

Spring 1-1-2011

Fault Detection and Diagnosis Using a Probabilistic Modeling Approach

Jordan Mann

University of Colorado at Boulder, jordan.mann@colorado.edu

Follow this and additional works at: https://scholar.colorado.edu/cven_gradetds



Part of the [Civil Engineering Commons](#)

Recommended Citation

Mann, Jordan, "Fault Detection and Diagnosis Using a Probabilistic Modeling Approach" (2011). *Civil Engineering Graduate Theses & Dissertations*. 301.

https://scholar.colorado.edu/cven_gradetds/301

This Thesis is brought to you for free and open access by Civil, Environmental, and Architectural Engineering at CU Scholar. It has been accepted for inclusion in Civil Engineering Graduate Theses & Dissertations by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.

FAULT DETECTION AND DIAGNOSIS USING A PROBABILISTIC
MODELING APPROACH

by

JORDAN MANN

B.S.M.E, University of Colorado Boulder, 2009

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Masters of Science
Department of Civil Engineering

2011

This thesis entitled:
Fault Detection and Diagnosis Using a Probabilistic Modeling Approach
written by Jordan Mann
has been approved for the Department of Civil Engineering

Prof. Gregor Henze, Ph.D., P.E.

Prof. Michael Brandemuehl, Ph.D., P.E.

Prof. David Bortz, Ph.D.

Prof. Moncef Krarti, Ph.D., P.E.

Date Oct 10th 2011

The final copy of this thesis has been examined by the signatories, and we
find that both the content and the form meet acceptable presentation
standards of scholarly work in the above mentioned discipline.

Mann, Jordan R. (M.S., Civil, Environmental and Architectural Engineering)
Fault Detection and Diagnosis Using a Probabilistic Modeling Approach
Thesis directed by Prof. Gregor Henze, PhD, P.E.

Abstract

The building sector in the U.S. accounts for approximately 40% of the national primary energy usage, 37% of which comes from space heating. Faulty systems and control schemes can increase energy usage by 15% to 20%. Quick and accurate fault detection and diagnosis will play a key role in reducing building energy consumption.

This thesis explores potential Bayesian fault detection, diagnosis and correction methods in commercial buildings. The first experiment investigated fault correction in a test building. A test-building model was calibrated to measured data. It was then found, that by implementing a nighttime setback, energy savings of approximately 20 percent could be achieved.

Next, experiments were carried out using surrogate model data to investigate a number of hydraulic system faults, such as an inefficient boiler, high valve leakage, valves with high hysteresis and heat exchanger fouling. It was determined, that experimentally with surrogate data, Bayesian methods are effective for detecting hydraulic heating system faults.

Bayesian methods were then used to examine heat exchanger fouling for two heat exchangers in a test building using measured data. The amount of propagated model uncertainty and measurement noise present made fault detection in this case more difficult. The heat transfer values in the heat exchangers were not determined to be low enough to be considered significantly faulty.

Lastly, an experiment was carried out to test if the number of models created could be minimized. The models were correct in predicting faults in many circumstances, however mischaracterized faults were not uncommon.

Acknowledgments

This thesis would have not been possible without the tremendous amount of help I received from colleagues, co-workers and friends. First off, I would like to thank my advisor Gregor Henze, who gave me the opportunity and connections that made it possible to conduct this research in Germany. Dr. Henze answered many questions and was willing to have regular teleconference meetings despite the difficulties of meeting across time zones. He provided me with the building models necessary to complete my research and was always willing to share his knowledge.

While in Germany, I was fortunate enough to be able to work at the Fraunhofer Institute for Solar Energy Systems where I worked with a large number of talented and knowledgeable individuals. In particular, I would like to express my gratitude to Dirk Jacob for his hospitality and help, not only within the scope of the workplace but also inviting me into his home, helping me with visa issues and devoting many hours to discussion and critical analyses of my work; and to Sebastian Burhenne whose depth of knowledge on the building industry and great attitude made for a productive and enjoyable working environment.

I would also like to extend my thanks to Anthony Florita. The work and expertise he openly contributed made this project possible. He shared code related to Bayesian parameter estimation, which helped me to learn the procedure and methods necessary for Bayesian analyses.

I would like to thank the thesis committee comprised of Michael Brandemuehl, Moncef Krarti and David Bortz for their helpful examination of the research presented and the time they are willing to contribute to make this thesis as thorough and accurate as possible.

Last but not least I would like to thank my family for all of their support in my decision to work in Germany and the pursuit of my education. Without their understanding and love none of this would have been possible.

TABLE OF CONTENTS

1	Introduction.....	1
1.1	Project Background	1
1.2	Introduction	1
1.3	Thesis Organization	4
2	Literature Review	6
2.1	FDD Overview	6
2.2	Fault Detection in Buildings.....	11
2.3	Bayesian Fault Detection.....	13
3	Methodology	14
3.1	Model Development for an Example Building.....	15
3.2	Bayesian Fault Analysis	17
3.2.1	Boiler Model.....	17
3.2.2	Heat Exchanger Fouling	18
3.2.3	Hydraulic Network	19
4	Test Building Model Development and Analysis	20
4.1	Building Description.....	20
4.2	Zone Modeling	22
4.3	Plant Modeling	26
4.4	Calibration.....	28
5	Bayesian Analysis.....	30
5.1	Background	30
5.2	Boiler Model.....	33

5.2.1	Model Description	33
5.2.2	Single Parameter	34
5.2.3	Multi Parameter with a Single Measurement	36
5.2.4	Multi Parameter with Multiple Measurements.....	38
5.3	Heat Exchanger Fouling	39
5.3.1	Model Description	39
5.3.2	Uncertainty Analysis.....	41
5.3.3	Single Parameter Analysis.....	46
5.4	Hydraulic Network.....	47
5.4.1	Model Selection	47
5.4.2	Bayesian Model Selection Analysis	50
6	Results.....	53
6.1	Test Building.....	53
6.2	Bayesian Fault Detection.....	58
6.2.1	Boiler	58
6.2.2	Heat Exchanger	69
6.2.3	Hydraulic Network	81
7	Conclusions and Further Work.....	84
7.1	Conclusions	84
7.2	Future Work.....	86
8	Works Cited.....	89
9	Appendix a	92
9.1	Model Selection Matlab Code.....	92

Figure 1 Fault Detection Methods	7
Figure 2 Fault Diagnosis Methods	8
Figure 3 BZRD Isometric View	21
Figure 5 BZRD Photo	22
Figure 6 Equivalent R-C Circuit	23
Figure 7 Simulink Diagram of Boiler Model	33
Figure 8 Efficiency Beta Distribution	35
Figure 9 Valve Let By	37
Figure 10 BZRD heat exchanger (Schmidt 2011)	40
Figure 11 Sigma Comparison	42
Figure 12 SIMBAD (left) and Henze (right) Comparison for 6000 W Internal Gain	50
Figure 13 Simulated vs. Measured Heat Usage	54
Figure 14 Simulated and Measured Heat Use	55
Figure 15 Cumulative Heat Use	55
Figure 16 Heating with Nighttime Setoff	56
Figure 17 Cumulative Energy with Nighttime setoff	57
Figure 18 Inside and Ambient Temperatures	58
Figure 19 Boiler Efficiency of 0.3 With a Non-Uniform Prior	59
Figure 20 Boiler Efficiency of 0.5 with a Non-Uniform Prior	60

Figure 21 Boiler Efficiency of 0.9 with a Non-Uniform Prior	61
Figure 22 Boiler Efficiency of 0.8 with a Uniform Prior	62
Figure 23 Multiple Parameters with a Single Measurement	63
Figure 24 Multiple Parameters with a Single Measurement	64
Figure 25 Efficiency Prior Distribution	66
Figure 26 Multiple Parameters with a Single Measurement and Prior	67
Figure 27 Multiple Parameters with Multiple Measurements	68
Figure 28 Multiple Parameters with Multiple Measurements	69
Figure 29 Uncertainty Analyses for $T_{1, out}$	70
Figure 30 Uncertainty Analyses for $T_{2, out}$	71
Figure 31 Sigma Uncertainty Analysis	72
Figure 32 Parameter Estimation with Uniform Prior	73
Figure 33 Parameter Estimation with Updated Prior	74
Figure 34 Uncertainty and Determined k Values	75
Figure 35 k Value Without Updating Prior 1 st Heat Exchanger	76
Figure 36 k Value Without Updating Prior for 2 nd Heat Exchanger	77
Figure 37 k Value With Updating Prior 1 st Heat Exchanger	78
Figure 38 k Value With Updating Prior 2 nd Heat Exchanger	79
Figure 39 Model Selection for Linear Valve	82
Figure 40 Model Selection for Proportional Valve	83

Table 1 BZRD Location and Usage	20
Table 2 BZRD Building Description	21
Table 3 Fault Combinations	51

1 INTRODUCTION

1.1 Project Background

The research presented in this thesis was carried out almost entirely in Freiburg, Germany at the Fraunhofer Institute for Solar Energy Systems. While in Freiburg, I worked in collaboration for 6 months on the ModQS project, and this thesis is a result of that research. ModQS is a project with the main goal of energy reduction for heating systems of commercial buildings in Europe achieved by ongoing commissioning and monitoring of at least 10 commercial buildings.

1.2 Introduction

The building sector in the U.S. accounts for approximately 40% of the national primary energy consumption. Space heating makes up the largest portion of this energy usage, accounting for 37% of the building energy usage (Buildings Energy Data Book). It is clear that in order to reduce global energy consumption, a reduction of building sector heating energy usage is necessary. Smarter building design, better control strategies and more energy efficient products will all play a key role in energy reduction. This thesis will focus on

potential energy savings that can be achieved through the detection of heating system faults in commercial buildings.

Fault detection and diagnosis (FDD) is a crucial component for realizing potential energy savings in commercial buildings. Improper control schemes, faulty equipment or degraded equipment can increase energy usage by 15% to 30% energy usage (Katipamula and Brambley 2005A). By quickly detecting, diagnosing and repairing potential faults energy savings are significant. Another key issue when faults go undetected is a reduction of thermal comfort. If a heating or cooling system is not functioning properly and the building temperature is not within the thermal comfort zone, the occupants may have decreased productivity. A study from the Technical University of Denmark found that an increase of 5 Kelvin could result in a 5% reduction of productivity (Technical University of Denmark 2009). This makes fault detection crucial from a financial perspective both for energy savings and worker productivity.

Typically, faults in commercial buildings are detected using rule-based methods (Katipamula and Brambley, Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems- A Review, Part 2 2005B). With the use of monitored data from commercial buildings, a system is determined as faulty by passing data through a series of if-then statements or simply detecting when a measured value crosses a certain threshold. If a condition is satisfied, for example a low level of heat exchange in a heat

exchanger, then the system is flagged as faulty. When creating these rule-sets a tolerance must be included to account for measurement error. This process often results in late fault detection as systems often do not instantly change from a non-faulty to faulty state but rather degrade over time (Isermann 1984). This thesis presents an alternative method for fault detection in which Bayesian inference is applied to yield a probability of a system state. The faulty state is then presented as a probability, which, later, using decision theory, could give recommendations for system maintenance or repair based on fault probabilities as well as cost benefit analyses for a given action. Using Bayesian fault detection leads to a deeper understanding of the system state being analyzed since it does not give simply a binary yes or no to a faulty state but rather a probability. In addition, Bayesian methods may lead to a more automated approach that can be applied to a wider spectrum of buildings since a set of rules would not need to be created for each individual building.

This thesis applies Bayesian fault detection methods to models and test buildings and thus expands the amount of current test cases in the building industry. Thus far very few have applied Bayesian methods to building fault detection. The Bayesian techniques applied to buildings thus far focus on parameter estimation to determine uncertain parameters which describe building physics such as (Booth et al 2011). Parameter estimation can also be used to estimate parameters describing faults. First a baseline is

established to test these methods using surrogate data. These techniques are then also applied to a test building.

1.3 Thesis Organization

The following thesis is organized into seven chapters that aim at giving insight into potential energy savings that can be realized by fault detection, diagnosis and correction with a heavy emphasis on Bayesian fault detection.

After the introduction, a literature review is presented that provides information on current fault detection methods in buildings as well as Bayesian fault detection in both the building industry as well as applications of Bayesian fault detection in other industries.

Then, a methodology section describing the basic background and procedure for the fault detection, diagnosis and correction methods is given. In this section, only introductory principles are introduced, more in depth descriptions of each method are given later in the chapter relating to the corresponding method.

The fourth section details the processes and methods used to analyze the fault of a missing nighttime zone temperature setback. A model for a test building is developed and the procedure for model calibration and comparison to measured data is detailed.

Next, Bayesian fault detection methods are presented. This section contains a variety of potential fault detection methods as well as offering examples that use surrogate data. Once the methods are proven, effective

with surrogate data, a fault detection analysis for a heat exchanger using measured data is provided.

Section 6 presents the results obtained from the nighttime setback analyses as well as the Bayesian fault detection analyses. The success of different methods, including shortcomings for surrogate and measured data is provided here.

The last section is the conclusions and future work section. Here general conclusions for fault detection, diagnosis and correction are discussed. There is a wealth of potential future work that could be carried out and some of the opportunities for future work are presented here.

2 LITERATURE REVIEW

2.1 FDD Overview

The automation of fault detection started in the 1960's with analog controllers used for limit checking (Isermann, Fault-Diagnosis Systems 2006). Later in the 1970's, with the advent of micro-controllers, automated fault detection gained popularity since model based fault detection involving more complex algorithms were made possible (Isermann 2006). Peer reviewed papers on fault detection started appearing in the late 60's and early seventies such as (Jones 1973) which produced a dissertation on fault detection for the aerospace industry, a fault is defined here as any failure reducing a systems effectiveness. Other publications include (Cohn and Ott 1972) where an algorithm for detecting a single fault, where a fault is considered to be any anomaly between an input vector and output vector, is developed and then connected with an estimated cost for finding the fault. Since these early developments, a broad body of literature has been developed including full books on fault detection and diagnosis methods. One such book is (Isermann 2006) which gives a large overview of fault detection and tolerance specification. Another, (Chen and Patton 1999), offers a comprehensive text that attempts to create a unified framework for model

based fault detection. The application of fault detection and diagnosis with respect to building systems will be discussed further in the next section.

FDD can be separated into two categories; the first being the phase in which the fault is detected and the next being where the fault is diagnosed. From here techniques can be sorted into a variety of categories depending on how the fault is detected and diagnosed. Figure 1 and Figure 2 taken from (Isermann 2006) show ways into which these processes can be separated and offer an overview of detection and diagnosis possibilities. For many of these processes, faults can be discovered in real time or by analysis of past data, although real time fault detection is preferable so that immediate action to repair the fault can be taken.

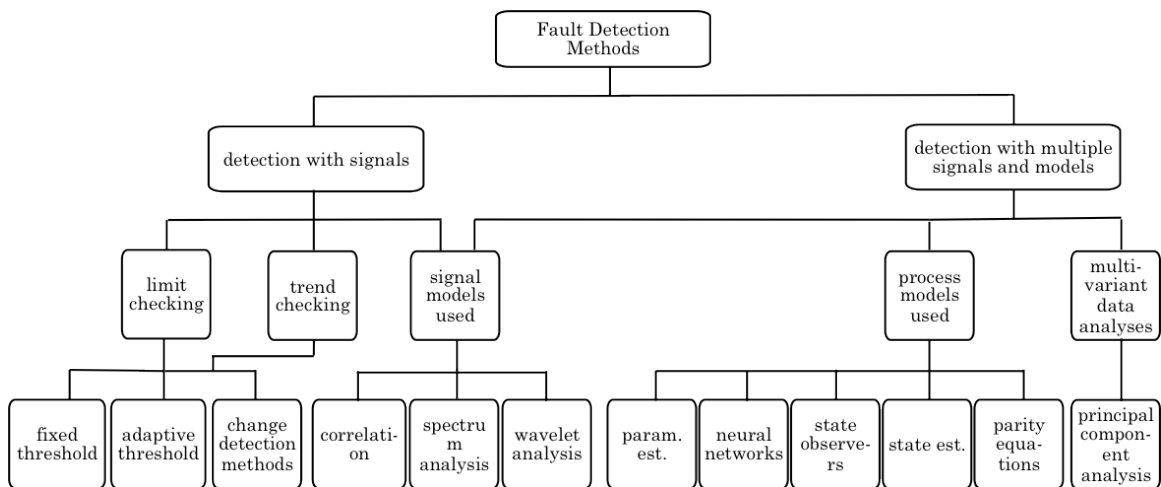


Figure 1 Fault Detection Methods

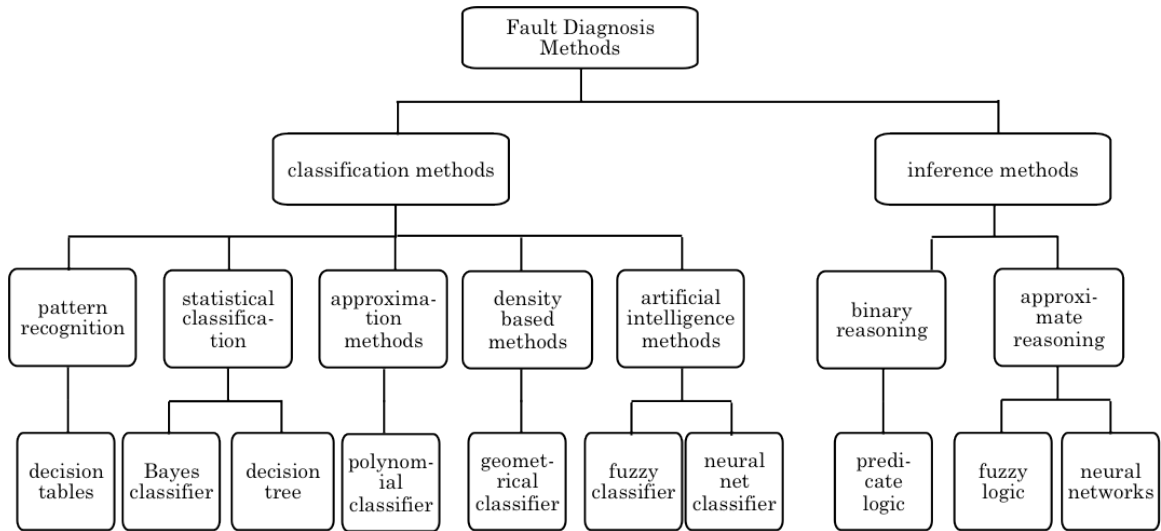


Figure 2 Fault Diagnosis Methods

For the process of fault detection, the most simple and common method is detection with single signals using limit checking and fixed thresholds, a method seen in Figure 1. The general process consists of measurement of a signal, then, when the measurement crosses a certain threshold an alarm is triggered, notifying relevant personnel, or some sort of automatic correction is implemented. The typical disadvantage of these systems are that they often detect faults late and that determining the correct threshold that specifies a fault can be difficult. For simpler applications where tolerances are not particularly crucial this may be a useful and simple method to detect faults.

Once a fault is detected the next step is diagnosis. This becomes exceedingly difficult when analyzing multiple faults, as the states in which the faults exist may overlap. The goal of fault diagnosis is to determine a fault with as much detail as possible; this could include the fault type,

severity, location or time. The typical method for fault diagnosis is to use prior knowledge of a physical system to reconstruct a system state based on measured data. As shown in Figure 2, diagnosis can be split into 2 general categories, classification and inference. Classification methods consist of comparing measured data to expected values. A common method of this is to use Bayesian or geometric classifiers that relate the difference between measured and expected values. Inference methods use fault trees or if-then statements to determine the likely cause of events. A review on fault diagnosis from (Leonhardt and Ayoubi 1997) compared classification, inference and hybrid systems. A typical issue found with classification methods is they require many data points. Inference systems can perform well for fault diagnosis but the binary nature of using decision trees or if-then statements often leads to delayed fault detection. The delayed reaction time is a result of the fault not being shown as an increasing probability but only being shown once a threshold has been completely crossed. With any fault diagnosis system, there is always the risk that an unexpected or underspecified fault may lead to data that is not previously considered by the fault detection system; therefore, for any fault diagnosis problem, specifying the faults that will be diagnosed is a challenge.

With the availability of inexpensive microprocessors, the ability to use models for FDD is common (Isermann, Fault-Diagnosis Systems 2006). By using models, it is possible to determine system states that are difficult or

expensive to measure, this leads to a larger amount of faults that can be detected. In model based fault detection, a model predicted value is typically compared to a measured value. If the measured value is outside an expected range then the system is considered faulty. The benefit of model-based systems is the ability to detect faults quicker and more precisely compared to simpler limit checking methods.

An example of model based FDD is given by an experiment from (Isermann 1984), where a leakage in a pipe is considered. With measurements of pressure and mass flow rate at the inlet and outlet of a pipe one could determine a leakage by simply comparing the in and out mass flow rate; when the outlet mass flow rate is smaller than the inlet mass flow rate it could be assumed a leak is present. However, because of dynamic effects and signal noise only large leaks of approximately 2% or more for liquids and 10% or more for gases can be found. Furthermore, even when a fault is detected, the location of the fault is unknown. However, with the use of a model, a fault signature in the residuals of the mass and momentum balances can be obtained. Using such a model the experiment carried out by (Isermann 1984) resulted in the ability to detect a leakage of only 0.2% for a liquid and 2% for a gas. For a liquid system, the location of the leak was found within 0.7% of the total pipe length. This example clearly demonstrates the power of model based fault detection.

2.2 Fault Detection in Buildings

Fault detection and diagnosis in the building sector, despite growing research efforts, is still an area where much work is needed. FDD has been actively researched in other fields such as aerospace, process controls, automotive industries, manufacturing, nuclear, and national defense but the building industry has lagged these other industries by about a decade (Katipamula and Brambley 2005A). Of the developments from the HVAC industry, a survey of building system fault detection methods conducted by (Katipamula and Brambley 2005A) found no fault detection system that was commercialized and widely used. Most of the current research remains in universities and laboratories and focuses on chillers, air handling units and refrigerators (Katipamula and Brambley 2005B). Of the methods being used in buildings, FDD expert rule based approaches are most common. Although Bayesian fault detection is used in other industries such as aerospace, a literature survey shows that HVAC applications using this method are undeveloped.

The most comprehensive review on FDD for building systems is a two-part study from (Katipamula and Brambley 2005A). In the first part of this review, major studies such as the Annex 25 study are presented (Hyvärinen and Kääri 1996). In the Annex 25 study, many major faults are listed for a wide array of systems, detection and diagnosis methods are discussed, and a background for FDD is presented The second part of this survey focuses on

research specific applications in the HVAC industry focusing on a large number of individual studies. Key findings from these reviews were, that although progress is being made, the state of current research doesn't allow for the development of commercial products. Main problems preventing commercial development are the cost of developing customized systems for each individual building, a lack of research around the decision support systems when a failure is detected, the high cost of reliable sensors and lack of methods for predictive fault detection as opposed to corrective fault detection.

Since this literature review performed by (Katipamula and Brambley 2005A), most studies for FDD in building systems focus on fault detection for AHU's such as (H. Wang et al. 2011; Du and Jin 2007b; Du, Jin, and Wu 2007; Du, Jin, and Yang 2009). Common analyzed faults include errors in sensor readings, damper failure, excessive air supply, high supply temperature, incorrectly sized components and inappropriate flow rates of water and air. Many current studies for fault detection also analyze the potential of sensor faults that can lead to non-functioning control schemes (Chen and Lan 2010; Du and Jin 2007a; S. Wang, Zhou, and Xiao 2010). Studies focusing primarily on heating system secondary circuits fault detection were not found; however, some of the studies that focus on AHU or cooling systems deal with similar principles such as incorrect outlet water

temperature cooling coil fouling or heating coil fouling (Liang and Du 2007; Lalot and Pálsson 2010; Persin and Tovornik 2005).

2.3 Bayesian Fault Detection

Fault detection using a Bayes classifier is one of the most well known fault classification schemes (Isermann 2006). Typically the likelihood of a certain parameter is determined by comparing a fault free state with measurement data. A Gaussian distribution is typically assumed in determining the likelihood of a measured point being outside of an acceptable range. From here, a probability of a fault given an observation can be determined. In a study by (House, Lee, and Shin 1999) it was found that for fault detection of AHU's a Bayesian classifier outperformed all other classifiers tested. However, for fault diagnosis all methods performed approximately the same. An experiment from (Xu and Tang 2011) analyzes the time to failure rate of industrial heaters and compares the effect of the prior on posterior distributions, finding that good choice of prior distributions outperformed the use of likelihood functions alone.

Although, not directly used for fault detection, (Booth et al 2011) uses Bayesian parameter estimation do determine uncertain building physics parameters. In this study uncertain parameters are given prior distributions then with the use of typical parameter estimation techniques posterior distributions for parameters such as internal heating set point, air leakage and window to wall ratio are determined. This results in a better model fit

when determining daily energy consumption than the pre-calibration model prediction.

Another paper that exemplifies the scalability of a Bayesian approach is a parameter estimation problem for non-domestic buildings in London (Choudhary 2011). This paper focuses using building stock models and other available information such as mapping databases and floor space statistics to estimate energy consumption. One could imagine a similar study being performed on similar but non-identical potential faults. With a rule based method unique rules would have to be developed for each non-unique system. With a Bayesian approach parameters describing a system could be estimated with measurement data. Then potential faults in this system could be determined.

Currently, there is a lack of application of Bayesian methods to the building sector. Quick and accurate fault detection can lead to large energy savings, equipment failure avoidance and provide the additional incentive of increased worker productivity due to room temperature remaining within an acceptable thermal comfort range. The goal of this thesis is to further develop and test Bayesian inference methods for the building environment

3 METHODOLOGY

3.1 Model Development for an Example Building

The focus of this research in terms of experimental data source and validation lies on a government building in Germany. The purpose of investigating the Bezirksregierung Düsseldorf (BZRD) is to analyze the consequences of the fault of missing nighttime and weekend temperature setback on the heating temperature setpoint. The procedure for analyzing this fault will include development of a building model, calibration of the latter model using Bayesian parameter estimation, and finally an analysis of the potential savings available by implementing a nighttime setback.

The building model developed was created in the Matlab Simulink environment. The room is modeled as a thermal network model using two capacitors and three resistors for each wall. The model inputs are solar radiation, outside air temperature and temperature of hot water for the circuit, which produce the outputs; resulting zone temperature and energy consumption. The model constructed for this experiment consists of a single zone. This is adequate model complexity to estimate total potential nighttime and weekend temperature setback energy saving potential.

The method used for determining uncertain parameters for this building model is a Monte Carlo simulation with Bayesian parameter estimation, again using the Matlab Simulink environment. The simulation is run 5000 times, each time with a different set of input parameters. The varied parameters chosen included air change rate, thermal capacity of walls, thermal capacity of floors, thermal resistance of walls and thermal resistance of windows. For each simulation, each parameter value is chosen from a normal distribution for that parameter with a mean value of the expected value and a standard deviation that spans the most likely values for that parameter. After running all 5000 simulations with these parameter sets, code developed by Anthony Florita was used to post-process this data and determine the most likely parameter values by comparing the residuals of the measurement values to the model output for monthly data. Following this comparison, the best set of parameters was chosen for further analysis.

To analyze the available savings, the model with the best matching parameters was used for a yearly simulation. The simulation was run with nighttime, nighttime plus weekend setbacks and with no setbacks. The model predicted energy is validated with the actual energy consumption and the overall energy savings with setbacks is analyzed.

3.2 Bayesian Fault Analysis

To analyze a variety of potential building faults using the Bayesian method, a group of three models were constructed: a model for boiler energy consumption and a leaky valve, one for the heat exchanger from the BZRD and a model of a more complex hydronic network. Different Bayesian inference methods are developed and explored for each of these three models.

3.2.1 Boiler Model

The first Bayesian fault detection is the analysis of boiler efficiency in a simple hydraulic network. The model consists of a single zone and an HVAC plant to distribute heating energy. The zone modeled is for demonstrative purposes only and is therefore modeled as a simple single lumped resistance capacitance zone. The zone has internal heat gains and heat losses to the environment through building surfaces. In this model the heat from the plant into the zone is calculated over a one-week period; the overall heat consumption is determined by changing the boiler efficiency. To test these methods, surrogate data was first produced for a given boiler efficiency. The most likely boiler efficiency was then calculated by running Monte Carlo simulations as part of Bayesian parameter estimation. In addition to calculating the most likely boiler efficiency, a probability distribution can be generated showing the certainty of the estimate. The probability distribution is then used to determine the likelihood of the system being in a faulty state.

After analyzing a single fault, multiple faults are analyzed. The addition of a potentially leaky valve is added to the possible faults. The leakiness of the valve is represented by the let-by in the valve, which is the amount of flow that passes when the valve is in the closed position. Valve let-by also has an effect on energy consumption since the room is still heated even when the valve should be closed. This results in both an overheating of the room and additional energy consumption. The two faults are first compared to one measurement series, thermal power, and then to two measurement series, thermal power and hot water mass flow rate.

3.2.2 Heat Exchanger Fouling

After validating the Bayesian methods for fault detection using surrogate data, an experiment using the model of a real system with measured data was performed. The BZRD heat exchanger model was adapted from Max Schmidt (Schmidt 2011). Again, by running Monte Carlo based Bayesian analysis techniques, the heat transfer coefficient for the heat exchanger is analyzed. When the value for heat transfer degrades to a value that is lower than acceptable, then the heat exchanger is experiencing fouling and can be considered to be faulty.

For this portion of the analysis, a closer look at the uncertainty is performed. To determine the model uncertainty an analytical solution for uncertainty is calculated. Then, a Monte Carlo model output uncertainty analysis is performed. This is done to test whether model uncertainty can be

determined using a Monte Carlo approach when an analytical solution is no longer possible.

3.2.3 Hydraulic Network

Finally, fault detection of a more complex six-zone hydraulic network is analyzed. First, two separate modeling methods are compared and cross-validated. Then the faster of the two approaches is used for further analyses. Sets of fault combinations are detected using a model selection approach. For model selection, as few models as are determined necessary to accurately detect faults are created. Each model corresponds to a fault combination and the most likely fault combination is determined using Bayesian analyses. In this way the most likely fault set is determined instead of the most likely parameter set. So instead of a valve let by, for example, being determined as a discrete value, it is instead classified as faulty or non faulty.

4 TEST BUILDING MODEL DEVELOPMENT AND ANALYSIS

4.1 Building Description

The first issue investigated for the Bezirksregierung Düsseldorf (BZRD) building is the potential energy savings and feasibility of a nighttime setback. A nighttime setback for the BZRD would be achieved by turning off the heating system. Since users in individual offices control the heating with manually controlled radiators, it is not practical to do a temperature setback by adjusting the room setpoint temperatures. To achieve a nighttime setback either the pump must be shut off or the heating curve for the system must be adjusted.

The building chosen to model is the BZRD, a government building and ModQS demo building located in Düsseldorf, Germany. The building is approximately 16 meters tall and has a footprint of 7,000 square meters. The heating system is a hot water system using district heating as a heat source. Table 1 and Table 2 give further information on the building characteristics.

Table 1 BZRD Location and Usage

Location, Climate and Building Use	
City	Düsseldorf

Latitude	51°13'
Longitude	6°46'
Elevation	38 m
Heating days	269 d
Heating degree days	3058 Kd
Average Temperature	8.8°C
Monthly average max Temperature	12.9°C
Monthly average min Temperature	6.4°C
Use	Office
Occupied hours	8:00am to 6:00pm
Offices	480 people

Table 2 BZRD

Building

Description

Error!

Building Physics and Heating	
Floors	5
Footprint	7000 m ²
Total floor area	19500 m ²
Shading	Venetian Blinds, Shutters
Blind position	Outside
Blind operation	Manuel
Outer wall U value	1 W/m ² K
Inside wall U value	1.31 W/m ² K
Roof U value	1.6 W/m ² K
Window U value	2.3 W/m ² K
Energy	District Heating
Heat generation	Gas
Heat transfer	Radiators

Reference source not found.

shows a model of the buildings footprint and isometric view, Figure 4 shows a photo of the building.

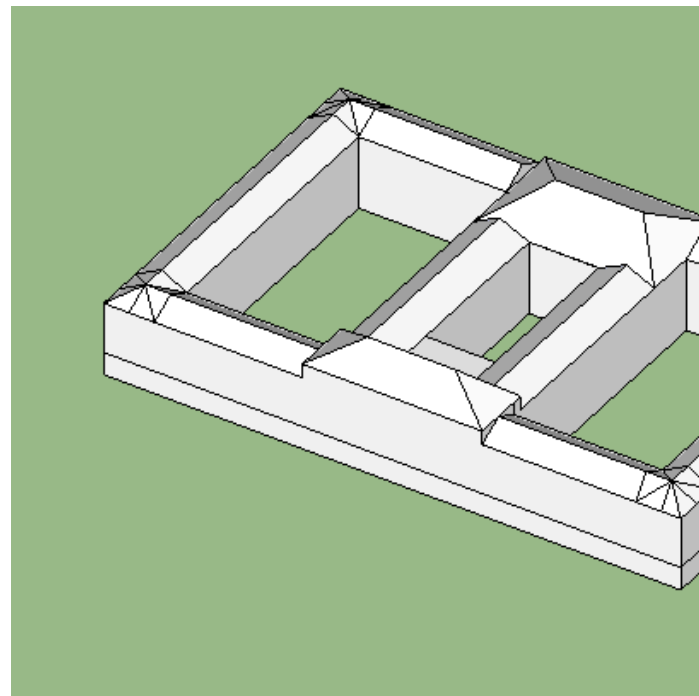


Figure 3 BZRD Isometric View



Figure 4 BZRD Photo

4.2 Zone Modeling

To analyze the possibility of a nighttime setback, a simple single-zone model was constructed in Matlab. This model analyzes a winter month in Düsseldorf. District heating water temperature, ambient temperature and solar data are measured; the resulting heating energy needed and indoor temperature can be calculated.

The zone constructed is modeled as a 2nd order resistance capacitance equivalent circuit, adapted from a model developed by Gregor Henze. Solar gains are calculated using a simple transmission model and processing of

solar data from a customized solar function that translates global horizontal radiation into components striking a particular surface.

Figure 5 (Henze) shows the equivalent RC circuit representing a construction element in the zone such as a wall, ceiling, floor or internal partition.

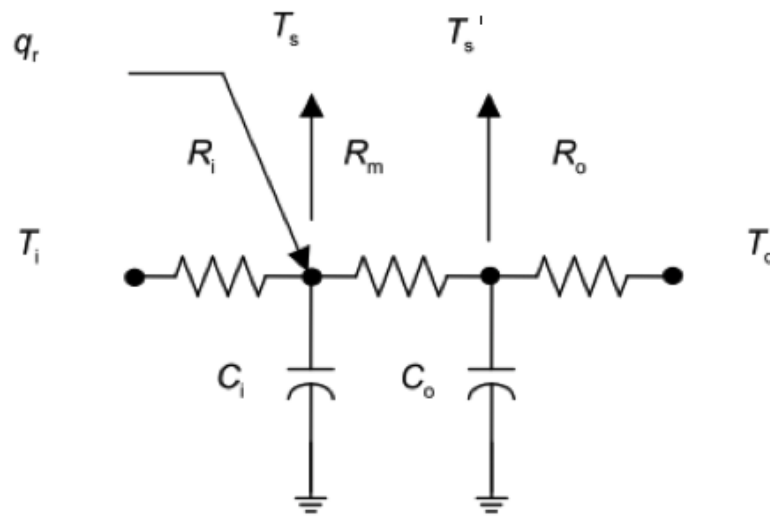


Figure 5 Equivalent R-C Circuit

The percentage of each resistance and capacitance element from the total resistance (R_T) is shown below. The chosen percentages are typical values when describing construction elements (Henze).

$$R_i = 0.1R_T; R_m = 0.4R_T; R_o = 0.5R_T$$

$$C_i = 0.15C_T; C_o = 0.85C_T$$

By solving the energy balances at the nodes it is possible to solve for the internal temperature values. The state variable derivatives \dot{T}_s and \dot{T}'_s are described in equations [4-1] and [4-2].

$$\dot{T}_s = \frac{q_r}{0.15C_T} + \frac{T_i}{0.015R_T C_T} - \frac{0.075T_s}{0.0009R_T C_T} + \frac{T'_s}{0.06R_T C_T} \quad [4-1]$$

$$\dot{T}'_s = \frac{T_s}{0.34R_T C_T} - \frac{0.765T'_s}{0.1445R_T C_T} + \frac{T_o}{0.425R_T C_T} \quad [4-2]$$

It is often useful to express equations in state space notation when using Matlab Simulink. State space notation allows the equations to be expressed by matrices. By arranging these equations in state space notation the output variables T_s and T'_s are solved for in the Matlab Simulink environment. With the construction elements fully defined it is now possible to determine the inside surface temperature given an outside temperature, inside temperature and incident solar radiation as input variables. The initial condition for the room temperature is also needed and is fixed as the setpoint for the control system.

To determine the temperature inside the room it is necessary to perform an energy balance that includes the heat transfer through construction elements as previously developed as well as internal gains, solar transmission gains through windows, infiltration, and heating gains from the HVAC system.

The solar transmission into the room, \dot{q}_r , is given in equation [4-3] where A_g , τ_g , and q_{r-surf} , are the surface area of the glass, transmissivity and incident solar radiation respectively.

$$\dot{Q}_r = A_g \tau_g \dot{Q}_{r-surf} \quad [4-3]$$

The infiltration into the space is determined by equation [4-4] where N , the total air change rate, is influenced by N_s (air change per hour), the standard air change rate, and v_w , T_o and T_i which are the wind speed velocity (m/s), external temperature and internal temperature (degrees C) respectively. Since the BZRD also has operable windows, an additional term was added to equation [4-4] that increased the total air change rate during occupied hours.

$$N = \frac{N_s [0.15 + (0.013)(2.2369)v_w + (0.005)(1.8) \times |T_o - T_i|]}{0.695} \quad [4-4]$$

The air temperature rate of change can now be determined by performing an energy balance as described in equation [4-5] where C_i is the thermal capacity of the air in the room, \dot{q}_{cc} is the convective heat gain, \dot{q}_{cp} is the convective plant heat inputs which will be further described in section 4.3, A is surface area, T is temperature, R_T is overall thermal resistance, V is the room volume, and U is the U-value. The subscripts ew, iw, f, r, g and o represent the external wall, internal wall, floor, roof, glass and external respectively.

$$C_i \frac{dT_i}{dt} = q_{cc} + q_{cp} - \left[\frac{A_{ew}(T_i - T_{sew})}{0.1RT_{ew}} + \frac{A_{iw}(T_i - T_{siw})}{0.1RT_{iw}} + \frac{A_f(T_i - T_{sf})}{0.1RT_f} + \dots \right. \\ \left. \dots \frac{A_r(T_i - T_{sr})}{0.1RT_r} + \frac{NV(T_i - T_o)}{3} + A_g U_g (T_i - T_o) \right] \quad [4-5]$$

4.3 Plant Modeling

To model heating input into the zone, a third-order model is used to model the heat emitted from a hot water radiator. In this approach, the heat exchanger is analyzed in three lumped capacitance segments that are connected in series. The heat transfer in each section is related to the difference between water and air temperature raised to a constant exponent, which for convective radiators is chosen as 1.3. Equations [4-6], [4-7] and [4-8] describe the energy balance at each of the three nodes. C_w , m_w , c_p , K and n refer to the overall water and material thermal capacity, flow rate of water, thermal capacity of water, emission constant, and emission index. T_{wi} , T_{w1} , T_{w2} , T_{wr} , are the inlet water temperatures for input, section one, section two and section three respectively, the water temperature at section three is the output water temperature. T_i is the internal zone air temperature.

$$C_w \frac{dT_{w1}}{dt} = 3m_w c_p (T_{wi} - T_{w1}) - K(T_{w1} - T_i)^n \quad [4-6]$$

$$C_w \frac{dT_{w2}}{dt} = 3m_w c_p (T_{w1} - T_{w2}) - K(T_{w2} - T_i)^n \quad [4-7]$$

$$C_w \frac{dT_{wr}}{dt} = 3m_w c_p (T_{w2} - T_{wr}) - K(T_{wr} - T_i)^n \quad [4-8]$$

In order to determine the mass flow rate a control system that measures internal temperature and gives a response signal to a valve actuator is needed. A temperature sensor in the room is a first order model having a rate of change determined by a time constant. The controller is then modeled using Matlab's built in proportional-integral-derivative controller. The output signal from the controller is sent to the valve having characteristics that result in a mass flow m_w described by equation [4-9]. m_{wd} is the design mass flow, f_o is the let by, p is the stem position, and A is the valve authority.

$$m_w = m_{wd} f_o^{(1-p)} \left[(f_o^{(1-p)})^2 (1-A) + A \right]^{\frac{1}{2}} \quad [4-9]$$

The last term needed to determine the room air temperature is the heat input from the HVAC plant. The heat output of the heat exchanger is shown in equation [4-10] where the water temperature at each node is solved for using a Simulink integrator block to integrate equations [4-6], [4-7] and [4-8].

$$q_{cp} = \frac{K}{3} \left[(T_{w1} - T_i)^n + (T_{w2} - T_i)^n + (T_{wr} - T_i)^n \right] \quad [4-10]$$

With all of the zone components adequately described, simulation of resulting zone temperature and heat use is possible with measured input data for input water temperature, outside air temperature and solar radiation.

4.4 Calibration

In order to calibrate the model, a combination of manual adjustment, given building construction information, and Bayesian parameter inference was used. Resistances of walls and windows were taken from construction data sheets for the building. Other parameters such as air change rate and thermal capacitance were adjusted manually or with parameter estimation techniques to try to minimize the root mean square difference between the measured and simulated data.

Since measurements for inside air temperature were unavailable, the model was calibrated to measured heating input into the zone. Measurements of supply temperature, return temperature and mass flow for the main circuit allow for the total heat into the building to be easily calculated, these values are then compared instantaneously and cumulatively to the model output.

At first, parameters were adjusted from expected values to reduce the root mean square of the measurement and model differences. It was then decided that a more sophisticated approach might lead to better model parameter classification. To better estimate uncertain parameters, code developed by Anthony Florita for Bayesian parameter estimation was applied to the existing zone model. The zone simulation was run 5000 times using Monte Carlo techniques, which sampled five parameters, each from a normal distribution. Then, each of these simulations is compared to the

measurement and the best resulting parameter set can be used for further analysis.

5 BAYESIAN ANALYSIS

5.1 Background

Bayesian analysis is used for fault detection in this thesis by the method of parameter estimation. The idea behind Bayesian statistical methods is to combine prior knowledge and current knowledge to determine the most likely state of a system. A Bayesian application that helps to exemplify its usage is the use of Bayesian statistics to examine if a person carries a certain disease. Let's assume that one out of every 5000 people is a carrier of the disease. Our prior assumption would be that a random person would have a $1/5000$ chance of being a carrier, or a 0.02% chance of being a carrier. Now let us assume the probability of a test correctly predicting the disease is 99.9% true positives and 99.9% true negatives. If we want to determine the probability of someone who has tested positive of having the disease we apply Bayes theorem. The prior is multiplied by the likelihood and then divided by all possibilities, the two possibilities is the person is a carrier and tested positive or the person is not a carrier and tested positive. The equation below shows how even someone who tested positive is still only 17% likely to be a carrier.

$$P(\text{Disease} | \text{Positive}) = \frac{\left(\frac{1}{5000}\right)(0.999)}{\left(\frac{1}{5000}\right)(0.999) + \left(\frac{4999}{5000}\right)(0.001)} = 17\% \quad [5-1]$$

The value describing the assumption for the system state is called the posterior, which is described by equation [5-2] for a continuous state or [5-3] for a discrete state. β is the parameter being examined and D is the data set. The posterior $p(\beta|D)$ is calculated as the probability for a system state based on prior information about the parameter $p(\beta)$, and the likelihood of the system state $p(D|\beta)$. The denominator is used to normalize the function so that all possible posteriors integrate to one in a continuous case or add to one in a discrete case.

$$p(\beta|D) = \frac{p(D|\beta)p(\beta)}{\int p(D|\beta)p(\beta)} \quad [5-2]$$

$$p(\beta|D) = \frac{p(D|\beta)p(\beta)}{\sum p(D|\beta)p(\beta)} \quad [5-3]$$

The posterior will always be a combination of prior assumptions as well as the likelihood function; thus, it is crucial to use correct prior information. When no or little prior information is available, an uninformative prior is used. An uninformative prior can still give information about the system; for example, that the number is positive or between certain intervals. A uniform prior is a type of uninformative prior in

which all possibilities between certain intervals are considered equally likely. An example of this could be the unknown efficiency of some system. If no information about the system is available a priori, then a good prior would be equal probabilities between the interval of 0 to 1.

The likelihood of a measurement given a parameter β is determined by equation [5-4] when assuming a normal model. This model determines how likely it is that a single measurement matches a model prediction. When determining likelihoods, it is useful to compute the likelihood of an entire time series instead of only a single point. In case the data points are independent of each other, the likelihood of a series of data points is the product of the likelihoods of each point in the time series where n is the number of data points [5-5].

$$p(D_i | \beta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(D_i - \mu)^2}{2\sigma^2}\right] \quad [5-4]$$

$$p(D | \beta) = \prod_i^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(D_i - \mu)^2}{2\sigma^2}\right] \quad [5-5]$$

The variable σ is the standard deviation of the normal distribution and was determined as the measurement and model uncertainty. Development of this variable is further discussed in section 5.3.2.

To combine the likelihood function as a product the error of measurement is considered to be independent. This helps to simplify math and is a fair assumption in most circumstances. Another form of measurement error is bias. Having a bias or auto correlated errors are also

treated much in the same way, and measurements are just adjusted according to there bias and likelihoods can be combined still as a product of individual likelihoods.

5.2 Boiler Model

5.2.1 Model Description

The zone and heat plant model used for this analysis are the same as those described in sections 4.2 and 4.3 respectively. Figure 6 shows a layout of the model in the Simulink environment.

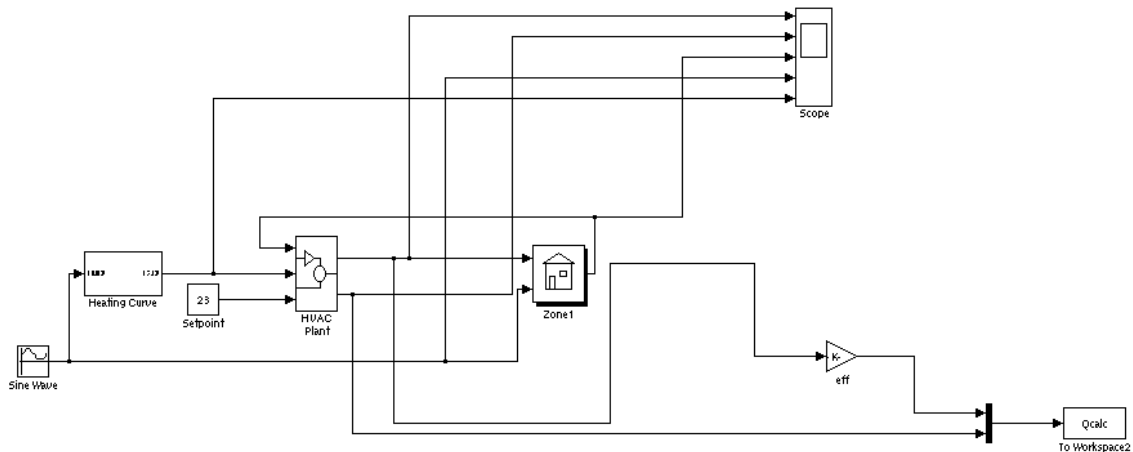


Figure 6 Simulink Diagram of Zone Model with a Boiler

This is the model at its most outer level and the plant, zone, and heating curve blocks can all be explored further.

5.2.2 *Single Parameter*

Consider a boiler supplying heat to a one-zone building. The efficiency of the boiler can be determined as the heat supplied to the zone, P_{out} , divided by the power consumption of the boiler, P_{in} as seen in equation [5-6].

$$\eta = \frac{P_{\text{out}}}{P_{\text{in}}} \quad [5-6]$$

Using the power input as a measured variable, a model can be constructed which compares the modeled power input and the measured power input as a function of the parameter η . To generate the “measured data”, a set of surrogate data is constructed by entering a known efficiency into the model. The model is then run with this efficiency and the value is determined using Bayesian parameter estimation.

The first step to estimating the parameter η is constructing a number of models with different η values, which are generated by performing a Monte Carlo simulation. It is assumed that the boiler is operating in accordance with the second law of thermodynamics and therefore the parameter space that is sampled is from zero to unity. It is also likely that the boiler is operating somewhere around normal conditions so it is best to have a higher sampling density around the more likely efficiencies. Other parameter values are also sampled, but by having a higher sampling rate near expected values the resolution can be increased. To perform the Monte Carlo

simulation, an efficiency is taken at random from the beta distribution shown in Figure 7.

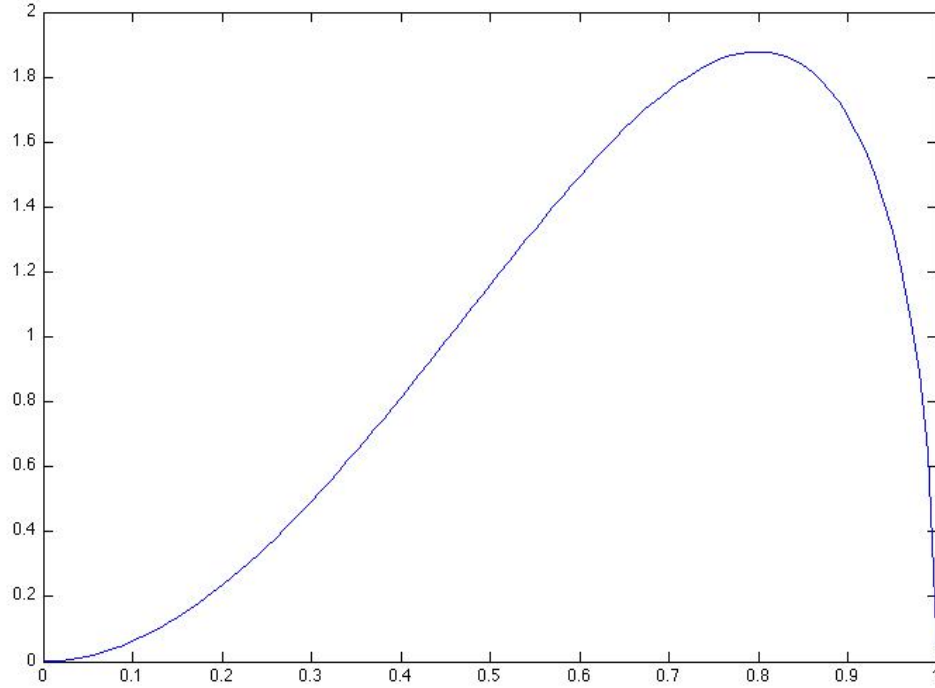


Figure 7 Efficiency Beta Distribution

This process of using a randomly chosen efficiency is carried out 100 times, resulting in 100 data sets containing power inputs for a one-week period for different values of η . The likelihood of the modeled η value is given in equation [5-7]. The prior needed for Bayesian inference is first tested as a normal distribution around an expected value then is assumed to be a uniform prior between the intervals 0 to 1. The posterior can be determined using equation [5-3].

$$p(D|\eta) = \prod_i^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(D_i - \mu_i)^2}{2\sigma^2}\right] \quad [5-7]$$

5.2.3 Multi Parameter with a Single Measurement

The next problem investigated deals with the potential for multiple faults within a system, still using the power output as the data source. The two potential faults analyzed here are a leaky valve and the efficiency of the boiler. Both faults will have an effect on the power consumption of the boiler. The leaky valve will result in an increase in energy consumption and overheating of the room since heated water will still be circulated when the valve should be closed. An inefficient boiler will also result in increased energy consumption. Investigating multiple faults can become difficult because of the possibility of tradeoff effects, i.e., an increase in power consumption could be a result of either of these faults. Since the Bayesian analysis is looking at many points in a discrete time period, it is still possible that the correct parameter values can be determined since a fault signature can be detected. The most closely fitting parameters should result in a response that matches the surrogate data.

Another issue that arises when detecting multiple faults is adequately defining the parameter space for the Monte Carlo simulation. More parameters result in a larger space that needs to be sampled, which means more Monte Carlo simulations and more comparisons between the generated

models and surrogate data are necessary. This inevitably leads to higher computational costs.

The procedure for multiple parameter estimation is the same as for a single parameter except that each data series is now connected to a pair of parameters η and K_v , the valve let by as a percentage of the design mass flow rate. The parameter space is now sampled with 1000 Monte Carlo runs. η is sampled the same as in the single fault example. The valve let by is sampled as a log normal distribution as shown in Figure 8. A log normal distribution was chosen, as the valve let by cannot be negative.

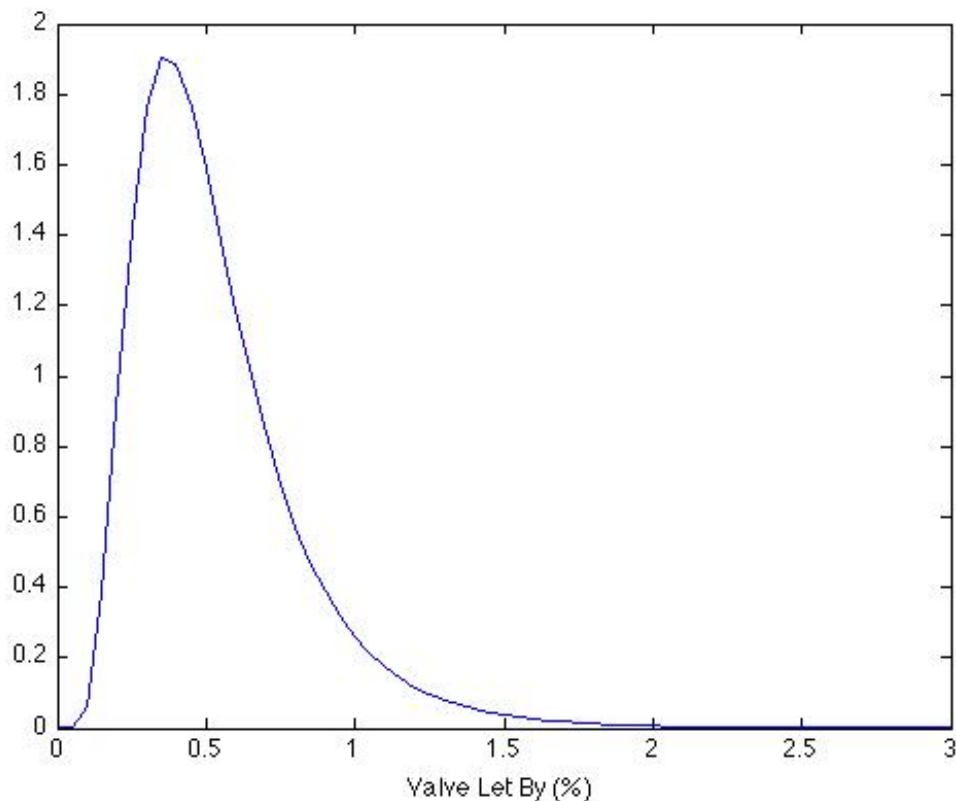


Figure 8 Valve Let By

5.2.4 Multi Parameter with Multiple Measurements

Since tradeoff effects often present themselves when examining multiple parameters, it is useful to compare the models to the surrogate data using multiple measurements. With more information available it becomes easier to define the system being investigated. Preferably the measurements being chosen should aim to isolate the investigated faults. For this example a good second measurement is the flow rate, as this isolates the valve let-by.

To perform Bayesian parameter estimation with multiple measurements, the analysis is first run comparing one of the measurement time series with model output. The posterior from this initial analysis becomes the prior for the proceeding analysis with the other measurement series. In this way all of the knowledge of the system can be used to estimate the faults that are being investigated. This could be repeated for any number of measurements and the more faults being investigated the more important this process of sequential prior updating will become.

The parameter space investigated for this problem was the same as in section 5.2.3, however, this time both power consumption and mass flow rate of water was recorded for both surrogate data and the model data, which results in 1000 data sets, each with two modeled predicted data series and two parameters.

5.3 Heat Exchanger Fouling

The next step with parameter estimation was to extend the previous developed techniques to estimate faults with measured data instead of surrogate data. There are two heat exchangers, one of which is suspected to have fouling. The goal is to estimate the heat transfer coefficient to determine if fouling is present.

5.3.1 Model Description

The heat exchanger model is a steady-state effectiveness NTU model. The term k , used as a measurement heat transfer, is defined as $1/U$ where U is the standard heat transfer coefficient. The value k is determined through equation [5-8] where m_2 is a mass flow rate, T is temperature, c_p is the specific heat of water, and ΔT_m is the log mean temperature difference [5-9].

$$k = \frac{c_p}{A\Delta T_m} (m_2 T_{2,out} - m_2 (T_{2,in} - T_{2,out})) \quad [5-8]$$

$$\Delta T_m = \frac{(T_{2,out} - T_{2,in}) - (T_{1,in} - T_{1,out})}{\ln \frac{(T_{1,out} - T_{2,in})}{(T_{1,in} - T_{2,out})}} \quad [5-9]$$

Figure 9 shows a diagram of the two heat exchangers in the BZRD, note that the subscript “ein” is in and the subscript “aus” is out.

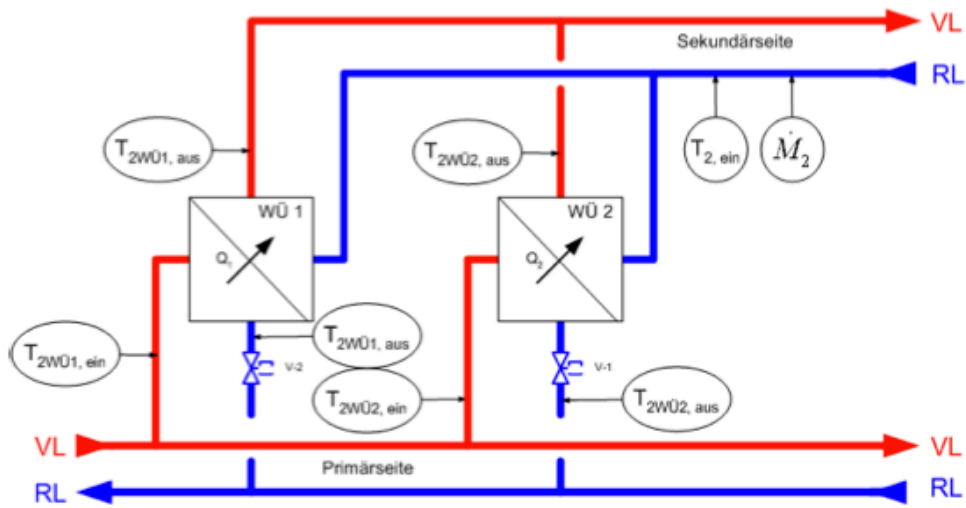


Figure 9 BZRD heat exchanger (Schmidt 2011)

The mass flow rate and inlet temperatures are known and used in equations [5-10] and [5-11] to solve for the outlet temperatures.

$$T_{2,\text{out}} = (T_{2,\text{in}} - T_{1,\text{in}}) \frac{1 - \frac{C_1}{C_2}}{1 - \frac{C_1}{C_2} \alpha} + T_{1,\text{in}} \quad [5-10]$$

$$T_{1,\text{out}} = (T_{2,\text{in}} - T_{1,\text{in}}) \frac{1 - \alpha}{1 - \frac{C_1}{C_2} \alpha} + T_{1,\text{in}} \quad [5-11]$$

The variables α , C_1 and C_2 are defined by equations [5-12], [5-13] and [5-14] respectively.

$$\alpha = \exp \left[\left(\frac{1}{C_2} - \frac{1}{C_1} \right) kA \right] \quad [5-12]$$

$$C_1 = c_p \dot{m}_1 \quad [5-13]$$

$$C_2 = c_p \dot{m}_2 \quad [5-14]$$

5.3.2 Uncertainty Analysis

An issue that has not yet received extensive attention is the determination of the parameter σ . At each point in a time series, the measured data is compared to model data. How closely the data matches the model is determined by the probability of the model being the “true” value. The certainty of the true value is based on a normal distribution around the true value. The width of the normal distribution is defined by the variance, σ . When σ is large, the distribution is wide, and certain model values are not particularly more likely than other model values even when they are far apart numerically. When the true value is more certain the distribution is narrow and small numerical value differences in the model can create large differences in probability. Figure 10 shows a potential comparison of different sigma values. The lines represent the distribution surrounding the measured true value of 22.5 degrees Celsius. The two points represent a model output. So far, the way in which the sigma epsilon value is determined is by the measurement error. Errors are considered to be distributed normally and without bias.

An issue that this creates is that model error is not considered. Measurement values are used as inputs for a model, these measurement values are not exact but normally distributed around a true value. Therefore model output contains these measurement values that are propagated through the model. When model output is compared to measurements it was

assumed previously to be a correct value when actually it is a value with uncertainty. Therefore, reported results are reported as more certain than they actually are when model uncertainty is not included.

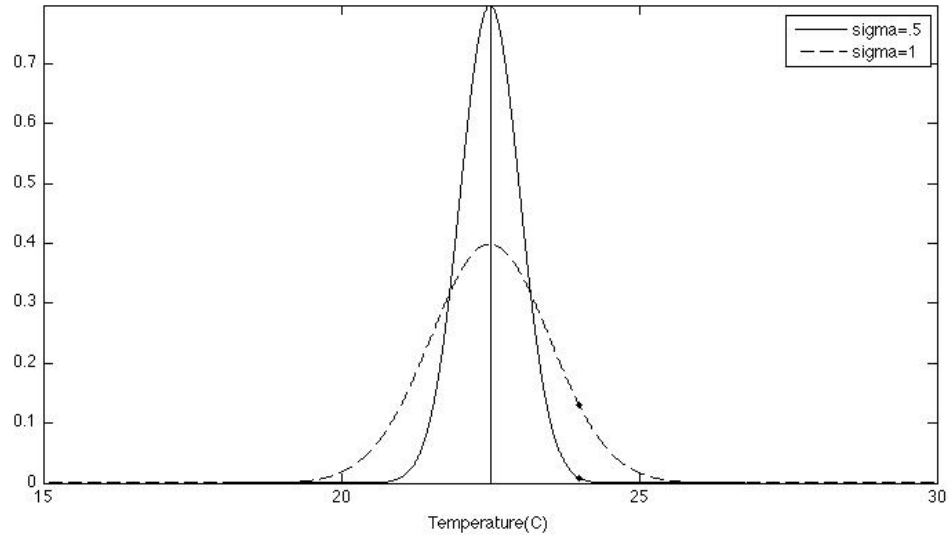


Figure 10 Certainty of Simulated Value Being the “True” Value

A proposed solution to this underestimated uncertainty is to determine the sigma value as a combination of model error and measurement error. There were two methods in which this revised uncertainty was calculated for the BZRD heat exchanger: Kline-McClintock uncertainty analysis and a Monte Carlo uncertainty analysis. The uncertainty in the model output in both of these analyses is the result of uncertainty propagation from the measurements rather than a result of the model structure or underlying physical assumptions.

For the Kline-McClintock uncertainty analysis, the partial derivatives are taken for both of the model outputs which were defined previously in equations [5-10] and [5-11]. The resulting partial derivatives for [5-10] are shown in equations [5-15] through [5-18].

$$\frac{\partial T_{2,out}}{T_{1,in}} = \frac{\dot{m} - \dot{m}_2}{\dot{m} - \exp\left(\frac{(\dot{m}_2 - \dot{m})UA}{c_p \dot{m} \dot{m}_2}\right) \dot{m}_2} \quad [5-15]$$

$$\frac{\partial T_{2,out}}{T_{2,in}} = \dot{m}_2 \dot{m} + \frac{\dot{m} \dot{m}_2 (\dot{m}_2 - \dot{m})}{\exp\left(\frac{(\dot{m} - \dot{m}_2)UA}{c_p \dot{m}_2 \dot{m}}\right) \dot{m} - \dot{m}_2} \quad [5-16]$$

$$\frac{\partial T_{2,out}}{\dot{m}} = \frac{\left(\exp\left(\frac{(\dot{m} - \dot{m}_2)UA}{c_p \dot{m} \dot{m}_2}\right) \dot{m}_2 (T_{1,in} - T_{2,in}) \left(c_p \dot{m}^2 \left(\exp\left(\frac{(\dot{m} - \dot{m}_2)UA}{c_p \dot{m} \dot{m}_2}\right) - 1 \right) \right) + (\dot{m}_2 - \dot{m})UA \right)}{\left(c_p \dot{m}^2 \left(-\exp\left(\frac{(\dot{m} - \dot{m}_2)UA}{c_p \dot{m} \dot{m}_2}\right) \dot{m} + \dot{m}_2 \right) \right)^2} \quad [5-17]$$

$$\frac{\partial T_{2,out}}{\dot{m}_2} = \frac{\left(\exp\left(\frac{(\dot{m} - \dot{m}_2)UA}{c_p \dot{m} \dot{m}_2}\right) (T_{1,in} - T_{2,in}) \left(c_p \dot{m} \dot{m}_2 \left(\exp\left(\frac{(\dot{m} - \dot{m}_2)UA}{c_p \dot{m} \dot{m}_2}\right) - 1 \right) + (\dot{m}_2 - \dot{m})UA \right) \right)}{\left(c_p \dot{m}_2^2 \left(-\exp\left(\frac{(\dot{m} - \dot{m}_2)UA}{c_p \dot{m} \dot{m}_2}\right) \dot{m} + \dot{m}_2 \right) \right)^2} \quad [5-18]$$

After taking the partial derivatives the uncertainty can be calculated with equation [5-19] using the measurement uncertainties for mass flow rate as well as temperature in combination with the partial derivatives.

$$\sigma_{T_{2,out}} = \sqrt{\left(\sigma_{T_{1,in}}\right)^2 \left(\frac{\partial T_{2,out}}{\partial T_{1,in}}\right)^2 + \left(\sigma_{T_{2,in}}\right)^2 \left(\frac{\partial T_{2,out}}{\partial T_{2,in}}\right)^2 + \left(\sigma_{\dot{m}}\right)^2 \left(\frac{\partial T_{2,out}}{\partial \dot{m}}\right)^2 + \left(\sigma_{\dot{m}_2}\right)^2 \left(\frac{\partial T_{2,out}}{\partial \dot{m}_2}\right)^2} \quad [5-19]$$

The uncertainty of $T_{1,out}$ is calculated in a similar fashion as described by equations [5-20] through [5-24].

$$\frac{\partial T_{1,out}}{\partial T_{1,in}} = 1 + \frac{\dot{m}_2 - \dot{m}}{\exp\left(\frac{(\dot{m} - \dot{m}_2)UA}{c_p \dot{m} \dot{m}_2}\right) \dot{m} - \dot{m}_2} \quad [5-20]$$

$$\frac{\partial T_{1,out}}{\partial T_{2,in}} = \frac{(\dot{m} - \dot{m}_2)}{\exp\left(\frac{(\dot{m} - \dot{m}_2)UA}{c_p \dot{m}_2 \dot{m}}\right) \dot{m} - \dot{m}_2} \quad [5-21]$$

$$\frac{\partial T_{1,out}}{\partial \dot{m}} = \frac{\left(\exp\left(-\frac{UA}{c_p \dot{m}}\right) \dot{m}_2 (T_{1,in} - T_{2,in}) \left(c_p \dot{m} \dot{m}_2 \left(\exp\left(\frac{UA}{c_p \dot{m}}\right)\right) - \exp\left(\frac{UA}{c_p \dot{m}_2}\right) \left(c_p \dot{m} \dot{m}_2 + (\dot{m}_2 - \dot{m})UA\right)\right)\right)}{\left(c_p \dot{m} \left(-\exp\left(\frac{(\dot{m} - \dot{m}_2)UA}{c_p \dot{m} \dot{m}_2}\right) \dot{m} + \dot{m}_2\right)\right)^2} \quad [5-22]$$

$$\frac{\partial T_{1,out}}{\dot{m}_2} = - \frac{\left(\exp\left(\frac{UA}{c_p \dot{m}}\right) (T_{1,in} - T_{2,in}) \left(-c_p \dot{m}_2 \exp\left(\frac{UA}{c_p \dot{m}}\right) + \exp\left(\frac{UA}{c_p \dot{m}_2}\right) \left(-\dot{m} UA + \dot{m}_2 (c_p \dot{m}_2 + UA) \right) \right) \right)}{\left(c_p \dot{m}_2 \left(-\exp\left(\frac{(m-m_2)UA}{c_p \dot{m} \dot{m}_2}\right) \dot{m} + \dot{m}_2 \right) \right)^2} \quad [5-23]$$

$$\sigma_{T_{1,out}} = \sqrt{\left(\sigma_{T_{1,in}} \right)^2 \left(\frac{\partial T_{1,out}}{T_{1,in}} \right)^2 + \left(\sigma_{T_{2,in}} \right)^2 \left(\frac{\partial T_{1,out}}{T_{2,in}} \right)^2 + \left(\sigma_{\dot{m}} \right)^2 \left(\frac{\partial T_{1,out}}{\dot{m}} \right)^2 + \left(\sigma_{\dot{m}_2} \right)^2 \left(\frac{\partial T_{1,out}}{\dot{m}_2} \right)^2} \quad [5-24]$$

The total combined model and measurement uncertainty is then described by equations [5-25] and [5-26], where σ_c denotes the combined measurement uncertainty and σ_T is the measurement uncertainty.

$$\sigma_{cT_{2,out}} = \sqrt{\sigma_{T_{2,out}}^2 + \sigma_T^2} \quad [5-25]$$

$$\sigma_{cT_{1,out}} = \sqrt{\sigma_{T_{1,out}}^2 + \sigma_T^2} \quad [5-26]$$

For this particular problem there is an analytical solution and taking partial derivatives is possible. However, this is not always the case for more complex models. Another approach to determine the model uncertainty from propagation of errors is a Monte Carlo propagation of errors uncertainty analysis.

A Monte Carlo propagation of errors uncertainty analysis is performed by running the simulation many times while varying the inputs values, then recording the distribution of the output values. Again, the measurement uncertainty is considered to be normally distributed. For each run, input parameters are selected randomly from a normal distribution centered

around the measured value. After each run the output is recorded and $\sigma_{T_{2,out}}$ and $\sigma_{T_{1,out}}$ are determined as the variance from the resulting distribution of output values.

5.3.3 Single Parameter Analysis

The goal of this experiment is to analyze a fault as a continuously. The previous analyses used time series calibration, where entire time series were compared all at once, instead of at each time step. When analyzing faults in a real hydraulic heating circuit, it is useful to detect a fault as it occurs. The way that this problem was analyzed was by using equation [5-4] to determine the likelihood that a model matches the measurement at a single point. Since the fault being analyzed is an effect of degradation and not instantaneous, the previously determined k value distribution is used as the prior distribution. The aim of this is to help reduce measurement noise between short time steps.

The process was developed and tested first using surrogate data with a linearly decreasing heat transfer coefficient. Surrogate data was altered by introducing normally distributed measurement noise centered around the model output values.

After testing and confirming the process using surrogate data, measurements from the BZRD for a three-year period were used to attempt to detect faults in the heat exchangers.

5.4 Hydraulic Network

5.4.1 Model Selection

In order to have the ability to investigate a more in-depth set of faults, a more complex Simulink model was used. Gregor Henze developed this model, which applies the physics described in section 4.2 and 4.3. Another model investigating the same set of faults was created using the French SIMBAD toolbox, which also utilizes the Matlab Simulink environment. The purpose of this was to cross-validate the Henze model and to explore the possibility of using the SIMBAD toolbox.

The SIMBAD toolbox is a library of HVAC components created for the Simulink environment, with capabilities of modeling complex hydronic systems with pressure dependent components. When modeling hydronic systems in the Simulink environment, SIMBAD components are arranged in a way that approximately corresponds to the layout of an actual physical system. This approach is not more or less valid than other arrangements, but the systems were arranged in this way to create ease of understanding when sharing models. Properties of water are always carried in a 3 element vector with temperature, pressure and mass flow rate. Describing the water flow this way is useful because it can then easily be checked with the measurements in demo buildings.

Calculating the pressure in a hydronic system is not a simple task, especially when considering complex dynamics involving many branches and

heating system control valves. A “check” element is implemented at every diverting and corresponding converging “t branch” in the hydraulic network. The purpose of this check vector is to ensure the pressure is balanced throughout the system by using a check and iterate method with the diverting and converging element to ensure the pressure values are correct. If the pressures are balanced then the corresponding mass flow rates through each branch can be determined.

The model created is a six-zone model with one pump and air terminals with reheat coils placed in each zone. Each zone contains a hydraulic circuit consisting of a control valve, heat exchanger and balancing valve. The overall hydraulic circuit resistance consists of these six zone resistances as well as supply piping, return piping and pump resistance. The hot water supply temperature is given as the negative of the external temperature in Celsius plus 50 degrees, as this is a typical control scheme for hot water heating temperature.

The SIMBAD model created attempts to match the results from Gregor Henze’s six-zone model, also developed in the Matlab Simulink environment. The zone models used were identical but the way in which the hydraulic system behavior is determined is different.

Henze’s model uses Kirchhoff’s 1st and 2nd laws to determine the overall pressure drop of the system by adding series components first. Then, starting with the outermost circuit, the parallel resistances are added. This

information is given to the pump, which then determines the corresponding total flow rate. The resistances of each individual circuit in relation to the total combined resistance weights the flow rate and flow rates through each corresponding circuit are determined.

Results between both modeling approaches were compared for steady state and non steady state external conditions, constant set point temperature and nighttime setback, situations with internal gains from 8:00 am to 6:00 pm, correct/incorrect valve sizing and valve characteristics. Examples of the results are shown in Figure 11. Results were found to be nearly identical with the largest variations occurring with the combination of non-steady state, large internal gains and linear valve characteristics. This caused longer lasting fluctuations in the supply and return temperature of the water in the SIMBAD models.

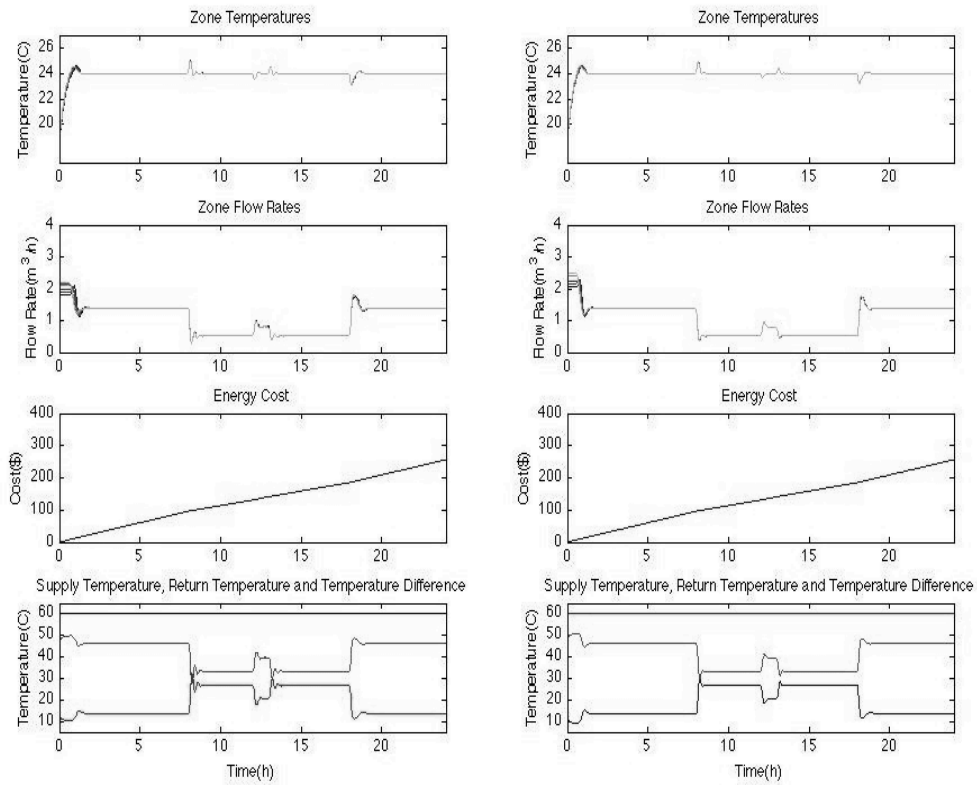


Figure 11 SIMBAD (left) and Henze (right) Comparison for 6000 W Internal Gain

Moving forward, it was decided to use Henze’s model for further analysis of hydraulic systems. This was based on the computational speed being considerably faster for this model. For future work SIMBAD may prove a useful tool because of its high level of adaptability.

5.4.2 Bayesian Model Selection Analysis

Using the Henze model, a group of binary faults was analyzed. The set of faults analyzed were valve type, valve let-by and high valve hysteresis. Each fault is considered to be binary, either in a faulty state or a fault-free

state or in the case of valve type linear or proportional. This results in 8 possible states for each valve as shown below in Table 3.

Table 3 Fault Combinations

Fault Combination Number	Valve Type	Valve Let-By (%)	Valve Hysteresis
1	Linear	0.025	0.025
2	Linear	0.025	0.075
3	Linear	0.075	0.025
4	Linear	0.075	0.075
5	Proportional	0.025	0.025
6	Proportional	0.025	0.075
7	Proportional	0.075	0.025
8	Proportional	0.075	0.075

All 8 models are created and time series results are stored for all of them for a one-week period for measurements of mass flow, zone temperature and return water temperature. The models are then tested with 8 sets of surrogate data using the exact values as described by the model to ensure functionality. Next, 1000 sets of surrogate data were generated using random values for input parameters. Valve let by and valve hysteresis were both uniformly sampled on the intervals of 0 to 0.1 and valve type was selected randomly as linear or proportional. Noise is then added to the

surrogate data by selecting a value from a normal distribution centered on the model-produced value.

The advantage of this method is the reduction in total models. If a fault can be detected with only 8 simulations as opposed to before when 1000 simulations were run, simulation time is greatly reduced. One disadvantage is that the result in this case is binary. That is, the system is either faulty or not, but the value is not determined, whereas with parameter estimation values of the fault parameters were determined. The system does not have to be binary, however; any number of models could be created, but as the number of faults and fault states increases, the amount of models grows exponentially. The goal is to determine faulty system states with as few models as is necessary.

6 RESULTS

6.1 Test Building

To determine how well the model created matched the measured values, measured heating energy was plotted against the simulated heating energy. Ideally this should result in a linear relationship with a slope of unity going through the origin without much variance on either side of this line. As shown in Figure 12, the relationship between measured and simulated data is close to expected. However, the slope is slightly less than 1 and the variance is somewhat consistent. A more analytical way to determine how well the data fits is the coefficient of determination, which compares the sum of the squares of the error and the total. A coefficient of determination of 1.0 would mean the data matches perfectly, while a value of 0 means the data is completely uncorrelated. The coefficient of determination for this data was determined to be 0.84.

Figure 13 and Figure 14 shows the simulated and measured instantaneous and cumulative heat use, respectively. The measured and simulated data are somewhat consistent, however, the simulated data does have larger dips when large amounts of solar energy are present and slightly over predicts heat usage during times of very cold weather. A possible

explanation for this could be the model uses an air change rate (ACH) that predicts ventilation based on wind speed and infiltration and not the likelihood of occupants using operable windows. Users may be less likely to open windows when the weather is cold. Therefore, the model may overpredict the ACH for cold weather resulting in a higher prediction of heating energy for cold temperatures.

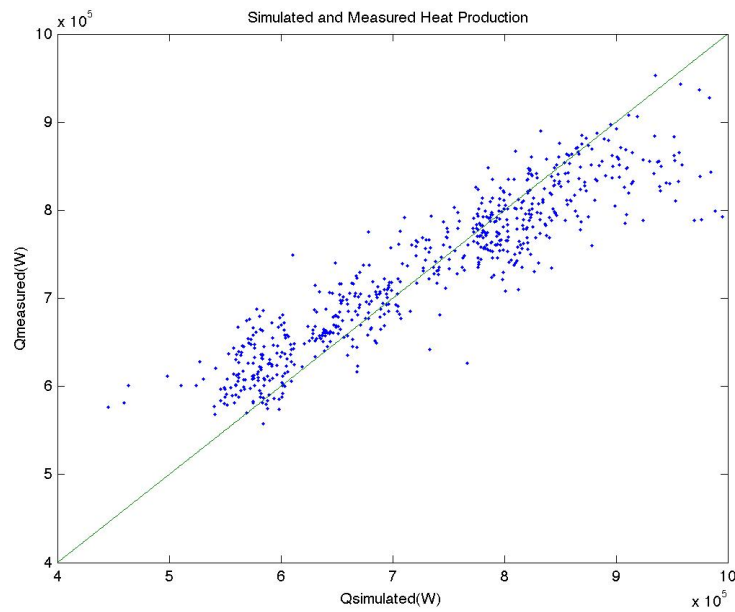


Figure 12 Simulated vs. Measured Heat Usage

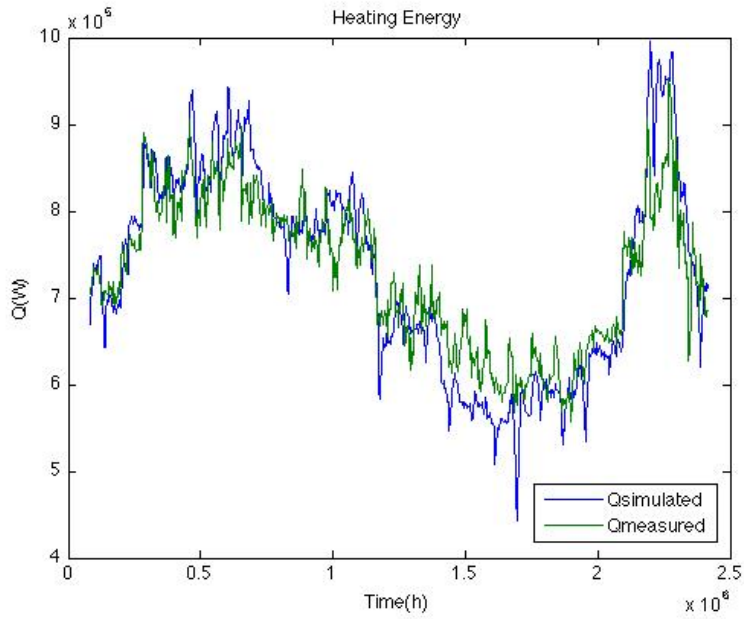


Figure 13 Simulated and Measured Heat Use

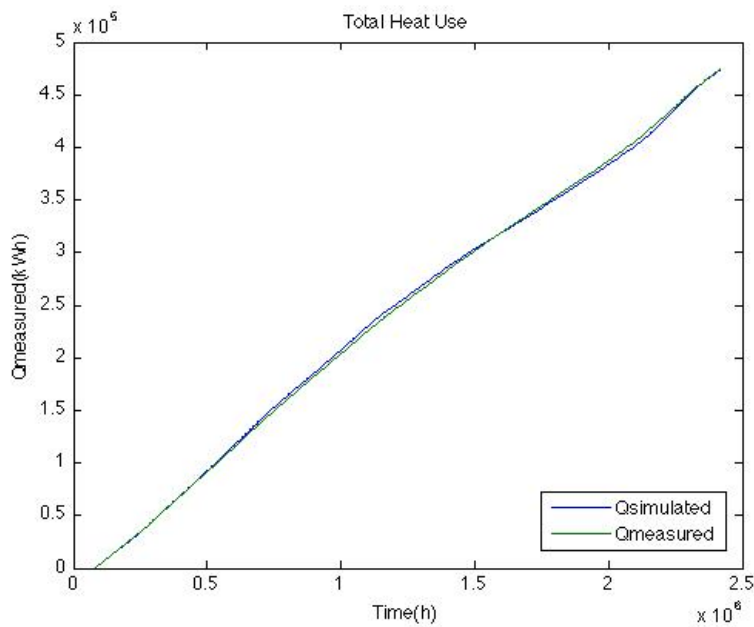


Figure 14 Cumulative Heat Use

After validating the model, the next step that was performed was to analyze the energy savings available from creating a nighttime shutoff. In

order for a nighttime shutoff to be effective, the temperature of the zone must be within the thermal comfort zone during occupied hours. In order to ensure the zone is always warm enough, the heating system is shut off at 6pm and turned back on at 5am. Figure 15 and Figure 16 show the instantaneous and cumulative heating energy, respectively, for both simulated and measured data. Over the 4-week period, a nighttime shutoff reduces the total energy consumption by about 20%, saving approximately 93,000 kWh of heating energy.

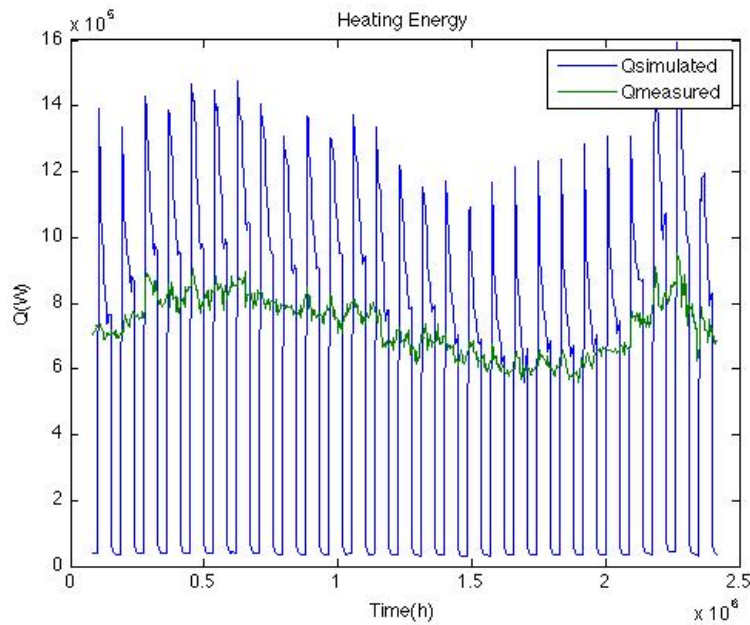


Figure 15 Heating with Nighttime Setoff

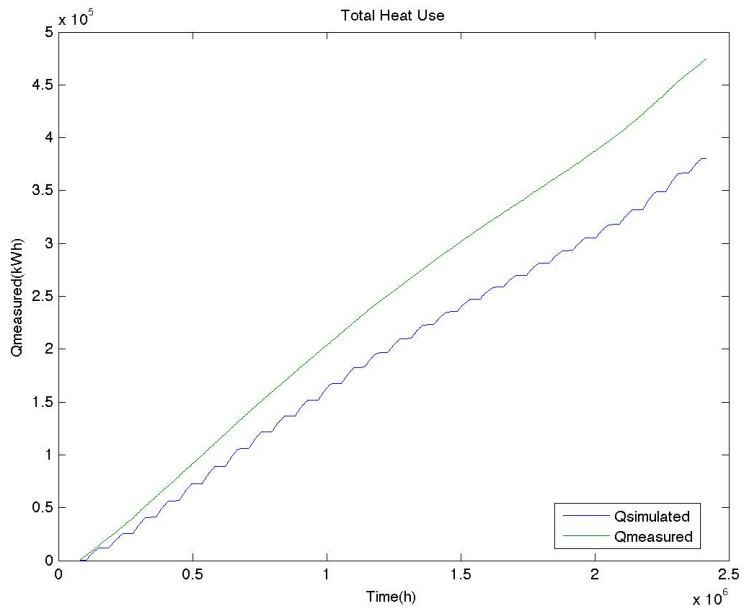


Figure 16 Cumulative Energy with Nighttime setoff

One reason a building operator may choose not to implement a nighttime shutoff is fear of pipes freezing and breaking. However, in a building with such high thermal mass, this is highly unlikely even with a window left open or with very cold ambient temperatures. Figure 17 shows a comparison between the zone temperature and outside temperature. As shown the temperature in the zone never drops below 14°C even when the ambient temperature goes as low as -4°C .

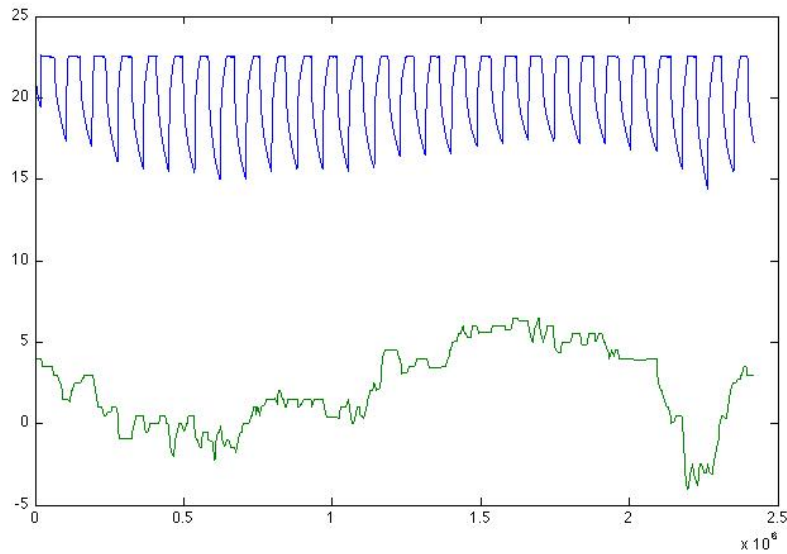


Figure 17 Inside and Ambient Temperatures

6.2 Bayesian Fault Detection

6.2.1 Boiler

6.2.1.1 Single Parameter

The results from the first test performed with a non-uniform prior are shown in Figure 18, Figure 19 and Figure 20. It is revealed that the posterior distribution is always a combination of the likelihood and prior functions. Since the fault explored here is only using surrogate data, the likelihood function is a very narrow distribution centered around the actual value. This means when it is combined with a wide prior distribution the posterior distribution is more heavily influenced by the likelihood function. The use of a prior distribution can be useful when there is a fairly high certainty of a certain parameter and multiple parameters are being estimated as shown in

the next section. However, for this test case it becomes inappropriate because it skews the results without any added benefit.

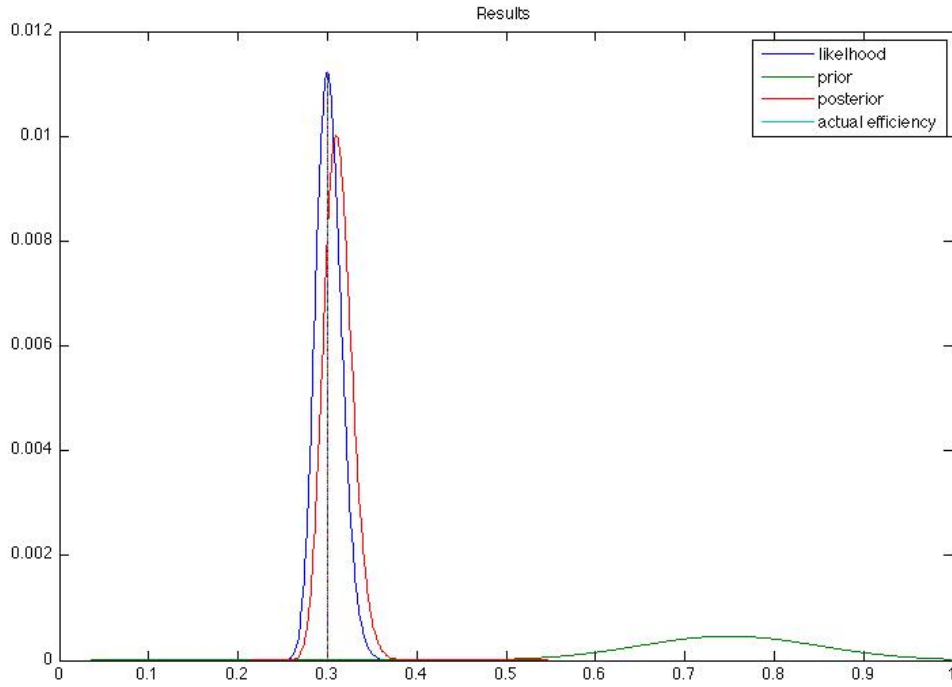


Figure 18 Boiler Efficiency of 0.3 With a Non-Uniform Prior

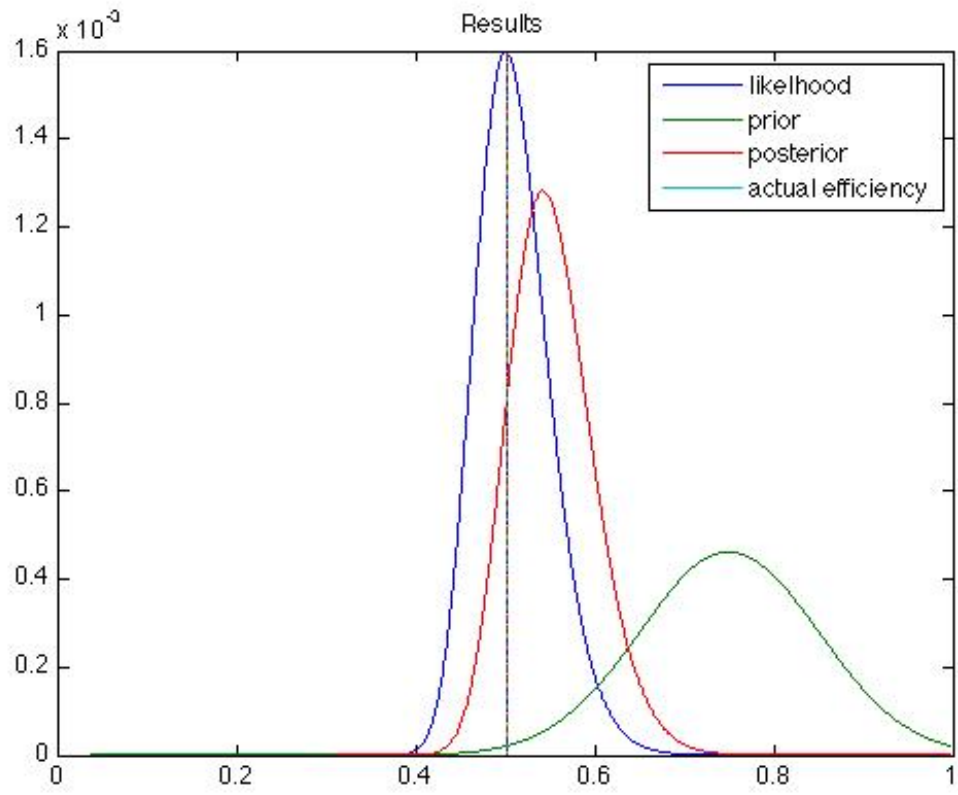


Figure 19 Boiler Efficiency of 0.5 with a Non-Uniform Prior

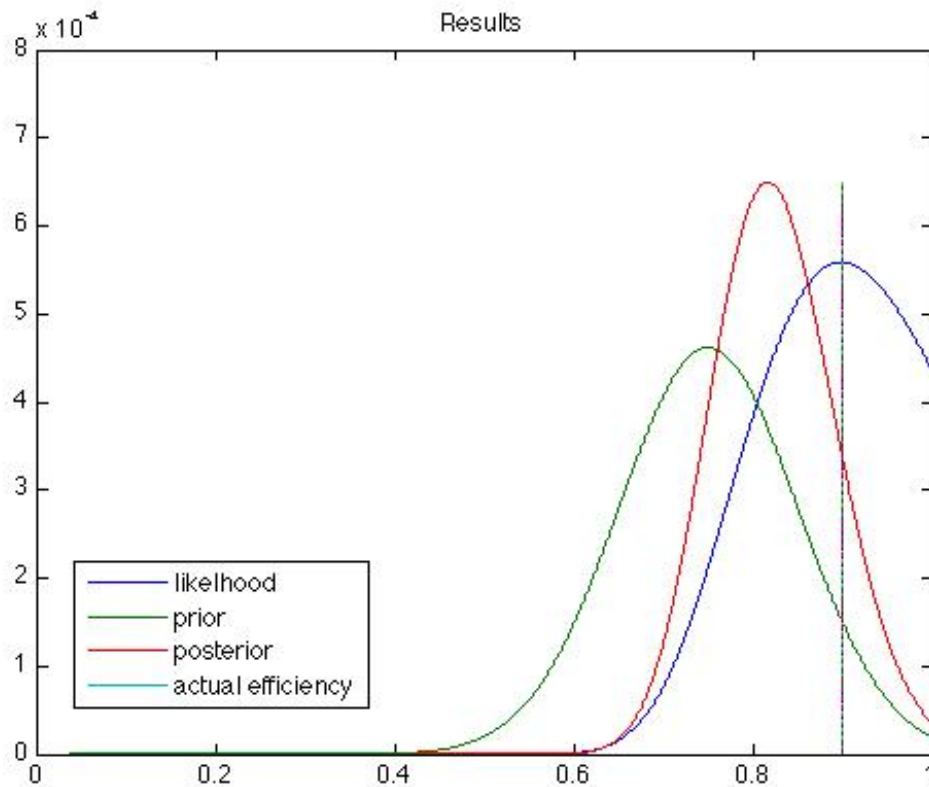


Figure 20 Boiler Efficiency of 0.9 with a Non-Uniform Prior

Next, the test was performed using a uniform prior as would be expected when all of the priors are considered to be equally likely and the posterior function is normalized the likelihood function becomes the posterior. Figure 21 shows that the program is capable of correctly detecting the surrogate data's efficiency.

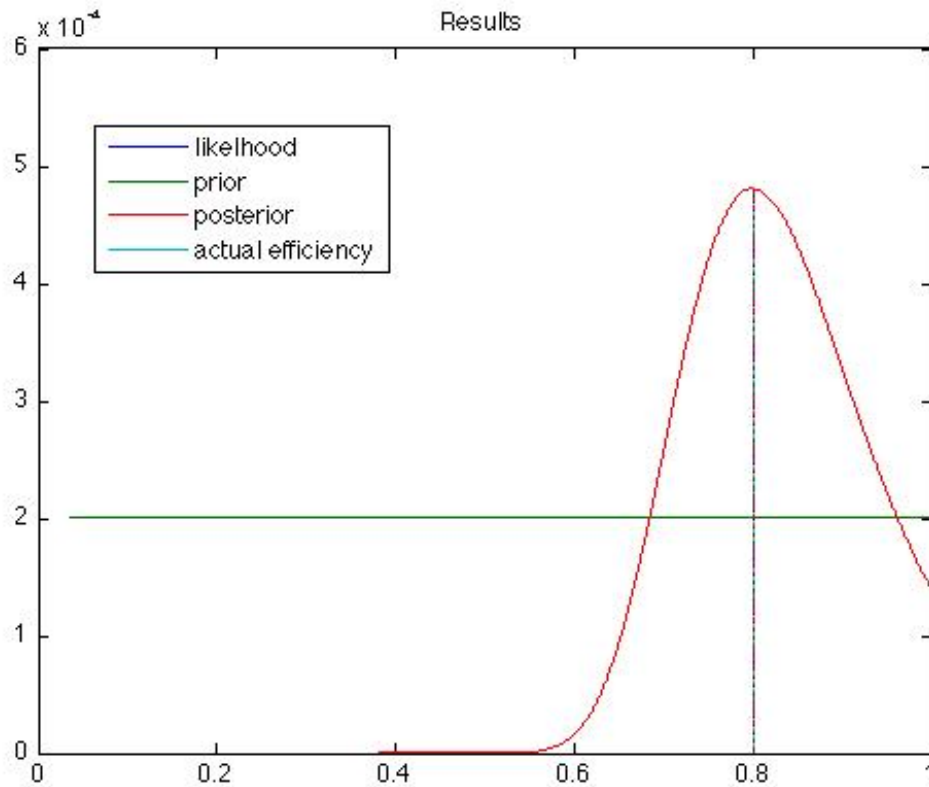


Figure 21 Boiler Efficiency of 0.8 with a Uniform Prior

6.2.1.2 Multiple Parameters with a Single Measurement

This section investigates the ability to determine the values of multiple parameters, still using the same measurements as before. This becomes more difficult since tradeoff effects make it difficult to isolate and determine the fault. Figure 22, Figure 23 and Figure 24 exemplify the tradeoff effects that emerges in this scenario. In both figures, there are two areas with higher probabilities or a long area where posterior probabilities are essentially equal and determining the true value can be difficult. It should

also be noted that none of the values are much more likely than the others presented as can be seen by the scale of the figures.

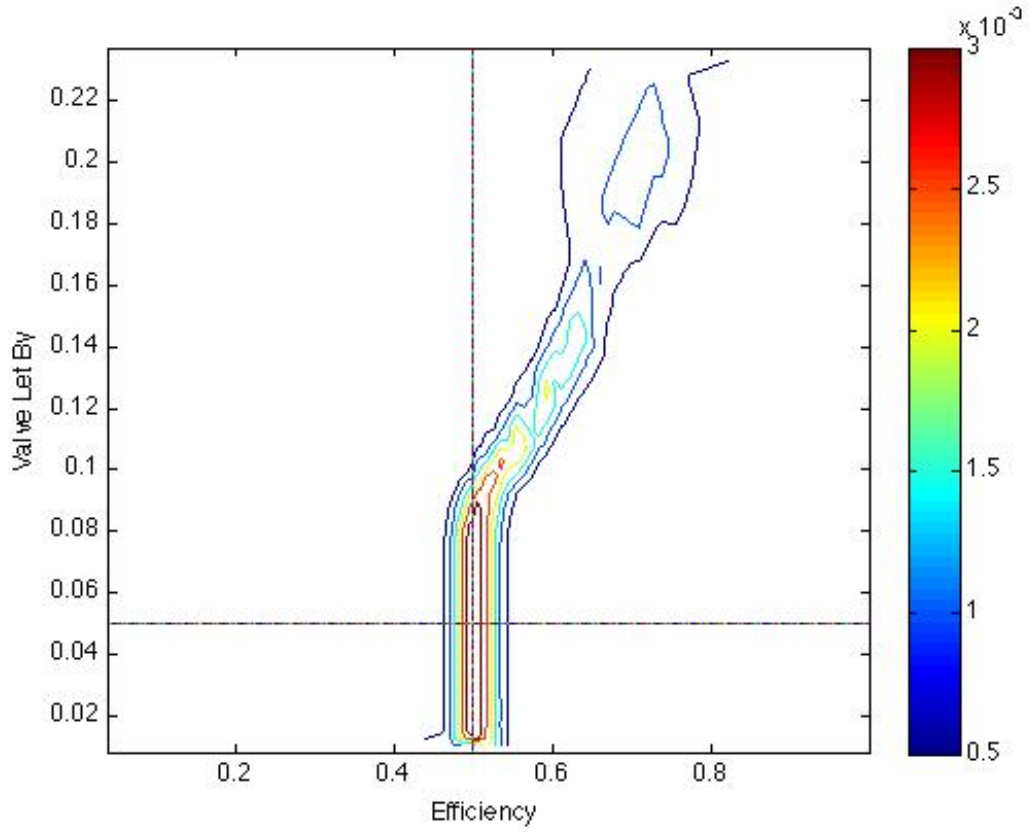


Figure 22 Multiple Parameters with a Single Measurement

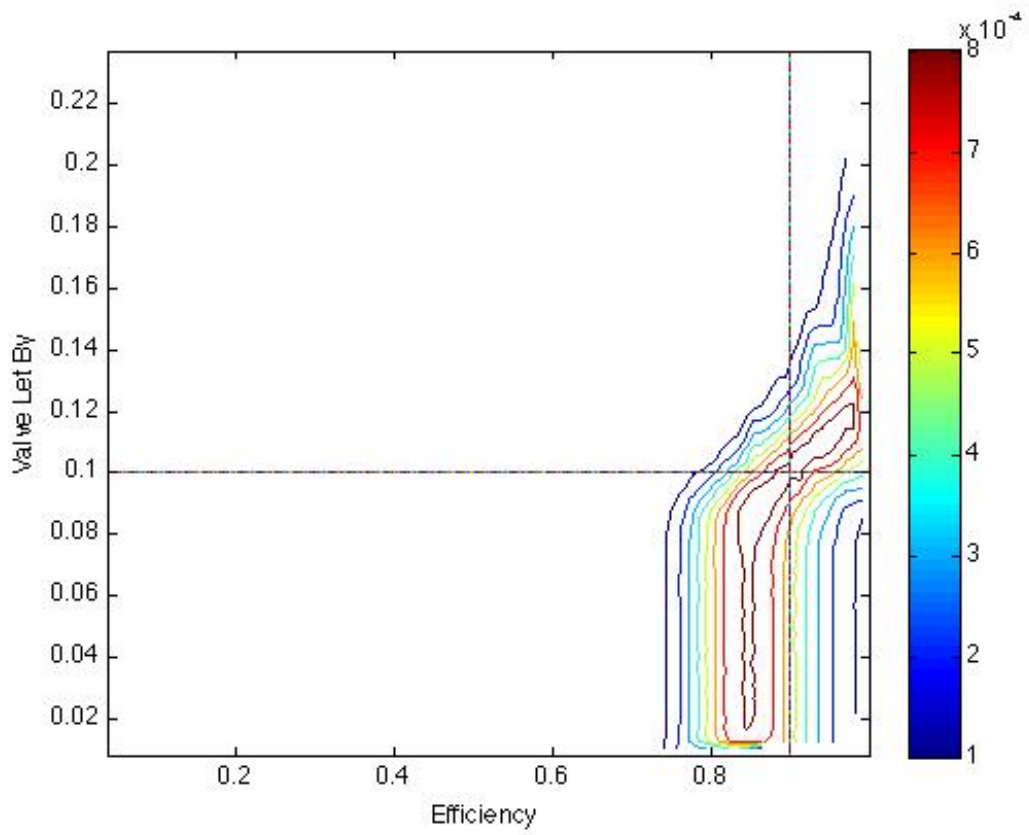


Figure 23 Multiple Parameters with a Single Measurement in Winter

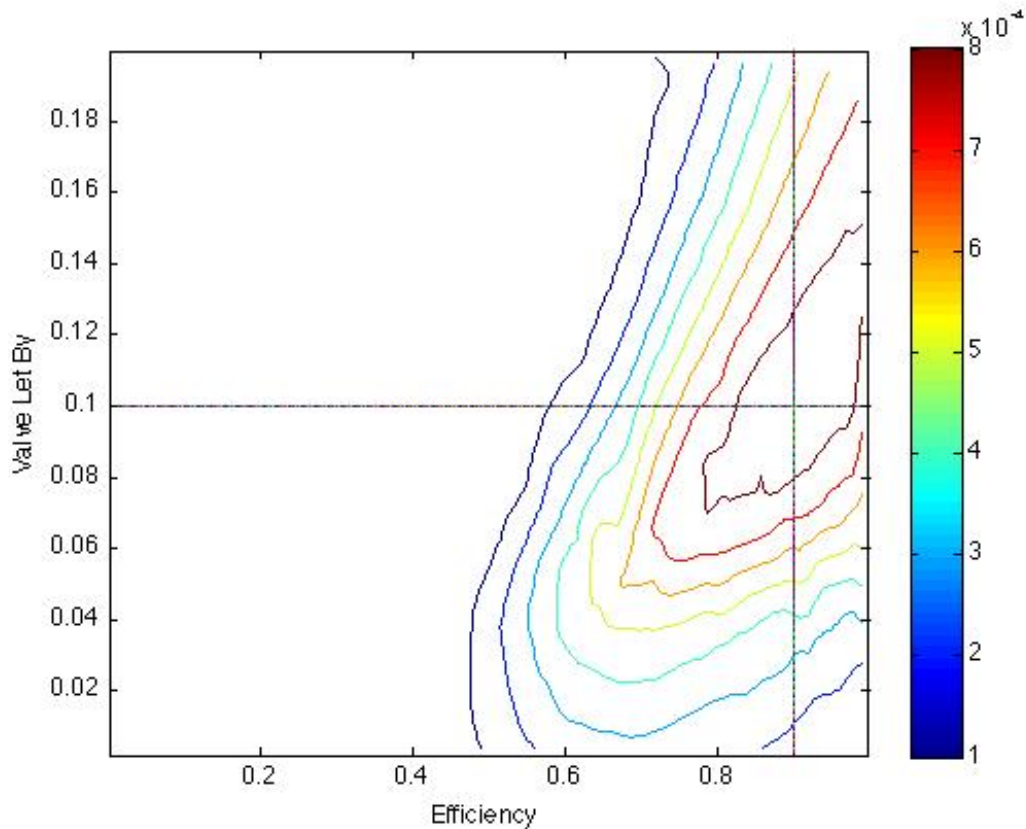


Figure 24 Multiple Parameters with a Single Measurement in Spring

A possible solution to isolate one of the faults and thus gain certainty is to establish a prior distribution as more information becomes available. If, for example, a building owner had installed a new boiler and was fairly certain it was operating at nominal conditions that assumption could be added as prior information. If it is assumed that the boiler is likely to be operating at 90 percent efficiency a prior distribution can be created as shown in Figure 25. The prior distribution for the valve let by is still considered uniform. This results in Figure 26, revealing that there is now only one region clear region where the fault exists. This is a result of a high prior

probability in the region where there is a higher likelihood of a fault being present.

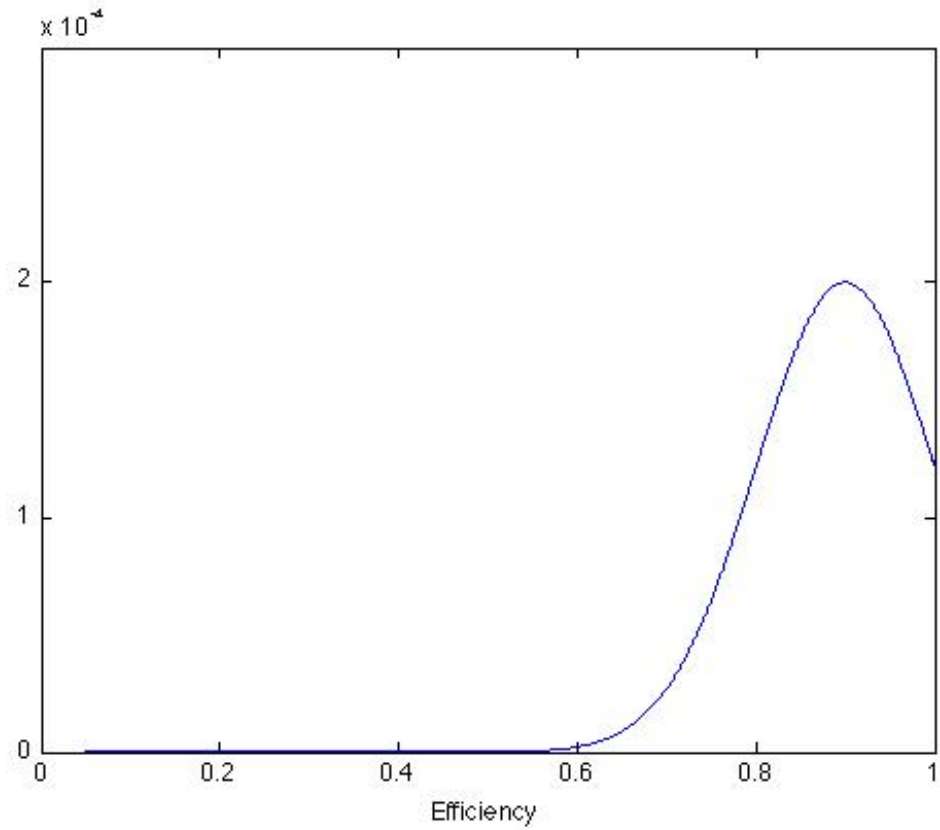


Figure 25 Efficiency Prior Distribution

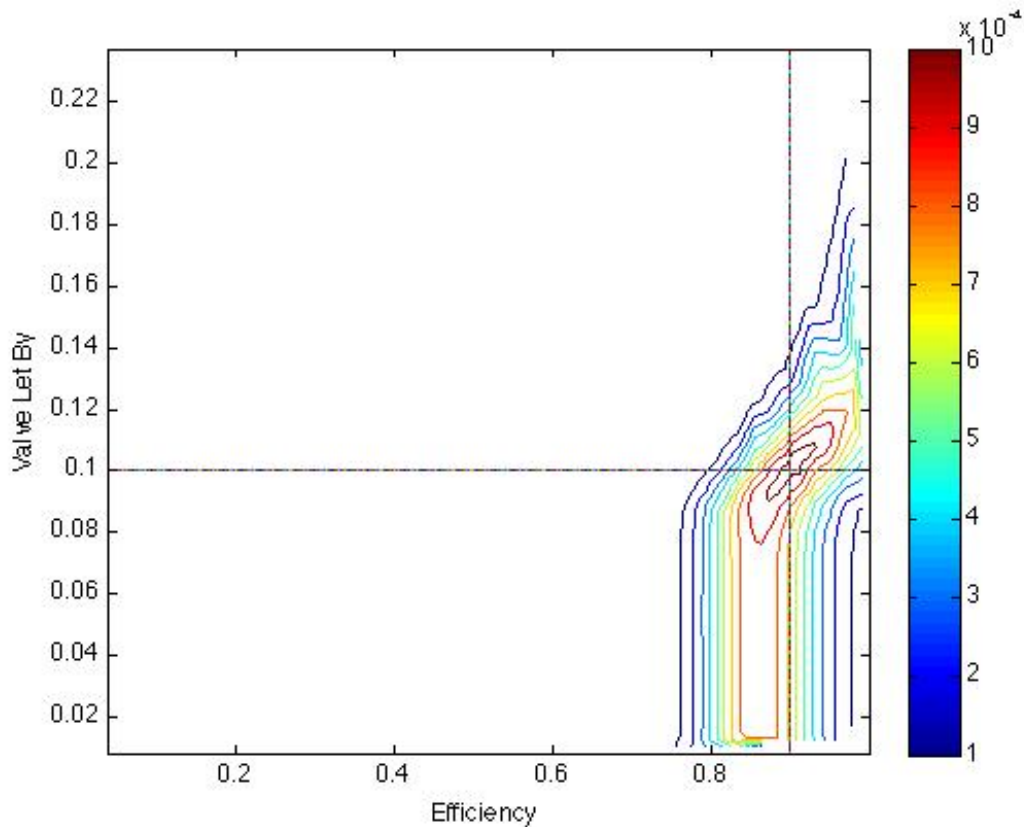


Figure 26 Multiple Parameters with a Single Measurement and Prior

6.2.1.3 Multiple Parameters with Multiple Measurements

If multiple measurements are available, then isolating a fault becomes much simpler, especially if interaction between the measurements is low. For this case, adding a flow rate measurement helps to isolate the valve let-by. Figure 27 and Figure 28 show how much more accurate the faults are detected with the second measurement available. It should be noted that the scale is now 10^{-3} as opposed to the figures in the previous section where the scale is 10^{-4} , with both showing the normalized posterior probability.

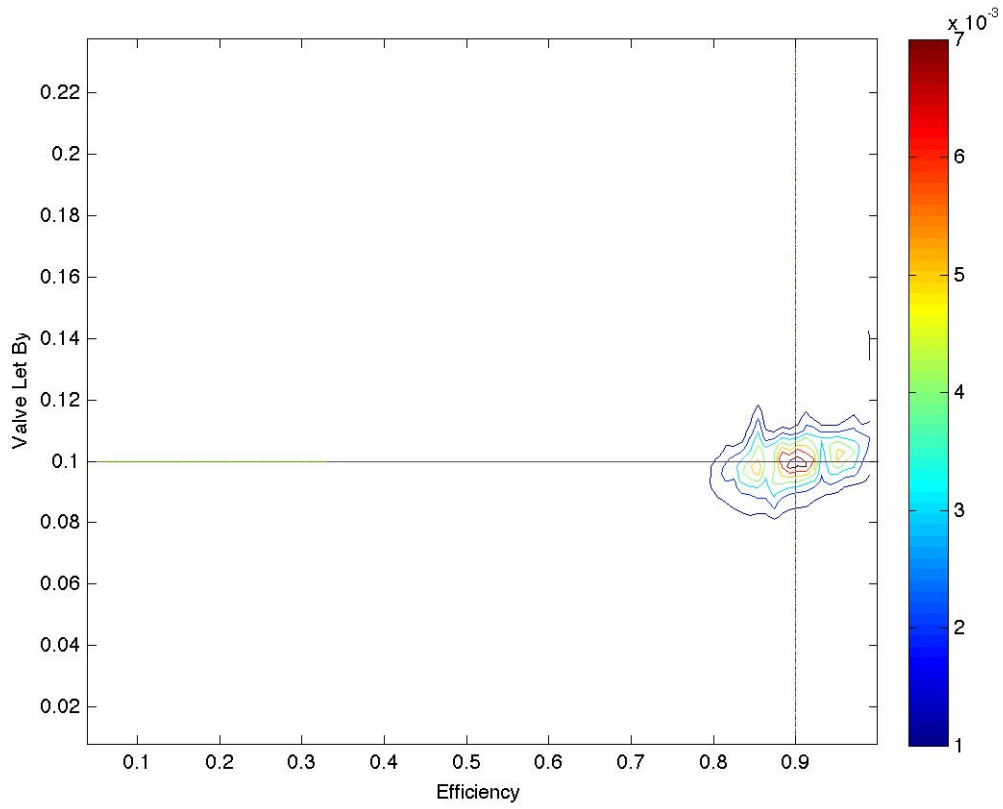


Figure 27 Multiple Parameters with Multiple Measurements

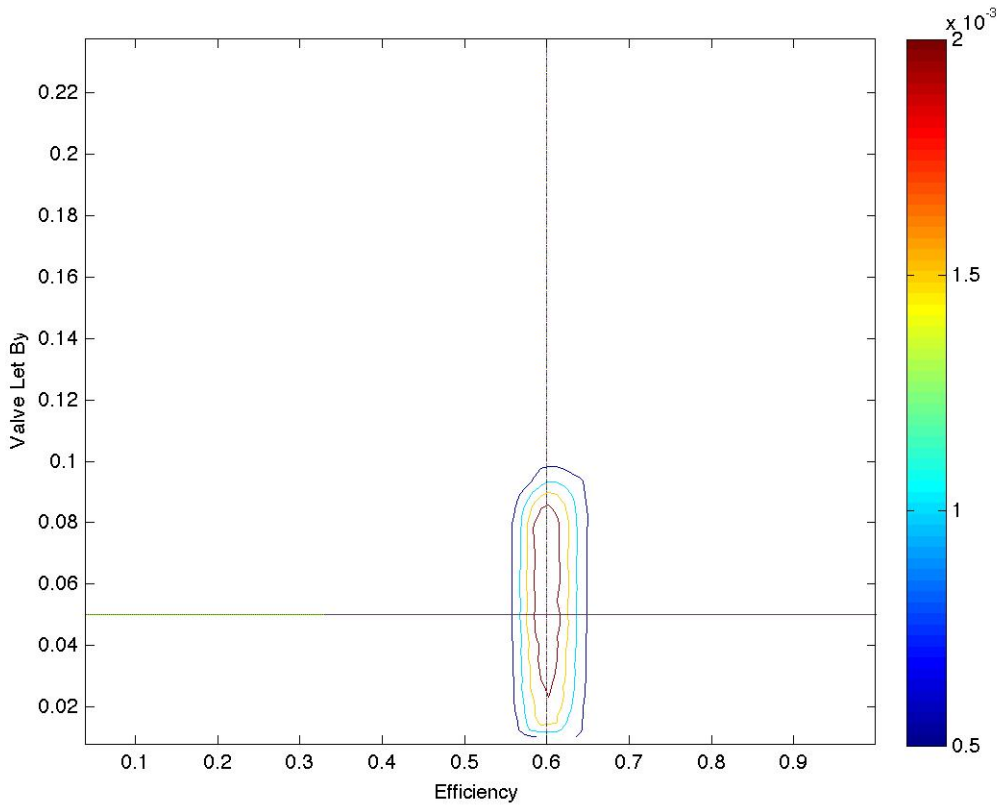


Figure 28 Multiple Parameters with Multiple Measurements

6.2.2 Heat Exchanger

6.2.2.1 Uncertainty Analysis

Both methods of uncertainty analyses result in a higher sigma value than just measurement error, as they are a combination of both measurement and propagated model uncertainty. Figure 29 and Figure 30 show an uncertainty comparison for a three-year period between the Kline-McClintock method and Monte Carlo method for measured data. The largest discrepancies between the two methods occur in the summer time when the flow rate through the heat exchanger approaches zero.

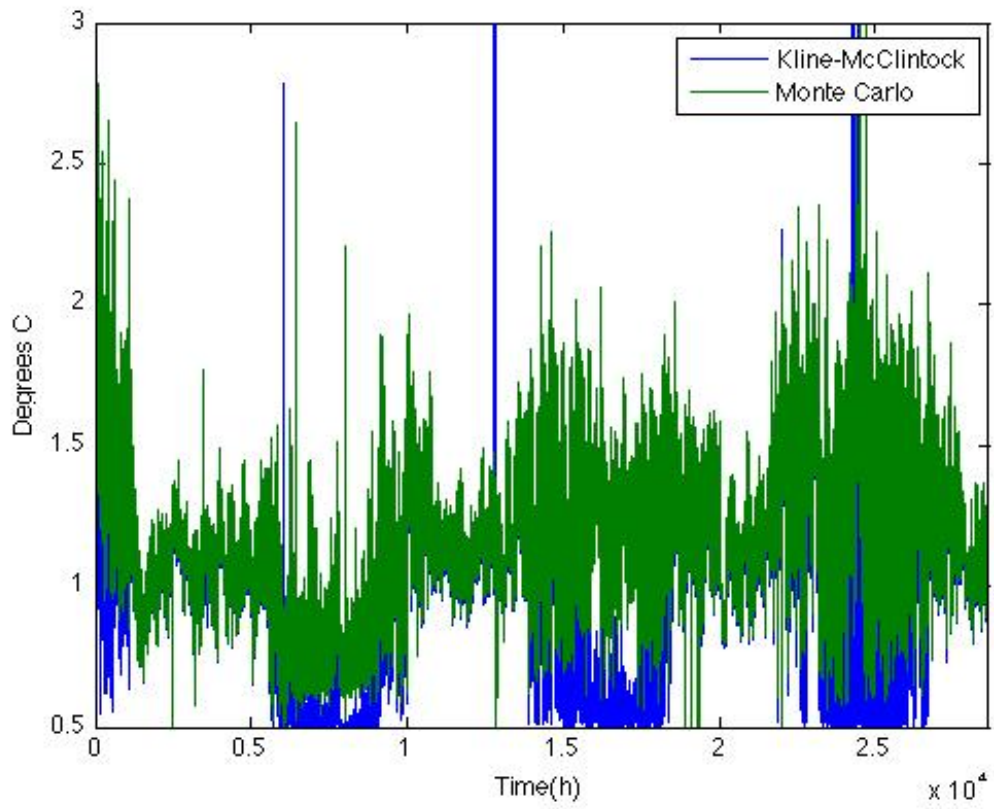


Figure 29 Uncertainty Analyses for $T_{1, out}$

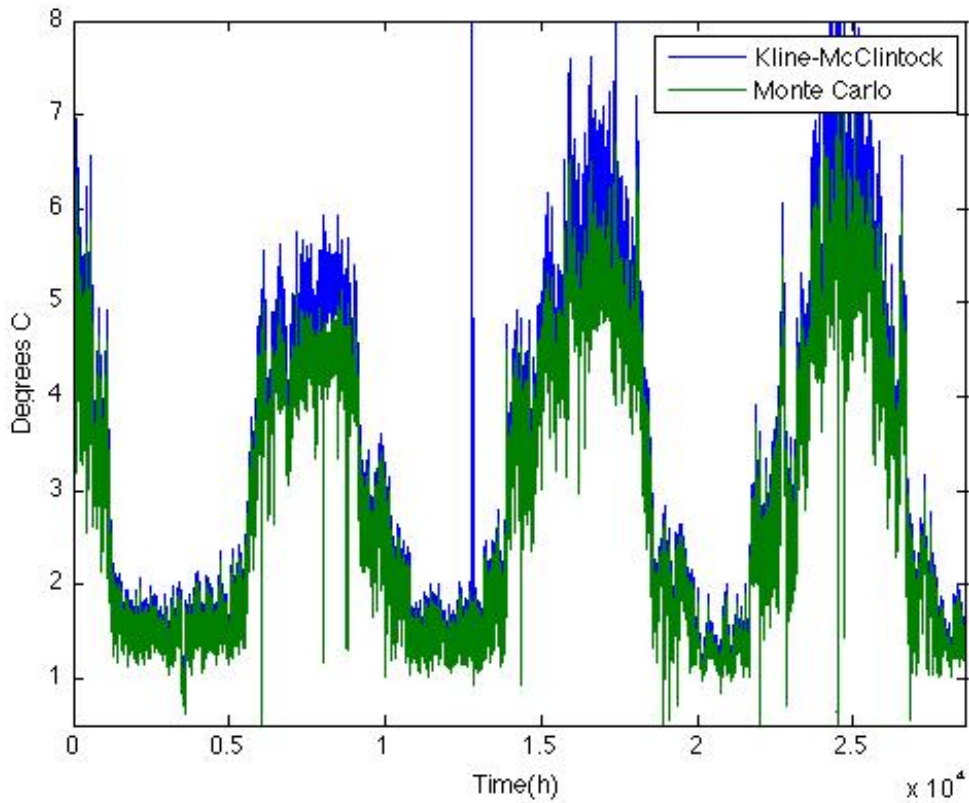


Figure 30 Uncertainty Analyses for $T_{2,out}$

Although both methods show similar results, issues do arise with the Bayesian analysis. As shown in equations [5-15] through [5-23], the uncertainty of the output is dependent on the UA value. This means that different k values have different sigma values and all k values are not being compared equally. Even if one k value were to produce better matching data to the surrogate or measured data, it may end up with a lower likelihood if it belongs to a wider distribution. Figure 31 shows how one point is closer to the true value but because it belongs to a wider distribution, it has a lower likelihood.

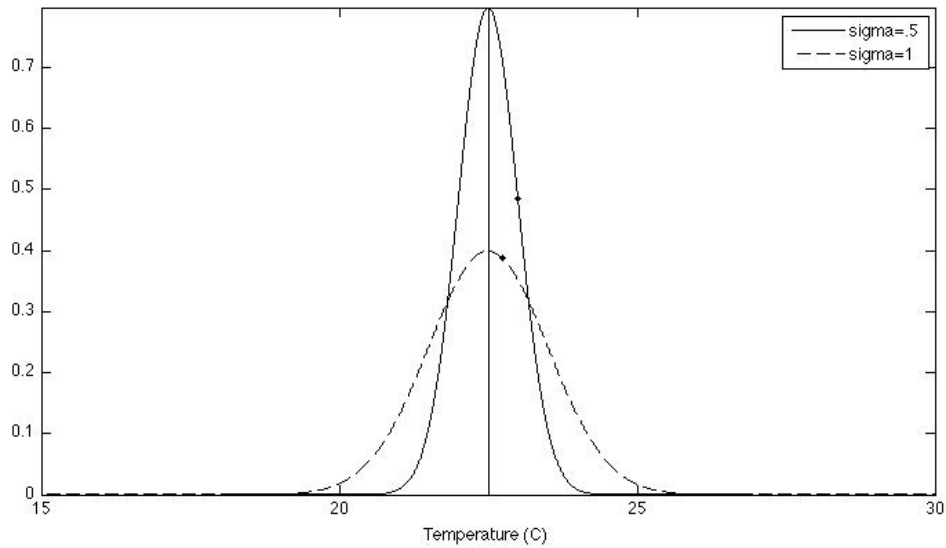


Figure 31 Sigma Uncertainty Analysis

A possible solution to this could be to take a single k value and calculate the sigma value using this and have that be the only sigma value used during the corresponding time step. Another solution could be to calculate the k value using a constant sigma value and simply report the uncertainty of the detected k value at that time-step.

6.2.2.2 Single Parameter Analysis

The first test performed was a comparison using surrogate data and parameter estimation with a uniform prior and two measurements. A k value that is declining linearly is used to model fouling in the heat exchanger. The parameter is accurately determined, as shown in Figure 32, during winter months, where the green line is the actual value that was used as an input for the surrogate data and the blue line is the k value determined with parameter estimation.

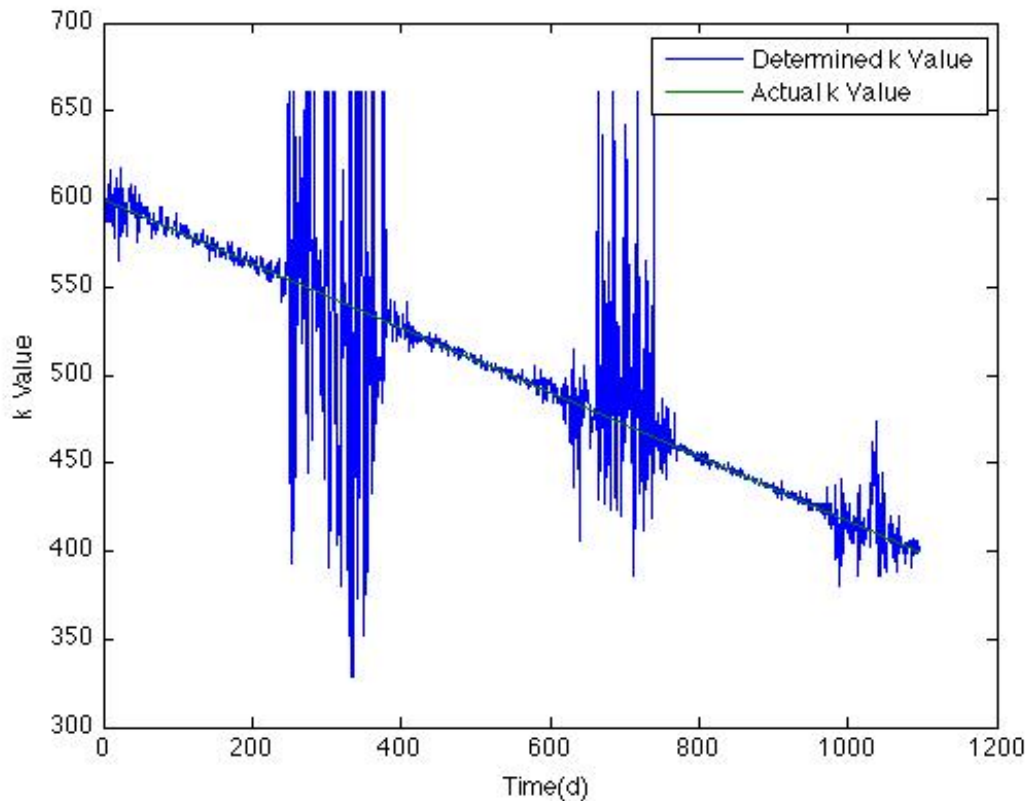


Figure 32 Parameter Estimation with Uniform Prior

Next, the prior was changed from being uniform to being the posterior from the previous time step to eliminate measurement noise. This method, at first, was somewhat ineffective. One key issue is that Matlab can only work with numbers greater than $1 \cdot 10^{-308}$. Often the likelihood function has extremely small numbers and when these are multiplied together each time step they can become zero. With this method, once a prior becomes zero, the corresponding k value will remain at zero even if it has a high likelihood. This results in certain k values no longer being anything but zero and therefore the k value can get stuck as a constant value especially at later points in the time series. To remedy this issue, the prior value became a

combination of the normalized posterior function and a uniform prior. This ensured that all k values could still be most likely. Figure 33 shows the analyses of parameter estimation with a posterior becoming the next time steps prior. This method was found to be effective at reducing noise from the previous analyses with a constant and uniform prior.

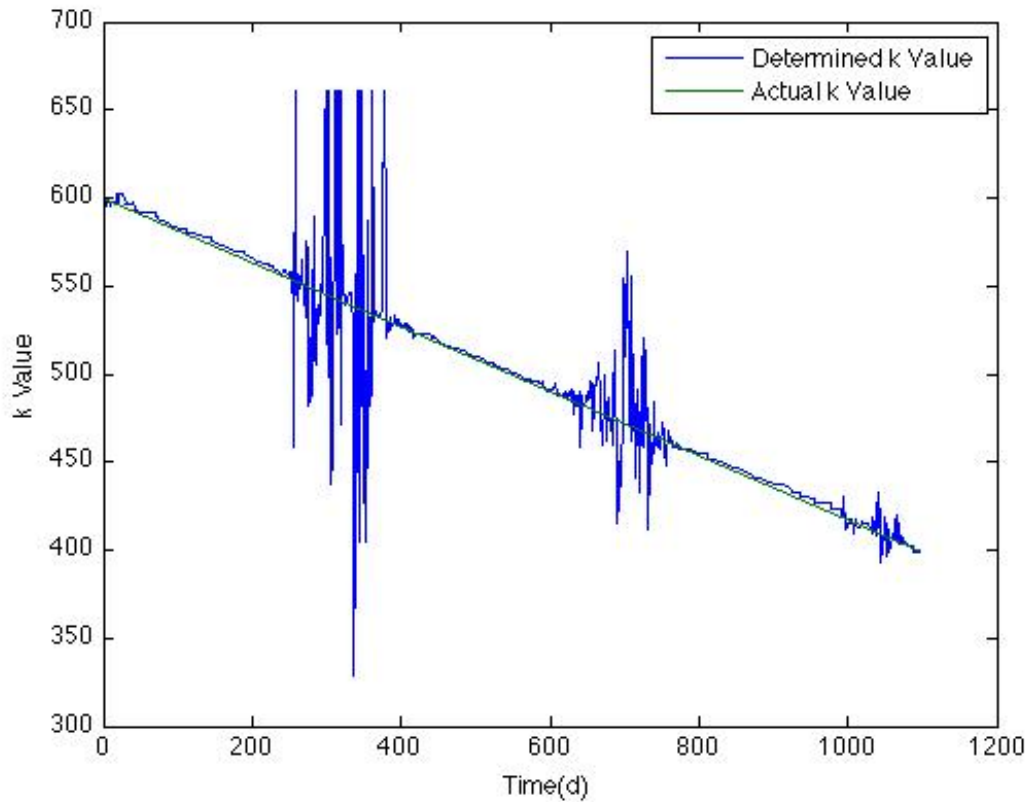


Figure 33 Parameter Estimation with Updated Prior

To try and explain the noise present, the determined k value is plotted with propagated measurement output uncertainty, as shown in Figure 34. In summer months when the heat exchanger is less likely to be used low flow rates cause relatively high uncertainty. This makes the heat transfer coefficient unpredictable. The importance of uncertainty analyses is

highlighted here, as a building manager may interpret the heat exchanger as crossing into faulty states as the k value determined drops to very low values. This however, is not because the k value is actually low, but rather almost all likelihoods become equally probable and a k value is undeterminable.

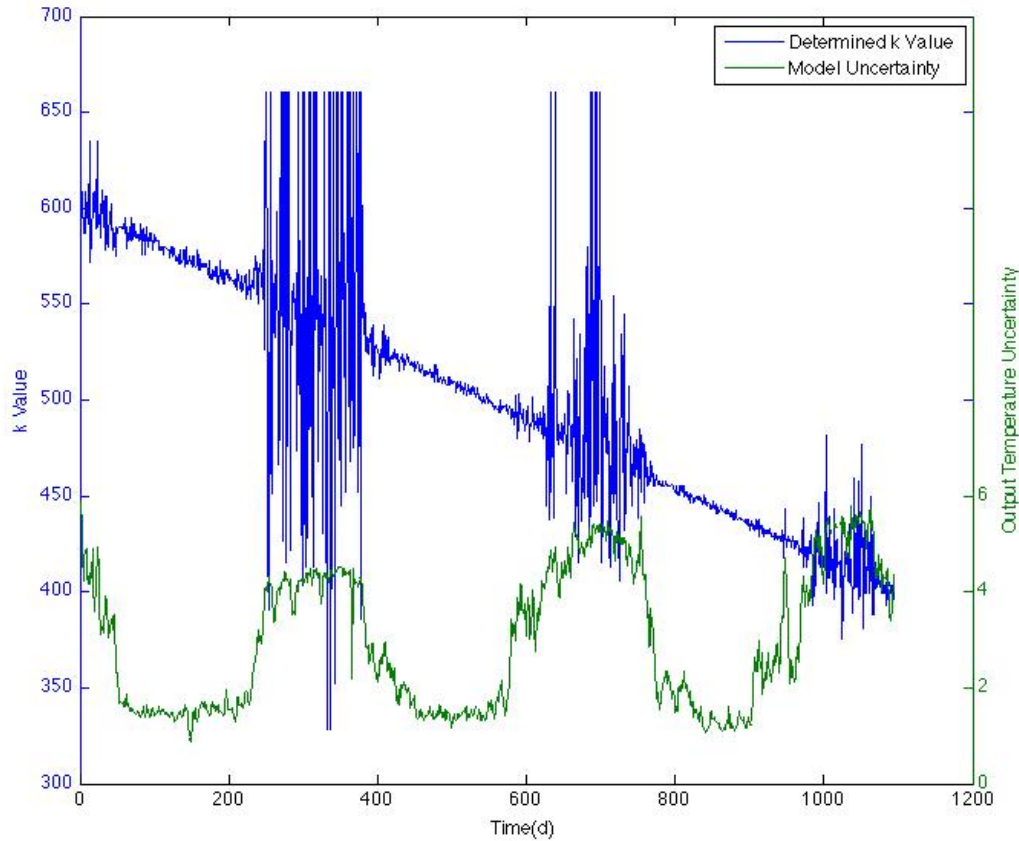


Figure 34 Uncertainty and Determined k Values

Next, the Bayesian methods developed with surrogate data were applied to measured data for the BZRD heat exchanger.

Figure 35 and Figure 36 show the k value for a three-year period for the first heat exchanger and second heat exchanger respectively. The figures clearly

show that the amount of noise present makes fault detection difficult for this system.

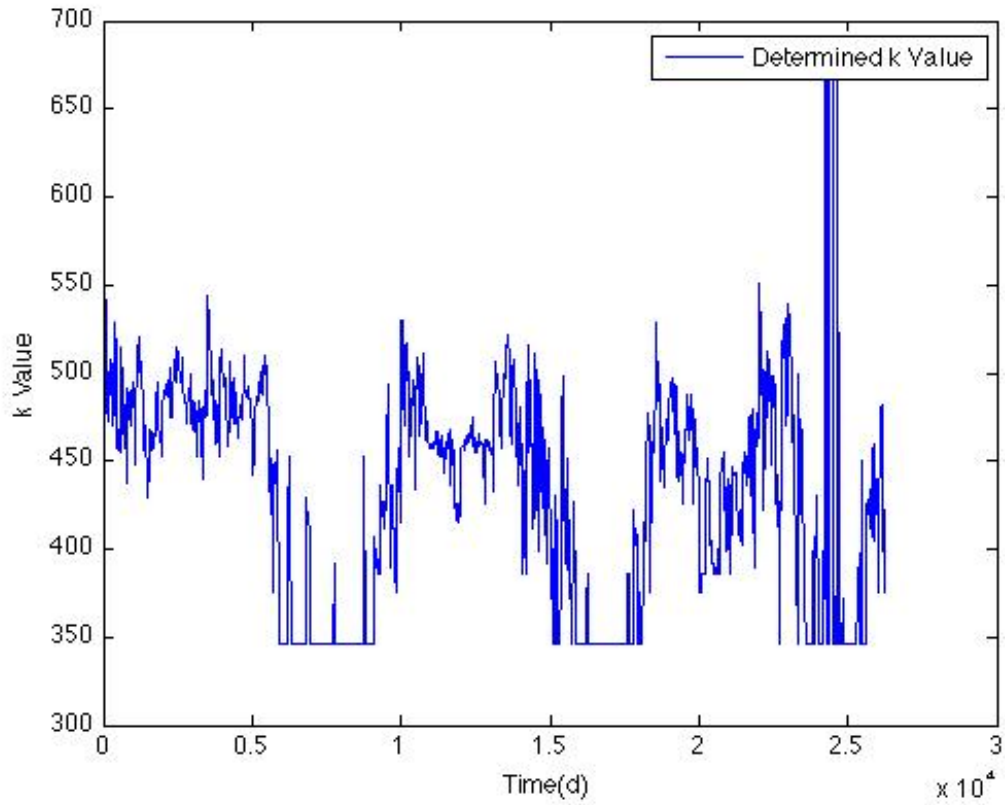


Figure 35 k Value Without Updating Prior 1st Heat Exchanger

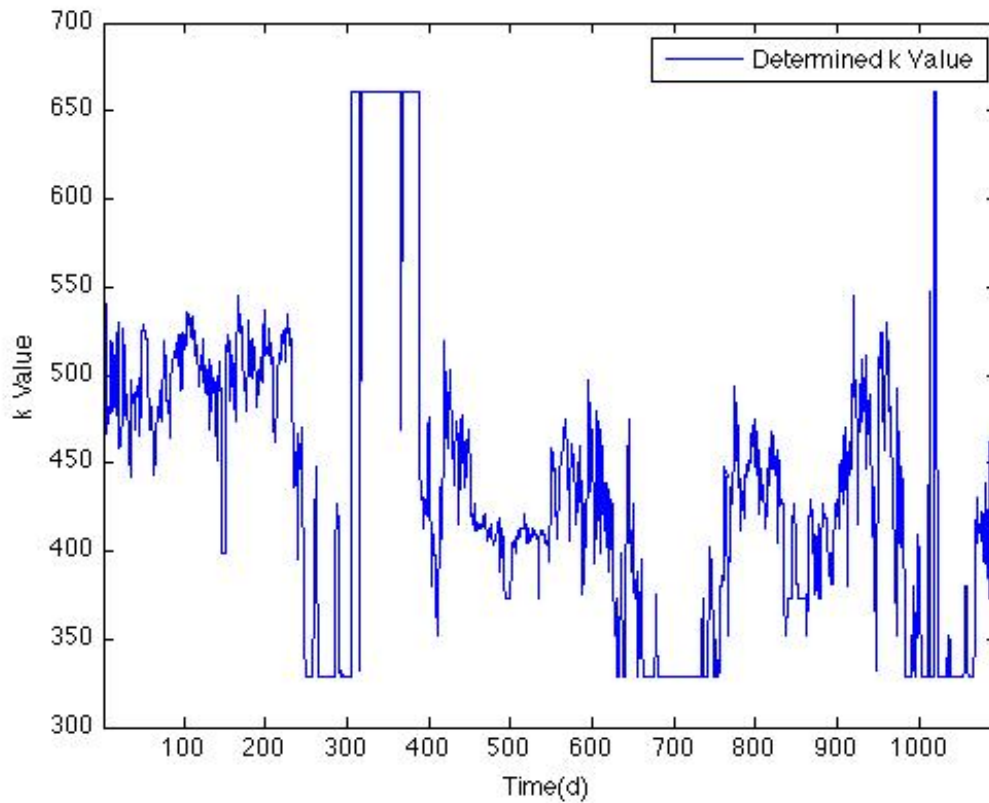


Figure 36 k Value Without Updating Prior for 2nd Heat Exchanger

Next, an updating prior is applied to reduce measurement noise. Figure 37 and Figure 38 show a three-year period for the first heat exchanger and second heat exchanger respectively. The prior is set at the previous time steps posterior, again combined with a uniform prior to ensure posteriors do not become fixed at zero.

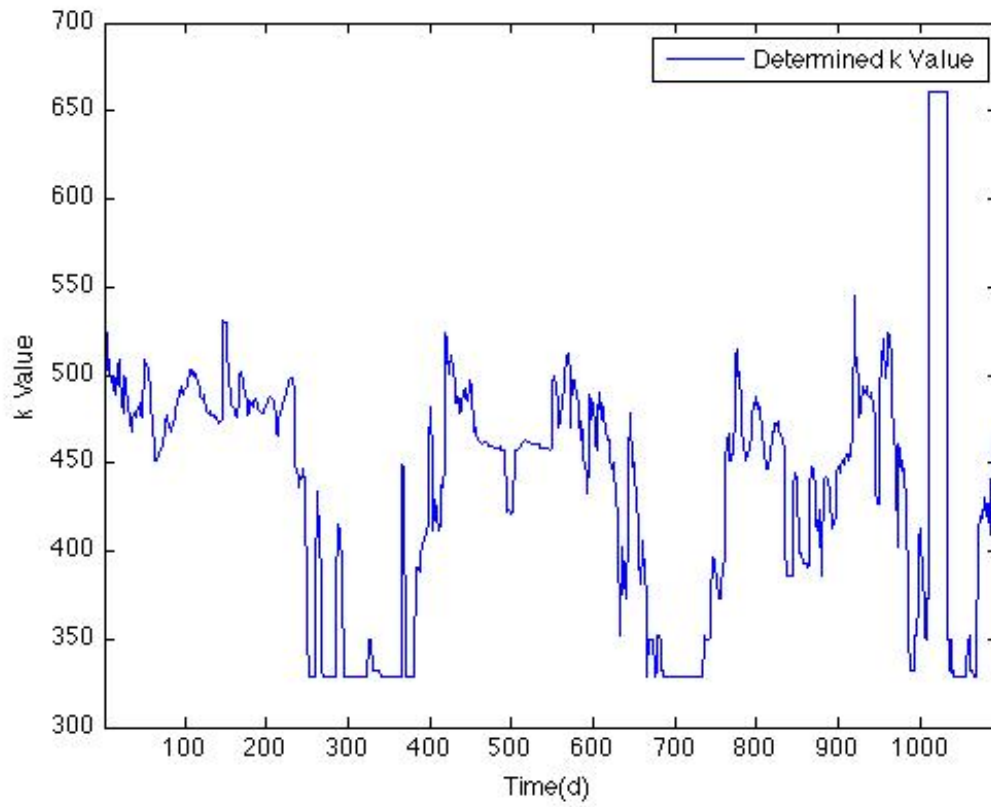


Figure 37 k Value With Updating Prior 1st Heat Exchanger

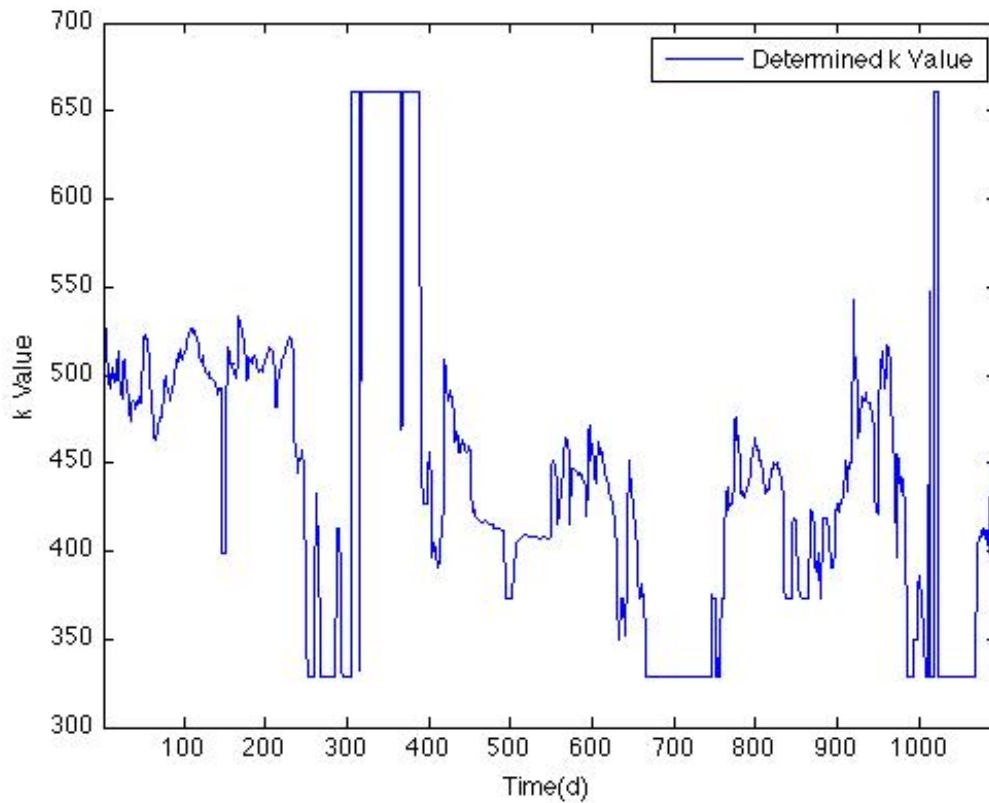
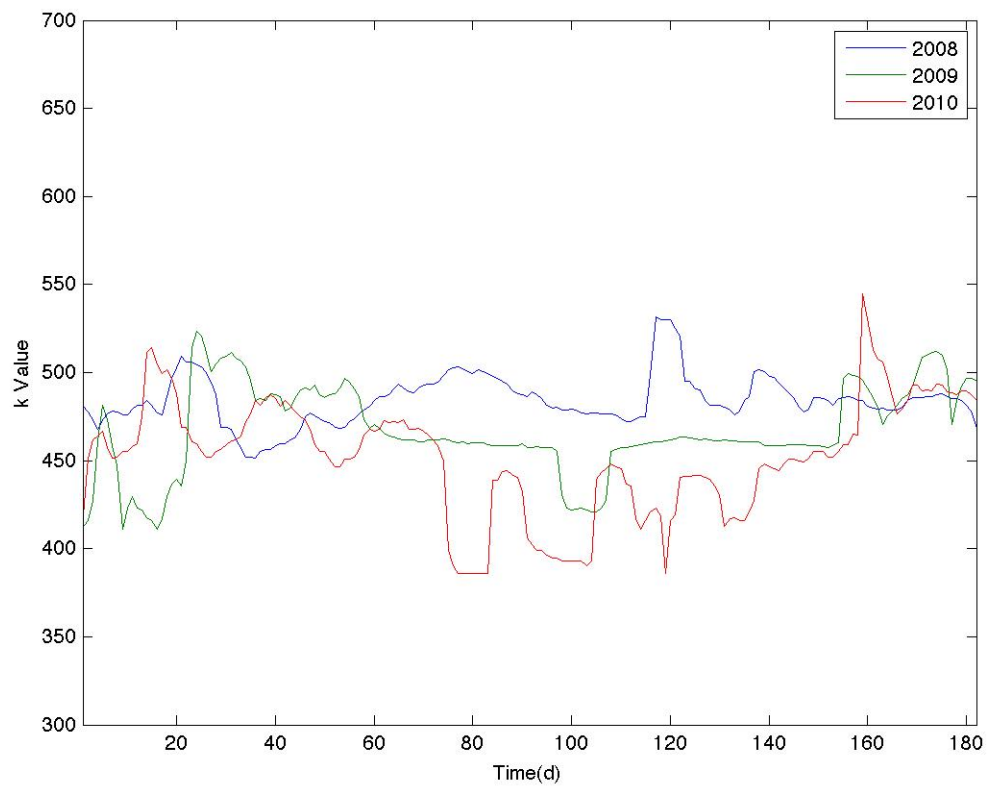
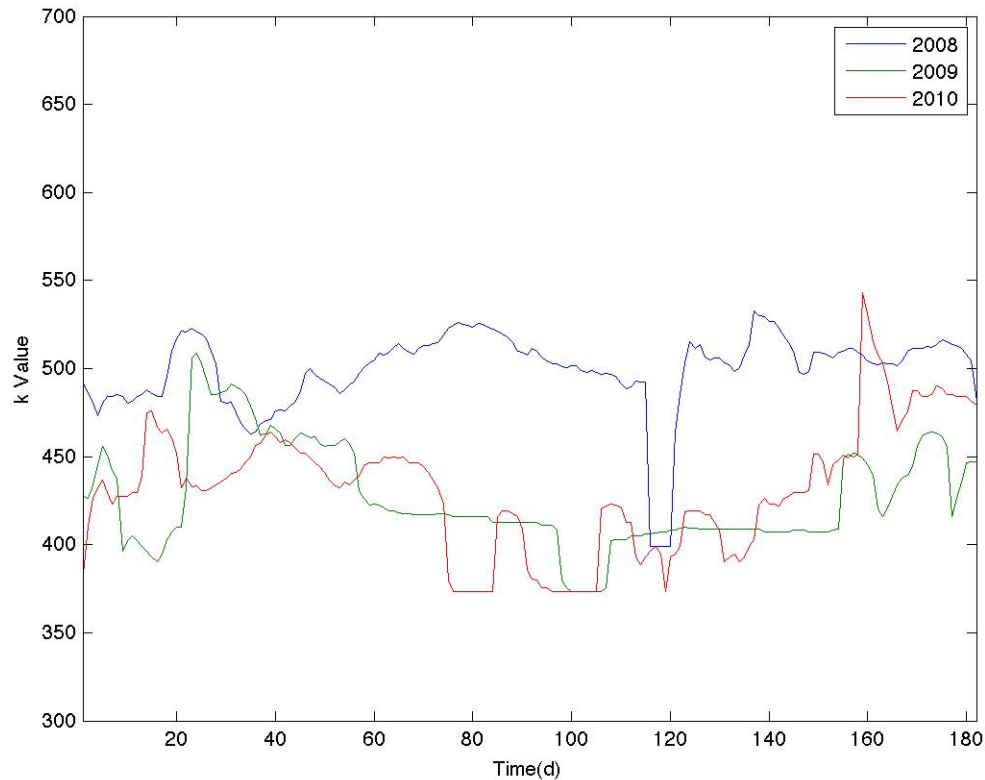


Figure 38 k Value With Updating Prior 2nd Heat Exchanger

To better analyze the heat exchange rate the heat exchangers are looked at only during times when the mass flow rate is high enough that uncertainty allows for k value detection. This is the case in winter months. K values are determined for each year from October to March for both heat exchangers as shown in





Both heat exchangers have a fluctuating k value during winter months, which makes it difficult to say that a fault clearly exists. Both heat exchangers operate at a level that is lower in the second and third year as compared with the first and cleaning them could be a possible solution for increased heat transfer.

6.2.3 Hydraulic Network

6.2.3.1 Bayesian Model Selection Analysis

For this analysis, eight models were compared to surrogate data sets to examine the capability of model selection. Figure 39 and Figure 40 display the results found for this analysis for linear and proportional valves

respectively. The four lines in the graph break the space into separate fault regions, a fault model was created at each intersection. The blue markers represent a fault free state; the green markers are for a high hysteresis and fault free state; the black markers represent both faults; and the red markers represent a high valve let by with low hysteresis. These parameters occupy a three-dimensional space existing on two separate planes, one for each valve type, however, to make the plot more easily readable values are plotted on two separate graphs. If the marker is solid this signifies that the valve type was correctly determined, where as a hollow marker represents the incorrect valve type being detected.

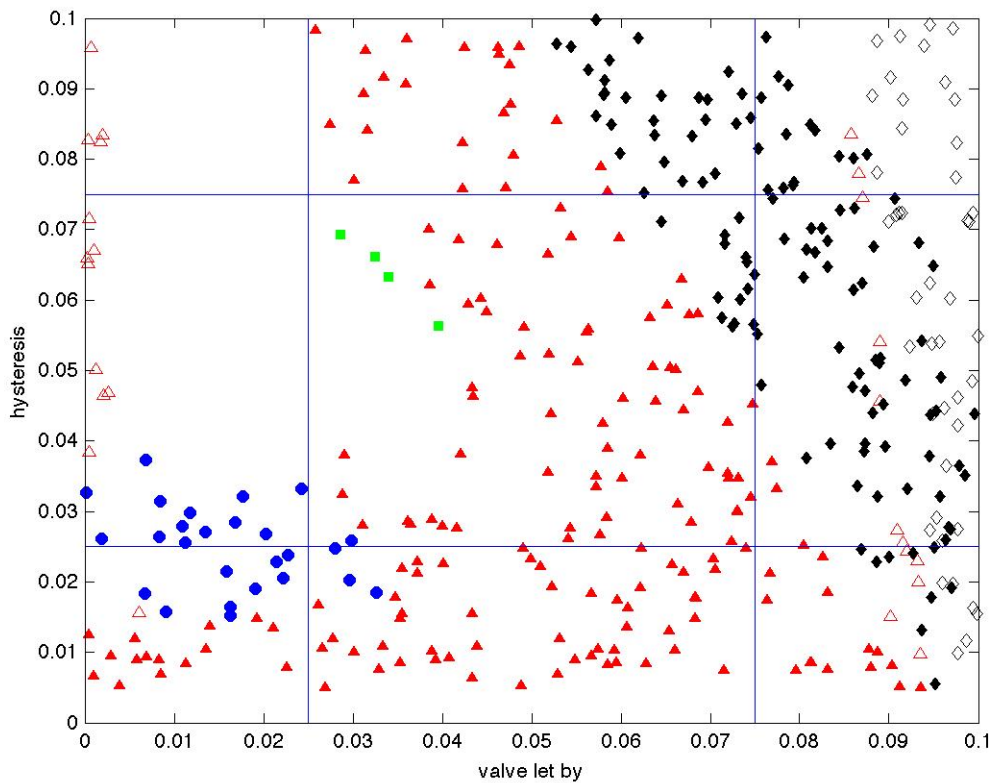


Figure 39 Model Selection for Linear Valve

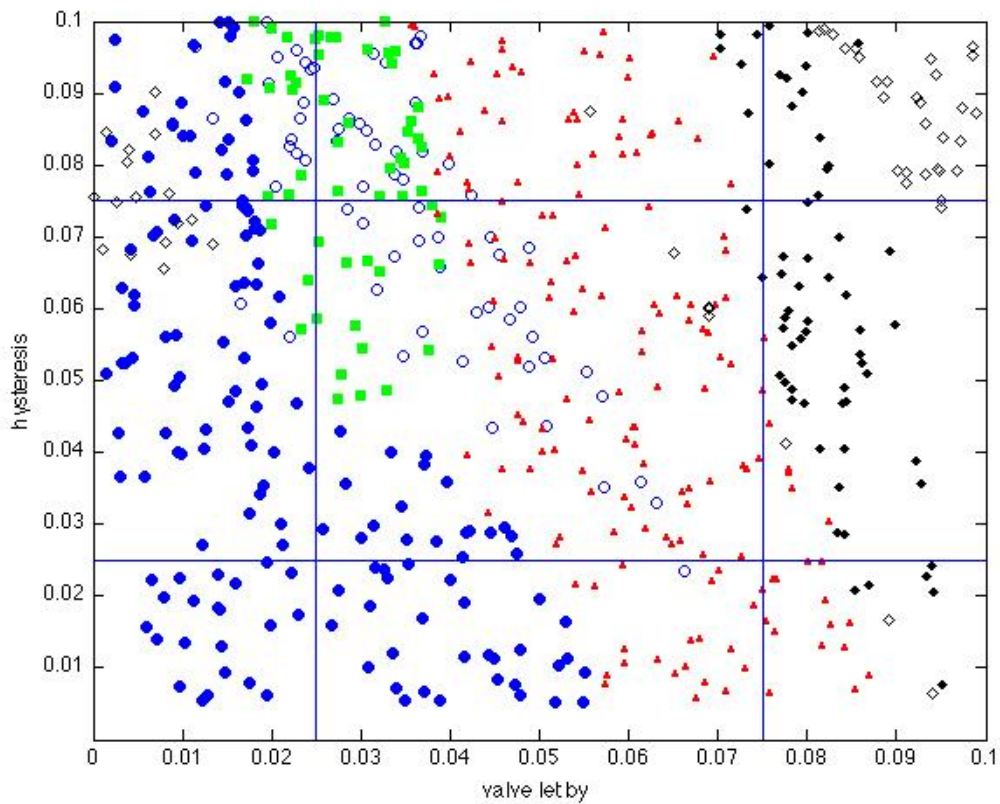


Figure 40 Model Selection for Proportional Valve

As shown in the previous figures, the most difficult detection is for high hysteresis. Ideally the plot would have points centered on each of the faults radiating out occupying $\frac{1}{4}$ of the parameter space.

7 CONCLUSIONS AND FURTHER WORK

7.1 Conclusions

The building sector is responsible for a large portion of energy consumption. The motivation of this thesis is potential energy reduction and increased thermal comfort of commercial buildings by detecting and diagnosing operational faults in hydronic heating systems. The goal of energy reduction and increased thermal comfort is examined with the help of models created in the Matlab Simulink environment to detect, diagnose and correct faults in the heating systems of commercial buildings.

A combination of surrogate data generated from building models, and measurements for a test building provided by the German ModQS project were used for this investigation.

Analysis of the Bezirksregierung Düsseldorf (BZRD) showed that often simple first steps such as using a nighttime and weekend temperature result in significant energy savings. A nighttime setback at the BZRD would result in a 20% energy savings during winter months. Model based analysis provides insight such as this, which can help show building managers quantitatively, potential energy and financial savings.

Many faults are much more difficult to detect and diagnose than improper control schemes, which can be sometimes detected by simply looking at measurements of room temperature. Current methods for fault detection systems in buildings are typically rule based. Creating rule based fault detection systems in commercial buildings is typically cost prohibitive, as it is time intensive and not scalable. This motivated research for Bayesian fault detection methods, which could be potentially scalable and take less time to develop a fault detection scheme for an individual building.

When detecting faults using surrogate data, Bayesian fault detection proved effective for a variety of examples including boiler efficiency, heat exchanger fouling and hydraulic valve faults. The effectiveness of fault detection was found to increase when more measurements were available and with more models to compare to. When only a single, or insufficient measurement values are available tradeoff effects become apparent and fault detection becomes difficult. Possible faults should be isolated using measurement values such as temperature and mass flow rate, which both offer unique information about the system.

Although Bayesian fault detection in the building sector has some promising results, many complications still exist. One such example is determining the certainty of a fault correctly: The posterior distribution may allow a user to determine the most likely parameter value or model, but describing the certainty of the most likely value can be challenging. The

posterior distribution is shaped by the likelihood function as well as the prior distribution. The width of likelihood function will define how likely one parameter is compared to another. Having this as accurate as possible is critical because it informs the user of the certainty of the detected parameter or model. This is also important when decision theory is applied since the course of action may differ if a building manager is informed that a fault is 10% certain or 80% certain. In this thesis, a method of combining model and measurement uncertainty to determine the width of the likelihood function was developed. This method was found to be unsuccessful for parameter estimation as the varying distribution width, caused by a varying sigma value, resulted in the values with the least model uncertainty to be predicted even if they were not the most likely.

Bayesian model selection was also found to be somewhat successful, although some faults were mischaracterized. The advantage of this method is model speed, both in developing the models, which are compared to measured or surrogate data and in computing the associated posterior values.

7.2 Future Work

Currently, this work helps provide a background for some of the potential methods for Bayesian fault detection applied to commercial buildings. However, there is still much work that is necessary for this to advance from pure research to something that could be useful in the field.

One key area of interest is developing a more automated and accurate way of determining σ , the standard deviation for the likelihood distribution. Research of the uncertainty associated with a detected fault will be critical in making Bayesian parameter estimation and model selection a useful fault detection method.

Another area that should be further explored is which measurements are most useful for determining different faults. This is important because the cost purchasing of installing measurement equipment in buildings should be reduced as much as possible. Only installing measurement devices that are necessary will help make this technology cost effective. Certain fault sets may require different measurement sets to be correctly detected. On future test buildings, experiments could be set up to ensure adequate but not extraneous measurements are taken for the faults being investigated.

Another issue that could be investigated further is model selection. In this thesis model complexity vs. computational resource was not addressed systematically. It is possible to alter model complexity or model structure then, with Bayesian analyses, determine which model best fits measurement data. This is called calculating the model odds. The basic process consists of running each model, calculating the posterior for each model, and then comparing the posteriors for each different model structure or complexity. This addresses how well a model represents measured data but does not take into account computational resource. To address computational resource one

could apply methods from Akaike information criterion(AIC). Here the number of parameters is combined with the likelihood function to include model complexity as well as goodness of fit.

Numerical issues can also be an obstacle to fault detection. When analyzing large time series data sets it is often the case that Matlab does not have adequate memory to store all necessary values or values become very small and approach zero. Enhancing the numerical techniques and coding used to make a more robust system is still a high priority.

Once the code is made more robust, and some of the techniques used are further refined, then it will be necessary to apply this to more test buildings. The building analyzed for this thesis, the BZRD, gave some useful insight, but may have lacked measurements that could lead to a more in-depth investigation of faults.

Within the scope of this project, only faults related to hydronic heating systems were analyzed. Bayesian methods should also be tested and developed for other building faults related to cooling and ventilation.

8 WORKS CITED

A.T. Booth, R. Choudhary, D.J. Spiegelhalter, Handling uncertainty in housing stock models, *Building and Environment*, Volume 48, February 2012, Pages 35-47, ISSN 0360-1323, 10.1016/j.buildenv.2011.08.016. (<http://www.sciencedirect.com/science/article/pii/S0360132311002599>)
Keywords: Uncertainty; Energy; Housing; Stock; Bayesian; Calibration

“Buildings Energy Data Book”, n.d.,
<http://buildingsdatabook.eere.energy.gov/ChapterIntro1.aspx?1#1>.

D&R International, Ltd. . *2010 Buildings Energy Data Book*. U.S. Department of Energy, 2011.

Chen, Jie, and R.J. Patton. *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Norwell: Kluwer Academic Publishers, 1999.

Cohn, Martin, and Gene Ott. "Design of Adaptive Procedures for Fault Detection and Isolation ." *Microelectronics Reliability* , April 1972: 115.

Isermann, Rolf. *Fault-Diagnosis Systems*. Darmstadt, 2006.

Isermann, Rolf. "Process Fault Detection Based on Modeling and Estimation Method-A Survey." *Automatica*, 1984: 387-404.

Hyvärinen, J., and S. Käarki. *Final Report Vol 1: Building Optimization and Fault Diagnosis Source Book*. Finland, Technical Research Centre of Finland, Building Technology, 1996.

Jones, Harold Lee. *Failure detection in linear systems*. Boston: Massachusetts Institute of Technology, 1973.

Katipamula, Srinivas, and Michael R. Brambley. "Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems - A Review, Part 1." *HVAC&R Research* 11, no. 2 (January 2005A).

Katipamula, Srinivas, and Michael R. Brambley. "Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems- A Review, Part 2." *HVAC&R Research* 11, no. 2 (April 2005B).

Leonhardt, S., and M. Ayoubi. "Methods of Fault Diagnosis ." *Control Eng. Practice* 5, no. 5 (1997): 683-692.

Li, Shun. "A Model-Based Fault Detection and Diagnostic Methodology for Secondary HVAC Systems." 2009.

Choudhary, Ruchi. 2011. "Energy Analysis of the Non-Domestic Building Stock of Greater London." *Building and Environment* (September 15).

Schmidt, Max. "Modellierung des Wärmetauschers des Bezirksregierungsgebäudes in Düsseldorf." 2011.

Technical University of Denmark. "Thermal Comfort in Buildings with Low-Energy Cooling." 2009.

Haitao Wang et al., "A robust fault detection and diagnosis strategy for pressure-independent VAV terminals of real office buildings," *Energy and Buildings* 43, no. 7 (July 2011): 1774-1783.

Haitao Wang et al., "A robust fault detection and diagnosis strategy for pressure-independent VAV terminals of real office buildings," *Energy and Buildings* 43, no. 7 (July 2011): 1774-1783.

House, John M., Lee, Won Yong, and Shin, Dong Ryul, "Classification Techniques for Fault Detection and Diagnosis of an Air-Handling Unit" (ASHRAE AMERICAN SOCIETY HEATING REFRIGERATING, 1999).

Zhimin Du and Xinqiao Jin, "Detection and diagnosis for multiple faults in VAV systems," *Energy and Buildings* 39, no. 8 (August 2007): 923-934.

Zhimin Du and Xinqiao Jin, "Detection and diagnosis for multiple faults in VAV systems," *Energy and Buildings* 39, no. 8 (August 2007): 923-934.

Sylvain Lalot and Halldór Pálsson, "Detection of fouling in a cross-flow heat exchanger using a neural network based technique," *International Journal of Thermal Sciences* 49, no. 4 (April 2010): 675-679.

Zhimin Du and Xinqiao Jin, "Detection and diagnosis for multiple faults in VAV systems," *Energy and Buildings* 39, no. 8 (August 2007): 923-934.

Youming Chen and Lili Lan, "Fault detection, diagnosis and data recovery for a real building heating/cooling billing system," *Energy Conversion and Management* 51, no. 5 (May 2010): 1015-1024.

Zhimin Du and Xinqiao Jin, "Detection and diagnosis for multiple faults in VAV systems," *Energy and Buildings* 39, no. 8 (August 2007): 923-934.

Zhimin Du and Xinqiao Jin, "Detection and diagnosis for multiple faults in VAV systems," *Energy and Buildings* 39, no. 8 (August 2007): 923-934.

Ancha Xu and Yincai Tang, "Objective Bayesian analysis of accelerated competing failure models under Type-I censoring," *Computational Statistics & Data Analysis* 55, no. 10 (October 1, 2011): 2830-2839.

Stojan Persin and Boris Tovornik, "Real-time implementation of fault diagnosis to a heat exchanger," *Control Engineering Practice* 13, no. 8 (August 2005): 1061-1069.

9 APPENDIX A

9.1 Model Selection Matlab

Code

OuterModel.m

```
clear all
clc
global i
% run the parameter definition script to
initialize parameter values
param_def_model
% run thermdat to initialize the model
geometry and description of model
% parameters
thermdat

%run through varying model parameters to
generate models
for i = 1:8
    i

    % run the parameter definition script to
determine model number, valve
    % type, valve let by and valve
hysteresis. the global i variable
    % causes these to change as described
by "param_def_model"
    param_def_model

    %store values in a vector
samps=[M Vt1 Vlb1 Vh1];

    %call the bmodel function and take the
samples and run the simulation
    [mdl_out] = bmodel(samps);

    % store sample information with the
model out
```

```
mdl_out.samps = samps;

% save the data as the output
containing all info on parameters and
% measurement values from the model
output(i) = mdl_out;

end

%Generate the surrogate data to compare
with
for i=1:1000
    i
    %run the surrogate data generator
surrogate_data_generator

    %Store the original measurement
values here along with the randomly
%sampled parameters that describe the
model
    Parameter(i,:)={Vt1 Vlb1 Vh1
data_surrogate};
    %Store measurement values centered
around a normal distribution for the
%three chosen methods
    Parameternew(i,:)={Vt1 Vlb1 Vh1
dataalt(:,1) dataalt(:,2) dataalt(:,3)};
end

%Manually save Parameter,
Parameternew and output
```



```

param_def_model.m

Vt1=1;
Vlb1=.025;
Vh1=.075;

global i
% the number of models?
num_models = 8;

%Define the other valves in the hydraulic
system
Vt2=2;
Vlb2=.01;
Vh2=.01;

Vt3=2;
Vlb3=.01;
Vh3=.01;

Vt4=2;
Vlb4=.01;
Vh4=.01;

Vt5=2;
Vlb5=.01;
Vh5=.01;

Vt6=2;
Vlb6=.01;
Vh6=.01;

%Cycle through pre defined models

if i==1
    %000
    % model 1
    %Vt1=1
    %Vlb1=.01
    %Vh1=.005
    M=1;
    Vt1=1;
    Vlb1=.025;
    Vh1=.025;

elseif i==2
    %001
    % model 2
    %Vt1=1
    %Vlb1=.01
    %Vh1=.005
    M=2;

elseif i==3
    %010
    % model 3
    %Vt1=1
    %Vlb1=.01
    %Vh1=.005
    M=3;
    Vt1=1;
    Vlb1=.075;
    Vh1=.025;

elseif i==4
    %011
    % model 4
    %Vt1=1
    %Vlb1=.01
    %Vh1=.005
    M=4;
    Vt1=1;
    Vlb1=.075;
    Vh1=.075;

elseif i==5
    %100
    % model 5
    %Vt1=1
    %Vlb1=.01
    %Vh1=.005
    M=5;
    Vt1=2;
    Vlb1=.025;
    Vh1=.025;

elseif i==6
    %101
    % model 6
    %Vt1=1
    %Vlb1=.01
    %Vh1=.005
    M=6;

```

```

Vt1=2;
Vlb1=.025;
Vh1=.075;

elseif i==7
    %110
    % model 7
    %Vt1=1
    %Vlb1=.01
    %Vh1=.005
    M=7;
    Vt1=2;
    Vlb1=.075;
    Vh1=.025;

else
    %111
    % model 8
    %Vt1=1
    %Vlb1=.01
    %Vh1=.005
    M=8;
    Vt1=2;
    Vlb1=.075;
    Vh1=.075;

end

bmodel.m

function [mdl_out]=bmodel(samps)

%initialize valves for the other circuits
Vt2=2;
Vlb2=.01;
Vh2=.01;

Vt3=2;
Vlb3=.01;
Vh3=.01;

Vt4=2;
Vlb4=.01;
Vh4=.01;

Vt5=2;
Vlb5=.01;
Vh5=.01;

Vt6=2;
Vlb6=.01;
Vh6=.01;

%define the necessary parameters for the
first circuit

M = samps(1)

Vt1=samps(2)

Vlb1=samps(3)

Vh1=samps(4)

% options to take simulation parameters
from function workspace
options =
simset('SrcWorkspace','current');
% run the simulink model

sim('HydSim_1_DeltaTDegr_2_HydVar_R
2008b_model',[],options);
% function output becomes necessary
measurements from model out block in

```

```

% simulink
mdl_out = Measurements;

end

```

SurrogateData.m

```

%% Create Surrogate Data

%Determine valve values for surrogate
data using rand
%valve type can be 1 or 2
Vt1=round(rand)+1;
%valve let by is in the interval 0 to .01
Vlb1=rand/10;
%valve hysteresis is on the interval .005 to 1
Vh1=.005 + (.1-.005) * rand;

%initialize valves
Vt2=2;
Vlb2=.01;
Vh2=.01;

Vt3=2;
Vlb3=.01;
Vh3=.01;

Vt4=2;
Vlb4=.01;
Vh4=.01;

Vt5=2;
Vlb5=.01;
Vh5=.01;

Vt6=2;
Vlb6=.01;
Vh6=.01;

%options to take simulation parameters
from workspace
options = simset('SrcWorkspace','current');
% run the simulink model
sim('HydSim_1_DeltaTDegr_2_HydVar_R
2008b_model',[],options);
%store the data from the simulation
data_surrogate = Measurements;

%store variable data
data=data_surrogate.signals.values;

%create an empty matrix for data alt
dataalt(1:length(data),1:3)=zeros;

%create new surrogate values by sampling
from a normal distribution, data

```

```

%stored her is for room temp, circuit mass
flow rate, and combined
%return temperature from all zones,
values are only allowed to be positive
%since it is assumed the flow is not ever
backwards through the circuit

```

```

dataalt(1:length(data),1)=abs(randn(lengt
h(data),1)*.1+data(:,1));
dataalt(1:length(data),2)=abs(randn(lengt
h(data),1)*.1+data(:,7));
dataalt(1:length(data),3)=abs(randn(lengt
h(data),1)*.05+data(:,8));

```

surrogate_data_generator.m

```

%%Create Surrogate Data

```

```

%Determine valve values for surrogate
data using rand
%valve type can be 1 or 2
Vt1=round(rand)+1;
%valve let by is in the interval 0 to .01
Vlb1=rand/10;
%valve hysteresis is on the interval .005 to 1
Vh1=.005 + (.1-.005) * rand;

```

```

%initialize valves

```

```

Vt2=2;
Vlb2=.01;
Vh2=.01;

```

```

Vt3=2;
Vlb3=.01;
Vh3=.01;

```

```

Vt4=2;
Vlb4=.01;
Vh4=.01;

```

```

Vt5=2;
Vlb5=.01;
Vh5=.01;

```

```

Vt6=2;
Vlb6=.01;
Vh6=.01;

```

```

%options to take simulation parameters
from workspace

```

```

options = simset('SrcWorkspace','current');
% run the simulink model
sim('HydSim_1_DeltaTDegr_2_HydVar_R
2008b_model',[],options);
%store the data from the simulation
data_surrogate = Measurements;

```

```

%store variable data

```

```

data=data_surrogate.signals.values;

```

```

%create an empty matrix for data alt
dataalt(1:length(data),1:3)=zeros;

%create new surrogate values by sampling
from a normal distribution, data
%stored her is for room temp, circuit mass
flow rate, and combined
%return temperature from all zones,
values are only allowed to be positive
%since it is assumed the flow is not ever
backwards through the circuit

dataalt(1:length(data),1)=abs(randn(lengt
h(data),1)*.1+data(:,1));
dataalt(1:length(data),2)=abs(randn(lengt
h(data),1)*.1+data(:,7));
dataalt(1:length(data),3)=abs(randn(lengt
h(data),1)*.05+data(:,8));

```

```

Bayes_a_model.m

clear all
clc
% initialization information
bayes_def_model

% load necessary measurements and
model
load output
load Parameter

%determine desired timestep to compare
values
t2=(0:15:2*24*3600);

%initialize normalization factor
normlikelihood(1:4)=0;

%create derivative matrix one digit longer
to make it the same length as
%other vectors
dydxmod(length(t2),1)=zeros;
dydxmeas(length(t2),1)=zeros;

% loop through all surrogate data
for k=1:1000
    k
    %loop through each model to compare
with surrogate data
    for i = 1:n_models

        %loop through the desired number of
measurements
        for j=1:n_measurements

            %run a interpolation function to
adjust model to desired time
            %step from output values which
are variable time step

tseries=interp1(output(1,i).time,output(1,i)
).signals.values,t2);

            %adjust surrogate data to the same
time step as in the model

```

```
M=interp1(Parameter{k,4}.time,Parameter
r{k,4}.signals.values,t2);
```

```

% calculate the squared-difference
between time series and the
% calibration target
% a variety of measurement values
are stored, the vector B
% allows the user to decide which
measurements to compare
% 1 7 and 8 correspond to zone
temperature, circuit flow rate
% and total return flow
temperature
B=[1 7 8];
```

```
diff_sq = (M(:,B) - tseries(:,B)).^2;
```

```

% a fourth measurement, the
derivative approximation of the
% mass flow rate is added as an
attempt to better capture valve
%characteristics
dydxmod(1:length(t2)-
1,1)=diff(tseries(:,8));
dydxmeas(1:length(t2)-
1,1)=diff(M(:,8));
```

```

%calculate diff of squares for this
diff_sq(:,4) = (dydxmod -
dydxmeas).^2;
```

```

%measurements are available for a
two day period, however only
%one day is analyzed as new
information is not available from
%the first day.
```

```

%Determine the likelihood from
this function:
```

```

%likelihood(i,j)=(1/(sigma_eps(j)*sqrt(2*pi)
)^n)...
%*exp(-
0.5*(1/sigma_eps(j)^2)*sum(diff_sq(1440:2
880,j)));
%the
(1/(sigma_eps(j)*sqrt(2*pi))^n) term can
be removed since
%the likelihood function is
```

normalized and this value is

```

%constant
likelihood(i,j)= exp(-
0.5*(1/sigma_eps(j)^2)*sum(diff_sq((round
(length(t2)/2):length(t2),j)));
```

```

%prior is uniform and in this case
doesn't effect the
%posterior, however it is left in so
it can be adjusted if
%desired
prior(i,j)=1/n_models;
```

```

%normalization factor is
determined for each measurement set
```

```
normlikelihood(j)=likelihood(i,j)+normlikel
ihood(j);
```

```
end
```

```
end
```

```

%the simulations are run through again
for comparison now that they are
% normalized
for i = 1:n_models
for j=1:n_measurements
```

```

%the normalized likelihood
function can have Nan values if the
%differnce of the squares is large,
if an entire measurement
%set is Nan values it is ignored by
setting all of the values
%equal to 1/n_models
```

```
likelihood(i,j)=likelihood(i,j)/normlikelihoo
d(j);
```

```

if isnan(likelihood(:,j)) ==1
likelihood(:,j)=1/n_models;
end
```

```
end
```

```

%posterior values are determined for
each measurement
```

```
posterior1(i,1)=likelihood(i,1).*(prior(i,1));
```

```
posterior2(i,1)=likelihood(i,2).*(prior(i,2));
```

```
posterior3(i,1)=likelihood(i,3).*(prior(i,3));
```

```
posterior4(i,1)=likelihood(i,4).*(prior(i,4));
```

```
%the posterior values are then  
combined,the various posterior  
%values are used to compare which  
method best describes the system
```

```
posteriora(i,1)=posterior3(i,1)*posterior1(i,  
1)*posterior4(i,1)*posterior2(i,1);
```

```
posteriorb(i,1)=posterior3(i,1)*posterior1(i,  
1)*posterior4(i,1);
```

```
posteriorc(i,1)=posterior3(i,1)*posterior1(i,  
1)*posterior2(i,1);
```

```
p(1:4,i)=[posterior1(i,1),  
posterior2(i,1), posterior3(i,1),  
posterior4(i,1)];
```

```
posteriord(i,1)=mean(p(:,i));
```

```
end
```

```
%the maximum posterior is found to  
decide which model best represents
```

```
%the surrogate data
```

```
[r1
```

```
c1]=find(max(posteriora)==posteriora);
```

```
[r2
```

```
c2]=find(max(posteriorb)==posteriorb);
```

```
[r3
```

```
c3]=find(max(posteriorc)==posteriorc);
```

```
[r4
```

```
c4]=find(max(posteriord)==posteriord);
```

```
Most_likely_model(1,k)=r1;
```

```
Most_likely_model(2,k)=r2;
```

```
Most_likely_model(3,k)=r3;
```

```
Most_likely_model(4,k)=r4;
```

```
%normalization constant is reset here
```

```
normlikelihood(1:4)=0;
```

```
end
```

Plotfunction.m

```
%plot function
```

```
clc
```

```
clear all
```

```
clear figures
```

```
count(:,8)=zeros;
```

```
load Parameter
```

```
load most_likely_matrix
```

```
%extract values from cell
```

```
for i=1:1000
```

```
P(i,1)=Parameter{i,1};
```

```
P(i,2)=Parameter{i,2};
```

```
P(i,3)=Parameter{i,3};
```

```
end
```

```
%compare different Posterior methods
```

```
for m=1:4
```

```
Most_likely_model =
```

```
Most_likely_matrix(m,:);
```

```
%cycle through all models and sort the  
parameter values with the connected
```

```
%predicted model
```

```
for i=1:1000
```

```
if Most_likely_model(i)==1
```

```
count(1)=count(1)+1;
```

```
M1(count(1),1:3)=P(i,1:3);
```

```
elseif Most_likely_model(i)==2
```

```
count(2)=count(2)+1;
```

```
M2(count(2),1:3)=P(i,1:3);
```

```
elseif Most_likely_model(i)==3
```

```
count(3)=count(3)+1;
```

```
M3(count(3),1:3)=P(i,1:3);
```

```
elseif Most_likely_model(i)==4
```

```
count(4)=count(4)+1;
```

```
M4(count(4),1:3)=P(i,1:3);
```

```
elseif Most_likely_model(i)==5
```

```
count(5)=count(5)+1;
```

```
M5(count(5),1:3)=P(i,1:3);
```

```
elseif Most_likely_model(i)==6
```

```

count(6)=count(6)+1;
M6(count(6),1:3)=P(i,1:3);

elseif Most_likely_model(i)==7
count(7)=count(7)+1;
M7(count(7),1:3)=P(i,1:3);

elseif Most_likely_model(i)==8
count(8)=count(8)+1;
M8(count(8),1:3)=P(i,1:3);

end
end

%plot scatter points for predicted model
with a filled marker when the
%valve type prediction is correct
for i =1:length(M1)
if M1(i,1)==1
scatter(M1(i,2),M1(i,3),'b','filled')
axis([0 .1 0 .1])

hold on
else
scatter(M1(i,2),M1(i,3),'b')
axis([0 .1 0 .1])
hold on
end
end

%
for i =1:length(M2)
if M2(i,1)==1
scatter(M2(i,2),M2(i,3),'g','s','filled')
axis([0 .1 0 .1])

hold on
else
scatter(M2(i,2),M2(i,3),'g','s')
axis([0 .1 0 .1])
hold on
end
end

%
for i =1:length(M3)
if M3(i,1)==1
scatter(M3(i,2),M3(i,3),'r','^','filled')
axis([0 .1 0 .1])

hold on
else
scatter(M3(i,2),M3(i,3),'r','^')
axis([0 .1 0 .1])
hold on
end
end

%
for i =1:length(M4)
if M4(i,1)==1
scatter(M4(i,2),M4(i,3),'k','d','filled')
axis([0 .1 0 .1])

hold on
else
scatter(M4(i,2),M4(i,3),'k','d')
axis([0 .1 0 .1])
ylabel('hysteresis')
xlabel('valve let by')
hold on
end
end

%%
figure()
for i =1:length(M5)
if M5(i,1)==2
scatter(M5(i,2),M5(i,3),'b','filled')
axis([0 .1 0 .1])

hold on
else
scatter(M5(i,2),M5(i,3),'b')
axis([0 .1 0 .1])
hold on
end
end

%
for i =1:length(M6)
if M6(i,1)==2
scatter(M6(i,2),M6(i,3),'g','s','filled')
axis([0 .1 0 .1])

hold on
else
scatter(M6(i,2),M6(i,3),'g','s')
axis([0 .1 0 .1])
hold on
end
end

%

```



```

for i =1:length(M7)
if M7(i,1)==2
scatter(M7(i,2),M7(i,3),'r','^','filled')
axis([0 .1 0 .1])

hold on
else
scatter(M7(i,2),M7(i,3),'r','^')
axis([0 .1 0 .1])
hold on
end
end

%
for i =1:length(M8)
if M8(i,1)==2
scatter(M8(i,2),M8(i,3),'k','d','filled')
axis([0 .1 0 .1])

hold on
else
scatter(M8(i,2),M8(i,3),'k','d')
axis([0 .1 0 .1])
ylabel('hysteresis')
xlabel('valve let by')
hold on
end
end
figure()
clear Most_likely_model
end

```