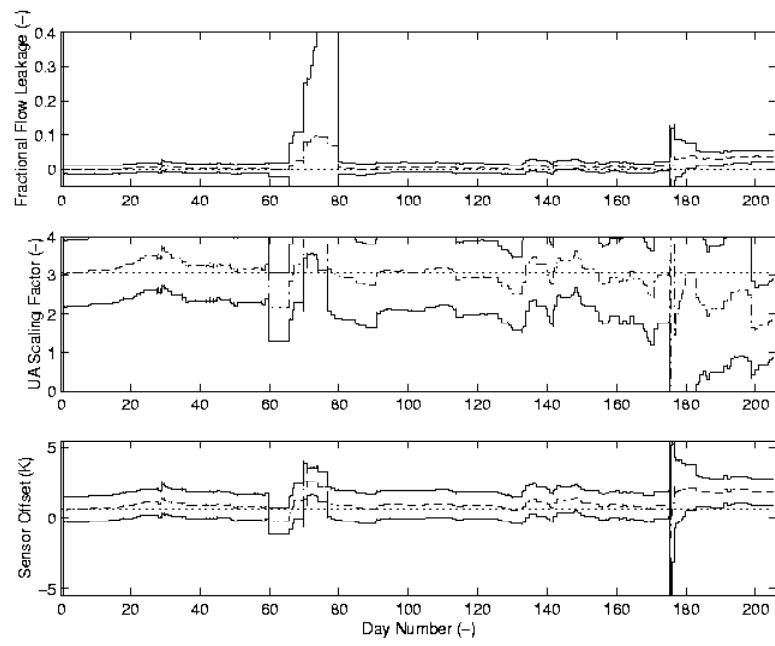


Figure 12: End of day parameter estimates for the test period.

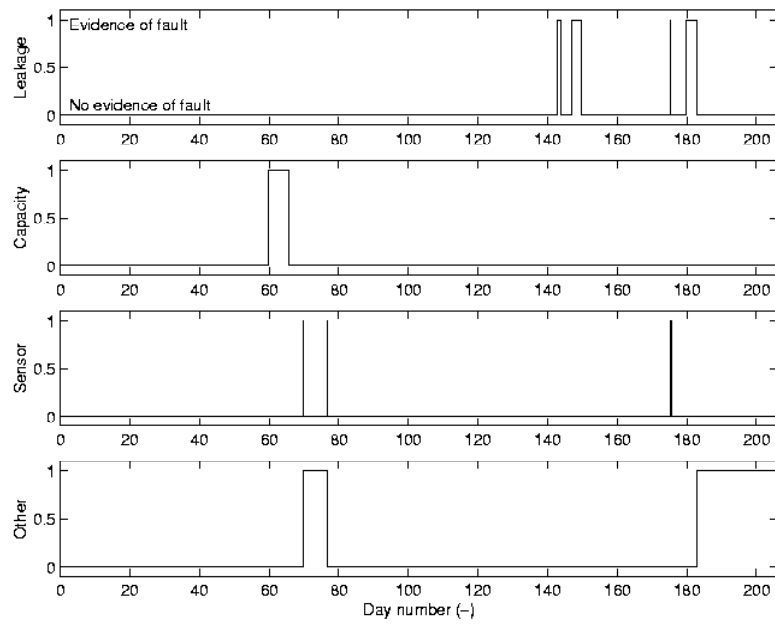


Figure 13: Fault evidence output from the RPE scheme.