MODELING THE THERMAL DYNAMICS OF A SINGLE ROOM IN COMMERCIAL BUILDINGS AND FAULT DETECTION

By

YASHEN LIN

A THESIS PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2012
To my parents and wife
ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Prabir Barooah for guiding me through my study. This thesis would have not been possible without his support and advice. He not only provided insightful advice in research topic, but also helped me become a rigorous and independent thinker. His emphasis and guidance on communication skills is especially beneficial to me as an international student. I feel very fortunate to have the opportunity to work with him and I would like to thank him for everything he has done for me.

I also want to extend my special gratitude to Dr. Timothy Middelkoop, who has been always supportive and helpful to me. I am grateful for his constructive advice and inspiring discussions as well as his crucial contributions to my research, especially in the experiments part. It is a pleasure to thank Dr. Herbert A. Ingley, for sharing his expertise. His patience, kindness, and knowledge, is extremely helpful in completing my work. I want to give special thanks to Peder Winkel, Skip Rockwell, and others from University of Florida (UF) Physical Plant Division who spared time helping with our experiments though they have many other commissions.

Also, I would like to thank my colleagues He Hao and Chenda Liao, who provided me much needed help and useful advice, both in research and personal life. I thank Siddharth Goyal for his help in the experiments, which would not have been possible otherwise. Last but not the least, I would like to thank my parents and wife. I am heartily thankful for their faith, devotion, love, support and encouragement.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>7</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>8</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>9</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td><strong>1</strong> INTRODUCTION</td>
<td>10</td>
</tr>
<tr>
<td>1.1 Motivation</td>
<td>10</td>
</tr>
<tr>
<td>1.2 Problem Formulation</td>
<td>11</td>
</tr>
<tr>
<td>1.2.1 Modeling A Single Room</td>
<td>11</td>
</tr>
<tr>
<td>1.2.1.1 Q1: Acceptable accuracy</td>
<td>12</td>
</tr>
<tr>
<td>1.2.1.2 Q2: Model structure</td>
<td>12</td>
</tr>
<tr>
<td>1.2.1.3 Q3: Parameter estimation</td>
<td>13</td>
</tr>
<tr>
<td>1.2.2 Fault Detection</td>
<td>14</td>
</tr>
<tr>
<td><strong>2</strong> MODEL STRUCTURE</td>
<td>16</td>
</tr>
<tr>
<td>2.1 General Settings</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Full-scale Model</td>
<td>18</td>
</tr>
<tr>
<td>2.3 Low-order Models</td>
<td>20</td>
</tr>
<tr>
<td>2.3.1 First-order Model</td>
<td>20</td>
</tr>
<tr>
<td>2.3.2 Second-order Model</td>
<td>20</td>
</tr>
<tr>
<td>2.3.3 The Linear Time Invariant (LTI) Case</td>
<td>21</td>
</tr>
<tr>
<td>2.4 Model Structure Comparison</td>
<td>21</td>
</tr>
<tr>
<td>2.4.1 ASHRAE Values</td>
<td>22</td>
</tr>
<tr>
<td>2.4.2 Time Domain Comparison</td>
<td>23</td>
</tr>
<tr>
<td>2.4.3 Frequency Domain Comparison</td>
<td>25</td>
</tr>
<tr>
<td><strong>3</strong> CALIBRATION AND VALIDATION</td>
<td>28</td>
</tr>
<tr>
<td>3.1 Field Data</td>
<td>28</td>
</tr>
<tr>
<td>3.1.1 Test Bed</td>
<td>28</td>
</tr>
<tr>
<td>3.1.2 Data Sets</td>
<td>30</td>
</tr>
<tr>
<td>3.2 Identification Methods</td>
<td>31</td>
</tr>
<tr>
<td>3.2.1 Least-squares</td>
<td>31</td>
</tr>
<tr>
<td>3.2.2 Maximum Likelihood (ML) Method</td>
<td>32</td>
</tr>
<tr>
<td>3.3 Model Calibration and Validation</td>
<td>35</td>
</tr>
<tr>
<td>3.3.1 Attempt 1: Apparent Success But Really A Failure</td>
<td>35</td>
</tr>
<tr>
<td>3.3.2 Attempt 2: A More Reliable Calibration</td>
<td>38</td>
</tr>
<tr>
<td>3.4 Effect of Open Door</td>
<td>40</td>
</tr>
<tr>
<td>3.5 Summary</td>
<td>44</td>
</tr>
<tr>
<td><strong>4</strong> FAULT DETECTION</td>
<td>45</td>
</tr>
<tr>
<td>4.1 Unknown Input Observer</td>
<td>45</td>
</tr>
<tr>
<td>4.2 Implementation Example</td>
<td>47</td>
</tr>
</tbody>
</table>
4.2.1 Room Model .............................................. 48
4.2.2 Implementation ........................................ 49
4.2.3 Analysis .................................................... 50
4.3 Door Status Detection ................................. 53
4.4 Summary ..................................................... 56

5 CONCLUSION AND FUTURE WORK ..................... 57

REFERENCES ...................................................... 58

BIOGRAPHICAL SKETCH ........................................ 61
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Best fit parameters when data set A is used for model calibration.</td>
<td>36</td>
</tr>
<tr>
<td>3-2</td>
<td>Best fit parameters when data set C is used for model calibration.</td>
<td>38</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2-1</td>
<td>Schematic figure of the room</td>
<td>17</td>
</tr>
<tr>
<td>2-2</td>
<td>Full-scale model structure</td>
<td>19</td>
</tr>
<tr>
<td>2-3</td>
<td>Simulation of Room 241 with ASHRAE values as parameters</td>
<td>23</td>
</tr>
<tr>
<td>2-4</td>
<td>Simulation of Room 243 with ASHRAE values as parameters</td>
<td>23</td>
</tr>
<tr>
<td>2-5</td>
<td>Time domain comparison of different models</td>
<td>24</td>
</tr>
<tr>
<td>2-6</td>
<td>Gain comparison of the three models</td>
<td>26</td>
</tr>
<tr>
<td>2-7</td>
<td>The error defined in (2–11) for the two inputs $T_s$ and $T_{sd}$</td>
<td>26</td>
</tr>
<tr>
<td>3-1</td>
<td>Screen shot of Insight Workstation</td>
<td>29</td>
</tr>
<tr>
<td>3-2</td>
<td>Floor plan of testing area</td>
<td>30</td>
</tr>
<tr>
<td>3-3</td>
<td>HOBO sensors</td>
<td>30</td>
</tr>
<tr>
<td>3-4</td>
<td>Three data sets collected in Pugh Hall.</td>
<td>32</td>
</tr>
<tr>
<td>3-5</td>
<td>Calibration/validation when data set A is used for calibration.</td>
<td>36</td>
</tr>
<tr>
<td>3-6</td>
<td>Contour of the cost $J$ defined in (3–2), which shows its non-convexity.</td>
<td>37</td>
</tr>
<tr>
<td>3-7</td>
<td>Calibration/validation when data set C is used for calibration.</td>
<td>38</td>
</tr>
<tr>
<td>3-8</td>
<td>DC gain comparison, left figure is for first attempt, right is for second attempt</td>
<td>39</td>
</tr>
<tr>
<td>3-9</td>
<td>Simulation with original model</td>
<td>41</td>
</tr>
<tr>
<td>3-10</td>
<td>Structure for the model between hallway and room</td>
<td>42</td>
</tr>
<tr>
<td>3-11</td>
<td>Simulation with moderated model</td>
<td>43</td>
</tr>
<tr>
<td>3-12</td>
<td>Prediction error with respect to $R_{od}$</td>
<td>43</td>
</tr>
<tr>
<td>3-13</td>
<td>Prediction error with respect to $R_{od}$</td>
<td>44</td>
</tr>
<tr>
<td>4-1</td>
<td>Simulation of UIO</td>
<td>52</td>
</tr>
<tr>
<td>4-2</td>
<td>Estimated states from UIO</td>
<td>53</td>
</tr>
<tr>
<td>4-3</td>
<td>Open door detection</td>
<td>55</td>
</tr>
</tbody>
</table>
Abstract of Thesis Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Master of Science

MODELING THE THERMAL DYNAMICS OF A SINGLE ROOM IN COMMERCIAL BUILDINGS AND FAULT DETECTION

By

Yashen Lin

August 2012

Chair: Prabir Barooah
Major: Mechanical Engineering

Modern control strategy in building automation control has large potential of saving energy. This thesis attempts to develop a model which describes the thermal dynamics of a single room that can be used for intelligent control, such as model predictive control. The use of the model in fault detection is discussed in the second part of this thesis.

In the first part, three important questions in the modeling problem are discussed: (1) How accurate should the model be to be useful for modern control methods? (2) What model structure should be used so that the desired accuracy can be achieved and yet it is not too complicated? (3) How should we determine the parameters in the model once the model structure is chosen? Various model structures are compared, and a second order model using electrical analogy is found to be a good choice. Experimental data from Pugh Hall on University of Florida are collected for model calibration and validation. It is found that the data is crucial. Normal closed loop operation data may lead to grossly wrong parameter estimation, even with rich excitation. The effect of open door on the model is also discussed, and a modified model is developed.

In the second part, the fault caused by occupants is studied. Unknown input observer method (UIO), a model based method, is implemented for fault detection. Using the model we got from the first part, we successfully detected the open door with UIO. However, implementation in another model structure shows that the technique has certain limitation. A tradeoff is found in such situation.
CHAPTER 1
INTRODUCTION

1.1 Motivation

Improving energy efficiency in buildings is a topic that garners much attention nowadays. As one of the primary energy consumer, buildings account for 34% of total energy use in the United States and heating, ventilation, and air conditioning (HVAC) system accounts for roughly half of that [11]. Inefficient technology in operating buildings, in particular, HVAC system, causes significant energy waste. Commercial buildings, usually large in size, have great potential to save energy if more efficient control strategy is implemented. Currently, HVAC is mostly operated under a pre-designed schedule with fixed temperature set points for zones in the building and local proportional-integral-derivative (PID) controller is often used to maintain such zone climate. This relatively simple operation strategy provides large room for improvement, and model-based control of buildings has generated excitement in the community of control researchers in recent years. The most popular candidate for control of building HVAC systems is MPC (Model Predictive Control), in which dynamical models are used to predict and optimize system performance. A slew of recent papers have been published on this topic [4, 6, 15, 18, 19, 23–25, 27].

In order to be able to implement MPC, or many other novel control algorithms, a good model is required. It is natural to start from the most basic unit, namely, a single room. Accuracy and complexity are two major metric to measure how good a model is. Clearly, the model has to have certain level of accuracy to be useful. If the model cannot predict the states accurately, we cannot use it for prediction or control. At the same time, the model should not be too complicated. For the building system, the model size can grow very fast. For example, if the model of a single room has 10 states, there will be 1000 states for a building with 100 rooms. Moreover, more intelligent control strategy tends to have higher computation cost. Thus it is vital to keep the model size small, so that such control algorithms can be implemented in real time. One of the objectives of this thesis is to develop a model with good balance between accuracy and complexity.
Fault diagnosis is an important task in building control system. From safety aspect, device failure in HVAC system may cause discomfort and danger to the occupants. For instance, the system may cool down or heat up the supply air too much, blow too little air into rooms, etc. It is critical to be able to detect these faults and take corresponding measures to solve the problem. From energy efficiency aspect, unnecessary energy may be spent due to malfunctioning devices. For example, if the reheat valve is stuck, the system may heat up the supply air while it is actually trying to cool down the room. Another type of fault is caused by occupants, such as windows or doors being left open when the A/C is on, which will also cause energy wastage. The latter part of this thesis discusses a fault diagnosis technique and its implementation to detect if a door is open or closed.

1.2 Problem Formulation

1.2.1 Modeling A Single Room

We focus our attention on the modeling of a single room in a multi-zone commercial building with VAV (variable air volume) system, since sophisticated control methods - such as MPC - are more likely to make an impact here than in residential buildings or older commercial buildings with CAV (constant air volume) systems. Some researchers attempted to model the building as a whole \[9, 17, 32\]. These models are useful for the analysis of total energy consumption, but fall short in zone level control. Others studied computational fluid dynamics (CFD) models for rooms in building \[3, 16\]. In CFD models, the room is divided into many zones, and the air flow between zones inside the room is studied. However, for most rooms in buildings, there are no mechanism to control different areas inside the room, especially for typical offices that do not have large space. Thus, from the perspective of room level control, a single room is the most basic unit. Once we have good understanding of the model of a single room, we can use it as a “building block” to build model for a zone with several rooms, or even the whole building.

There are three questions that are of importance: (1) How accurate should the model be so that it can be used for MPC? (2) What model structure should be used so that the desired accuracy can be achieved yet the model is not too complex? (3) How to determine the parameters in the model once the model structure is chosen? We will take a closer look at each of these questions.
1.2.1.1 Q1: Acceptable accuracy

Irrespective of what model structure is chosen or what parameter estimation method is used, there will always be discrepancy between the model’s prediction and what is measured. The task of modeling is to find a useful model, not a correct model. Thus the criterion to measure a model closely depends on how we are going to use the model. In this room modeling problem, our goal is to use it for MPC, so there are two major aspects that we concern: prediction error and rate of response. For the former, we consider the peak of prediction error less of $3^\circ F$ to be the acceptable accuracy. This is because the room temperature is not uniform in reality, the difference between different areas inside the room can be $3^\circ F$ or more, so asking for better accuracy in the model is unreasonable. The rate of response is also important. It is possible that a model have small peak error, but the temperature changes much slower then reality, which is clearly not good for control design. For example, MPC predicts the temperature using the model to determine the optimal inputs. If the model is slower, the controller may decides to blow conditioned air into the room for longer time since it thinks that is required to bring the room temperature to desired value.

1.2.1.2 Q2: Model structure

White-box model, black-box model, and grey-box model are three types of commonly used model. White-box model is the most detailed in the three. It starts from first principle in the physical process, and make as few assumptions as possible. It is the closest description of reality and is easy to change. However, the underlying physics is usually very complicated, resulting in high order and high dimension model which requires enormous computational resource when applying. Another problem with this approach is that pure white-box model does not exist. There are always some unknown and uncertainties, and some information in the physical process may not be available. Take the room modeling as example, the furniture and occupants activity inside a room is hard to get, and often time varying. Assumptions are needed in those cases, and they may not be good enough.

Black-box model, on the contrary, does not touch the physical process at all. Instead, it is data driven. With certain parameter estimation methods, Black-box model is obtained directly from input and output data, without any knowledge of the underlying physical
process. This frees us from the possibly very complex derivation from first principle, and unknown information will not cause problem. Also, black-box model is often small in size. However, black-box model does not have any physical interpretation of the real system, which makes it hard to make sense of the model parameters or to find error in the model. Lack of flexibility is another major disadvantage of black-box model. If the process which the model describes changed even just slightly, all the parameters have to be estimated again, which is a lot of work.

Grey-box model is a combination of the above two, which has a physical representation but with some approximation. Some of the parameters in the model need to be identified from data. Grey-box model is popular since it is not as complex as white-box model, yet it can provide physical insight of the system reasonably well. In this thesis, we will establish a grey-box model for the room using an electrical circuit analogy. The resistor and capacitor values are the parameters to be estimated in the model. The need for having R,C parameters (as opposed to obtaining a black-box model) in the model is manifold. First, R,C values have intuitive, physical meaning. Second, one can check if the identified parameters values are reasonable. This provides a sanity check and can help unearth potentially grossly inaccurate system identification. Third, since R,C modeling paradigm is prevalent in the HVAC/buildings community, it is useful simply as a common language. Many research results can be found on this topic [2, 12–14, 21]. However, various models of different structures, orders, and parameters are used. In this thesis, we attempt to develop an analysis to find good model structure and parameters.

1.2.1.3 Q3: Parameter estimation

The identification problem is not trivial. Powerful state-space identification methods (such as subspace methods [30]) cannot be directly employed as they do not lead to an identification of R,C parameters. Rather, they identify the system matrices in an arbitrary state space. Frequency domain and adaptive techniques [29] suffer similar problems due to the complex relationship between the R,C parameters and the coefficients in the polynomials that describe the transfer functions. Also, persistence of excitation may not be guaranteed, especially when many parameters have to be estimated.
We pick two parameter estimation techniques, least-squares method and maximum likelihood method, to identify the R,C values of low order models from data collected from a building in the University of Florida campus. Least-squares method can handle the nonlinear property in the system dynamics. With the prediction error as its cost function, it is straightforward and provides intuitive evaluation of how good the model prediction is. Maximum likelihood is another widely used method for parameter estimation in dynamical system, and it is great in dealing with measurement noise. A disadvantage of it is that it is only applicable to linear system. However, under some circumstances, which we will discuss later, our model will be linear. Thus maximum likelihood method is applicable.

Model validation is another important problem. A few authors have proposed methods for estimating R,C parameter values of building thermal models, e.g., [2, 8, 10, 20, 22]. Such papers show the effectiveness of their proposed methods by comparing the prediction of the identified model with measured data. However, our analysis shows for normal operated closed-loop data, such comparison is risky; grossly wrong models may reproduce such data quite accurately. Careful consideration have to be taken, and we will discuss this in detail in this thesis.

1.2.2 Fault Detection

As modern control systems grow more complex, reliability becomes a major concern [7]. For safety-critical systems (for example, aircraft), the results of faults can be extreme, or even catastrophic. The ability to diagnose faults on-line is thus obvious. For systems which are not safety-critical, fault diagnose can improve system efficiency, indicate need for maintenance, or prevent system breakdown when fault is developing. Fault diagnosis can be separated into three steps, namely fault detection, fault isolation, and fault correction. Fault detection is to use the available input/output measurement to tell whether the system is operating normally. This is a binary decision, i.e., to determine whether there is fault in the system or not. Fault isolation is to identify what kind of fault it is and where it is, for instance, which sensor or actuator is not working. Fault correction is rather self-explanatory. It is to fix the problem after the fault is located. It is often hard to accomplish the correction automatically, since the fault can be caused by a malfunctioning hardware, which require replacement.
In a HVAC system in a large building, fault diagnosis plays an important role. First of all, if devices break down in HVAC system, it could cause discomfort to occupants, or even bring danger in serious case. Timely detection of such device failure is critical for safety concern. In not so critical conditions, fault detection is also useful in finding deteriorating devices, improving efficiency, and intelligent maintenance scheduling. In this thesis, we will focus on the detection of the fault caused by occupants. For example, if the doors are left open, when the A/C is still on, energy will be wasted since the A/C is trying to cool down the whole hallway instead of the room. It would be desirable if the system can detect this kind of fault, and signal a message, so that people can fix the problem. From control perspective, this is also helpful since the model will be different depending on whether the door is open or closed. The fault detection technique can help us pick the right model to use in our control algorithm without the need of installing any sensors.

Fault detection methods can be broadly classified into three categories. They are quantitative model-based methods, qualitative model-based methods, and process history data based methods [31]. The advantages of the model-based methods are that they are often computationally light, and in many cases the input/output measurements required for control is also sufficient for fault diagnosis. These methods also have some drawbacks. They are limited to linear systems, have difficulties with multiplicative faults, requires accurate knowledge of the model. The process history data based methods, on the contrary, extract information from data, without knowledge of a model. The disadvantage of it is that large amount of data are needed to for training, which may not always be available. In this thesis, we pick one of the commonly used model-based method, unknown input observer, and implement it to detect the above mentioned fault in a single room.
In this section, we attempt to answer Q2, which is the question of what is the model structure of minimum complexity that can achieve our accuracy requirement. The underlying processes that govern the dynamics of temperature evolution are complex and uncertain, so grey-box models are usually used. The lumped parameter models that researchers usually use to model temperature dynamics is a combination of RC network models that capture inter-zone conduction, which is linear in structure, and a non-linear term that captures the effect of enthalpy exchange with the outside due to the supply and exhaust air. We restrict ourselves to this class of models, which is denoted by \( M_{RC}(n, q, p) \), where \( n \) refers to state dimension, \( q \) to the number of inputs, and \( p \) to the number of uncertain parameters. We first develop a very high state dimension model which has high accuracy. Then we propose various low order models of the same class, and compare with the high order one. It turns out that though a model in \( M_{RC}(n, q, p) \) is non-linear, it becomes LTI (linear time invariant) when the mass flow rate of supply air is held constant. Both time and frequency response comparison is possible in that case.

2.1 General Settings

In this section, we will discuss the states, inputs, and parameters of three different models of a zone in the class \( M_{RC}(n, q, p) \). Consider a typical zone in a multi-zone building with VAV system, which is separated from the outside through an external wall and a window and from five internal spaces (floor, ceiling, one room on each side, and a hallway) through internal walls and a door to the hallway. The major heat transfer mechanisms include the following: (1) heat conduction through external and internal walls, windows, roof, and ceiling; (2) radiation and convection between the surface and the air mass in contact with it; (3) heat convection with outside air due to the air supplied to and extracted from the room by the HVAC system; (4) solar radiation through the window and external wall; (5) casual heat gain from occupants and equipments; and (6) infiltration and exfiltration. A schematic figure of the room is shown in Figure 2-1.

The objective of building control is to maintain comfortable room climate. Naturally, it is our major concern when modeling. There are three aspects of it that are relatively
important, which are temperature, humidity, and contaminants. Contaminants, such as $CO_2$, have minimal impact on the thermal dynamics, and low level of contaminants can be assured by supplying some fixed amount of fresh air. Thus, we do not consider contaminants in this thesis. However, temperature and humidity are coupled. The heat gain from the supply air equals to the enthalpy of incoming air minus the enthalpy of exhaust air, i.e.:

$$Q_{AC} = m(h_{in} - h_{out}) \tag{2-1}$$

where

$$h_{in} = C_p T_s + W_s (h_{we} + C_{pw} T_s) \tag{2-2}$$

$$h_{out} = C_p T_i + W_i (h_{we} + C_{pw} T_i) \tag{2-3}$$

where $m$ is the supply air mass flow rate, $C_p$ is the specific heat of air, and $T_s$ is the supply air temperature, $T_i$ is the inside room temperature, $W_s$ is the humidity of the supply air, $W_i$ is the humidity inside the room, $h_{we}$ is the evaporation heat of water at $0^\circ C$, $C_{pw}$ is the specific heat of water vapour. Note that the main source other than supply air that affects humidity is the water generated by occupants. When there is no occupant, the humidity inside the room will be equal to that of the supply air. This provides an opportunity to break down and simplify the problem. Suppose we collect data when the room is
unoccupied, then $W_s = W_i = W$. We can rewrite equation (2-1) as:

$$Q_{AC} = mC_p(T_s - T_i) + WC_{pw}(T_s - T_i)$$

(2-4)

where the first term is separated from humidity. After calculation, we found that the second term is much smaller than the first one, thus we can ignore it and eliminate humidity dynamics from the problem.

To further simplify the problem, the following assumptions are made:

1. The air inside the room is well mixed, so that we have one uniform temperature in the room.

2. We ignore in/ex-filtration. Exfiltration has no effect on the temperature dynamics; it hardly matters if the air leaving the room is leaving through a return air grille or through cracks in an window. Most commercial buildings are maintained at positive pressure to preclude the possibility of infiltration from the outside through cracks, so assuming infiltration does not occur is reasonable. In the situation when door is open, there might be infiltration through the door from the hallway. This will be discussed separately in section 3.4.

3. We confine our study to night time data when there is no solar radiation and occupant heat gain, which often have large uncertainty and not easy to measure accurately.

Under these assumptions, the main variable of interest is the temperature of the space inside the zone, denoted as $T_i$. This is the first state, and depending on the model dimension, there may be other states. We now describe the various models.

### 2.2 Full-scale Model

The inputs that affect the room temperature are outside temperature, temperatures of the surrounding spaces, and heat gain inside the zone. We define the input vector to be

$$\bar{u} = [T_o, T_f, T_c, T_{n1}, T_{n2}, T_{hw}, T_s, m, Q]^T$$

where $T_o$ is the outside temperature, $T_f$ is the temperature of the space below the floor, $T_c$ is the temperature of the space above the ceiling, $T_{n1}$ and $T_{n2}$ are the temperature of the adjacent rooms to the side, $Q$ is additional casual heat gain from appliances etc., $m$ is the flow rate of supply air and $T_s$ is its temperature. With the no occupants assumption, humidity term is ignored, so the net heat gain due to the supply air equals:

$$Q_{AC} = mC_pT_s - mC_pT_i$$

(2-5)
where \( m \) is the supply air mass flow rate, \( C_p \) is the specific heat of air, and \( T_s \) is the supply air temperature. The conductive heat transfer through a solid surface separating two rooms can be modelled as lumped capacitance and resistance. We employ the commonly used choice of 3R-2C (3 resistors and 2 capacitors) \([13]\) in constructing the full-scale model. Since windows have very low heat capacitance, it is modelled as a single resistor. Each surface element is then connected to the room “node” to form a RC network model. An additional capacitor is included to model the heat stored by the air and other objects in zone. The structure of the full-scale model is shown in Figure 2-2.

![Figure 2-2. Full-scale model structure](image)

Since each wall is a 3R-2C component, there are two states for each wall. Together with \( T_i \), we have a 13-state vector: \( T = [T_i, T_{fw1}, T_{fw2}, \ldots] \), where \( T_{sw1}, T_{sw2} \) are the temperatures of two nodes associated with the surface \(*\). The dynamics of \( T_i \) can be then expressed as:

\[
C_r \dot{T}_i = (-\frac{1}{R_{win}} - \frac{1}{R_{ow1}} - \frac{1}{R_{fw1}} - \frac{1}{R_{cw1}} - \frac{1}{R_{n1w1}} \nonumber \\
- \frac{1}{R_{n2w1}} - \frac{1}{R_{hww1}})T_i + \frac{T_o}{R_{win}} + \frac{T_{ow1}}{R_{ow1}} + \frac{T_{fw1}}{R_{fw1}} \\
+ \frac{T_{cw1}}{R_{cw1}} + \frac{T_{n1w1}}{R_{n1w1}} + \frac{T_{n2w1}}{R_{n2w1}} + \frac{T_{hww1}}{R_{hww1}} + Q_{AC} + Q
\]

The dynamics of the wall nodes of each wall have similar structure:

\[
C_{sw1} \dot{T}_{sw1} = (-\frac{1}{R_{sw1}} - \frac{1}{R_{sw2}})T_{sw1} + \frac{T_f}{R_{sw1}} + \frac{T_{sw2}}{R_{sw2}} \\
C_{sw2} \dot{T}_{sw2} = (-\frac{1}{R_{sw2}} - \frac{1}{R_{sw3}})T_{sw2} + \frac{T_{sw1}}{R_{sw2}} + \frac{T_s}{R_{sw3}}
\]
For each surface, there are three resistances and two capacitances, which results in 30 parameters for all six surfaces. Together with the room capacitance and window resistance, there are a total of 32 parameters in the model. Thus, the full-scale model is of the class $\mathcal{M}_{RC}(13,8,32)$. If we assume that the resistors and capacitors are uniformly distributed in a surface, i.e., $R_{su_1} = R_{su_2} = R_{su_3}, C_{su_1}C_{su_2}$, the parameters of each surface are reduced to two, a total resistor and a total capacitor. In this case, we have 14 parameters, and the model is of class $\mathcal{M}_{RC}(13,8,14)$.

2.3 Low-order Models

We now consider two low order models that still employs the RC network analogy of conductive heat transfer as the full-scale model described above, but with fewer states and parameters.

2.3.1 First-order Model

We start with a model of lowest possible state dimension, namely, one. Here all the capacitive elements of a zone (walls, air, furniture) are aggregated into a single capacitor, so that the dynamics of $T_i$ become

$$C_r \dot{T}_i = \frac{T_o - T_i}{R_{win}} + \frac{T_{sd} - T_i}{R_w} + Q_{AC} + Q$$  (2-7)

where $T_{sd}$ is an average of all the surrounding space temperatures. This model has a single state $T_i$, five inputs $\bar{u} = [T_o, T_{sd}, T_s, m, Q]^T$, and three parameters: $\theta = [C_r, R_{win}, R_w]^T$. Thus, it is of the class $\mathcal{M}(1,5,3)$.

2.3.2 Second-order Model

The response of the room temperature $T_i$ to changes in mass flow rate and temperature of the supply air is usually faster than its response to changes in the surrounding temperatures. A natural idea is to use two capacitors to reproduce the two-time scales of the process. One capacitor (room capacitance $C_r$) is used for the low thermal mass of the air and other objects in the room, and the other ($C_w$) is used for the heat capacity of all the walls combined. Again, there are many possible choices of model structure for the integrated wall. We choose to model the integrated wall as a 2R-1C
element, which leads to:

\[
\begin{align*}
C_r \dot{T}_i &= \frac{T_o - T_i}{R_{\text{win}}} + \frac{T_w - T_i}{R_1} + Q_{AC} + Q \\
C_w \dot{T}_w &= \frac{T_i - T_w}{R_1} + \frac{T_{sd} - T_w}{R_2}
\end{align*}
\tag{2.8}
\]

where \(T_w\) is the temperature of the wall node. This model has 2 states \(T = [T_i, T_w]^T\), the same five inputs as the single-state model, and five parameters \(\theta = [C_r, C_w, R_{\text{win}}, R_1, R_2]^T\). Thus, it is of class \(\mathcal{M}(2, 5, 5)\).

### 2.3.3 The Linear Time Invariant (LTI) Case

Note that the \(Q_{AC}\) term in (2.6), (2.7), and (2.8) is the only nonlinear term in each of the three models described above. When the supply air flow rate, \(m\), is constant, the \(Q_{AC}\) term becomes linear in the state \(T_i\) and input \(T_s\). The system then becomes a LTI system:

\[
\dot{T} = AT + Bu, \quad z = CT, \tag{2.9}
\]

where the state \(T\) and input \(u\) varies depending on which of the three models is under consideration, the output \(z\) is \(T_i\). The matrix \(A\) depends on the parameter \(m\). The number of inputs is reduced by one for each class (since \(m\) has moved from being an input to a known fixed parameter).

### 2.4 Model Structure Comparison

Various models within the class \(\mathcal{M}_{RC}(n, q, p)\) can be compared among themselves to see how well they compare against one another in terms of speed of response, gains etc. When the mass flow rate is constant, the models are LTI, so it is possible to compare their frequency response as well. A comparison between their frequency response is useful to determine whether a low-order model within the class \(\mathcal{M}_{RC}(n, q, p)\) is powerful enough to characterize the temperature dynamics within a range of input frequencies as well as a high order model. If so, this provides justification for using that low order model in the interest of reduced complexity. In addition, examining the frequency response is also useful for model calibration. For instance, the time constant and DC gain with respect to controllable inputs (such as supply air temperature or surrounding room temperatures) can be obtained through forced response experiments without making assumption on the
model. Therefore, a model can be calibrated by tuning its parameters so that it has the same pre-specified speed of response or DC gain as those obtained experimentally.

We do not deal with calibration in this section. Instead, we choose the parameters of the full-scale model according to ASHRAE handbook. We call these values ASHRAE values, and they will be discussed in section 2.4.1. In the low-order models, however, we need a way to compute the effective resistance and capacitance of the integrated wall. We consider the surfaces that are aggregated to form the integrated wall to be parallel components, so that the total capacitance is the sum of capacitance of each surface, and the total conductance is the sum of conductance (inverse of resistance) of each surface.

2.4.1 ASHRAE Values

ASHRAE (American Society of Heating, Air conditioning and Refrigeration engineers) handbook [1] describes how to determine the R,C (resistance/capacitance) values for a solid surface given its material and construction type. This information is now available in software such as HAP [5]. The resistances and capacitances of a RC network model are carefully chosen to model the combined effect of conduction between the air masses separated by the surface, as well as long wave radiation and convection between the surface and the air mass in contact with it [1], [14, 26]. We refer to parameter values so obtained as “ASHRAE values”. However, there is uncertainty in these parameter values. First, information on wall and window construction and material are not always easy to obtain for existing buildings due to poor record keeping. Second, due to cracks in windows and walls, the effective resistance of an window and wall is likely to be lower than what is inferred from construction data. Finally, if a window or a door is open, the effective resistance could be far lower than the resistance estimated for a closed window or door. Therefore although being a valuable reference, the ASHRAE values may not be the best parameter values for a given room. Here we provides two simulations results using the full-scale model with ASHRAE values as their R,C parameter values. Figure 2-3 and Figure 2-4 are the simulations of two different rooms (Room 241 and Room 243) in Pugh Hall on UF campus.
Figure 2-3. Simulation of Room 241 with ASHRAE values as parameters

As shown in the figures, ASHRAE values provides good prediction for Room 241, but not for Room 243. Therefore, there is a need to identify/estimate R,C parameters from measured data, a process referred to as model calibration and/or system identification.

2.4.2 Time Domain Comparison

We pick Room 241 for the following comparison. The reason is that the ASHRAE values already have good prediction power for this room, thus we can compare the difference between model structures without calibrating the models. The time domain comparisons of the full-scale and reduced model (along with measured temperatures from
the three data sets described in Section 3.1) are shown in Figure 2-5. We see that the full-scale model and second order model have very similar prediction. Prediction errors are both smaller than $1^\circ F$. First order model under-predicted the temperature a little more than the other two models, but still within reasonable range, with prediction error under $2^\circ F$.

![Figure 2-5. Time domain comparison of different models](image)

**Remark 2.1.** In all time domain simulations in this thesis, we used a constant load of $Q(t) \equiv 50W$. Though the three data sets was collected at night with no occupants, initial comparison showed that all the models under-predicted the measured temperature, no matter what $R,C$ parameter values are chosen. We conjecture that an additional heat gain due to the computers in the room was present. An inactive desktop PC & monitor can produce 20-30 W of heat, and an idle desktop printer can produce 10-35 W [1]. So we picked 50W as a somewhat ad-hoc estimate.


2.4.3 Frequency Domain Comparison

Frequency domain comparisons reveal more than a single time-domain simulation comparison, and allows more general conclusions. Therefore comparing the frequency response among models is preferred, when possible. This is of course only possible for LTI systems.

By examining the data from Pugh Hall for several months, we find that the mass flow rate usually stays at a constant value for a duration of several hours. Thus, considering the case of constant \( m \) that makes the model LTI is not altogether unreasonable. By examining the data, we find that the supply air temperature has the fastest change rate among all input signals. It can change from its minimum value to maximum value in 5 minutes. Assuming that the largest period we consider is 5 hours, we choose the frequency of interest to be \( \frac{1}{5} \text{ hours} \times (10^{-5} \text{Hz}) \) to \( \frac{1}{5} \text{ mins} \times (10^{-3} \text{Hz}) \).

Let \( H_{\ell}(s) \) be the transfer function from the \( \ell \)-th input to the output \( T_i \). The supply air temperature is a common input in all the models. However, in the full-scale model, there are five different inputs for the five surrounding temperatures; while in the low-order models there is only one input \( T_{sd} \) for all the surrounding temperatures since the walls are integrated into one surface. For comparison, we define a transfer function from a single surrounding temperature \( T_{sd} \) to output in the full scale model as the sum of transfer functions of all five surrounding temperatures, i.e.,

\[
H_{sd}(s) := \sum_{i=1}^{5} H_{sd_i}(s)
\]  (2–10)

where \( H_{sd_i} \) are the transfer functions from each surrounding temperature to output.

Figure 2-6 shows the frequency response comparison (only the magnitude, not phase) of the three models: full-scale (13-th order), first order, and second order. The parameters are chosen as described earlier. The value of the mass flow rate is set to be the maximum possible value, \( m = 0.12 (Kg/s) \), as we have observed that the maximum error occurs at the maximum values of \( m \). We also observed that the outside temperature \( T_0 \) has small impact on the output compared to the other inputs, which is consistent with the fact that commercial buildings are usually well insulated from the outside. Therefore, we only show the result of the two inputs \( T_s \) and \( T_{sd} \) here. The following conclusions are drawn from the

---

25
figure. First, for the range of frequencies deemed of interest, the second order model is almost as accurate as the 13-th order model. Second, the 1st order model is also quite accurate in terms of predicting the response due to supply air temperature $T_s$, but less so in case of the surrounding temperature. However, since the surrounding temperature changes very slowly, the higher error that is seen in the higher frequencies may not result in significant error in the output prediction. In addition, it is possible to bring the first order model’s response closer to that of the higher order models by ‘tuning’ the parameters. Third, the gain from the two inputs are of similar magnitude. This is in opposition to the prevalent belief that supply air is the dominant input. One reason could be that all the five surrounding temperatures are combined into one. These conclusions are in consistent with what the time-domain simulations indicated.

The effect of an input on the output depends not only on the transfer function but also on the magnitude of input. The variation in $T_s$ and $T_{sd}$ can be significantly different from each other. The error in the prediction of $T_i$ may be different due to these inputs even though they might have the same gain. Therefore we define the error with respect to $i$-th
input as:

\[ e_i(\omega) = \left| |H_{fs_i}(j\omega)| - |H_{di}(j\omega)| \right| \delta u_i, \quad (2-11) \]

where \( H_{fs_i}(\cdot) \) and \( H_{di}(\cdot) \) correspond to full-scale and low-order models, and \( \delta u_i \) is the maximum variation in the \( i \)-th input. By examining the data from Pugh Hall, we observe that the largest variations in supply air temperature and surrounding room temperatures are \( 45^0F \) and \( 10^0F \).

Calculation have been done for different supply air flow rates. The error \( e_i \) as a function of frequency is shown in Figure 2-7. We see from the the figure that the prediction error is most sensitive to the input \( T_s \). For first order model, this maximum error is \( 3^0F \); for the second order model, it is less than \( 1^0F \). Therefore the 2nd order model, for the chosen parameter values, is closer to the 13-th order model by a factor if three, compared to the 1st order model, in terms of the prediction error across the range of frequencies of interest.
CHAPTER 3
CALIBRATION AND VALIDATION

3.1 Field Data

3.1.1 Test Bed

The test bed we use for experiments and data collection is Pugh Hall on University of Florida campus. Finished in 2008, Pugh Hall is a 40,000-square foot facility, includes a teaching auditorium and public space for lectures and events. As for HVAC part, it has three air handling units (AHUs), each of them mixes outside air and return air, cool the air down, maintain certain humidity level, then supplies the conditioned air into different rooms. The terminal devices are variable air volume (VAV) boxes, which distributes the conditioned air into one or a few rooms. VAV boxes can control the mass flow rate of air going into the target rooms, and can reheat the air if necessary. Newly built, Pugh Hall is equipped with many pre-installed sensors, which makes it an ideal test bed. Some very useful sensors for this particular research, such as room temperature, VAV box supply air temperature and mass flow rate, VAV damper and reheat valve position, and etc., are already installed in the building. The sensors are connected to building automation system (BAS), which keeps the building climate within a specific range. This is really helpful, since installation and wiring sensors into existing system is extremely difficult and inefficient.

Siemens Insight Workstation is also installed in Pugh Hall. It is a building monitoring, management, and control software, which has many good features. For example, it supports BACnet protocol and has friendly graphical interface, which makes monitoring and controlling the building easier. A screen shots of Insight Workstation is given in figures 3-1.

Although Insight Workstation is an useful tool, it is not easy to incorporate our own online algorithm. To record online data or implement modern control strategy, we need other methods. Fortunately, Dr. Timothy Middelkoop in Industrial Engineering department has built up PostgreSQL databases which records all the data points in the BAS. By sending queries to the databases, we are able to obtain large amount of online data at a fast speed.
The testing area is chosen to be a section on the second floor. It contains several typical offices, an internal room which is used as a small studio, a small rest area which has refrigerator and cooking appliances, a small meeting room, and a large conference room. Another advantage of this area is that, a few offices are each supplied by individual VAV box, which decreases the uncertainty in supply air distribution after air leave the VAV boxes. A floor plan is given in figure 3-2.

Hallway temperature is a variable of our interest, but there is no pre-installed sensors to measure it in the test area. To collect hallway temperature, we deployed HOBO U10 Datalogger to record data, and Telaire 7001 sensor as a display for the current measurement. The left photo in Figure 3-3 shows the sensors and dataloggers (white boxes), the right photo shows the sensors and dataloggers on test field.

The environment temperature outside the room is another variable we would like to measure. Due to the installation and safety concerns, we decided not to install outdoor sensors. Instead, we got the weather data from two websites, namely
Figure 3-2. Floor plan of testing area

Figure 3-3. HOBO sensors

http://www.wunderground.com/ and http://www.phys.ufl.edu/weather/. The former one has history weather records collected from local weather station, which we can refer to when we want to get data during a certain period of time. The later one is the website of UF Physics Department Weather Station. It shows the current weather information, and updates every 15 minutes. A Perl script is developed for exacting the useful information from the website in real time, such as current time, temperature, humidity etc. This enables us to get real time data which we can use in our control algorithm.

3.1.2 Data Sets

For model calibration and validation in this particular research, we choose a typical office, Room 241, in Pugh Hall. The room is connected to two adjacent offices and the
hallway through internal walls; and to the outside through an external wall and window. Three data sets are studied. All data are collected from 6pm in the evening to 6am next morning. The first data set, set A, was collected on a summer day from Aug. 2nd to Aug. 3rd, 2011, under normal closed loop control strategy. The second data set B, was collected on a winter day from Dec. 5th to Dec. 6th, 2011 under normal closed loop control strategy. These two data sets are chosen among data collected over several months to meet the criteria that the room and supply air temperatures should have as much variation as possible (the latter ensures persistence of excitation), while the mass flow rate should be constant for long periods of time (so that the models become LTI). The third data set, set C, was collected from Oct. 24th to Oct. 25th, 2011, under a forced response test. During the test, we first heated up the target room and cooled down the surrounding rooms, in order to generate a temperature difference between the room and the surrounding. Then we shut down the air supply to study how the surrounding affect the room temperature. All these data sets are collected with the door closed, and the windows cannot be open.

The room temperature, supply air temperature, and supply air flow rate of the three data sets are plotted in Figure 3-4.

Another two data sets are collected for studying the effect of opening the door. Data set D is collected on Apr. 20th, 2012, and set E is collected on May, 3rd, 2012. For both two sets, the building was operated under designed forced response experiment. More detail will be discussed in section 3.4.

3.2 Identification Methods

We pick two methods for parameter estimation, which are described below. In the sequel, \( \mathcal{P} \) denotes a region in the parameter space \( \mathbb{R}^p \) where \( \theta \) can lie.

3.2.1 Least-squares

For a given model structure with fixed parameters, we define a prediction error cost \( J \) as:

\[
J = \int_0^\tau (T_{im}(t) - T_{ip}(t))^2 dt
\]  

(3–1)

where \( T_{im}(t) \) is the measured room temperature at time \( t \), \( T_{ip}(t) \) is the room temperature at time \( t \) predicted by the model with a given set of parameter values, and \( \tau \) is a
user-specified time interval. Now the parameter estimation problem can be posed as minimizing the prediction error cost $J$:

$$
\hat{\theta} = \underset{\theta \in \mathcal{P}}{\arg\min} J 
$$  \hspace{1cm} (3-2)

The minimization can be performed by using an optimization algorithm such as gradient descent or by again a direct search. An advantage of this method is that identification of parameters of a non-linear model is possible.

### 3.2.2 Maximum Likelihood (ML) Method

The second parameter estimation method we use is the maximum likelihood method that is proposed in [20]. This method is applicable to LTI models in discrete time. The continuous time model (2-9) is first converted to a discrete time with additive process noise and measurement noise. Under some assumptions, which will be explained later, the likelihood function can be formulated with the help of Kalman filter. The procedure is as follows:
First, discretize the continuous dynamics:

\[
T(k + 1) = AT(k) + Bu(k) + \theta(k)
\]
\[
z(k) = CT(k) + \xi(k)
\]

where \(T(k)\) is \(n \times 1\) state vector, \(z(k)\) is \(m \times 1\) output vector, \(\theta(k) \in \mathbb{R}^{n \times 1}\) represents the uncertainty in states, \(\xi(k) \in \mathbb{R}^{m \times 1}\) represents the measurement error, \(A, B,\) and \(C\) are matrices containing unknown parameters.

Before moving on, it is helpful to make some Gaussian and independent assumptions about the noise and initial states, which is normal in noise characterization. The following assumptions are made:

1. The initial state is Gaussian, i.e., \(T_0 \sim N(0, \theta_0)\);
2. The noises are Gaussian, and independently distributed. The covariances of the noises are given as:
   \[
   \text{cov}(\theta_k, \theta_j) = \Theta_k \delta_{kj}
   \]
   \[
   \text{cov}(\xi_k, \xi_j) = \Xi_k \delta_{kj}
   \]
   where \(\Theta_k \in \mathbb{R}^{n \times n}\) and \(\Xi_k \in \mathbb{R}^{m \times m}\) are known, and \(\delta_{kj}\) is the Kronecker matrix defined by:
   \[
   \delta_{kj} = \begin{cases} 
   1, & k = j \\
   0, & k \neq j 
   \end{cases}
   \]
3. The initial state, disturbance and measurement noise are independent.

Let \(Z = [z(k), z(k-1), ..., z(0)]\) be the observations up to time \(k\). Then the likelihood function is given by the joint density function of all observations, i.e.,:

\[
L(\theta, Z(k)) = f(Z(k))
\]
\[
= f(z(k)|Z(k-1))f(Z(k-1))
\]
\[
: \quad = f(z(k)|Z(k-1)) f(z(k-1)|Z(k-2)) \cdots f(z(0))
\]

Note that \(L\) is a multiplication of conditional density functions. Under the above assumptions, it can be shown that \(z(k)\) and \(Z(k-1)\) are jointly normally distributed, since they are both functions of \(T_0, \theta_{(k-1)}, \ldots, \theta_0, \xi_{(k-1)}, \ldots, \xi_0\), which are jointly normally distributed. By property 5 and 6 in Chapter IV of [28], each individual conditional density
in $L$ is also normal, and the conditional mean and covariance can be calculated using Kalman filter in a recursive fashion. More precisely, let the conditional mean and variance at time step $k$ be $\hat{z}(k|k-1)$ and $R(k|k-1)$. Note that, $\hat{z}(k|k-1)$ is an one-step prediction and $R(k|k-1)$ is the corresponding variance. For notation simplicity, we define the estimation error to be:

$$e(k) = z(k) - \hat{z}(k|k-1)$$  \hspace{1cm} (3-8)

Since the conditional distribution is Gaussian, the likelihood function can be written as:

$$L(\theta, Z(N)) = \prod_{k=1}^{N} \{(2\pi)^{-m/2}|R(k|k-1)|^{-1/2} \exp\left(-\frac{1}{2}e(k)^T R(k|k-1)^{-1}e(k)\right)\}$$  \hspace{1cm} (3-9)

where $| \cdot |$ is the determinant operator. It is more convenient to use logarithm likelihood function:

$$\log(L(\theta, Z(N))) = -\frac{1}{2} \sum_{k=1}^{N} \{\log |R(k|k-1)| + e(k)^T R(k|k-1)^{-1}e(k)\} + \text{constant}$$  \hspace{1cm} (3-10)

The next step is to compute $\hat{z}(k|k-1)$ and $R(k|k-1)$. This is done using Kalman one-step predictor. The estimates of states $\hat{T}$ is updated by the equations:

$$\hat{T}(k+1|k) = A\hat{T}(k|k-1) + Bu(k) + L_k(z(k) - CT(k|k-1))$$  \hspace{1cm} (3-11)

where $L_k$ is the Kalman gain, which can be calculated by:

$$L_k = AP(k)C^T(CP(k)C^T + \Theta)^{-1}$$  \hspace{1cm} (3-12)

The estimation error covariance matrix $P$ is updated by:

$$P(k+1) = A[P(k) - P(k)C^T(CP(k)C^T + \Theta)^{-1}CP(k)]A^T$$  \hspace{1cm} (3-13)

Then the condition mean and variance we need to calculate the logarithm likelihood function can be calculated by:

$$\hat{z}(k+1|k) = C\hat{T}(k+1|k)$$  \hspace{1cm} (3-14)

$$R(k+1|k) = C(AP(k+1)A^T + \theta_k)C^T + \xi_k$$  \hspace{1cm} (3-15)
The estimation problem then becomes one of finding $\theta$ that maximizes the likelihood function within the set $\Theta$ of allowable parameter values:

$$\hat{\theta} = \arg \min_{\theta \in \mathcal{P}} - \log L,$$  \hspace{1cm} (3–16)

For each set of given parameter, the logarithm likelihood function can be calculated by equation (3–10). The minimization can be performed by direct search or by other optimization algorithms.

### 3.3 Model Calibration and Validation

Model calibration refers to estimating the R,C parameters. The full scale model in 2.2 has 32 parameters; it is quite difficult to estimate them. Furthermore, we saw in the previous section that the second order model is closer to the full-scale model than the first order model in terms of the frequency response when such a comparison is possible. For this reason (and due to lack of space), we limit our attention to the second order model $M_{RC}(2,3,p)$ for the purpose of identification. Also, we only consider the situation when door is closed in this section.

The following assumptions are made to reduce the number of uncertain parameters in the second order model: (i) The resistance of the wall is symmetric, i.e., $R_1 = R_2 = \frac{1}{2}R_w$. (ii) The wall capacitance $C_w$ and window resistance $R_{win}$ are assumed known (ASHRAE values). (iii) The surrounding temperature $T_{sd}$ is taken to be the average of all the surrounding room temperatures (above the ceiling, below the floor, two adjacent rooms and the hallway). The first assumption is made so that models of multiple zones can be combined to create a model of a building or section of it. The second assumption is made since the capacitance of a wall changes little due to wear and tear over time or the appearance of cracks. With these assumptions, the model we have to identify is $M(2,3,2)$: the uncertain parameters to be estimated are $C_r$ and $R_w$.

#### 3.3.1 Attempt 1: Apparent Success But Really A Failure

We use data set A for model calibration and data set B for model verification. For the ML method of Section 3.2.2, we pick a segment from the data set where the mass flow rate is constant in that interval. This is required to make the system LTI since the ML method is only applicable to LTI systems. The minimization of prediction error cost (3–2) and $-L$
Table 3-1. Best fit parameters when data set A is used for model calibration.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Least-squares</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_r (J/K)$</td>
<td>$5 \times 10^4$</td>
<td>$5 \times 10^4$</td>
</tr>
<tr>
<td>$R_w (K/W)$</td>
<td>$6.44 \times 10^{-4}$</td>
<td>$5.6 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

in (3–16) are both done with direct search. The resulting best-fit parameter values are shown in Table 3-1.

The room temperature predicted by the model with the best-fit parameter values and the measured room temperature are shown in Figure 3-5 for data set A. The fact that the prediction error is small is not surprising since the model is calibrated with this data set. Next we perform verification of the identified model by comparing its prediction for the data set B; see Figure 3-5. Again the prediction error is seen to be small. At this stage, we can claim that the model calibrated with summer data is able to predict winter data quite well, so we can declare we have a calibrated model and move on.

Figure 3-5. Calibration/validation when data set A is used for calibration.
However, when we try to predict the temperature of data set C with this “calibrated” model, which is shown in the third plot in Figure 3-5, we see that the model predictions are completely off. So what happened?

First, we note that the reason we estimated the parameters by direct search instead of using a more sophisticated search method is that the cost functions that are minimized to estimate the parameters are non-convex. This is clear from the contours of the cost function $J$ which are shown in Figure 3-6 for data set A. By performing direct search, possible hypotheses that blames the search method for getting stuck at a local minimum can be eliminated. Clearly the problem is not the search method, but either the model or the data used for calibration.

![Figure 3-6. Contour of the cost $J$ defined in (3–2), which shows its non-convexity.](image)

The reason for the failure was found to be the data. In the calibration data (data set A), the surrounding room temperatures are almost the same as the room we study. So the best fit resistance values are those that are so small that the room temperature essentially follows the temperature of the surroundings, leading to small prediction error. In validation data set B, though the room temperature profile is quite different from that in set A, the surrounding room temperatures are still close to the room temperature, so the model predicts well. However, in validation data set C, the surrounding room temperatures are significantly different from the room temperature, so it shows that the calibrated parameter values are incorrect.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>least-squares</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_r (\text{J/K})$</td>
<td>$7.8 \times 10^5$</td>
<td>$4.7 \times 10^5$</td>
</tr>
<tr>
<td>$R_w (\text{K/W})$</td>
<td>0.01</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 3-2. Best fit parameters when data set C is used for model calibration.

### 3.3.2 Attempt 2: A More Reliable Calibration

Now we use data set C for calibration, which leads to best-fit parameter values shown in Table 3-2. In this case, the best-fit parameters from the two methods are not so similar. The reason could be that, the data that can used for the ML method corresponds to a constant input (supply air temperature), which is not sufficiently exciting. Still, while in the previous case both methods yield a much lower value of the resistance compared to their ASHRAE values, this time the estimated resistance value is much closer to its ASHRAE value. The time-domain simulations for calibration and validation data sets are shown in Figure 3-7. From the figures, we can see that though the maximum prediction error is large ($30^\circ F$), this set of parameters predict the trend of the temperature well in all three data sets.

![Figure 3-7](image-url)
We can also look at this problem from another aspect. Assume that the supply air flow rate is constant, then the system becomes linear, we can then study the DC gain of the system with respect to each input. Write the dynamics in state space form, we get

\[
\dot{T} = \begin{bmatrix}
\frac{1}{R_{\text{win}}C_r} - \frac{mC_p}{C_r} & \frac{1}{R_1C_r} & -\frac{1}{R_1C_w} \\
\frac{1}{R_1C_w} & 0 & -\frac{1}{R_2C_w} \\
\frac{mC_p}{C_r} & \frac{1}{R_{\text{win}}C_r} & 0 \\
0 & \frac{1}{R_1C_r} & 0 \\
0 & 0 & \frac{1}{R_2C_w}
\end{bmatrix} \begin{bmatrix} T \\ u \end{bmatrix}
\]

\[y = \begin{bmatrix} 1 & 0 \end{bmatrix} T \tag{3–17}\]

where \(T = \begin{bmatrix} T_i & T_w \end{bmatrix}^T\), \(u = \begin{bmatrix} T_s & T_o & T_{sd} \end{bmatrix}^T\). Then we can compute the the DC gain for each input:

\[
K_s = \frac{mC_p R_{\text{win}}(R_1 + R_2)}{R_1 + R_2 + R_{\text{win}} + mC_p R_{\text{win}}(R_1 + R_2)}
\]

\[
K_o = \frac{R_1 + R_2}{R_1 + R_2 + R_{\text{win}} + mC_p R_{\text{win}}(R_1 + R_2)}
\]

\[
K_{sd} = \frac{R_{\text{win}}}{R_1 + R_2 + R_{\text{win}} + mC_p R_{\text{win}}(R_1 + R_2)}
\tag{3–18}\]

Note that the capacitance of the wall and room does not affect the DC gain, and for given set of \(R\) values, the DC gain is a function of the supply air flow rate. We plot the DC gains for the \(R\) values we get from the first attempt. It is shown in the left of Figure 3-8. We can see that with this set of parameters, even when the A/C is blowing maximum amount of air, the dominant effect is still the surrounding room temperature, which should not happen, as shown in Figure 3-5. If we instead use the parameters from second attempt and

![Figure 3-8. DC gain comparison, left figure is for first attempt, right is for second attempt](image-url)
plot the DC gain in the right of Figure 3-8. Now the supply air temperature will have larger effect than the surrounding room temperature when the air flow rate is high.

The following conclusions are drawn from these results:

(1). Calibration accuracy crucially depends on the data set chosen. This by itself is not surprising. What is surprising is that having large variation in the measured inputs (to have persistence of excitation) and outputs is not enough. The data should be chosen so that the output (room temperature) does not have a strong correlation with any of the surrounding temperatures. The features required in the data to ensure identification of parameters seem to be possible only through forced response experiments. If these conditions are not met, the usual method of calibration is not likely to yield useful results, even when the identified parameters are “validated” by using a dataset distinct from the calibration data set.

(2). It is important to keep the number of parameters to a minimum so that one is not forced to simply accept the best-fit parameters returned by some optimization algorithm. Instead, small number of parameters make it feasible to examine (non)-convexity of the cost function used to search for parameters. The cost function used to search for the best-fit parameters was observed to be non-convex for both least-squares and ML methods. With a model with higher number of parameters, this situation is likely to be exacerbated.

### 3.4 Effect of Open Door

More often than not, when an office is occupied, the door is kept open. This create a large open area, through which air exchange occurs between the hallway and the inside of the room. In that case, the model we developed earlier may not be applicable. Moreover, when MPC is implemented, there could be a large temperature difference between the hallway and inside the room, and the opening door may have a substantial effect on the room thermal dynamics. Thus, addition work for modeling the room when door is open is required.

To study this, we conducted an experiment in Pugh Hall. Again, we chose Room 241 as the target room. During the experiment, we would like to generate a temperature difference between the room and the hallway, and make it reasonably similar to the real practice. Thus, we first set the room temperature set point of Room 241 to $70^\circ F$, and
temperature set point of hallway to 75°F. After the temperature settled down, we opened the door, and wait for some time to observe its effect.

First, we study whether opening the door has a noticeable effect on the thermal dynamics of the room. To do this, we performed a simulation with the above mentioned calibrated model for the whole duration of the experiment. The simulation result is shown in Figure 3-9. During the experiment, the hallway temperature did not raise up to 75°F as we designed, however, the temperature difference is still good enough for us to study the effect of opening door. In the plot, the vertical line indicates the time door was opened. It is clear that before that time, the predicted temperature matches the measured temperature reasonably well, but after that time, it decreased while the measured temperature increased. The deviation indicates that the effect of the opening door is not negligible.

Naturally, the next question is how to moderate the model when door is open. One idea is to model it as natural convection, then when the door open, another heat transfer term is added, and it can be expressed as:

\[ Q_{hw} = m_{hw} C_p (T_{hw} - T_i) \]  

(3–19)
where \( m_{hw} \) is the mass flow between the hallway and the room. The value of \( m_{hw} \) depends on the density difference between the two spaces. If the air pressure and humidity are the same in the two spaces, it can be calculated by a nonlinear function of \( T_{hw} \) and \( T_i \) based on the geometry of the opening area.

Another simpler way is to add another RC component to model the heat transfer through the door. Since when the door is open, there is no solid surface separating the two areas, the capacitance is small. Thus we choose to use a single resistor. The structure is shown in Figure 3-10. In this case, the addition heat transfer term is:

\[
Q_{hw} = \frac{1}{R_{od}} (T_{hw} - T_i)
\]  

(3–20)

where \( R_{od} \) is the effective resistance of open door.

In both cases, the \( Q_{hw} \) term is in the form of a coefficient multiply with the temperature difference between the hallway and the room, i.e., \( Q_{hw} = \alpha(T_{hw} - T_i) \). The difference is that in the natural convection case, the coefficient is nonlinear, and is coupled with the \( T_{hw} \) and \( T_i \), while in the resistor case, the coefficient is just a constant parameter. In this thesis, we will take the second method because of its simplicity, and it turns out that it is sufficient to provide good prediction.

To determine the value or \( R_{od} \), we again use exhaustive search method. There is only one variable and we have a general idea of the range of it, so exhaustive search is handy. The result is shown in Figure 3-11 and 3-12. As shown in the figures, with the additional resistor, the prediction matched the measurement through whole experiment duration. The
cost function in the exhaustive search also shows a clear minimum. Also, we performed a validation using data from another experiment. The result is shown in Figure 3-13 and the $R_{od}$ value works well. So far, we got reasonable models for both door close and door open situations.
Figure 3-13. Prediction error with respect to $R_{cd}$

3.5 Summary

We examined two questions for models of HVAC zones that can be used for predictive control: required model complexity and parameter identification. We examined models of varying complexity within the popular class of non-linear RC network models. By comparing low order models with high order ones, we conclude that a second order model reproduces the input-output behavior of the full-scale, 13th order model quite accurately. Even a first order model is accurate enough that it may very well suffice for the purpose of predictive control. Thus, complex models with high state dimension and large number of resistance/capacitance models are not needed.

The work reported here on parameter identification of low-order models from experimental data has revealed some surprising results. The results indicate that calibrating the parameters of the R,C network model to closed-loop data from a building is likely to lead to grossly inaccurate parameter estimates, so that the resulting model is unlikely to be useful in predictive control. Even data with sufficiently exciting input is not enough. The results reveal the features that the data should have to enable correct identification. These features seem to be possible to ensure only through forced response tests.
CHAPTER 4
FAULT DETECTION

Fault diagnosis is an important task in many aspects of building control. With the model we established in the earlier chapters, we are able to exploit the use of fault diagnosis techniques. One of the problem with the model is that it changes significantly when the door is opened or closed. One can certainly put sensors to detect the door status, however it is expensive to install sensors for each room. Plus, sensors will fail sooner or later, so this also causes a higher maintenance cost. Thus, it is desired to be able to detect that from measurements from sensors that are already installed in a building. In this section, we will consider door closed status to be the normal operating mode, while door left open to be a faulty status. With this setup, fault diagnosis techniques can be utilized for detecting whether the door is open or closed.

4.1 Unknown Input Observer

Unknown input observer (UIO) is a model-based fault diagnosis method. It is a two-step procedure: residual generation and decision making. The basic idea is to estimate the states with input/output measurements, use the estimation error as residual. When the system is working under normal condition, the residual will be close to zero. When fault occurs, the assumed model is no longer accurate, so the estimated states can not follow the real states, causing the residual to be non-zero. A threshold for residual is set. By observing the change in residual, and comparing with the threshold, we can detect the fault. In real world, we will never get the exact model, due to all kinds of disturbances. In UIO method, we consider all disturbances as an unknown input, denoted by $d$. By carefully design the observer, UIO method is able to estimate the states even when disturbance is present. We will start from the system model.

Let the system dynamics with disturbance to be:

$$\dot{x} = Ax + Bu + Ed$$
$$y = Cx$$

(4–1)
where \( d \) is the disturbance and \( E \) is the disturbance distribution matrix. The observer is designed to be

\[
\dot{z} = Fz + TBu + Ky \\
\hat{x} = z + Hy
\]  

(4-2)

where \( F, T, K, H \) are to be designed.

Define the estimator error as \( e = x - \hat{x} \). After some algebraic manipulation, we get

\[
\dot{e} = (A - HCA - K_1C)e + (F - (A - HCA - K_1C))z + (K_2 - (A - HCA - K_1C)H)y \\
+ (T - (I - HC))Bu + (HC - I)Ed
\]  

(4-3)

where \( K_1 + K_2 = K \).

If we can designed the matrices, so that the following are satisfied:

\[
\begin{align*}
(HC - I)E &= 0 \\
T &= I - HC \\
F &= A - HCA - K_1C \\
K_2 &= FH
\end{align*}
\]  

(4-4)

Then we have \( \dot{e} = Fe \). Let \( A - HCA = A_1 \), then \( F = A_1 - K_1C \). If \( (A_1, K_1) \) is observable, we can make the observer error go to zero asymptotically by choosing \( K_1 \) properly.

Now the problem becomes solving the set of equations (4-4). Consider the first equation:

\[
(HC - I)E = 0.
\]  

(4-5)

It can be shown that if \( rank(CE) = rank(E) \), then a special solution of \( H \) is given by:

\[
H = E[(CE)^TCE]^{-1}(CE)^T
\]  

(4-6)

With this \( H \), we have \( T = I - HC \) and \( A_1 = TA \). The next task is to design \( K_1 \) so that \( F = A_1 - K_1C \) is Hurwitz. If the pair \( (A_1, C) \) is observable, than the eigenvalues of \( F \) can be assign to anywhere in the left half plane by standard pole placement technique. If \( (A_1, C) \) is unobservable, we can perform an observable canonical decomposition, which is a
similarity transformation to obtain observable canonical form. Suppose $n_1 < n$ is the rank of the observability matrix of $(A_1, C)$, let $P$ be the transformation matrix, then we have:

$$PA_1P^{-1} = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}$$ (4–7)

$$CP^{-1} = \begin{bmatrix} C_0 & 0 \end{bmatrix}$$ (4–8)

where $A_{11}$ is $n_1 \times n_1$ matrix. If $A_{22}$ have eigenvalues with non-negative real part, then the UIO does not exist. Otherwise, design $K_1$ to assign desirable eigenvalues to $A_{11} - K_1C_0$.

Let $K_2$ be a arbitrary matrix of dimension $(n - n_1) \times m$, where $m$ is number of rows of $C$. Then compute $K_1 = P^{-1}\left[(K_1^1)^T(K_2^2)^T\right]^T$. Finally $F = A_1 - K_1C$,

$K = K_1 + K_2 = K_1 + FH$. The design of UIO is completed.

Now add fault into the system. We consider two kinds of faults, actuator faults and sensor faults. Actuator fault is the discrepancy between the inputs commands and actual inputs into the system. For example, when a damper is stuck, whatever commands are given, the actual input will stay at certain constant value. Actuator fault is modelled as an additive term to the input. Sensor faults is the discrepancy between a sensor reading and the actual value of the variable, and can be modelled as an additive term to the output. Thus, we have

$$\dot{x} = Ax + B(u + f_a) + Ed$$

$$y = Cx + f_s$$

where $f_a$ is actuator fault and $f_s$ is sensor fault.

Apply the same design as (4–2) and (4–4) to the system with fault, we get

$$\dot{e} = F\dot{e} + TBf_a - K_1f_s - H\dot{f}_s$$ (4–10)

Define the residual to be $r = y - \dot{y}$, then can also be written as $r = Ce + f_s$. Ideally, $r$ is zero under normal condition, and is non-zero under faulty condition.

4.2 Implementation Example

In this section, we will implement UIO in an illustrative example, and discuss some of the issues encountered.
4.2.1 Room Model

In the example, we chose to use the room model from M.M.Gouda’s paper (2000). The room has five heat-exchanging surfaces, which are two types of external wall, one floor, one ceiling, and one internal wall. Each surface is modelled as 2R-1C component. Besides the heat exchange through all surrounding areas, there are three other heat sources, which are heat gain from A/C system, solar heat gain, and casual heat gain from occupants and appliances in the room. The states, inputs, and outputs are defined as follows.

There are 6 states,

\[ x = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_i \end{bmatrix} = \begin{bmatrix} \text{Temperature of external wall 1} \\ \text{Temperature of external wall 2} \\ \text{Temperature of the floor} \\ \text{Temperature of the ceiling} \\ \text{Temperature of the internal wall} \\ \text{Inside room temperature} \end{bmatrix} \] (4-11)

and 7 inputs,

\[ u = \begin{bmatrix} T_o \\ Q_s \\ Q_c \\ Q_g \\ T_{z1} \\ T_{z2} \\ T_{z3} \end{bmatrix} = \begin{bmatrix} \text{Outside temperature} \\ \text{Solar heat gain} \\ \text{AC unit heat gain} \\ \text{Casual heat gain} \\ \text{Ground temperature} \\ \text{Temperature above the ceiling} \\ \text{Temperature connected to the internal wall} \end{bmatrix} \] (4-12)

and one output, \( T_i \), which is the inside room temperature.

Then the room model can be expressed in state space form,

\[ \dot{x} = Ax + bu \] (4-13)
\[ y =Cx \]
where

\[
A = \begin{bmatrix}
\frac{(-U_1-U_2)}{C_1} & 0 & 0 & 0 & 0 & \frac{U_1}{C_1} \\
0 & \frac{(-U_3-U_4)}{C_2} & 0 & 0 & 0 & \frac{U_3}{C_2} \\
0 & 0 & \frac{(-U_5-U_6)}{C_3} & 0 & 0 & \frac{U_5}{C_3} \\
0 & 0 & 0 & \frac{(-U_7-U_8)}{C_4} & 0 & \frac{U_7}{C_4} \\
\frac{U_1}{C_6} & \frac{U_3}{C_6} & \frac{U_5}{C_6} & \frac{U_7}{C_6} & \frac{1}{C_6} & (-U_1 - U_3 - U_5 - U_6 - U_7 - U_8)
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
\frac{U_2}{C_1} & 0 & 0 & 0 & 0 & 0 \\
\frac{U_4}{C_2} & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{U_6}{C_3} & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{U_8}{C_3} & 0 & 0 & 0 \\
\frac{U_2}{C_5} & 0 & \frac{1}{C_6} & \frac{1}{C_6} & 0 & 0
\end{bmatrix}
\]

\[
C = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

and parameters \(C_i\) and \(U_i\) can be found in M. Gouda (2000).

### 4.2.2 Implementation

In this section, we will discuss the setup and assumptions we adopted in the implementation. The fault we are studying is the door being left open. When the fault occurs, there will be an additional heat exchange between the outside and the room, which can not be measured. This is modelled as an extra heat gain term into \(Q_c\). Thus we have

\[
f_a = \begin{bmatrix}
0 & 0 & Q_{\text{door}} & 0 & 0 & 0
\end{bmatrix}^T \tag{4-14}
\]

The value of \(Q_{\text{door}}\) can be computed by:

\[
Q_{\text{door}} = m_{\text{door}} C_p (T_o - T_i) \tag{4-15}
\]

where \(m_{\text{door}}\) is the air mass flow rate between outside and inside, whose value is determined by \(T_o, T_i\), and the geometry of the door.

For simplicity, we also assume the following:

1. A/C unit heat gain \(Q_c\), is measured; \(Q_s\) and \(Q_g\) are set to zero;
2. There is no sensor fault, i.e., $f_s = 0$.

3. There is only one disturbance, and it only affects the 6th state $T_i$, i.e.,
$$E = [0 \ 0 \ 0 \ 0 \ 0 \ 1]^T.$$ This is reasonable since the room temperature is more prone to be affected by uncertainties, while the wall nodes are not. This assumption ensures that solutions of $H$ exists.

4. The disturbance is normally distributed with zero mean.

**4.2.3 Analysis**

First, note that when fault occurs, the observer error dynamic (4–10) will be affected by the fault through the term $T B f_a$, and the error will in turn affect the residual $r = C e$.

Note that in this model, $C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$, which means only the last entry of the residual $r$ can be observed. Moreover, since $f_a$ is non-zero only on its third entry, $T B f_a = (T B)_{(:,3)}$.

Now examine the matrix $T$. From (4–4), we have
$$T = I - HC = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -h_1 \\ 0 & 1 & 0 & 0 & 0 & -h_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 - h_6 \end{bmatrix}$$

(4–16)

So
$$T B f_a = (T B)_{(:,3)} = Q_{door} \begin{bmatrix} -h_1 \frac{1}{C_6} \\ \vdots \\ -h_5 \frac{1}{C_6} \\ (1 - h_6) \frac{1}{C_6} \end{bmatrix}$$

(4–17)

Since only the last entry is of interest, we want $(1 - h_6) \frac{1}{C_6} \neq 0$, so $(1 - h_6) \neq 0$.

Also from (4–4), we know that
$$(I - HC)E = 0 \Rightarrow (1 - h_6)e_6 = 0$$

(4–18)

Here comes a conflict. Equation (4–18) says that if $e_6 \neq 0$, then $(1 - h_6) = 0$, which will make us unable to detect the fault.

In that case, we slightly modify the condition. Instead of making $(I - HC)E = 0$, we let $(I - HC)E = \begin{bmatrix} 0 & 0 & 0 & 0 & \beta \end{bmatrix}^T$. Then the residual we observe will be corrupted by
the disturbance. However by assumption 5, the disturbance is of zero mean. Thus, although corrupted by disturbance, the expected value of the residual will not be affected, i.e., it will be zero under normal condition, and non-zero under faulty condition.

With this change, we have

\[(1 - h_6)e_6 = \beta \Rightarrow (1 - h_6) = \frac{\beta}{e_6}\] (4–19)

Then, we have

\[
\dot{e} = Fe + TBf_a + (HC - I)Ed = (A_1 - K_1C)e + Q_{\text{door}}
\]

Apply Laplace transform, we get

\[
e(s) = (sI - (A_1 - K_1C))^{-1}(TBf_a(s) + (HC - I)Ed(s))\] (4–21)

Assume the fault to be a step signal, apply Final Value Theorem, we get

\[
\lim_{t \to \infty} e(t) = \lim_{s \to 0} se(s) = -(A_1 - K_1C)^{-1}(Q_{\text{door}}) + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \beta \end{bmatrix} \] (4–22)

Note that the residual equals to the last row. We see a tradeoff here: on one hand, we would like to make the effect of fault large, and \(\beta\) should be large; on the other hand, we would also like to make the effect of disturbance small, and \(\beta\) should be small. Thus, a tuning process is required. In this particular example, we choose \(\beta = 0.1\). An one hour simulation is performed to study the UIO, the door is opened at 30 minutes. Moreover, the outside temperature \((T_o)\) is set to 85°F, the temperature of ground \((T_{z1})\), ceiling \((T_{z2})\), and internal wall \((T_{z3})\) is set to 68°F, A/C heat gain is set to \(-1000\)W. The simulation result is shown in Figure 4-1 and 4-2. In the figures, the vertical line indicates the time when fault occurs. Figure 4-1 shows the room temperature and residual. As can be seen, before the
fault occurs, the residual remains close to zero, which indicates fault free status; after the fault occurs, the residual jumps up, which indicates the detection of fault. Figure 4-2 shows the estimated six states from UIO comparing with the actual states. It is clear that the observer estimates the states well when the system is free of fault but not when fault occurs. The simulation result shows that after tuning, the UIO is able to detect the door being left open.

We will give a short summary about the issues we encountered in this example. Due to the special structure of the room model, UIO can not be implemented directly to detect the fault. More generally, to detect the actuator fault appears in the $i^{th}$ row of $f_a$, the $i^{th}$ column of $TB$ must be non-zero, the general condition $TB \neq 0$ is not sufficient. In the case when the $i^{th}$ column of $TB$ does equal to zero, we developed a tradeoff method. Instead of eliminating the effect of disturbance from the observer, we relax the restriction a little, allowing $(I - HC)E$ to be non-zero. By doing this, the fault signal will appear in the estimation error dynamics, thus affects the residual. However, this move also bring disturbance into the observer. The larger we make the fault effect, the larger effect of disturbance occurs. Because of this tradeoff, tuning is required. Another design parameter is the poles of the observer. As can be seen from equation (4–22), the pole locations also
affect the final value of the estimation error, thus affect the value of residual when fault occurs. Normally, we would like the poles to be away from the imaginary axis so that the observer will converge faster, however, during our tuning, we observed that moving the poles to the left will decrease the final value of residual. This indicates there is another tradeoff we have to take into account when applying UIO method for fault diagnosis.

### 4.3 Door Status Detection

Now we attempt to detect opening door with UIO. We take the second order model for the room in equation (2–8), and we assume that the mass flow rate is constant for short period. Let $x = [T_i, T_w]^T$ be the states, $u = [T_{sd}, T_o, T_s, Q]$ be the input vector, and $y = T_i$ be the output, then the system can be write as:

$$\dot{x} = Ax + Bu$$

$$y = Cx$$

(4–23)
where

\[
A = \begin{bmatrix}
-\frac{1}{R_{\text{win}}C_r} & -\frac{1}{R_1C_r} & -\frac{mC_p}{C_r} & \frac{1}{R_1C_r} \\
\frac{1}{R_1C_r} & \frac{1}{R_{\text{win}}C_r} & \frac{1}{R_1C_w} & -\frac{1}{R_1C_w} & -\frac{1}{R_2C_w}
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
0 & \frac{1}{R_{\text{win}}C_r} & \frac{mC_p}{C_r} & \frac{1}{C_r} \\
\frac{1}{R_2C_w} & 0 & 0 & 0
\end{bmatrix}
\]

\[C = [1, 0]\]

(4–24)

Then we add disturbance and fault into the system. Again we assume the disturbance enters the system only through \(T_i\), i.e., the disturbance distribution matrix \(E = [1, 0]^T\). We have discussed in section 3.4, when door is open, the heat exchange can be modelled as an additional resistor between the hallway and the room. During this analysis, we make a little modification: consider the fault as an addition term into the input \(T_{sd}\). The reasoning is that when the door is open, more heat transfer occurs. By changing \(T_{sd}\), the same effect can be applied to the system. For example, when \(T_{sd} > T_i\), the fault can be considered to be a positive term added to \(T_{sd}\), which increases the temperature difference, and thus increase the heat transfer. By making this modification, opening the door can be modelled as an actuator fault. More precisely, we have \(f_a = [\alpha, 0, 0, 0]^T\). The whole system becomes:

\[
\dot{x} = Ax + B(u + f_a) + Ed
\]

\[y = Cx\]

(4–25)

We follow the procedure described earlier to design a UIO. First,

\[
H = E[(CE)^TCE]^{-1}(CE)^T = \begin{bmatrix}
1 \\
0
\end{bmatrix}
\]

(4–26)

and

\[
TB = (I - HC)B = \begin{bmatrix}
0 & 0 & 0 & 0 \\
\frac{1}{R_2C_w} & 0 & 0 & 0
\end{bmatrix}
\]

(4–27)

Note that the fault happened to be at the first input, so that

\[
TBf_a = \begin{bmatrix}
0 \\
\frac{\alpha}{R_2C_w}
\end{bmatrix}
\]

(4–28)
is nonzero, which means the fault will enter the error dynamics, and the standard UIO can be applied. All the design matrices, $F, T, K, H$, can be computed by the standard procedure in section 4.1.

To test the UIO we designed, we apply it to Pugh Hall data set E. We pick the time period from 9pm to 11pm, during which the supply air flow rate was kept close to a constant. The door was opened at 10:08pm. The result is given in Figure 4-3.

![Figure 4-3. Open door detection](image)

The vertical line in the figure indicates the time when door was opened. Before that time point, the observer estimation catches real temperature well, and the residual is close to zero. After the door was opened, the model changed due to the fault, the observer no longer catches the real temperature, which causes the residual to jump up.

In the door status detection case, standard UIO can be applied directly. The residual shows a clear change when the door is opened, which indicate the implementation of UIO to detect the door status is successful.
4.4 Summary

We discussed the design of unknown input observer, which estimates states with the presence of unknown inputs. Then a residual is generated to detect actuator faults and sensors fault. An example for implementation of UIO in fault diagnosis is presented, where it cannot be applied directly and tuning is required. More importantly, we applied this method to detect whether the door of a room is open. The experimental data collected from Pugh Hall shows the attempt is successful.
CHAPTER 5
CONCLUSION AND FUTURE WORK

In this thesis, we first developed a high order model (full scale model) with electrical analogy. Then we proposed a few low order models, starting with the lowest order possible (first order model). Comparison shows that second order model is a good choice for model structure. Higher order models do not have significant improvement. After the model structure is chosen, we use data from Pugh Hall on University of Florida campus to calibrate and validate the model. It is found that the data sets must be selected very carefully. Normal closed loop operation data may lead to grossly wrong parameter estimation. Specially designed forced response experiments are required. We also use an additional resistor to model the effect of open door, which provides good prediction in the door open scenario. With the model we developed, unknown input observer is successfully implemented to detect the open door.

Some directions for future work include: develop model of larger zone with multiple rooms; apply MPC to minimize energy consumption; transform the parameter estimation problem into a convex optimization problem, so that unique solution is guaranteed; study the identifiability of the problem, establish criterion for calibration data, so that enough information is provided in the data to find the correct parameters.
REFERENCES


BIOGRAPHICAL SKETCH

Yashen Lin was born in 1987 in Beijing, China. He received his Bachelor of Science degree in automation in 2009 from University of Science and Technology Beijing. He then joined the Distributed Control Group at University of Florida to pursue his doctoral degree under the advisement of Dr. Prabir Barooah. His research interest lies in the field of modeling and intelligent control of heating, ventilation, and air conditioning (HVAC) system in buildings.