Title
System-Level Monitoring and Diagnosis of Building HVAC System

Author
Wu, Siyu

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

SYSTEM-LEVEL MONITORING AND DIAGNOSIS OF BUILDING HVAC SYSTEM

by

Siyu Wu

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Mechanical Engineering

Committee in charge:
Professor Jian-Qiao Sun, Chair
Professor Gerardo Diaz
Professor Arnold Kim

The dissertation of Siyu Wu is approved:

Date

Date

Chair Date

University of California, Merced

To my parents.

ACKNOWLEDGEMENTS

The pursuit of the "love of wisdom" is certainly not a smooth sailing on a tranquil lake. It is a courageous voyage that explores a furious ocean full of uncertainties and hazards. This expedition is boundless due to the progressive nature of "wisdom", and timeless due to the infinitive depth of "love". It is an inquisitive perspective, a manifesto of lifelong learning, a fortitude to defeat difficulties, and a determination to exceed and to create. This overwhelming journey would not be an enjoyable experience for everyone. Fortunately, I am not alone in this expedition. Looking back, many mentors, friends, and families have helped, supported, encouraged, and inspired me. With their kindness, patience, and love, I have grown both intellectually and emotionally, obtained a better understanding of my strengths as well as weaknesses and, most importantly, broadened and enriched my view of life.

My deepest and earnest gratitude is indebted to my advisor, Professor Jian-Qiao Sun. Dr. Sun's immense and profound knowledge, uncompromised determination in the pursuit of knowledge, sagacity of grabbing the essences of any challenges, philosophical interpretation of research, and art of enlightening and encouragement are the most precious and indispensable beacon that has guided me through all these years. Uphill against a routine wind, he rides a bike to school everyday. That tremendous moment and willpower he demonstrates resonate as the Symphony No.5 of Beethoven, and educate and motivate me significantly. Dr. Sun carries the merits from both Chinese and American scholars, and has influenced me much beyond mere research by teaching me how to be a happy, humble, exploring, and creative person who is always inquiring while really enjoying the things he is doing. I am also thankful for the excellent example he has provided as both an accomplished scholar, a loving father, and a caring husband. Dr. Sun's first name in Chinese means "building the bridge". I will always be grateful for the bridge he has built in my life.

I would like to sincerely thank my thesis committee, Professor Gerardo Diaz, and Professor Arnold Kim for their time, expertise, insightful comments, and care that significantly enhanced and polished this thesis. Special thanks to Professor Arnold Kim, whose achievement and balance in both science and fine art have inspired me greatly.

I am grateful for previous and current UC Merced Facility Management staffs, Mr. John Elliott, Dr. Varick Erickson, Mr. Zuhair Mased, Mr. Julian Ho, and Mr. Ashby Shaw for helping our group, with promptness and care, obtain the

web accesses to the HVAC control system and database, and providing insights into existing maintenance problems. Without their help, it is impossible for me to conduct the thesis works.

It has been very privileged to get to know and collaborate with many talented, hard-working, and visionary researchers. Their technical excellences, years of experience, and visions have always inspired me to see broader and more insightful about the field the its future. I deeply appreciate the generous technical guidance and invaluable input from Dr. Michael Wetter, Dr. Xiufeng Pang, Dr. Philip Haves of Lawrence Berkeley National Lab, Mr. Joseph Deringer of Institute for the Sustainable Performance of Buildings, and Professor David Auslander of UC Berkeley.

I must not forget those anonymous reviewers of my both accepted and rejected research articles. Their expertise, meticulousness, challenging questions, and rigorous scientific standards have contributed significantly to my academic training in terms of logics and scientific writing.

For the past five years, our lab has been a source of friendships as well as good advice and collaborations. I would like to thank all previous and current lab mates and visiting scholars. And I am especially grateful for our summer research undergraduate students, Phiet Doan, Steven Lin, Liliana Mendoza, Edlyn Covarrubias, and Christopher Harrison for their tremendous efforts of collecting and pre-processing the data that provides the foundation of my research. I sincerely wish all of you a great career ahead after graduation.

I also want to express appreciation to the scholars whom I have been lucky to work with or got to know along this journey. I would like especially thank Professor Michael Modest (UC Merced) for enlightening me by his meticulous attention to details, former HUST chancellor Shuhuai Huang, Academician Honglie Sun (CAS), and Professor Raymond Chiao (UC Merced) for inspiring me with their sense of academic mission and responsibility, rigorous scientific spirit, and lifelong curiosity of science, and my advisors back in China, Professor Chunsheng Ye (HUST) and Professor Zheng Geng (IACAS) for trusting and guiding me through my multidisciplinary research.

I am grateful for the CITRIS Seed Grant project and California Energy Commission that have funded my research through all these years. I must also thank UC Merced staffs, especially Ms. Shannon Anderson and Ms. Tomiko Hales for their professional and continuous administrative support from the very beginning.

I am enormously indebted to my friends, who are resources of joy, advice, and encouragement. Especially, I would like to thank Mentar Mahmudi and Oktar Ozgen for illuminating me with their courageous and adventurous perspectives toward life; Nicola Basilico for enlightening me with his extraordinary talent to seize and enjoy happiness out of the pursuit of knowledge; Zhijiang Ye for his intuitive understanding of aesthetics and enthusiasm to embrace new experiences. I am

also thankful to Jun Hu, Hongyu Gao, and Tiffany McConnell for motivating me with their uncompromised determinations, diligences, perseverance, and optimisms towards their dreams.

I want to express my wholehearted gratitude to my families for their unconditional love, trust, encouragement, and sacrifice. Especially, I want to thank my father Yifan Wu for raising me with a sense of responsibility and an appreciation of music; I want to thank my mother Xiaofen Xiong for nurturing me with idealism, compassion, and a sense of justice. I am grateful for my paternal grandparents for those liberal conversations that enlighten and push forward my intellectual pursuit, and my maternal grandparents for managing to get me into the best kindergarten in my hometown, where I built a great health and formed the majority of my good habits, from which I benefit significantly afterwards. Special thanks also go to my aunt Xiaoxia Xiong, who was the dominant factor that pushed me into the Olympic mathematics competition during my elementary school years, which had challenged my laziness and changed the trajectory of my life.

Last but not least, I want to thank these beautiful Californians for their sunny, cheerful, and innovative spirits, which ignite and fuel the flame of optimism in my heart. To end this acknowledgement, and to encourage and brave myself for the journey ahead, I would like to quote part of a speech from such an inspiring Californian inventor:

> "Your time is limited, so don't waste it living someone else's life. Don't be trapped by dogma – which is living with the results of other people's thinking. Don't let the noise of others' opinions drown out your own inner voice. And most important, have the courage to follow your heart and intuition. They somehow already know what you truly want to become. Everything else is secondary."
>
> - Steve Jobs

CURRICULUM VITAE

Education

B.S. in Materials Processing and Control, Huazhong University of Science and Technology (Wuhan, China), 2006.

M.S. in Numeric Materials Processing, Huazhong University of Science and Technology (Wuhan, China), 2008.

Honors

Yearly Recognition of Achievement in Graduate Research and Teaching (2011, and 2012), University of California at Merced.

UC Merced Energy Research Institute Research Fellowship (2010), University of California at Merced.

Gradaute Tuition Scholarship (2008, 2009, 2010, 2011, and 2012), University of California at Merced.

Publications

S. Wu and J.Q. Sun (2012): "Two-Stage Regression Model of Thermal Comfort in Office Buildings". *Building and Environment.* 57, 88-96.

S. Wu and J.Q. Sun (2012): "Multi-Stage Regression Linear Parametric Models of Room Temperature in Office Buildings". *Building and Environment.* 56, 69-77.

S. Wu and J.Q. Sun (2012): "A Physics-Based Linear Parametric Model of Room Temperature in Office Buildings". *Building and Environment.* 50, 1-9.

S. Wu and J.Q. Sun (2011): "Cross-Level Fault Detection and Diagnosis on Building HVAC Systems". *Building and Environment.* 46(8), 1558-1566.

S. Wu and J.Q. Sun (2011): "A Top-Down Strategy with Temporal and Spatial Partition for Fault Detection and Diagnosis of Building HVAC Systems". *Energy and Buildings.* 43(9), 2134–2139.

S. Wu and J. Q. Sun (2010): "Multilevel Fault Detection and Diagnosis on Building HVAC Systems". *Proceedings of 2010 ACEEE Summer Study on Energy Efficiency in Buildings.*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Heating, ventilation, and air conditioning (HVAC) is an indoor environmental technology that is extensively instrumented for large-scale buildings. Among all subsystems of buildings, the HVAC system dominates the energy consumption and accounts for 57% of the energy used in U.S. commercial and residential buildings. Unfortunately, the HVAC system may fail to meet the performance expectations due to various faults, including not only complete hardware failures, but also non-optimal operations. These faults waste more than 20% of the energy HVAC consumes. Therefore, it is of great potential to develop automatic, quick-responding, intelligent, and reliable monitoring and diagnosis tools to ensure the normal operations of HVAC and increase the energy efficiency of buildings.

To achieve these goals, increasing attentions have been attracted to two research areas, i.e., models that monitor the indoor thermal environment, and fault detection and diagnosis (FDD) tools that capture abnormal HVAC performance. Despite contributions of the existing works, there are still many challenges in these two areas. For the thermal models, the major concerns lie in 1) most of the models are determined empirically, 2) optimal structures and orders of the models are often determined through simulations, 3) the predictions of the models degrade quickly over longer time intervals, and 4) a lack of studies to incorporate architectural parameters and control variables into the models. For the FDD, we face the challenges of 1) the inherent complexity, coupled hardware and software, and increasing scale of HVAC significantly complicate the nature of faults, 2) faults occur at different levels with various degrees of impacts on upper-level HVAC units, 3) practical FDD tools at the system-level are scarce, and 4) the computational efficiency and calibration onerousness of the simulation-based FDD is a concern.

In this thesis, we address these challenges by innovating a system-level monitoring and diagnosis tool for HVAC. For the monitoring, we study and establish a parametric modeling approach to present indoor air temperature and thermal comfort. The resulting models take advantages of both analytical and numerical modeling techniques. These models have a two-stage regression structure, and explicitly include both architectural parameters and control variables as its predictors. As a result, they allow parametric studies of influence of the building envelope on indoor thermal behavior, serve as an efficient foundation for intelligent HVAC control design, and help optimize the design of and the material selection for office

buildings. For the diagnosis, we innovate and develop a system-level FDD architecture for detecting faults across different levels of the HVAC system. Specifically, this architecture monitors and detects faulty HVAC units in a top-down manner. By monitoring HVAC units at higher level, instead of lower level components, the proposed FDD strategy reduces the computational effort in real-time monitoring of the HVAC system, obtains a system-level view of the HVAC operation, and provides a way to integrate the existing methods for component fault detection when needed. Based on extensive data collected from an office building on the campus of the University of California at Merced, numerical validations of the models, and examples of detected faults demonstrate the effectiveness of the proposed monitoring and diagnosis tool.

Chapter 1

INTRODUCTION

1.1 Background

1.1.1 What is Building HVAC

Heating, ventilation, and air conditioning (HVAC) is a technology of indoor environmental comfort. HVAC is usually instrumented for large office and industrial buildings such as university classrooms, hospitals, skyscrapers to maintain thermal comfort, and create healthy indoor air quality with relative low cost and high energy efficiency.

Ever since the invention of its components during the industrial revolution, HVAC has gradually evolved to a highly interdisciplinary and complex system. Numerous new components, advanced sensing technologies, advanced control algorithms, and artificial intelligence have been introduced into HVAC to meet operational objectives in different types of buildings worldwide. Consequently, HVAC systems have been extensively deployed in the fields of both developed and developing countries. Take the United States as an example. HVAC systems are currently conditioning a total area of nearly 3.1 billion square feet in buildings [1]. Many regulating and standards organizations such as HARDI, ASHRAE, SMACNA, ACCA, Uniform Mechanical Code, International Mechanical Code, and AMCA have been established to support the HVAC industry, maintain high standards, and promote innovations.

1.1.2 HVAC Energy Consumption

Due to rapid global population growth and urbanization, more and more large-scale buildings are being built. Buildings have become one of the fastest growing energy consuming facilities on the earth. According to the U.S. Department of Energy 2009 Building Energy Databook, buildings use 72% of nation's electricity, and 38.9% of nation's total energy consumption, valued at $392 billion [2–4]. Currently, HVAC systems account for 57% of the energy used in U.S. commercial and residential buildings [2–4].

1.1.3 HVAC Faults

Unfortunately, as its scale and complexity evolve, HVAC becomes more fault-prone. Here, faults include not only complete equipment failures, but also non-optimal operations, e.g., occupant discomfort, energy inefficiency, poor choice of operating targets, sensor calibration errors, poor controller tuning. HVAC may fail to meet the performance expectations due to these faults, thus wasting more than 20% of the energy it consumes. Therefore, it is of great potential to develop automatic, quick-responding, accurate and reliable fault detection and diagnosis (FDD) schemes to ensure the normal operations of HVAC in order to save energy. According to the National Institute of Standards and Technology (NIST), FDD methods have a potential to save 10-40% of HVAC energy consumption [5], valued at $233 billion annually in the United States [2–4]. An FDD package for HVAC can help establish construction and renovation standards for new and existing buildings. In light of the worldwide energy crisis and increasing environmental awareness, and the current limited usage of renewable energy in buildings [6], FDD for HVAC is critical to increase the energy efficiency of buildings. Since the 1980s, HVAC researchers and manufacturers worldwide have been striving to improve the energy efficiency of the system due to rising energy cost and increased awareness of environment.

1.1.4 Thermal Modeling

Mathematical models of HVAC units provide bases for detecting faults, upgrading control strategies, and improving commissioning, and thus have a potential to reduce energy consumption of HVAC systems by 20% to 30% [7]. Consequently, thermal modeling has been a very active field and attracted many researchers for the past two decades. Among all topics, the models of indoor temperature and thermal comfort have been particularly popular.

The indoor temperature characterizes the thermal condition of a room, as is the case in the predicted mean vote (PMV) model [8, 9], the ASHRAE Standard 55 [10] and the ISO Standard 7730 [11]. Ascione *et al.* need precise room temperature estimation to evaluate the strict thermo-hygrometric environment in museums [12]. Balaras *et al.* require the accurate knowledge of the room temperature to examine the indoor environment in hospitals [13]. The room temperature is also required for the analysis of thermal performance, indoor air quality [14], and for the evaluation of the indoor environment in naturally ventilated buildings [15]. Furthermore, a robust model that accurately predicts the room temperature is important for control design. Engdahl and Johansson have studied optimal supply air temperature in a variable air volume (VAV) system [16]. Yang and Kim predict the time of room temperature variations [17]. Orosa has investigated the thermal comfort based control strategy [18]. Tanimoto and Hagishima have developed a Markov

model for the on-off cooling schedule control [19]. Accurate, robust, simple-to-implement indoor temperature models therefore become indispensable for detecting faults, upgrading control strategies, and improving commissioning.

According to the studies conducted by Wyon [20], Djongyang and colleagues [21], and the World Health Organization [22], an unpleasant thermal sensation distracts people from work by reducing their satisfaction, concentration, and motivation, which leads to decreased productivity. In the developed countries, more than 90% of people spend most of their time indoor. Levels of pollutants indoor may be two to five times higher than those outdoor [23]. For certain places like hospitals, well regulated and maintained thermal comfort becomes even more vital in order to provide patients an ideal environment for recovery. Ensuring the comfort and health of the occupants has therefore been one of the primary goals for HVAC operation. A fundamental issue of the thermal comfort control is to establish indices that relate human comfort to physical parameters of the indoor environment.

1.1.5 Current FDD Practice of HVAC

The challenges for widespread deployment of FDD systems for HVAC lie in: 1) insufficient information on the possible energy and labor saving from automated FDD leads to weak demand from the building operation and maintenance community, 2) lack of adequate sensors installed in buildings due to the high cost leads to significant barriers to thoroughly assessing the operation of the HVAC system, 3) lack of easy access to real-time data because FDD is not yet integrated into building automation systems, particularly for old buildings, 4) lack of infrastructure to gather data from existing building automation systems (BASs) for add-on applications, and finally 5) lack of a top-down energy based FDD strategy.

Due to these difficulties, the current practice of HVAC FDD in the field involves a fair amount of manual inspection. Most manually detectable faults are obvious failures of certain equipment [24]. The collected data are overwhelming, as there is little effort to consolidate the information into a coherent searchable database describing the equipment status. For large-scale HVAC systems, which contain thousands of sensors and actuators, a complete inspection cycle takes a long time and the labor cost is too high. Non-optimal operations caused by inadequate operating targets, poor controller tuning, inaccurate sensor calibration, software error, etc., could also be faults and are even more difficult to detect by manual inspection.

1.2 Challenges

1.2.1 Inherent Complexity of HVAC

HVAC is a complex and highly integrated system. For example, the HVAC system for buildings on the UC Merced campus has a three-level structure with a top level bridge, a middle level air handling unit (AHU) and air distributors at lower

level, called variable air volume (VAV) unit. Each level includes a large number of components, such as sensors, controllers, actuators and air dampers. The hardware and software of the HVAC system are coupled across different levels. Environmental and architectural factors lead to different energy consumptions among units at the same level. A single component may have a fault which degrades the entire system performance [25]. Even more challenging, HVAC faults include not only complete equipment failures, but also non-optimal operations, e.g., occupant discomfort, energy inefficiency, poor choice of operating targets, sensor calibration errors, poor controller tuning. The inherent complexity and various natures of faults greatly increase the difficulty for FDD of the HVAC system.

1.2.2 Lack of System-Level FDD of HVAC

Over the last two decades, there has been considerable research and development on developing FDD methods for HVAC equipment at the component level.

Within VAV, Qin and Wang adopted a hybrid approach utilizing expert rules to detect ten types of faults in VAVs [26]. Song *et al.* invented a handy tool to detect faults in VAVs based on indoor temperature fluctuations [27]. Yang *et al.* used fractal correlation dimension to detect relatively small bias component faults under noisy conditions [28]. Cho *et al.* considered the transient pattern of fan, sensor, and damper faults [29]. By setting proper thresholds learned from trainings, Du *et al.* could detect flow sensor faults in air dampers and VAV terminals [30, 31]. As for AHU, Schein *et al.* found twenty-eight rules to detect five typical faults in AHUs [5, 32]. Ghiaus developed a bond graph to detect faults in the air conditioning system [33]. Chen *et al.* developed an easy-to-implement FDD method for detecting faults in rooftop air conditioners [34]. Wang and Xiao applied the principal component analysis (PCA) to detect sensor faults in AHUs [35–37]. There are also relevant researches such as Carling and Haves compared three fault detection methods with the field data [38]. Norford *et al.* evaluated fault detection methods on the basis of their sensitivity, robustness, the number of required sensors, and ease of implementation [39]. Research on FDD for other HVAC units such as chillers and cooling towers can be also found in the work by Namburu *et al.* [40], Haves and Khalsa [41], Castro and Remington [42], Han *et al.* [43], Navarro *et al.* [44], Bailey and Kreider [45], Khan *et al.* [46] and Weyer *et al.* [47].

Simulations can also be used to develop FDD strategies for component faults. Wang developed a recurrent cerebellar model articulation controller (RCMAC) to learn the normal characteristics of heating and cooling coil valves [48, 49]. Researchers from Lawrence Berkeley National Laboratory have been working on Modelica and EnergyPlus simulation tools for years [50].

Nevertheless, deployed FDD packages at the system-level are still few. Two most successful packages are the Automated FDD (AFDD) developed by NIST [51] and the Whole Building Diagnostician (WBD) innovated by the Pacific Northwest

National Laboratory (PNNL) [52]. The AFDD detects faults by evaluating a set of rules. It is sufficiently simple that it can be embedded into the local controllers, and operates within the processor and memory limitations of commercial HVAC systems [53]. However, the AFDD lacks an effective means to establish the fault detection thresholds and statistical parameters. The WBD implements the decision tree for the diagnosis. This approach can detect abnormal behaviors of the outdoor-air economizer and whole-building energy consumption. However, the WBD does not have the ability to detect faults across different levels, and needs an accurate baseline performance of the building energy consumption, which is difficult to obtain. When the building contains numerous components of various types, the complexity, instability, and cost of the decision tree approach become a concern [54].

1.2.3 Limitations of Simulation-Based FDD

The simulation approach calculates and predicts the normal operations. The predication forms the basis for fault detection. Clarke et al. developed a simulation-assisted control to simulate and test the response of building energy management systems (BEMS) [55]. Djuric et al. reviewed the possibilities and necessities for building lifetime commissioning and estimated the heating system performance using optimization tools and BEMS data [56, 57]. Song et al. investigated the feasibility of implanting the simplified simulation program into the energy management and control system (EMCS) [58]. Pedrini et al. applied the EnergyPlus simulation tool to develop a methodology for monitoring the energy performance of a commercial building in Brazil [59]. Researchers from Lawrence Berkeley National Laboratory have been working on Modelica and EnergyPlus simulation tools for years [50]. Since the HVAC system is inherently complex, and some thermal models need further optimization, the computational efficiency and calibration onerousness of simulation approach is a concern.

1.2.4 Limitations of Existing Temperature Models

As discussed previously, accurate, robust, simple-to-implement models of HVAC units provide bases for detecting faults, upgrading control strategies, and improving commissioning. Modeling HVAC thermal properties, e.g., temperature, has consequently been an active research field for decades. Popular software tools such as EnergyPlus [59], DOE-2, TRANSYS, Modelica [50], *etc.* make use of the thermodynamics models, incorporate architectural parameters, and provide the most comprehensive description of the thermal process in the building with accurate estimation of various system outputs. The analytical models allow parametric studies of influence of the building envelope on indoor thermal behavior. This is very important because architectural and material parameters of a building greatly influence its thermal performance. Extensive parametric studies will help us optimize the design

of and the material selection for buildings. However, various simplifying assumptions to deal with complexity of thermal interactions, unmeasured disturbances, uncertainty in thermal properties of structural elements and other parameters make it quite a challenge to obtain reliable analytical models [60].

The difficulties in thermodynamics models motivates the development of black-box models, e.g., ARMAX and neural networks, which build a relationship between inputs and outputs of the system by using measurements only. These numerical models can also help us understand the behavior of HVAC systems, simulate and predict the HVAC response, and conduct fault detection and diagnosis. Lowry and Lee investigated the response of internal temperature in an office building and discovered that the output-error (OE) model provides the best prediction [60]. Frausto and colleagues found that the linear auto regression model with external input (ARX) and the ARMAX model are suitable to describe the greenhouse behavior [61]. Ríos-Moreno *et al.* demonstrated that the ARX model can be adopted to predict classroom indoor air temperature with very high coefficients of determination [62]. Boaventura *et al.* [63] used a recursive identification technique to obtain a second-order model to predict the inside air temperature of a greenhouse. Mustafaraj *et al.* compared various numerical models of room temperature in an office, and found that the Box-Jenkins (BJ) model outperforms ARX and ARMAX [64]. Loveday and Craggs developed a BJ model including the external disturbances such as outside temperature and solar irradiance to analyze the thermal behavior of a full-scale occupied building [65]. Yiu and Wang studied the optimal order of a multiple input multiple output (MIMO) ARMAX model to forecast the performance of an air conditioning system of an office building in Hong Kong [66]. Peitsman and Bakker developed a multiple input single output (MISO) ARX model to evaluate the performance of a VAV unit [67]. Yoshida and Kumar studied the ARX model for off-line FDD of an AHU. They concluded that the model can be used to detect most of the faults [68]. Kumar *et al.* improved the single input single output (SISO) recursive auto regressive with exogenous input (RARX) model to detect faults in an AHU [69]. Jiménez *et al.* constructed an ARMAX model and obtained the physical interpretation of model parameters by comparing the ARMAX model with an equivalent physical model [70]. Although these numerical models are computationally efficient due to their simple structures, they are heavily dependent on the measurements, which implies poor generalization of the model in some parameter space.

To further improve flexibility, semi-physical models or grey-box models are constructed based on both insight into the underlying system physics and experimental data. Ghiaus *et al.* conducted the grey-box identification of constant air AHU elements [71]. Braun and Chaturvedi [72] developed a grey-box model to investigate the transient building load. Wang and Xu [73] implemented the genetic algorithm to identify the parameters of a simplified building model for building

thermal performance estimation in different weather conditions. Zhou *et al.* [74] integrated air temperature, relative humidity, and solar radiation prediction modules within a grey-box model to predict the building thermal load in the next day. Zheng and Li developed a physics-based model of a fan to relate its power consumption to fan speed and airflow [75]. Déqué and colleagues have implemented the grey-box technique to simulate the temperature variation in a ground floor flat [76]. Based on the physical laws, Wen and Smith have applied the grey-box approach to model the room temperature in VAVs [77]. Leephakpreeda has combined the grey prediction model and the adaptive comfort theory model to estimate the indoor comfort temperature [78]. Nevertheless, the grey-box models are numerical and therefore don't include architectural parameters.

Building architectural parameters impose strong impacts on building thermal performance. Tsilingiris has studied the influence of structural parameters of a wall on the heat loss [79]. Devgan *et al.* [80] and Aste *et al.* [81] have investigated the influence of external wall and window construction parameters on the building overall heat transfer and energy performance of well insulated buildings. By exploring the optimal dimensions of walls and glasses, Hwang and Shu [82], and Sozer [83] have tried to improve the thermal comfort, and the energy efficiency of buildings. Korolija *et al.* has concluded that it is not possible to form a reliable judgement about the building energy performance without considering the architectural parameters [84]. Numerical models without including the building parameters do not reveal the relationship between the building thermal performance and the parameters.

Some other features of the modeling research also emerge from the literature studies. First, most numerical models of HVAC systems are determined empirically. Choices of input and output variables are dependent on their temporal, geographical, and architectural changes of effect on room temperature [61–64]. Second, optimal structure and order of the model are often determined through intensive numerical simulations [61, 64]. Third, the prediction error of the model can increase quickly as the prediction extends over a longer time interval [60, 62, 64]. Fourth, although some researchers have developed their models using long-period data collected from a real building [61, 64, 85], the majority of numerical models are trained using either simulation data or limited measurements from a small number of HVAC units [60, 62, 63, 65–69].

1.2.5 Limitations of Existing Thermal Comfort Indices

Common thermal comfort indices include the PMV index [8,9], the simplified thermal comfort model [86] and the neural computing thermal comfort index [87], etc. Among those indices, the PMV is widely adopted as the principal index [21], and has multiple adaptations for different circumstances. Fanger and Toftum established an extension of the original PMV index to non-air-conditioned buildings in warm

climates [88]. Kulkarni and Hong employed the PMV model to evaluate the comfort level in transient pull downs [89]. Brager and de Dear concluded that, during the winter, a slight positive adjustment is needed when the original PMV index is being applied to office buildings in San Francisco [90]. There is a lack of studies of the PMV model to relate it to architectural parameters and control variables of office buildings.

The PMV model was initially established by Fanger in 1970 to determine the condition of human internal thermal neutrality, i.e. feeling neither too cold or too hot [8]. It was developed by considering heat conduction, convection, radiation, and evaporative heat loss between the occupant and the surrounding. It predicts the mean response of a large group of people according to the ASHRAE thermal sensation scale. The original PMV model includes many variables in two categories: environmental variables and personal variables. The environmental variables consist of air temperature, mean radiant temperature, relative air velocity, air relative humidity, convectional heat transfer coefficient. The personal variables include metabolism, external work, thermal resistance of clothing, ratio of fully clothed body surface area to nude body surface area, surface temperature of clothing, etc. Due to the inherent complexity of the heat transfer process and unavailability of certain variables, the original PMV model is not desirable for design purposes [91], nor is efficient for real-time control systems [92].

This difficulty leads to several empirical models that express the PMV in terms of variables easily measurable from the environment. An empirical equation describing the PMV index as a function of only the temperature and partial vapor pressure is developed at the Kansas State University and adopted by ASHRAE [93]. The study was based on a field study of 1600 school-age students. Hamdi *et al.* developed a predictive thermal sensation index that avoids the iterative solution process and serves as an objective function for feedback control of HVAC systems [92]. Atthajariyakul and Leephakpreeda studied the explicit functional relationship between the PMV index to four accessible control variables, namely, indoor air temperature, humidity, velocity, and outdoor airflow rate and two personal variables [87,94]. Chen and colleagues applied the fuzzy adaptive network to express the PMV index as a function of air temperature and relative humidity [95].

The architectural parameters of a building and control variables of HVAC greatly influence the thermal performance of the building. Hwang and Shu found that glazing types, depths of overhang, and glazing areas have significant effect on thermal comfort [82]. Wang and colleagues discovered that the window-to-wall ratio plays an important role of determining thermal comfort in naturally ventilated residential buildings in Singapore [96]. By using neural networks and computational fluid dynamics, Stavrakakis *et al.* investigated the optimal sizes of windows, and building direction for maximizing the occupational comfort and hygiene [97–99]. Yu *et al.* conducted a systematic evaluation on energy and thermal performance for

residential envelops index, and concluded that the architectural parameters, such as room volume, areas and heat transfer coefficients of external wall, window, and roof, shape coefficient of the building, influence greatly the energy consumption and thermal comfort of the building [100]. With regard to the control variables, Ghaddar *et al.* [101], Prianto and Depecker [102], and Ho *et al.* [103] studied buildings located in different climates and found that by regulating the circulation speed of the supply air, the thermal comfort can be increased. Clearly, there is a need for developing an empirical PMV model that relates to the architectural parameters and the control variables.

1.3 Our Approach

There is room for improvement and a need to achieve a system-level monitoring and diagnosis of HVAC. The proposed method should be computationally efficient and less onerous from a calibration perspective compared to most computational model based approaches. It should be able to monitor and detect both hardware failures and non-optimal operations. To accomplish these objectives, two parts of work are studied in this thesis: 1) thermal models for monitoring; 2) a system-level FDD architecture for diagnosis.

For the monitoring, we propose and study a parametric modeling approach to present indoor air temperature and thermal comfort. The resulting models take advantages of both analytical and numerical modeling techniques. And they have a two-stage regression structure, and explicitly include both the architectural parameters and control variables as its predictors. As a result, these models allow parametric studies of influence of the building envelope on indoor thermal behavior, serve as an efficient foundation for intelligent HVAC control design, and will also help us optimize the design of and the material selection for office buildings.

For the diagnosis, there is a need to develop a FDD architecture for detecting faults across different levels of the HVAC system with a focus on energy consumption. Specifically, this architecture monitors and detects faulty HVAC units in a top-down manner. By monitoring HVAC units such as bridges and AHUs at higher level, instead of lower level components such as fans and motors, the proposed FDD strategy reduces the computational effort in real-time monitoring of the HVAC system, obtains a system-level view of the HVAC operation, and provides a way to integrate the existing methods for component fault detection when needed.

Several key technologies developed in this thesis include 1) an energy flow model that can capture abnormal energy consumption irrespective to the nature of faults, 2) a spatial and temporal partition strategy that enables top-down monitoring and diagnosis of HVAC, 3) a top-down FDD structure, 4) a physics-based modeling approach of indoor temperature, 5) a multi-stage regression modeling approach of indoor temperature; and 6) a parametric modeling approach of thermal comfort.

1.3.1 Energy Flow Model

HVAC faults trigger abnormal energy density of the flow across an AHU or a VAV. A stuck damper or a leaking air duct in a VAV unit, for example, requires more cooling or heating supply air from its upper-level AHU. A frozen zone temperature sensor in a VAV below the cooling setpoint may demand less cooling supply air. A fouled air pipe in an AHU may require more supply air to compensate the reduced volume capacity. Since faults in HVAC systems trigger abnormal energy consumption, we develop an energy flow model that can capture these abnormal energy consumption irrespective to the nature of faults.

1.3.2 Spatial and Temporal Partition

HVAC units have certain patterns of flow energy consumption over time domain varying with environmental conditions, architectural factors, human occupancy and control setpoints. We create a spatial and temporal partition strategy [104] to group these units such that the ones in a group share a similar pattern of flow energy consumption over different time intervals. The units in the a group can be monitored with the same and tighter threshold for fault detection.

1.3.3 A Top-Down FDD Architecture

With the energy flow model and spatial and temporal partition strategy, we develop a system-level FDD structure that can capture abnormal energy consumption irrespective to the nature of faults. By monitoring energy flow and consumption, the FDD strategy can detect faulty HVAC units in a top-down manner. By monitoring the energy flow fluctuations of HVAC units such as bridges and AHUs at higher level, instead of lower level components such as fans and motors, the proposed FDD strategy reduces the computational effort in real-time monitoring of the HVAC system, obtains a system-level view of the HVAC operation, and provides a way to integrate the existing methods for component fault detection when needed.

1.3.4 Physics-Based ARMAX Modeling Approach

A physics-based ARMAX (pbARMAX) modeling approach is proposed to present the room temperature. Thermodynamic equations help us decide the structure and order of this proposed pbARMAX model. We have access to extensive measurements over a long period of time involving a large number of HVAC units. The data are used to develop and validate the pbARMAX model. The parameters of the numerical model are related to architectural and environmental factors. We propose to use the pbARMAX model to predict the thermal behavior of the rooms by room temperature prediction. The pbARMAX model also provides an analytical foundation of the spatial and temporal partition strategy.

1.3.5 Multi-Stage Regression ARMAX Modeling Approach

We propose two new multi-stage regression, physics-based linear parametric (mpbARMAX) models of the room temperature to take the advantages of both analytical and numerical modeling approaches [105, 106]. In other words, we build a two-stage regression structure into the mpbARMAX model to explicitly include building geometries and control variables.

1.3.6 Parametric PMV Model

It is a difficult task to incorporate the architectural parameters and control variables in the PMV model in an analytical manner. We propose a grey-box numerical modeling approach to develop such an empirical PMV model by following the regression method in [107]. The resulting empirical parametric PMV model (epPMV) has a two-stage regression structure, and explicitly includes both the architectural parameters and control variables as its predictors. As a result, it allows parametric studies of influence of the building envelope on indoor thermal behavior, serves as an efficient foundation for intelligent HVAC control design, and will also help us optimize the design of and the material selection for office buildings.

1.3.7 Highlights

The research in this thesis utilizes existing sensors and controller hardware, takes advantage of data-rich performance surveillance in buildings, and employs artificial intelligence, deductive modeling, and statistical methods to automatically detect and diagnose deviations between actual and optimal HVAC system performance.

Potential benefits of the research in this thesis include 1) reduction of energy consumption, 2) reduction of maintenance cost, 3) improvement of occupant comfort, and 4) longer equipment life and reduced unscheduled equipment down time. This system-level monitoring and diagnosis tool of HVAC systems will contribute to the long term goal of net zero energy, net zero waste, and net zero carbon footprint for office buildings.

Highlights of the research in this thesis include:

- A holistic data-driven approach at the system-level to monitor and diagnose building HVAC operations;
- Able to detect hardware failures as well as non-optimal operations;
- Integratibility with existing FDD works;
- Expandability to other machine learning and data mining techniques;
- Adaptability to hardware widely deployed in the field;

- Complies with industrial standards;
- Relatively low computational demand.

The rest of the thesis consists of six chapters. Chapter 2 presents the parametric modeling approach of temperature and thermal comfort. Chapter 3 demonstrates the model validation based on real data collected from a building on the campus of the University of California at Merced. Chapter 4 implements sensitivity analysis to further investigate the statistical reliability of the thermal comfort model. Chapter 5 presents the system-level FDD scheme, application examples of which are demonstrated in Chapter 6. Finally, Chapter 7 concludes the thesis. Nomenclature and model parameters are defined in Appendices A and B respectively.

Chapter 2

THERMAL MODELS

2.1 Energy Flow Model

2.1.1 Introduction

In order to capture different faults in a uniform manner, we define a feature that is capable of reflecting various faults at the component level, and is a common characteristic shared by all HVAC units.

2.1.2 Thermodynamical Foundation

HVAC faults trigger abnormal energy density of the flow across an AHU or a VAV. A stuck damper or a leaking air duct in a VAV unit, for example, requires more cooling or heating supply air from its upper-level AHU. A frozen zone temperature sensor in a VAV below the cooling setpoint may demand less cooling supply air. A fouled air pipe in an AHU may require more supply air to compensate the reduced volume capacity. Since faults in HVAC systems trigger abnormal energy consumption, we adopt the energy consumption of HVAC units as the feature. The normal operations of the HVAC system preserve certain patterns of energy consumption as a function of time and location. We may detect faulty HVAC units using the energy consumption feature.

Consider the energy transfer and consumption of hardware at different levels. The boiler and condenser supply the building with heating and chilling water. The pump transfers the water. The VFD fan produces a pressure difference in the supply air. The energy supplied to the HVAC system is eventually converted to flow of either air or water with a certain temperature, velocity, static pressure and humidity. All HVAC units share a common property: they have input and output flows as shown in Figure 2.1. According to Bernoulli's principle, the energy of the flow through a unit is given by

$$E = V(C_v T + \frac{1}{2}\rho v^2 + SP), \tag{2.1}$$

where the C_v and ρ are both known functions of temperature and pressure. This is a well-established model in thermal fluids [108]. The difference between input and output energies of the unit represents its energy 'consumption'.

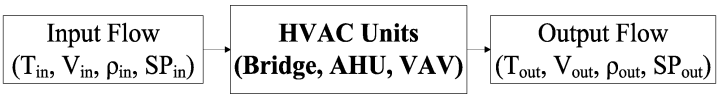

Figure 2.1: HVAC units with input and output flows of air or water.

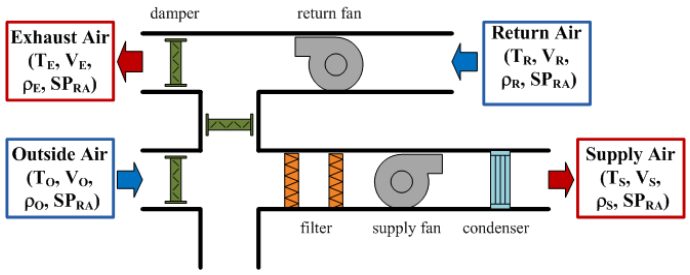

Figure 2.2: A typical AHU with two input and output air flows respectively.

2.1.3 The Model

A VAV takes the supply air from an AHU as input and return air as output. The energy change across the VAV is given by

$$\begin{aligned}\Delta E_{VAV} &= E_{RA} - E_{SA} \\ &= V_{RA}(C_v T_{RA} + \frac{1}{2}\rho v_{RA}^2 + SP_{RA}) - V_{SA}(C_v T_{SA} + \frac{1}{2}\rho v_{SA}^2 + SP_{SA}). \quad (2.2)\end{aligned}$$

In terms of the measured data, the above energy change can be rewritten as

$$\begin{aligned}\Delta E_{VAV} = Q_{SA}t[a(C_v T_{RM} + \frac{1}{2}\rho(aQ_{SA}/A_{RA})^2 + SP_{BLD}) \\ - (C_v T_{SA} + \frac{1}{2}\rho(Q_{SA}/A_{SA})^2 + \frac{1}{4}\sum_{i=1}^{4} SP_{SA}^i)], \quad (2.3)\end{aligned}$$

where

$$a = Q_{RF}/Q_{SF}. \quad (2.4)$$

Following the same approach, we can compute the energy change across an

AHU unit, as shown in Figure 2.2, with two inputs and two outputs

$$\begin{aligned}\Delta E_{AHU} = {} & Q_{RA}t(C_vT_{RM} + \frac{1}{2}\rho(Q_{RA}/A_{RA})^2 + SP_{RA}) + (Q_{SA} - (1-b)Q_{RA}) \\ & \cdot t(C_vT_{OA} + \frac{1}{2}\rho((Q_{SA} - (1-b)Q_{RA})/A_{OA})^2 + SP_{OA}) \\ & - Q_{SA}t(C_vT_{SA} + \frac{1}{2}\rho(Q_{SA}/A_{SA})^2 + \frac{1}{4}\sum_{i=1}^{4} SP_{SA}^i) \\ & - bQ_{RA}t(C_vT_{EA} + \frac{1}{2}\rho(bQ_{RA}/A_{EA})^2 + SP_{RA}),\end{aligned} \tag{2.5}$$

where

$$b = D_{EA}/(D_{RA} + D_{EA}). \tag{2.6}$$

2.2 Physics-Based Temperature Model

2.2.1 Introduction

The change of the room temperature T_{rm} is caused by both internal and external sources, e.g., cooling and heating air, solar radiation, heat conduction through walls, and human factors. To develop the pbARMAX model, we start with a well-studied ordinary differential equation (ODE) governing the evolution of room temperature under the influence of all these factors.

2.2.2 Thermodynamical Foundation

The energy balance of a room in steady-state can be expressed as [63, 70, 71],

$$\begin{aligned}\rho_i V C_v \frac{dT_{rm}}{dt} = {} & \dot{m}C_p(T_{disch} - T_{rm}) + h_{wa_i}S_{wa}(T_{wa_i} - T_{rm}) \\ & + h_{wd_i}S_{wd}(T_{wd_i} - T_{rm}) + \phi + \Delta E_{h,d},\end{aligned} \tag{2.7}$$

where T_{rm}, T_{disch}, T_{wa_i}, and T_{wd_i} represent the room, the discharge air, the inside wall surface, and the inside window surface temperatures respectively; V, S_{wa}, and S_{wd} denote the room geometries, i.e., room volume, surface area of the wall, and surface area of the window; ρ_i, C_v, C_p, h_{wa_i}, and h_{wd_i} are discharge air density, volumetric heat capacity of air, and heat capacity of air at constant pressure; ϕ, and $\Delta E_{h,d}$ represent the heat gain from solar flux, and from internal human and device. All the temperatures should be viewed as an average of the media. The equation states that the internal thermal energy change of a room equals to the summation of its heat gain from the discharged air from the HVAC system, through the walls, through the windows, from the solar flux, and from the heat generated by human and devices. Note that in this equation, we don't include the latent heat because there is a very low level of variations in humidity during the period of experiments reported later. However, the proposed model can readily take the latent heat into

consideration when the humidity data are available. Another simplification in the model is that the heat transfer through walls between rooms controlled by the same VAV with the same thermostat setpoint is neglected. This is based on the fact that, according to the data collected, the temperature difference between the room and the outside air is 2200% larger than the difference between two rooms on average. Therefore, the heat convection between the room and the outside air dominates the thermal interaction between the room and its surrounding. Nevertheless, our approach allows incorporating the heat convection between rooms into Eq. 2.7.

In the following, we make several simplifying assumptions about the heat transfer through walls and windows. Take the wall as an example. We assume that the thermal conductivity of the wall is constant, and that the heat conduction through the wall is one dimensional in steady-state. The temperature in the wall T_{wa} is a linear function of the thickness coordinate,

$$T_{wa} = \frac{T_{wa_o} - T_{wa_i}}{L_{wa}} x_{wa} + T_{wa_i}, \tag{2.8}$$

where T_{wa_o} is the outside wall surface temperature, L_{wa} is the thickness of wall, and x_{wa} is the distance from inside wall surface. Thus, the gradient of T_{wa} along the thickness direction is a constant.

We further assume that in the thermal equilibrium, the heat convection on both the inside and outside surfaces of the wall is equal to the heat conduction through the wall [108]. This gives us

$$-k_{wa}\frac{dT_{wa}}{dx_{wa}} = h_{wa_i}(T_{rm} - T_{wa_i}) = h_{wa_o}(T_{wa_o} - T_{oa}), \tag{2.9}$$

where k_{wa} is the thermal conductivity coefficient of wall, and T_{oa} is the outside air temperature.

Eqs. 2.8 and 2.9 suggest that T_{wa_i} can be expressed as a linear function of T_{rm} and T_{oa}, i.e.,

$$T_{wa_i} = P_1 T_{rm} + P_2 T_{oa}, \tag{2.10}$$

where the coefficients P_1 and P_2 are defined in Appendix B.

Following the same arguments, we obtain an expression for T_{wd_i} as

$$T_{wd_i} = P_3 T_{rm} + P_4 T_{oa}, \tag{2.11}$$

where P_3 and P_4 are defined in Appendix B.

With these simplifications, Eq. 2.7 now reads,

$$\frac{dT_{rm}}{dt} = P_{rm} T_{rm} + P_{disch} T_{disch} + P_{oa} T_{oa} + \Delta E, \tag{2.12}$$

where P_{rm}, P_{disch}, P_{oa} and ΔE are defined in Appendix B.

2.2.3 The Model

Let Δt be a sample interval and $T_{rm}(n)$ denote the sampled value of $T_{rm}(t)$ at time $t = n\Delta t$ where n is a positive integer. Applying the central difference scheme to Eq. 2.12 leads to the pbARMAX model as

$$T_{rm}(n+1) = T_{rm}(n-1) + 2\Delta t \left[P_{rm}T_{rm}(n) + P_{disch}T_{disch}(n) + P_{oa}T_{oa}(n) + \Delta E(n)\right]. \tag{2.13}$$

The coefficients of the pbARMAX model are determined by the properties of the architectural and control variables. The architectural parameters include the room volume, wall and window thickness, window-to-wall ratio, and heat conduction and convection constants. Since the control variables include the air mass flow rate and discharge temperature, the proposed pbARMAX model enables us to study the effect of these architectural and control variables in a numerical model. This aspect, however, is not discussed further in this section, and will be addressed in the following sections.

From the computational point of view, the pbARMAX model has the simplest structure with the least number of parameters when compared to a general ARMAX model without quantitative reference to thermodynamics [64]. According to the determination and validation criteria suggested by Norlén, a model with simplicity is preferred in order to increase the robustness of models [109].

We should also point out that the thermal inertia of the room temperature introduces certain time delay from model input to output [61, 110]. The proposed pbARMAX model can accommodate such a time delay by incorporating certain previous steps of measurements of inputs, e.g., T_{rm}, T_{disch}, T_{oa}, into the right part of Eq. 2.13.

2.3 Multi-Stage Regression Temperature Model

2.3.1 Introduction

In order to enable parametric study of building envelope parameters' impact on room temperature, we further enhance the pbARMAX model to have a multi-stage regression structure to include not only building envelope parameters but also control variables.

2.3.2 First-Stage Regression Model

In the physics-based temperature model presented in Eq. 2.13, there are four independent coefficients to determine, i.e., P_{rm}, P_{disch}, P_{oa} and ΔE. The model does not explicitly include all local control variables such as the air flow rate $\dot{m}$. This makes it difficult to use the model for control design. We can improve the

pbARMAX model by rearranging the coefficients so that $\dot{m}$ explicitly appears in the model. Referring to the definition of P_{rm} in Appendix B, we rewrite it as

$$\begin{aligned} P_{rm} &= -P_5 S_{wa}(h_{wa_i} P_2 + h_{wd_i}\zeta P_4) - P_5 \dot{m} C_p \\ &= -P_{oa} - P_{disch}, \end{aligned} \tag{2.14}$$

and substitute Eq. 2.14 into Eq. 2.13. We have

$$\begin{aligned} T_{rm}(n+1) = T_{rm}(n-1) &+ 2\Delta t a \cdot (T_{oa}(n) - T_{rm}(n)) \\ &+ b \cdot \dot{m}(n)(T_{disch}(n) - T_{rm}(n)) + \Delta E(n)], \end{aligned} \tag{2.15}$$

where a and b are in general functions of architectural parameters. Another way to arrange parameters can lead to a different form of the same model as

$$T_{rm}(n+1) = T_{rm}(n-1) + 2\Delta t \left[c \cdot (T_{disch}(n) - T_{rm}(n)) + d \cdot T_{oa}(n) + \Delta E(n)\right], \tag{2.16}$$

where c and d are also functions of architectural parameters.

For brevity, we shall refer to the models in Eqs. 2.15 and 2.16 as pbARMAX$_1$ and pbARMAX$_2$, respectively. We shall use a large set of measurements from VAVs involving a wide range of architectural parameters to determine the pbARMAX$_{1,2}$ models. This results in a collection of coefficients a, b, c and d.

2.3.3 Second-Stage Regression Model

With the help of the expressions in Appendix B, we can derive the following linear relationships for the coefficients a, b, c and d,

$$a = p_{11}/V + p_{12} \cdot S_{wa}/V + p_{13} \cdot S_{wd}/V, \tag{2.17a}$$
$$b = p_{21}/V + p_{22} \cdot S_{wa}/V + p_{23} \cdot S_{wd}/V, \tag{2.17b}$$
$$c = p_{31}/V + p_{32} \cdot S_{wa}/V + p_{33} \cdot S_{wd}/V, \tag{2.17c}$$
$$d = p_{41}/V + p_{42} \cdot S_{wa}/V + p_{43} \cdot S_{wd}/V, \tag{2.17d}$$

where V is the volume of the room, S_{wa} is the area of the wall, S_{wd} is the area of the windows, and p_{ij} $(1 \le i \le 4, 1 \le j \le 3)$ are constants. It should be noted that these coefficients are also dependent on other geometrical and material parameters. In the present work, we shall focus on $1/V$, S_{wa}/V and S_{wd}/V. In fact, the surface-area-to-volume ratios S_{wa}/V and S_{wd}/V are the driving forces to the thermodynamic processes that minimize the free energy [111, 112].

Substituting these linear relationships into Eqs. 2.15 and 2.16, we obtain two-stage regression models of the room temperature, denoted as mpbARMAX$_{1,2}$.

$$\begin{aligned} T_{rm}(n+1) = T_{rm}(n-1) &+ 2\Delta t[(p_{11}/V + p_{12} \cdot S_{wa}/V + p_{13} \cdot S_{wd}/V) \\ &\cdot (T_{oa}(n) - T_{rm}(n)) + (p_{21}/V + p_{22} \cdot S_{wa}/V + p_{23} \cdot S_{wd}/V) \\ &\cdot \dot{m}(n)(T_{disch}(n) - T_{rm}(n)) + \Delta E(n)], \end{aligned} \tag{2.18}$$

and

$$T_{rm}(n+1) = T_{rm}(n-1) + 2\Delta t[(p_{31}/V + p_{32} \cdot S_{wa}/V + p_{33} \cdot S_{wd}/V) \cdot (T_{disch}(n) - T_{rm}(n)) + (p_{41}/V + p_{42} \cdot S_{wa}/V + p_{43} \cdot S_{wd}/V) \cdot T_{oa}(n) + \Delta E(n)]. \quad (2.19)$$

It should be noted that the mpbARMAX$_{1,2}$ models can be identified with one-stage regression by directly using Eqs. 2.18 and 2.19. However, the two-stage regression has its merits.

1. According to Baguley [113], Ahn [114] and Kristjansson *et al.* [115], it is beneficial to separate regression modeling of fixed and random effects. Here, the architectural parameters are fixed, and the temperature variations are random. The variability of temperature measurements from different VAVs is usually greater than the variability of the measurements from the same VAV. The regressions in two stages tackle these two variabilities separately and effectively [116].

2. The number of independent data sets of $1/V$, S_{wa}/V and S_{wd}/V is significantly smaller than that of the temperatures $T_{disch}(n)$, $T_{oa}(n)$ and $T_{rm}(n)$. By separating these variables in two stages of regression, the samples in each stage have similar distributions, thus improving the so-called compound symmetry of samples as discussed in [117, 118]. This usually leads to better predictions.

3. Compared to the one-stage regression, the ratio of training samples to predictors in the two-stage regression increases more than 200%. This reduces the potential of overfitting [119, 120].

2.4 Parametric PMV Model

2.4.1 Introduction

We first present a further improved energy balance equation in a room with consideration of both sensible and latent heat. Then, we develop a regression representation of an ASHRAE empirical PMV model [93]. The regression model has a two-stage structure with coefficients that are explicit functions of architectural parameters and control variables.

2.4.2 Thermodynamical Foundation

Consider the rate of energy change in a room with the effects of latent heat, heat conduction, convection, and radiation.

$$\rho_i V C_v \frac{dT_{rm}}{dt} = \dot{m}(H_{disch} - H_{rm}) + h_{wa_i} S_{wa}(T_{wa_i} - T_{rm}) + h_{wd_i} S_{wd}(T_{wd_i} - T_{rm}) + \phi + \Delta E_{h,d}, \quad (2.20)$$

where T_{rm}, T_{wa_i} and T_{wd_i} represent the temperature of the room, inside wall surface, and inside window surface; V, S_{wa} and S_{wd} denote the room volume, surface area of the wall, and surface area of the window; ρ_i, C_v, h_{wa_i} and h_{wd_i} are the indoor air density, volumetric heat capacity of the indoor air, and heat convection coefficients on the inside surfaces of wall and window; ϕ, and $\Delta E_{h,d}$ represent the heat gain from solar flux, and from internal human and device, H_{disch}, and H_{rm} denote the specific enthalpy of discharge air and room air. Since enthalpy includes both sensible and latent heat, this energy balance equation is able to handle the dehumidification process which covers a significant portion of energy consumed by many HVAC systems [121]. This equation represents an improvement of the pbARMAX and mpbARMAX models without latent heat by describing the thermal process of HVAC systems more comprehensively.

Several assumptions are made for the energy balance equation including a) All the temperatures should be viewed as an average of the media; b) The heat transfer through walls between rooms controlled by the same VAV with the same thermostat setpoint is neglected. This is based on the fact that, according to the data collected, the average temperature difference between the rooms and the outside air is 2.49 °C. While the average temperature difference between adjacent rooms is merely 0.11 °C. Therefore, the heat convection between the room and the outside air dominates the thermal interaction between the room and its surrounding.

The specific enthalpy of the humid air can be calculated as [122–124]

$$H = C_{pa}T + \alpha(C_{pw}T + H_{we}), \tag{2.21}$$

where C_{pa} is the specific heat capacity of air at constant pressure, C_{pw} is the specific heat of water vapor at constant pressure, and H_{we} is the evaporation heat of water. C_{pa}, C_{pw} and H_{we} can be taken as constants 1.006 $kJkg^{-1}\,^{\circ}\mathrm{C}^{-1}$, 1.84 $kJkg^{-1}\,^{\circ}\mathrm{C}^{-1}$ and 2501 $kJkg^{-1}$ respectively. α is the humidity ratio given by

$$\alpha = 0.62198\frac{p_w}{p_a - p_w}, \tag{2.22}$$

where p_w is the partial pressure of water vapor in the moist air, and p_a is the atmospheric pressure of the moist air and can be taken as a constant of 101325 Pa. p_w can be calculated as

$$p_w = \frac{R}{10^3}e_d(T), \tag{2.23}$$

where R is the relative humidity of the moist air in percentage, and the function $e_d(T)$ represents the partial vapor pressure at dew point given by [125]

$$\begin{aligned} e_d(T) = \exp\Bigg[&-\frac{6096.9385}{T} + 21.2409642 \\ &- \frac{2.711193}{100}T + \frac{1.673952}{10^5}T^2 + 2.433502\log T\Bigg]. \end{aligned} \tag{2.24}$$

From Eqs. 2.21, 2.22, 2.23, and 2.24, we obtain the specific enthalpy as a function of the temperature and relative humidity

$$H = C_{pa}T + 0.62198\frac{p_w(C_{pw}T + H_{we})}{p_a - p_w}. \tag{2.25}$$

Hence, Eq. 2.20 reads

$$\begin{aligned}\rho_i V C_v \frac{dT_{rm}}{dt} = \dot{m}\Bigg[C_{pa}(T_{disch} - T_{rm}) + \frac{p_{w_{disch}}(C_{pw}T_{disch} + H_{we})}{p_a - p_{w_{disch}}} \\ - \frac{p_{w_{rm}}(C_{pw}T_{rm} + H_{we})}{p_a - p_{w_{rm}}}\Bigg] + h_{wa_i}S_{wa}(T_{wa_i} - T_{rm}) \\ + h_{wd_i}S_{wd}(T_{wd_i} - T_{rm}) + \phi + \Delta E_{h,d}.\end{aligned} \tag{2.26}$$

We assume that the thermal conductivity of the wall and window is constant, the heat conduction through the wall and window is one dimensional in steady-state, and the heat convection on both the inside and outside surfaces of the wall and window is equal to the heat conduction through the wall [107], the inside surface temperatures of wall and windows can be expressed as

$$\begin{aligned}T_{wa_i} &= P_1 T_{rm} + P_2 T_{oa}, \\ T_{wd_i} &= P_3 T_{rm} + P_4 T_{oa},\end{aligned} \tag{2.27}$$

where P_1, P_2, P_3, and P_4 are listed in Appendix B.

With Eqs. 2.27, Eq. 2.26 eventually arrives at

$$\begin{aligned}\frac{dT_{rm}}{dt} &= P_{disch}(T_{disch} - T_{rm}) + P_{oa}(T_{oa} - T_{rm}) \\ &\quad + P_f[F(T_{disch}, R_{disch}) - F(T_{rm}, R_{rm})] + \Delta E.\end{aligned} \tag{2.28}$$

where P_{rm}, P_{disch}, P_{oa}, P_f, and $F(T, R)$ are listed in Appendix B.

Let Δt be a sample interval, and X^n denote the sampled value of $X(t)$ at time $t = n\Delta t$ where n is a positive integer. We apply central difference scheme to the left side of Eq. 2.28 leading to

$$\begin{aligned}T_{rm}^{n+1} = T_{rm}^{n-1} + 2\Delta t\{a(T_{oa}^n - T_{rm}^n) + b(T_{disch}^n - T_{rm}^n) + c\dot{m}^n(T_{disch}^n - T_{rm}^n) \\ + d[F(T_{disch}^n, R_{disch}^n) - F(T_{rm}^n, R_{rm}^n)] + \Delta E^n\},\end{aligned} \tag{2.29}$$

where coefficients a, b, c, and d are defined in Appendix B.

It should be noted that other finite difference schemes, e.g., forward difference, can also be adopted to discretize the derivative of room temperature. However, as briefly discussed later in Section 3.4.3.2, central difference produces slightly more accurate prediction results than forward difference for our model.

Table 2.1: Coefficients of the ASHRAE empricial PMV model in Eq. 2.30.

Period	Occupants	β_1	β_2	β_3
1 hour	men	0.220	0.233	−6.673
	women	0.272	0.248	−7.245
	both	0.245	0.248	−6.475
2 hours	men	0.221	0.270	−6.024
	women	0.283	0.210	−7.694
	both	0.252	0.240	−6.859
3 hours	men	0.212	0.293	−5.949
	women	0.275	0.255	−8.620
	both	0.243	0.278	−8.802

2.4.3 The Model

We adopt an empirical PMV model from the ASHRAE Handbook [93],

$$PMV = \beta_1 T + \beta_2 p_w + \beta_3, \tag{2.30}$$

where β_1, β_2 and β_3 are constants listed in Table 2.1. Since p_w is a nonlinear function of T given in Eq. 2.23, PMV is also nonlinear in T. Here, we use Eq. 2.29 to develop a parametric model of the PMV. In particular, we employ the two-stage regression method discussed in Section 2.3 to relate the PMV to the envelope parameters and control variables of the building.

2.4.3.1 First-Stage Regression

We rearrange the function $F(T, R)$ in Appendix B as an explicit function of the PMV as,

$$F(T, R) = F_1 PMV - F_2, \tag{2.31}$$

where $F_{1,2}$ are defined in Appendix B.

By substituting Eq. 2.31 into Eq. 2.29, we have

$$\begin{aligned} T_{rm}^{n+1} = T_{rm}^{n-1} + 2\Delta t[a(T_{oa}^n - T_{rm}^n) + b(T_{disch}^n - T_{rm}^n) + c\dot{m}^n(T_{disch}^n - T_{rm}^n) \\ + d(F_{1_{disch}}^n PMV_{disch}^n - F_{1_{rm}}^n PMV_{rm}^n + F_{2_{rm}}^n - F_{2_{disch}}^n) + \Delta E^n]. \end{aligned} \tag{2.32}$$

Then, we substitute Eq. 2.32 into Eq. 2.30. After rearranging the terms, we obtain an equation governing the evolution of PMV, which we shall refer to as the epPMV model,

$$\begin{aligned} PMV_{rm}^{n} = {} & \beta_1 T_{rm}^{n-2} + \beta_2 p_{w_{rm}}^{n} + x_1 \frac{PMV_{rm}^{n-1}}{p_a - p_{w_{rm}}^{n-1}} + x_2 (T_{oa}^{n-1} - T_{rm}^{n-1}) \\ & + x_3 (T_{disch}^{n-1} - T_{rm}^{n-1}) + x_4 \dot{m}^{n-1} (T_{disch}^{n-1} - T_{rm}^{n-1}) + x_5 \frac{PMV_{disch}^{n-1}}{p_a - p_{w_{disch}}^{n-1}} \\ & + x_6 \frac{1}{p_a - p_{w_{rm}}^{n-1}} + x_7 \frac{p_{w_{rm}}^{n-1}}{p_a - p_{w_{rm}}^{n-1}} + x_8 \frac{1}{p_a - p_{w_{disch}}^{n-1}} + x_9 \frac{p_{w_{disch}}^{n-1}}{p_a - p_{w_{disch}}^{n-1}}, \end{aligned} \tag{2.33}$$

where x_i $(i = 1 \cdots 9)$ are functions of architectural parameters and material constants. The terms multiplied x_i excluding PMV are available measurements. Hence, the format of Eq. 2.33 is good for regression when x_i are taken as the regression coefficients. We shall use the collected data to obtain x_i with the least square method. A collection of the regression coefficients x_i will be obtained for a large number of rooms in the building with different architectural parameters. This step is called the first-stage regression. All the local control variables such as discharge air temperature, relative humidity, and air flow rate are the predictors of the epPMV model.

2.4.3.2 Second-Stage Regression

With the collection of the regression coefficients x_i on hand for a large number of rooms in the building with different architectural parameters, we seek for the functional relationships between the regression coefficients and the architectural parameters. The analytical expressions listed in Appendix B suggest the following linear functions in three terms $\frac{1}{V}$, $\frac{S_{wa}}{V}$ and $\frac{S_{wd}}{V}$,

$$\hat{x}_i = p_{i1} \frac{1}{V} + p_{i2} \frac{S_{wa}}{V} + p_{i3} \frac{S_{wd}}{V}, \tag{2.34}$$

where $\hat{x}_i$ are the estimate of x_i, V is the volume of the room, S_{wa} is the area of the wall, S_{wd} is the area of the windows, and p_{i1}, p_{i2}, and p_{i3} $(i = 1 \cdots 9)$ are functions of the material and air properties. p_{i1}, p_{i2}, and p_{i3} are treated as constant for a given building, and are determined with the collection of coefficients x_i for various parameter combinations of V, S_{wa} and S_{wd} by the least square method. This is the second-stage of regression.

In the present work, we shall focus on the three architectural terms $1/V$, S_{wa}/V and S_{wd}/V. In fact, the surface-to-volume ratios S_{wa}/V and S_{wd}/V are the driving forces to the thermodynamic processes that minimize the free energy [111, 112].

Eq. 2.34 together with Eq. 2.33 defines the empirical parametric PMV (epPMV) model as an explicit function of architectural parameters, control variables, and environmental conditions.

Chapter 3

MODEL EVALUATION

3.1 Building Description

Extensive past and current measurements of the HVAC system in the Science and Engineering (SE) building of UC Merced are available to us. The SE building is instrumented with a network of sensors and controls for the HVAC system. Such a highly instrumented building serves as an ideal living laboratory to support the research on energy efficiency. The SE building is a four-floor, southwest orientation, $19,666$ gross square meter building with primary use as office and laboratory. Two heating and cooling water bridges, nine unit heaters and ten AHUs work together with sixty-one VAVs to control and regulate 374 rooms and spaces in the building.

There are two major types of rooms, i.e., the faculty office and the conference room. The geometrical dimensions of the faculty office, and the conference room are $3.2004 \times 4.4714 \times 3.0480$ (length$\times$width$\times$height) cubic meters, and $9.6012 \times 5.1816 \times 3.0480$ cubic meters respectively. Figure 3.1 shows the third floor map of the SE building with detailed distribution of VAV units. Table 3.1 presents the geometries of the rooms regulated by VAVs under AHU 10 ($A10$). It should be noted that the volumes are calculated based on the room heights measured from the floor to the suspended ceiling. Some rooms have no surface exposed to the outside.

Merced, located in the San Joaquin Valley of California, is in the Mediterranean Steppe eco-region. Its climate exhibits rich variations during a year. The summer is hot and dry from June to August, and the winter is cold and rainy from November to April. The measurements of the HVAC system have rich dynamics.

As for the local control systems, the chilling plant has a lead-lag-standby setup of three chillers, and the heating plant consists of three, dual fuel boilers. The pump speeds of both the plants are modulated by PID controls to maintain differential pressures at discharge. The AHU units apply PID controls to regulate the supply fan variable-frequency drive (VFD), return fan VFD, and supply air temperature. These control loops maintain the duct pressure, return fan discharge pressure, and supply air temperature according to the setpoint. The VAV units implement two separate control loops, i.e., the cooling loop and the heating loop, to keep the temperature at setpoint. Both of the loops apply PI controls with three operational modes: heating, cooling and deadband.

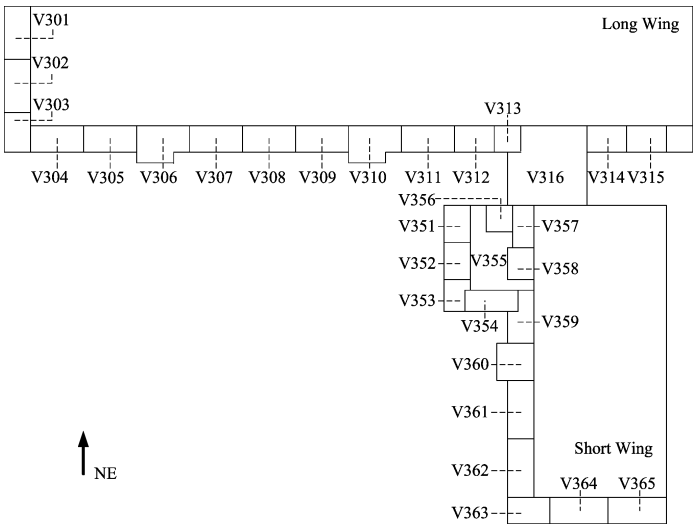

Figure 3.1: The distribution of VAVs on the third floor of the SE building on UC Merced campus.

Table 3.1: The geometriies of the rooms or spaces controlled by VAVs of AHU 10.

VAV	S_{wa} (m^2)	S_{wd} (m^2)	V (m^3)	VAV	S_{wa} (m^2)	S_{wd} (m^2)	V (m^3)
$V351$	11.9	19.0	78.7	$V251$	7.6	50.5	151.6
$V352$	3.0	15.5	78.7	$V252$	1.2	20.2	80.8
$V353$	8.1	24.3	75.1	$V253$	5.2	24.3	43.6
$V354$	13.5	9.4	73.9	$V254$	8.2	9.4	58.1
$V355$	1.4	4.7	181.9	$V255$	2.0	1.8	208.7
$V356$	6.7	4.7	50.2	$V256$	0	0	94.9
$V357$	0	0	97.4	$V257$	0	0	61.9
$V359$	13.6	6.8	132.4	$V258$	13.6	6.8	132.4
$V360$	8.1	48.6	244.7	$V259$	2.4	35.9	165.0
$V361$	19.0	10.2	130.8	$V260$	19.0	10.2	130.8
$V362$	19.0	10.2	130.8	$V261$	19.0	10.2	130.8
$V363$	8.8	30.4	101.0	$V262$	8.8	30.4	101.0
$V364$	4.8	24.5	151.5	$V263$	4.8	24.5	151.5
$V365$	0.4	46.7	151.5	$V264$	0.4	46.7	151.5

3.2 Physics-Based Temperature Model

3.2.1 Data Preprocessing

Measurements from sixty-one VAVs controlled by two AHUs are available for the development of the pbARMAX model. From May 31 to September 16, 2010, measurements of T_{rm}, T_{disch}, and T_{oa} from each VAV have been collected over 109 days with a sampling interval of five minutes, resulting in $31,392$ samples. The samples are divided into a training set and a validation set.

3.2.2 Model Validation

We identify the three parameters P_{rm}, P_{disch}, and P_{oa} with the method of least squares by using Eq. 2.13 where ΔE is treated as an unknown disturbance. The least square solution of those three parameters is obtained by using the back-slash operator in MATLAB [126]. This process is commonly referred to as training of the model. We have numerically determined the optimal number of samples for parameter identification by following steps: 1) For a particular VAV, we identify different sets of the regression parameters P_{rm}, P_{disch}, and P_{oa} with training sets by an increment of days; 2) For each set of regression parameters, we form a corresponding pbARMAX model, and use it to predict temperature of the following six weeks after the training set; 3) We evaluate the prediction residuals for all pbARMAX models obtained in the last step. The training length yielding the least residual determines the optimal number of samples for parameter identification of the VAV.

The local building energy management system applies different thermostat setpoints during the day from 5am to 1am of the next day, and the night from 1am to 5am. Most of the occupants work during the day. There is no significant difference in occupancy between weekdays and weekends. We divide the data by day and night. This is consistent with the temporal partition that will be discussed in Chapter 5. Figure 3.2 gives an example in which the optimal training length for a VAV ($V351$) is determined. The change of the training length leads to varied sets of parameters P_{rm}, P_{disch}, and P_{oa}. The overall effect of such parameter changes influences the prediction of the pbARMAX model. It is clear that as the training length increases, prediction errors in both the day and night drop gradually until reaching their minimum at Day 26 and Day 25 respectively. After reaching their minimums, the prediction errors increase due to overfitting [119]. This pattern is clearly shared by all VAVs involved. Also, the order of magnitude of prediction error during the day is on average 10^2 times smaller than that during the night. This is because: a) samples in the day time are six times of those at night, and b) due to a 87% tighter thermostat setpoints, the HVAC system operates more often during the day than at night. This yields much richer dynamics in the data in the day time, which is better for numerical modeling.

To validate the model, we first select a metric for evaluating the accuracy of prediction. In the literature [61, 127–130], it is common to use the mean absolute

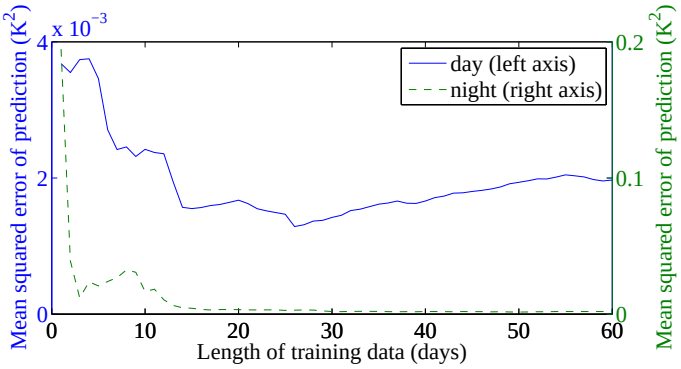

Figure 3.2: The mean square prediction vs. the length of the training data. For $V351$, and most of the other sixty VAVs, the optimal length of the training data during the summer is between twenty to thirty days.

error (MAE), mean squared error (MSE), the root mean squared error ($RMSE$), coefficient of determination (r^2) and maximum absolute error ($MaxAE$). These are defined as

$$MAE = \frac{1}{m}\sum_{i=1}^{m} |X_i - X_i^*|\,, \tag{3.1}$$

$$MSE = \frac{1}{m}\sum_{i=1}^{m} (X_i - X_i^*)^2, \tag{3.2}$$

$$RMSE = [\frac{1}{m}\sum_{i=1}^{m} (X_i - X_i^*)^2]^{1/2}, \tag{3.3}$$

$$r^2 = \frac{(m\sum X_i X_i^* - (\sum X_i)(\sum X_i^*))^2}{(m(\sum X_i^2) - (\sum X_i)^2)(m(\sum X_i^{*2}) - (\sum X_i^*)^2)}, \tag{3.4}$$

$$MaxAE = \max_i |X_i - X_i^*|\,, \tag{3.5}$$

where X_i, and X_i^* are the estimated values and measurements of a particular attribute in time domain respectively, and m is the total number of measurements.

Figures 3.3 and 3.4 present three-day room temperature prediction of $V351$. The optimal training lengths, i.e., 26 days for the day partition, and 25 days for the night partition are used to train the models respectively during the day and night. And the resulting models are implemented to predict temperature in the following three days after training sets. It is observed that, the prediction not only matches well with the measured T_{rm} during the night when the HVAC system has less dynamics, but also captures abrupt fluctuations during the day. The residual

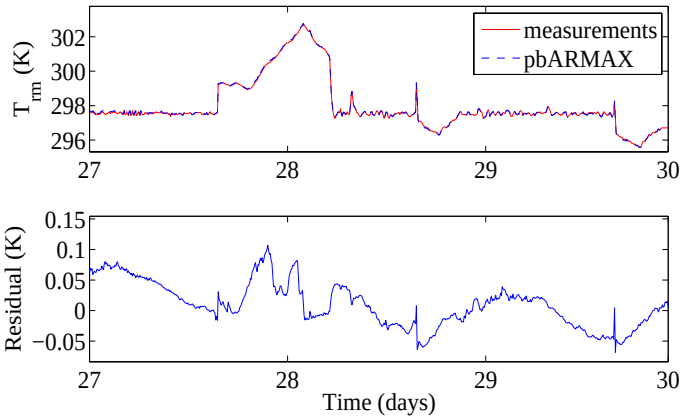

Figure 3.3: The short-term temperature prediction in room $V351$ during the day with the pbARMAX model. The prediction tracks well the measured temperature both when it is relatively smooth or when it changes rapidly.

of prediction remains within a small range with no significant increase as prediction length extends from one day to three days.

We also conduct parameter identifications for the other sixty VAVs. The residuals of three-day predictions for all VAV models are of the same order of magnitude as $V351$. In fact, all the VAV models can predict the room temperature over more than ten weeks with $MaxAE$ less than 0.08. Take $V352$ as an example. The optimal training length for this VAV is eleven days. The model predictions of the following ninety-four days keep a $MaxAE$ less than 0.065. Figure 3.5 shows part of this long-term prediction. It is clear that while the room temperature demonstrates different fluctuation patterns, the residual of prediction remains consistent and does not increase significantly.

To investigate the relationship between the prediction error and time duration quantitatively, we take $V201$ as an example. The model is trained with the data over one day only. The predictions over one to five weeks are shown in Table 3.2. It can be seen that as the prediction extends over time, the error increases slowly. However, the order of magnitude of the error remains small.

We find that the models trained with one day of data for all VAVs during the day can still achieve predictions over as long as ten weeks with an optimal $MaxAE$ less than 0.08. Tables 3.3 and 3.4 show the model parameters and validation results for fourteen VAVs located on the third floor of the SE building. When compared

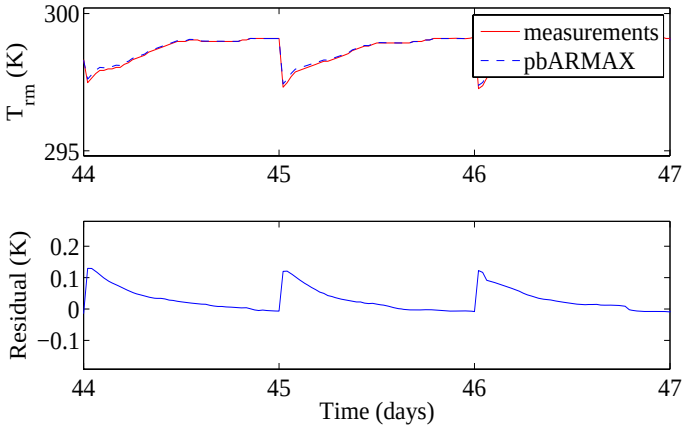

Figure 3.4: The short-term temperature prediction in room $V351$ at night with the pbARMAX model. The prediction tracks well the measured temperature despite there is poor dynamics present in the measurements at night.

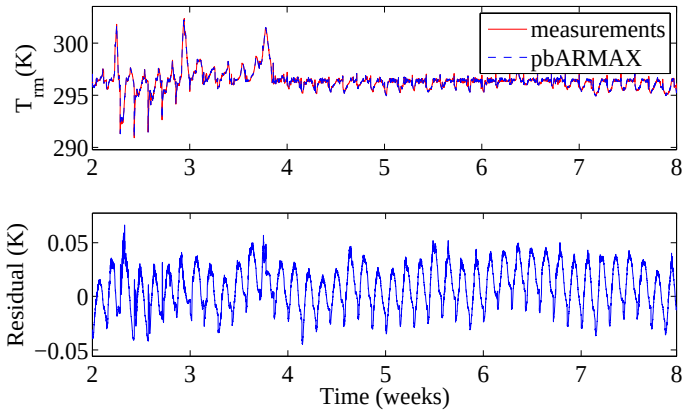

Figure 3.5: The long-term temperature prediction in room $V352$ with the pbARMAX model. The room temperature exhibits different patterns and fluctuations. Nevertheless, the prediction residual remains consistently small and does not increase significantly.

Table 3.2: The errors of the room temperature prediction over five weeks with the pbARMAX model for $V201$. The data over one day is used for training.

Prediction (weeks)	MSE	MAE	r^2	$MaxAE$
1	3.4344e-4	0.0155	0.9998	0.0444
2	4.2113e-4	0.0169	0.9998	0.0568
3	4.8388e-4	0.0184	0.9980	0.0519
4	5.8464e-4	0.0205	0.9981	0.0520
5	5.8800e-4	0.0205	0.9980	0.0501

to the prediction of the model with the optimal training data set as shown in Table 3.2, the prediction errors in terms of all validation metrics increase, but remain low: the average MSE is 0.019; the average MAE is 0.097; and the average r^2 is 0.985 during the day. The prediction errors are further reduced if optimal training data are applied.

3.2.3 Comparison with an ARMAX Model

Black-box linear parametric models such as ARX, ARMAX, BJ, OE, *etc.* have been successfully implemented to model HVAC units. In a recent work [64], ARMAX has been claimed to be very effective to model room temperature with relatively small prediction errors in the summer and fall; the same periods when our data were collected.

We select the ARMAX of order $[n_a = 1, n_b = 1, n_c = 1, n_k = 3]$, or ARMAX_{1113} for short. The parameters n_a, n_b, and n_c are defined in Appendix B. This model is the same as one of the examples studied in [64]. The structure of the model ARMAX_{1113} is

$$\begin{aligned} T_{rm}(n) = &\frac{B_1(q)}{A(q)}\dot{m}(n-n_k) + \frac{B_2(q)}{A(q)}T_{disch}(n-n_k) + \frac{B_3(q)}{A(q)}T_{clg}(n-n_k) \\ &+ \frac{B_4(q)}{A(q)}T_{oa}(n-n_k) + \frac{B_5(q)}{A(q)}T_{htg}(n-n_k) + \frac{B_6(q)}{A(q)}H_{disch}(n-n_k) \\ &+ \frac{B_7(q)}{A(q)}H_{oa}(n-n_k) + \frac{C(q)}{A(q)}e(n), \end{aligned} \tag{3.6}$$

where T_{clg}, and T_{htg} present the temperatures of cooling water, and heating water; H_{disch}, and H_{oa} denote the relative humidities of the discharge air, and outside air;

Table 3.3: Model parameters and prediction error over ten weeks for fourteen VAVs during the day. The data over one day is used for training.

VAV	P_{rm}	P_{disch}	P_{oa}	MSE	MAE	r^2	$MaxAE$
$V351$	-0.0075	0.0044	0.0031	0.0025	0.0420	0.9973	0.2018
$V352$	-0.0073	0.0052	0.0021	0.0010	0.0264	0.9992	0.1231
$V353$	-0.0085	0.0048	0.0037	0.0032	0.0482	0.9973	0.1672
$V354$	-0.0116	0.0094	0.0022	0.0218	0.1228	0.9898	0.8632
$V355$	-0.0017	0.0008	0.0007	0.0001	0.0108	0.9996	0.0313
$V356$	-0.0093	0.0078	0.0016	0.0115	0.0671	0.9767	0.4189
$V357$	-0.0090	0.0079	0.0011	0.0423	0.1784	0.9757	0.5129
$V359$	-0.0070	0.0060	0.0010	0.0079	0.0768	0.9920	0.2542
$V360$	-0.0039	0.0028	0.0011	0.0021	0.0397	0.9985	0.1207
$V361$	-0.0067	0.0055	0.0012	0.0028	0.0437	0.9948	0.1586
$V362$	-0.0085	0.0071	0.0015	0.0416	0.1797	0.9769	0.5203
$V363$	-0.0145	0.0129	0.0017	0.0612	0.2106	0.9200	0.6651
$V364$	-0.0067	0.0054	0.0014	0.0431	0.1785	0.9785	0.7097
$V365$	-0.0112	0.0088	0.0026	0.0272	0.1422	0.9897	1.6662

Table 3.4: Model parameters and validation results of ten-week prediction with one-day training for fourteen VAVs at night.

VAV	P_{rm}	P_{disch}	P_{oa}	MSE	MAE	r^2	$MaxAE$
$V351$	-0.0143	0.0083	0.0060	0.0271	0.1475	0.9726	0.3740
$V352$	-0.0115	0.0095	0.0019	0.0040	0.0429	0.9969	0.2971
$V353$	-0.0097	0.0056	0.0041	0.0061	0.0467	0.9891	0.4040
$V354$	-0.0152	0.0148	0.0003	0.0023	0.0323	0.9957	0.2352
$V355$	-0.0041	0.0022	0.0019	0.0048	0.0407	0.9886	0.2628
$V356$	-0.0143	0.0116	0.0027	0.0072	0.0589	0.9667	0.4656
$V357$	-0.0142	0.0155	-0.0013	0.0103	0.0793	0.9767	0.3527
$V359$	-0.0126	0.0129	-0.0002	0.0007	0.0213	0.9990	0.1152
$V360$	-0.0094	0.0084	0.0010	0.0013	0.0232	0.9967	0.1916
$V361$	-0.0110	0.0104	0.0005	0.0008	0.0202	0.9985	0.1392
$V362$	-0.0162	0.0164	-0.0001	0.0299	0.1471	0.9950	0.3508
$V363$	-0.0203	0.0184	0.0019	0.0032	0.0393	0.9936	0.2734
$V364$	-0.0180	0.0168	0.0012	0.0033	0.0461	0.9936	0.1711
$V365$	-0.0015	0.0032	-0.0018	0.0012	0.0255	0.9986	0.1604

Table 3.5: The short-term prediction errors by the pbARMAX model and the ARMAX_{1113} model. The pbARMAX model has much smaller errors.

Models	MSE	MAE	r^2	$MaxAE$
ARMAX (day)	6.2e-3	0.0529	0.9846	0.6922
pbARMAX (day)	6.9233e-5	0.0064	0.9998	0.0588
ARMAX (night)	2.28e-2	0.0946	0.9421	1.2653
pbARMAX (night)	3.5470e-4	0.0150	0.9995	0.0537

$A(q)$, $B_i(q)$, and $C(q)$ are polynomials in q defined in Appendix B. Within Eq. 3.6, variables not included in the pbARMAX model are dropped when we compare the pbARMAX model with the ARMAX_{1113} model.

First, we compare the short-term predictions. We use one-week data to train both ARMAX_{1113} and pbARMAX models and then apply them to predict the room temperature in the following three days. Table 3.5 takes V355 as an example, and compares the prediction results from both models. The pbARMAX model shows better validation results in all criteria selected. For instance, the orders of magnitude of MSE of the pbARMAX model are 10^{-2} smaller than those of the ARMAX_{1113} model both during the day and the night. For the rest sixty VAVs, although the ARMAX_{1113} model performs pretty well in terms of the validation metrics, the pbARMAX model performs better than the ARMAX_{1113} model. For all the VAVs involved, when compared to the ARMAX_{1113} model, the prediction of the pbARMAX model shows an average 95.3% decrease in MSE, 84.1% decrease in MAE, 91.0% decrease in $MaxAE$, and 1.6% increase in r^2 during the day, while 96.5% decrease in MSE, 85.6% decrease in MAE, 95.8% decrease in $MaxAE$, and 6.5% increase in r^2 at night. Take $V255$ as an example. Figure 3.6 shows that the pbARMAX model outperforms the ARMAX_{1113} model during the day when the room temperature fluctuates a lot, and at night when the system dynamics is poorly represented in the data and the room temperature is smooth.

Next, we study the long-term predictions. We use the data of an optimal training length to obtain the model parameters for both the pbARMAX and ARMAX_{1113} model, and use them to predict the room temperature over a duration of ten weeks. Figure 3.7 and Table 3.6 show the comparison results for $V255$. Figure 3.7 suggests that while the prediction residual of the ARMAX_{1113} model starts to increase as the prediction extends over time, the pbARMAX model is still able to maintain its residual at a much lower level. Table 3.6 indicates that compared

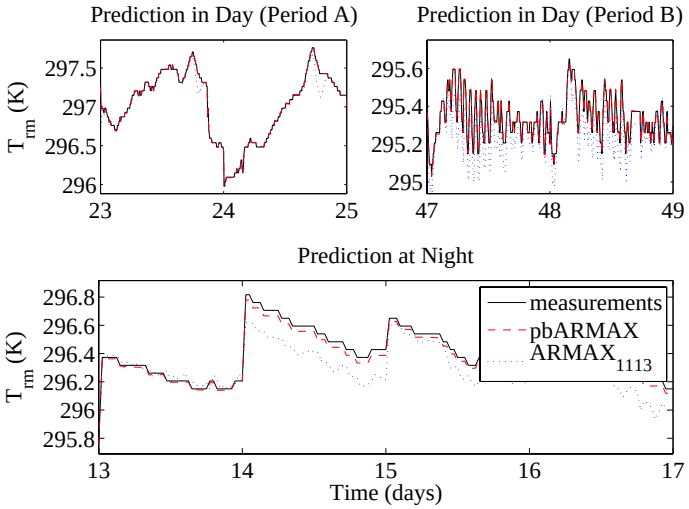

Figure 3.6: Comparison of the short-term predictions with the pbARMAX model and the ARMAX_{1133} model. The pbARMAX model consistently outperforms the ARMAX_{1133} model. during the day (upper subplots), and at night (lower subplot).

to the ARMAX_{1113} model, the pbARMAX model prediction has a 93% smaller increase in MSE over time, a 90% smaller increase in MAE, a 77% smaller increase in $MaxAE$, while a 99.6% less decrease in r^2. This observation of model accuracy is similar for the rest of sixty VAVs.

It should be noted that other sampling intervals, i.e., one minute, ten minutes, fifteen minutes, and thirty minutes, are also tested for developing the pbARMAX model. The one-minute sampling interval is achieved by linearly interpolating the original data, while the rest sampling intervals are obtained by deleting extra samples accordingly. Sampling intervals longer than thirty minutes are not tested in order to obtain enough training samples for the night partition, which is of only four hours per day. On one hand, for tested sample intervals no less than five minutes, the corresponding pbARMAX models demonstrate insignificant difference (less than 5%) in validation results of all metrics selected. The pbARMAX model developed with the five-minute sampling interval produces, on average, the best validation results for all VAVs involved. On the other hand, the pbARMAX model developed with one-minute sampling interval produces worse results than all the rest sampling intervals. This is mostly because the linear interpolation has altered the underlying system physics recorded in the original data. Considering all these facts,

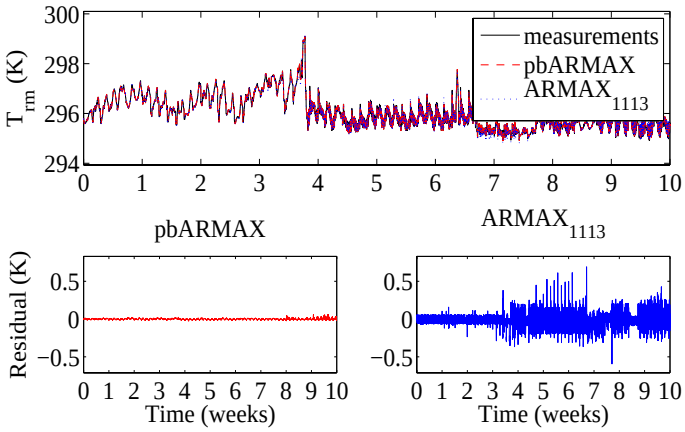

Figure 3.7: Comparison of long-term predictions with the pbARMAX model and the ARMAX_{1113} model. While the prediction residual of the ARMAX_{1113} model starts to increase as the prediction time becomes longer, the pbARMAX model still maintains its residual at a lower level.

Table 3.6: The long-term prediction errors over six weeks by the pbARMAX model (PB) and the ARMAX_{1113} model (NPB). The pbARMAX model demonstrates better performance consistently.

Weeks	MSE PB / NPB	MAE PB/ NPB	r^2 PB / NPB	$MaxAE$ PB / NPB
1	2.9898e-5/6.4025e-4	0.0046/0.0153	0.9999/0.9954	0.0138/0.1182
2	5.3093e-5/7.7538e-4	0.0058/0.0167	0.9998/0.9948	0.0213/0.2327
3	5.4892e-5/8.8781e-4	0.0059/0.0179	0.9998/0.9965	0.0181/0.2867
4	5.7200e-5/0.0075	0.0060/0.0596	0.9996/0.9859	0.0187/0.3781
5	5.9827e-5/0.0089	0.0062/0.0669	0.9995/0.9140	0.0204/0.5245
6	6.0906e-5/0.0091	0.0063/0.0719	0.9994/0.8878	0.0276/0.6149
	104% ↑/1416% ↑	37% ↑/369% ↑	0.04% ↓/10% ↓	100% ↑/420% ↑

we choose the five-minute sampling interval for developing the pbARMAX model for the SE building.

3.2.4 On Spatial and Temporal Partition

The potential applications of the pbARMAX model include building operation simulation and prediction, intelligent control design, HVAC fault detection and diagnosis, etc. Since the model includes architectural factors, it can be used for building design and optimization as well. Here, we discuss the implications of the present model to the proposed spatial and temporal partition strategy [104], which will be discussed in Chapter 5. The spatial and temporal partition strategy proposes to group VAVs with similar energy performance patterns to achieve tighter thresholds for fault detection. We can use the pbARMAX model to quantitatively assess the partition strategy.

Recall the solar flux term ϕ that is included in ΔE in Eq. 2.13. By neglecting the effect of clouds, and heat loss of the window, the heat gain ϕ can be approximately expressed as

$$\phi = \psi \sin\theta_s \zeta S_{wa}, \tag{3.7}$$

where ψ is the solar heat flux density, θ_s is the local solar elevation angle, and ζ is the window-to-wall ratio. It should note that Eq. 3.7 is a highly simplified expression for the perpendicular heat flux projection onto a window. It is proven later to be adequate for providing an analytical foundation for the spatial and temporal partition strategy.

To investigate the relationship between $\psi \sin\theta_s$ and T_{oa}, we first collect solar heat flux measurements from the California Irrigation Management Information System [131] during the same period when the building data are collected. The station providing the data is located only 4.7 miles away from the SE building. Then, by neglecting the effect of atmospheric refraction, we approximate $\sin\theta_s$ based on the equation [132]

$$\sin\theta_s = \cos\theta_h \cos\delta \cos\Phi + \sin\delta \sin\Phi, \tag{3.8}$$

where θ_h is the hour angle in the local solar time, δ is the sun declination, and Φ is the local latitude. With the measurements of T_{oa} collected from the campus, we evaluate the correlation between $\psi \sin\theta_s$ and T_{oa}. The results shows that $\psi \sin\theta_s$ is linearly correlated with T_{oa} during the day with a correlation coefficient around 0.8. We therefore assume

$$\psi \sin\theta_s = aT_{oa} + b, \tag{3.9}$$

where a and b are in general functions of time. Substituting Eq. 3.9 into Eq. 2.13, we obtain a modified pbARMAX model as

$$\begin{aligned} T_{rm}(n+1) = T_{rm}(n-1) + 2\Delta t(P_{rm}T_{rm}(n) + P_{disch}T_{disch}(n) + (P_{oa} + P_5\zeta S_{wa}a)T_{oa}(n) \\ + (\Delta E + P_5 b\zeta S_{wa})), \end{aligned} \tag{3.10}$$

where P_5 is defined in Appendix. Eq. 3.10 indicates that the solar flux influences the coefficient of $T_{oa}(n)$ in the pbARMAX model. For a room without direct sun exposure, the coefficient of $T_{oa}(n)$ is smaller. The measurements support this observation.

Consider the rooms ($V356$-$V359$) next to the hallway without direct sun exposure in the dean's suite shown in Figure 3.1. The coefficient P_{oa} is on average 65% smaller than that for the rooms ($V351$-$V354$) directly exposed to the sun in the same suite. This difference in the coefficient P_{oa}, due to the architectural feature of the room, provides a basis for spatial partition of VAVs in the building. Tables 3.3 and 3.4 show the model parameters of forty-seven VAVs during the day and at night. It is observed that the average of P_{oa} at night drops 62% from its value during the day. This variation of the coefficient P_{oa} is of course due to the change of sun exposure over time. The temporal partition proposed is based on this fact.

In summary, with the present modeling approach, the explicit functional relationship of the system parameters in Eq. 3.10 to various architectural properties and environmental factors can help, to some extend, create spatial and temporal partitions of VAVs. Hence, the present modeling method provides an analytical foundation for the spatial and temporal partition strategy.

3.3 Multi-Stage Regression Temperature Model

3.3.1 Data Preprocessing

We use measurements collected from the same two AHUs and during the same period of time as indicated in Section 3.2. We add the supply air flow rate $\dot{m}$ to the original data set of T_{rm}, T_{disch}, and T_{oa}. We also exclude three VAVs with intermittent sensor measurements, i.e., fifty-eight VAVs are involved in data collection. This results in $31,392$ samples with 232 attributes.

The samples are divided into a training set and a validation set. The identified regression model is verified by the validation set of the VAVs not included in the training set. The training set includes all VAVs controlled by AHU 9 ($A9$) and VAVs controlled by $A10$ on the third floor. The validation set includes VAVs controlled by $A10$ on the second floor. This yields 44 VAVs for training, and 14 VAVs for validation with 31 and 12 different combinations of the parameters $1/V$, S_{wa}/V and S_{wd}/V in each set.

3.3.2 Model Identification

3.3.2.1 First-Stage Regression Model

We use $RMSE$, MAE, r^2, and $MaxAE$ to quantify the modeling and prediction accuracy of the multi-stage regression models [120]. These metrics are defined in the previous section.

An optimal length of time duration exists for a given training set. We determine the optimal number of training samples for each VAV by changing the training

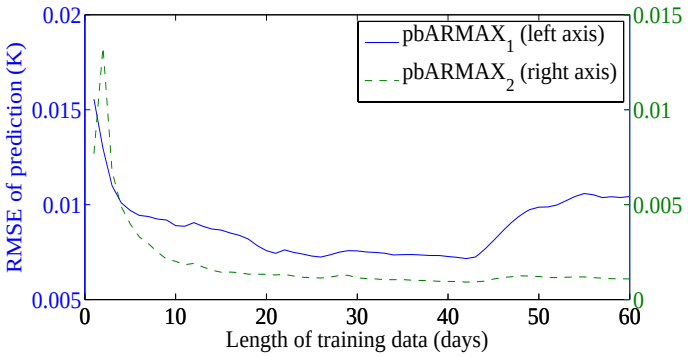

Figure 3.8: The root mean square error ($RMSE$) of prediction vs. the length of the training data. For $V201$, and most of the other fifty-seven VAVs, the optimal training length for the proposed pbARMAX models during the summer is between forty to fifty days.

length in the increment of days to find the smallest prediction error over the longest predictive period. Figure 3.8 shows an example of how the optimal training length during the day for a VAV ($V201$) is determined. With these optimal training lengths, we determine the regression models in Eqs. 2.15 and 2.16 with the method of least squares, resulting in a set of model coefficients a, b, c and d. These model coefficients are listed in Tables 3.7 and 3.8.

Figure 3.9 shows an example of the modeling of the room temperature over two days by the pbARMAX$_1$ model for V351. The modeling error is negligible with $RMSE$ and r^2 being 0.0057 and 0.9998 during the day, and 0.0378 and 0.9983 during the night. We have found that both the pbARMAX$_1$ and pbARMAX$_2$ models have a similar modeling accuracy for all the VAVs.

3.3.2.2 Second-Stage Regression Model

The first stage of regression produces 31 sets of model coefficients a, b, c and d. The second stage of regression finds the linear expressions in Eqs. 2.17 that relate the coefficients a, b, c and d to the architectural parameters $1/V$, S_{wa}/V and S_{wd}/V. The second stage continues to use VAVs in the training set. The method of least squares is used to determine the regression coefficients p_{ij} which are listed in Table 3.9. Table 3.10 lists the goodness-of-fit of the second stage regression. The average $RMSE$ of all four model coefficients is as small as 1.0291e-4.

Table 3.7: The coefficients of the pbARMAX$_1$ model and prediction errors over four weeks during the day for 14 VAVs in the training set.

VAV	a	b	$RMSE$	MAE	r^2	$MaxAE$
$V351$	0.0040	−2.6179e-5	0.1442	0.1010	0.9797	1.7669
$V352$	4.6931e-4	−3.6583e-5	0.1238	0.0898	0.9837	0.8897
$V353$	9.0462e-5	−1.7770e-5	0.1377	0.0859	0.9811	1.4507
$V354$	3.8409e-5	−1.0488e-5	0.1014	0.0678	0.9910	1.0675
$V355$	−7.8922e-7	−1.7027e-7	0.0599	0.0358	0.9877	0.5005
$V356$	−1.8222e-5	1.1515e-5	0.1529	0.0881	0.9552	1.4381
$V357$	−2.5264e-4	6.9368e-5	0.1391	0.1090	0.9683	1.1035
$V359$	−6.3169e-5	3.3502e-5	0.1040	0.0735	0.9793	1.2149
$V360$	−1.3155e-5	6.6516e-6	0.0812	0.0532	0.9904	0.6809
$V361$	−1.5193e-5	9.5484e-6	0.0725	0.0398	0.9830	1.3269
$V362$	−3.6004e-5	1.1201e-5	0.0888	0.0586	0.9939	0.8215
$V363$	−5.2251e-5	1.1045e-5	0.1593	0.0865	0.9077	2.5716
$V364$	−1.8862e-4	3.1082e-5	0.1285	0.0934	0.9816	1.2174
$V365$	−9.1219e-5	8.8934e-6	0.1725	0.1346	0.9747	1.7140

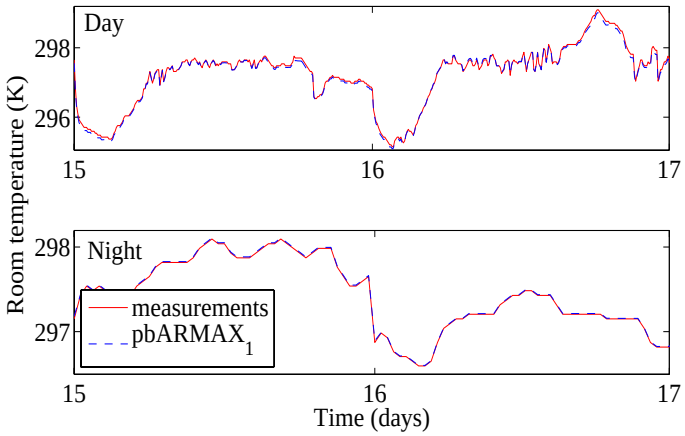

Figure 3.9: The pbARMAX$_1$ model has small modeling errors for one example VAV during both the day and night. The $RMSE$ and r^2 are 0.0057 and 0.9998 during the day, and 0.0378 and 0.9983 during the night.

Table 3.8: The coefficients of the pbARMAX$_2$ model and prediction errors over four weeks during the day for 14 VAVs in the training set.

VAV	c	d	$RMSE$	MAE	r^2	$MaxAE$
$V351$	-2.2778e-4	1.3034e-4	0.1432	0.1004	0.9785	2.0819
$V352$	-0.0023	0.0019	0.1088	0.0706	0.9867	1.0351
$V353$	-0.0011	8.7864e-4	0.1355	0.0866	0.9822	1.4643
$V354$	-7.1563e-4	5.7748e-4	0.1005	0.0670	0.9910	1.0565
$V355$	0.0038	-0.0033	0.0874	0.0643	0.9781	0.5594
$V356$	-0.0033	0.0028	0.1092	0.0696	0.9770	1.4521
$V357$	-6.4945e-4	5.3243e-4	0.0758	0.0478	0.9843	0.9054
$V359$	-7.4673e-4	6.1395e-4	0.0801	0.0512	0.9837	1.1594
$V360$	-0.0020	0.0016	0.0781	0.0524	0.9909	0.6482
$V361$	-0.0011	8.9718e-4	0.0677	0.0384	0.9844	1.3188
$V362$	8.7483e-5	-8.4688e-5	0.0749	0.0443	0.9945	0.7711
$V363$	-4.3001e-4	3.4604e-4	0.1486	0.0791	0.9145	2.3908
$V364$	-0.0019	0.0016	0.0886	0.0547	0.9865	0.9269
$V365$	-4.0561e-5	1.1577e-5	0.1565	0.1214	0.9774	1.5981

Table 3.9: The second stage regression coefficients identified from VAVs in the training set.

$p_{i,j}$	$j=1$	$j=2$	$j=3$
$i=1$	0.0283	0.0005	0.0008
$i=2$	-0.1200	0.0003	0.0008
$i=3$	1.7832	-0.0329	-0.0097
$i=4$	-1.4488	-0.0269	-0.0080

Table 3.10: Goodness-of-fit of the second stage regression of the model coefficients.

	$\widehat{a}$	$\widehat{b}$	$\widehat{c}$	$\widehat{d}$
$RMSE$	3.6497e-5	5.6382e-6	2.0870e-4	1.7075e-4
$MaxAE$	1.0639e-4	8.9216e-6	3.6774e-4	3.0310e-4

Table 3.11: The four-week prediction errors generated by the mpbARMAX$_1$ model for the example VAVs in the training set.

	$RMSE$	MAE	r^2	$MaxAE$
$V351$	0.3541	0.3204	0.9752	2.0186
$V352$	0.1273	0.0890	0.9894	1.2364
$V353$	0.2591	0.2330	0.9888	1.5983
$V354$	0.4786	0.4695	0.9916	1.3024

For the VAVs in the training set, we conduct four-week predictions by using the mpbARMAX models in Eqs. 2.18 and 2.19. The coefficients p_{ij} and the architectural parameters in Table 3.1 are used in the models. Take the mpbARMAX$_1$ model as an example, Table 3.11 shows the goodness-of-fit of such predictions for four VAVs in the dean's suite on the third floor, which have distinct combinations of the architectural parameters $1/V$, S_{wa}/V and S_{wd}/V. We can also compare the predictions of the two-stage and one-stage regression models by examining Table 3.11 and Table 3.7 (Table 3.7 and 3.8 also include the goodness-of-fits of long-term predictions of pbARMAX models which will be explained in Section 3.3.3. It is clear that the mpbARMAX$_1$ model is slightly less accurate than the one-stage model. The mpbARMAX$_2$ model demonstrates the similar performance in accurate long-term predictions.

3.3.3 Model Validation

We first validate the pbARMAX$_{1,2}$ models, which merely serves to assure that the appropriate models are chosen for the room temperature so that the mpbARMAX$_{1,2}$ models are based on a solid foundation.

We use the pbARMAX$_{1,2}$ models to conduct four-week long-term predictions with the data from the training set. Take fourteen VAVs in the short wing on the third floor as examples. Tables 3.7 and 3.8 in the previous Section 3.3.2.1 present the goodness-of-fits for the pbARMAX$_{1,2}$ models. The average $RMSE$, MAE, r^2, and $MaxAE$ of the predictions are 0.12, 0.07, 0.97 and 1.26 for the pbARMAX$_1$ model, while 0.10, 0.06, 0.98 and 1.24 for the pbARMAX$_2$ model. Compared to the modeling errors presented in Section 3.3.2.1, the prediction demonstrates an increase in $RMSE$, and a slight decrease in r^2. Nevertheless, the overall predictions of the pbARMAX$_{1,2}$ models are accurate.

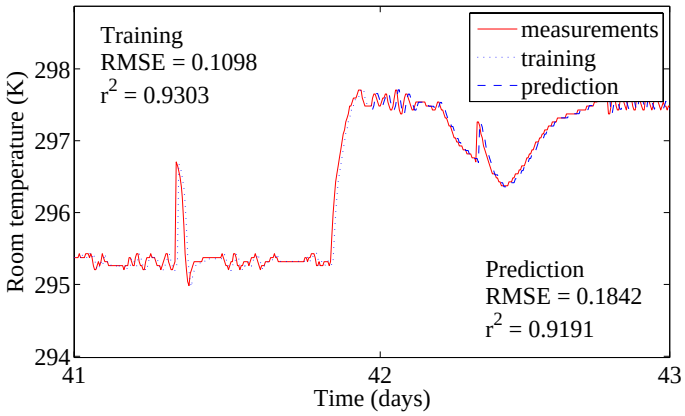

Figure 3.10: The short-term prediction of the two-stage regression mpbARMAX$_2$ model over the validation set of VAVs demonstrates goodness-of-fits nearly as good as those over the training set.

We can now validate the mpbARMAX models with the validation set. We form the models by using the coefficients $\widehat{a}$, $\widehat{b}$, $\widehat{c}$, and $\widehat{d}$ from Eqs. 2.17 with only p_{ij} and architectural parameters, and use them to conduct both short-term and long-term predictions.

As an example, Figure 3.10 shows the one-day prediction of V255 with the mpbARMAX$_2$ model. The prediction not only matches well the measured temperature $\widetilde{T}_{rm}$ when it fluctuates within the thermostat setpoints, but also tracks nicely when the VAV brings the room temperature back within the setpoints.

Goodness-of-fits of the four-week long-term predictions are presented in Table 3.12. For the four VAVs in the dean's suite on the second floor, the average $RMSE$ and r^2 of the predictions are 0.2572 and 0.9568, respectively. Compared to the single-stage regression models, the validation errors of the two-stage regression models demonstrate no significant difference. For all fourteen VAVs in the validation set, the average $RMSE$ and r^2 of the predictions of the one-stage regression models are 15% more and 3% less than those of the two-stage regression models. These results show that the proposed models can indeed capture the underlying thermodynamics of the system.

3.3.4 Comparison with an ARMAX Model

According to Schunn and Wallach [120], one cannot apply absolute standards in assessing the quality of a particular goodness-of-fit value. Each goodness-of-fit

Table 3.12: The four-week prediction errors generated by the mpbARMAX$_2$ model for the example VAVs in the validation set.

	$RMSE$	MAE	r^2	$MaxAE$
$V251$	0.3431	0.3009	0.9281	1.4529
$V252$	0.1462	0.1035	0.9800	1.1708
$V253$	0.2557	0.1986	0.9264	2.1889
$V254$	0.2839	0.2755	0.9930	0.9272

should be compared to those obtained by previous models in the same domain. To this end, we shall compare the mpbARMAX$_{1,2}$ models with existing models of the room temperature. Many linear parametric models such as ARX, ARMAX, BJ, OE, *etc.* have been successfully implemented to model the room temperature in HVAC units. In a recent work [64], an ARMAX model has been claimed to be very effective to predict the room temperature with relatively small errors in the summer and fall, the same periods when our data were collected.

Following the modeling approach in [64], we test different combinations of n_a, n_b, n_c and n_k, in which n_a, n_b and n_c range from 1 to 10 and n_k ranges from 1 to 6. And we find that the ARMAX of order $[n_a = 1, n_b = 1, n_c = 1, n_k = 3]$, or ARMAX$_{1113}$ for short, produces the best validation results on the SE building. This model is the same as one of the examples studied in [64]. The structure of the model ARMAX$_{1113}$ is

$$\begin{aligned} T_{rm}(n) = &\frac{B_1(q)}{A(q)}\dot{m}(n-n_k) + \frac{B_2(q)}{A(q)}T_{disch}(n-n_k) + \frac{B_3(q)}{A(q)}T_{clg}(n-n_k) \\ &+ \frac{B_4(q)}{A(q)}T_{oa}(n-n_k) + \frac{B_5(q)}{A(q)}T_{htg}(n-n_k) + \frac{B_6(q)}{A(q)}RH_{disch}(n-n_k) \\ &+ \frac{B_7(q)}{A(q)}RH_{oa}(n-n_k) + \frac{C(q)}{A(q)}e(n), \end{aligned}$$

where T_{clg} and T_{htg} present the temperatures of cooling water, and heating water; RH_{disch} and RH_{oa} denote the relative humidities of the discharge air and outside air; $A(q)$, $B_i(q)$ and $C(q)$ are polynomials in q defined in Appendix B. The variables in Eq. 3.6 that don't appear in the mpbARMAX models are dropped in the comparison.

The VAVs in the validation set are used for the comparison. The mpbARMAX$_{1,2}$ models are formed as discussed previously. The ARMAX$_{1113}$ model uses the historical data to train the model and obtains the optimal parameters. We apply the optimal training lengths for both approaches.

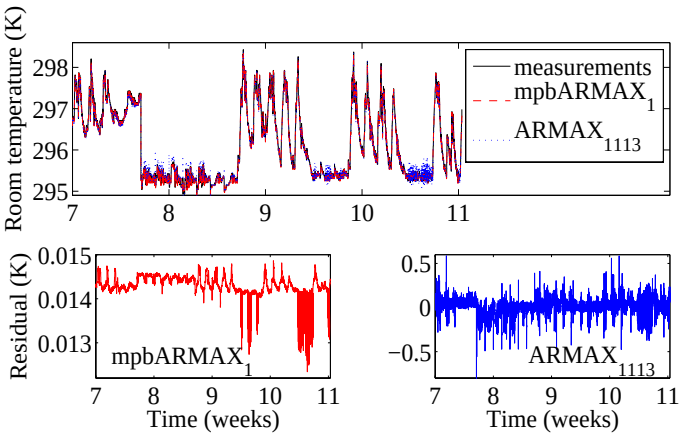

Figure 3.11: The mpbARMAX$_1$ model in the four-week long-term predictions over an example VAV outperforms the ARMAX$_{1113}$ model in terms of all four validation metrics.

We consider the prediction of the room temperature over a period of four weeks. Take V259 as an example. Figure 3.11 compares the predictions from both models. The ARMAX$_{1113}$ model performs well by providing a small $RMSE$ as 0.0992, and a large r^2 as 0.9712. The mpbARMAX$_1$ model demonstrates an even better prediction with a 85.5% decrease in $RMSE$ to 0.0143, and a 1.7% increase in r^2 to 0.9878. Furthermore, the mpbARMAX$_1$ model behaves better especially when the temperature is oscillating around the cooling set point. The observed MAE of the mpbARMAX$_1$ model is 79.9% smaller than that of the ARMAX$_{1113}$ model. For other thirteen VAVs in the validation set as well as the mpbARMAX$_2$ model, we have observed the similar results.

Table 3.13 compares the average goodness-of-fit of both the mpbARMAX$_{1,2}$ model and the ARMAX$_{1113}$ model for all VAVs in the validation set. Compared to the ARMAX$_{1113}$ model, the mpbARMAX$_{1,2}$ models have a 19.5% smaller increase in $RMSE$ over the four weeks, a 29.6% smaller increase in MAE, a 48.0% smaller increase in $MaxAE$, while a 90.9% less decrease in r^2. This illustrates that the mpbARMAX$_{1,2}$ models produce more accurate long-term predictions than the ARMAX$_{1113}$ model.

Table 3.13: The long-term prediction errors over four weeks by the mpbARMAX models (MR) and the ARMAX_{1113} model (AR). The mpbARMAX models demonstrate better performances consistently.

Weeks	$RMSE$	MAE	r^2	$MaxAE$
	MR / AR	MR / AR	MR / AR	MR / AR
1	0.0420/0.0672	0.0382/0.0351	0.9995/0.9876	0.2414/0.6020
2	0.0541/0.1373	0.0384/0.0612	0.9980/0.9807	0.3335/1.1166
3	0.0591/0.1498	0.0431/0.0700	0.9974/0.9724	0.3984/1.1812
4	0.0908/0.1643	0.0806/0.0895	0.9959/0.9587	0.4853/1.7723
	116% ↑/144% ↑	110% ↑/154% ↑	0.36% ↓/3.92% ↓	101% ↑/194% ↑

3.4 Parametric PMV Model

3.4.1 Data Preprocessing

Based on the measurements collected in Section 3.3, we further include twelve measurements of mixture air temperature, return air temperature, supply air temperature, return air damper position, outside air damper position, exhaust air damper position, mixture air flow rate, return air flow rate, return air relative humidity, outside air relative humidity, outside air temperature, preliminary filter partial pressure drop of moisture, and final filter partial pressure drop of moisture of two AHUs at the upper level over the same 109 days. This results in $31,392$ samples with 928 attributes. The measurements collected at the AHU level are used to estimate the supply air relative humidity R_{sa}.

The local building energy management system (BEMS) applies different thermostat setpoints during the day from 5am to 1am of the next day, and the night from 1am to 5am. Most of the occupants work during the day. There is no significant difference in occupancy between weekdays and weekends. We divide the data by day and night. This is consistent with the idea of temporal partition proposed in [133].

The samples are divided into a training set and a validation set in the same way as in Section 3.3.

3.4.2 Model Identification

3.4.2.1 First-Stage Regression Model

An optimal length of time duration exists for a given training set. We determine the optimal number of training samples for each VAV by changing the training length in the increment of days to find the smallest prediction error over the longest

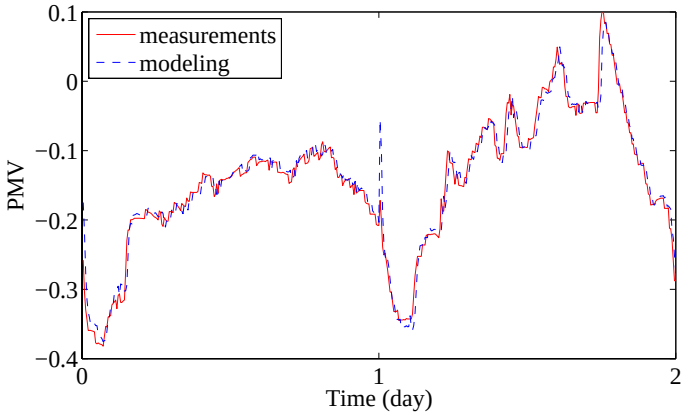

Figure 3.12: Comparison of the measured PMV of $V364$ during the day with the modeled PMV using the epPMV model obtained after the first-step regression. The overall goodness-of-fit of the modeling is excellent in terms of both root mean square error and coefficient of determination.

predictive period. With these optimal training lengths, we determine the regression models in Eq 2.33 with the method of least squares, resulting in a set of the coefficients x_i.

Take $V364$ as an example. Figures 3.12 and 3.13 show the modeling results over two days by the epPMV model after the first-stage regression. The modeling errors are negligible with $RMSE$ and r^2 being 0.0234 and 0.9883 during the day, and 0.0066 and 0.9967 during the night. We have found that the epPMV model after the first-stage regression has a similar modeling accuracy for all the VAVs.

3.4.2.2 Second-Stage Regression Model

The first stage of regression produces 31 sets of the coefficients x_i. The second stage of regression finds the linear expressions in Eqs. 2.34 that relate x_i to the architectural parameters $1/V$, S_{wa}/V and S_{wd}/V. The second stage continues to use VAVs in the training set. The method of least squares is used to determine the regression coefficients p_{ij} which are listed in Table 3.14. The average $RMSE$ of all nine model coefficients x_i and $\hat{x}_i$ is as small as 0.0881.

Figure 3.14 shows the model fitting of the PMV data for $V352$ over four days. The $RMSE$ and r^2 are 0.0215 and 0.9786 respectively. Other VAVs in the training set demonstrate similar modeling accuracy.

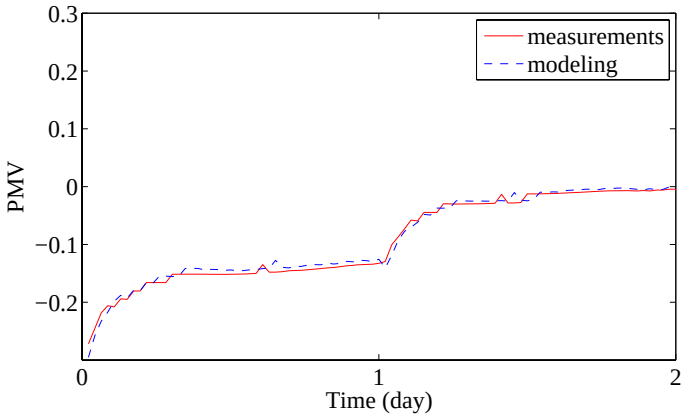

Figure 3.13: Comparison of the measured PMV of $V364$ during the night with the modeled PMV using the epPMV model obtained after the first-step regression. The overall goodness-of-fit of the modeling is as excellent as during the day.

Table 3.14: The second stage regression coefficients identified from VAVs in the training set.

p_{ij}	$j=1$	$j=2$	$j=3$
$i=1$	6.4778e+5	0.1590e+5	0.1889e+5
$i=2$	0.1661	0.0020	0.0111
$i=3$	7.6342	0.0519	−0.0305
$i=4$	−0.0055	0.0001	0.0001
$i=5$	−3.1559e+5	−0.0980e+5	0.1024e+5
$i=6$	−4.8772e+5	−0.1250e+5	−0.1473e+5
$i=7$	−2.9799e+8	−0.0592e+8	−0.0013e+8
$i=8$	1.1356e+6	0.0251e+6	0.0132e+6
$i=9$	−1.1554e+9	0.0039e+9	−0.0017e+9

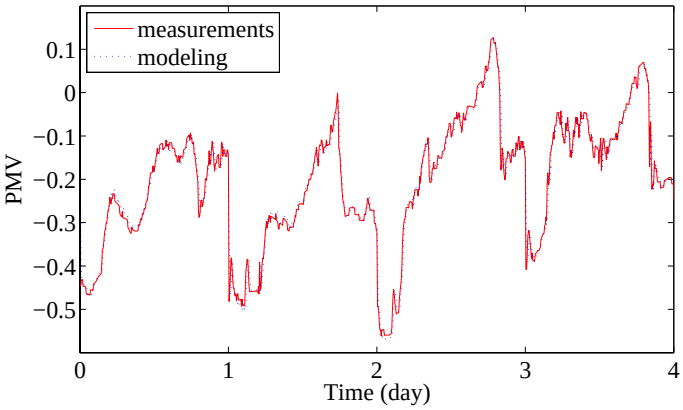

Figure 3.14: Comparison of the measured PMV of $V352$ during the day with the modeled PMV using the epPMV model obtained after the second-step regression. The overall goodness-of-fit of the modeling is excellent with RMSE as 0.0215 and r^2 as 0.9786.

3.4.3 Model Validation

3.4.3.1 First Stage

We first validate the epPMV model obtained from the first-stage regression with the validation data set. This merely serves to assure that the appropriate models are chosen for the indoor PMV so that the final epPMV model is based on a solid foundation.

Take $V364$ as an example. Figure 3.15 shows three-day prediction generated by the model obtained in the first-step regression. Compared to the model fitting with the training set, the prediction has a minor increase in $RMSE$ and decrease in r^2.

For long-term predictions over four weeks, Table 3.15 lists goodness-of-fits of fourteen VAVs in the training set. The average of $RMSE$ is as small as 0.0241, and the average of r^2 is as large as 0.9728. The goodness-of-fit of prediction (denoted as Pr in the table) is almost as good as the training result (denoted as Tr in the table).

From the practical point of view, the daily average value of PMV provides operators a system-level and also more useful view of the thermal comfort in the building. The proposed model has even better goodness-of-fit of the prediction of daily averages of PMV. On average, the order of magnitude of $RMSE$ is reduced by a factor of 10^2, while r^2 is increased by 5%. Figure 3.16 shows an example of three-week daily average PMV prediction of $V351$.

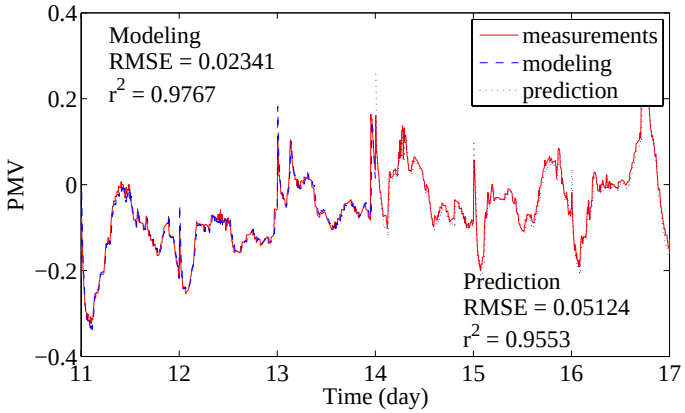

Figure 3.15: Comparison of the measured PMV of $V364$ during the day with the predicted PMV using the epPMV model obtained after the first-step regression. The overall goodness-of-fit of the prediction is as excellent as the modeling in terms of both root mean square error and coefficient of determination.

Table 3.15: The $RMSE$ and r^2 of the epPMV model for fourteen VAVs in the training set after the first-stage regression.

Tr/Pr	$RMSE$	r^2
$V351$	0.0353/0.0369	0.9747/0.9743
$V352$	0.0263/0.0468	0.9792/0.9896
$V353$	0.0265/0.0296	0.9914/0.9906
$V354$	0.0198/0.0202	0.9827/0.9874
$V355$	0.0124/0.0161	0.9841/0.9888
$V356$	0.0234/0.0299	0.9720/0.9776
$V357$	0.0141/0.0189	0.9782/0.9744
$V359$	0.0124/0.0148	0.9910/0.9756
$V360$	0.0145/0.0168	0.9937/0.9920
$V361$	0.0180/0.0140	0.9882/0.9647
$V362$	0.0148/0.0148	0.9933/0.9922
$V363$	0.0239/0.0252	0.9461/0.8710
$V364$	0.0176/0.0196	0.9767/0.9553
$V365$	0.0276/0.0340	0.9669/0.9860
$Mean$	**0.0205/0.0241**	**0.9799/0.9728**

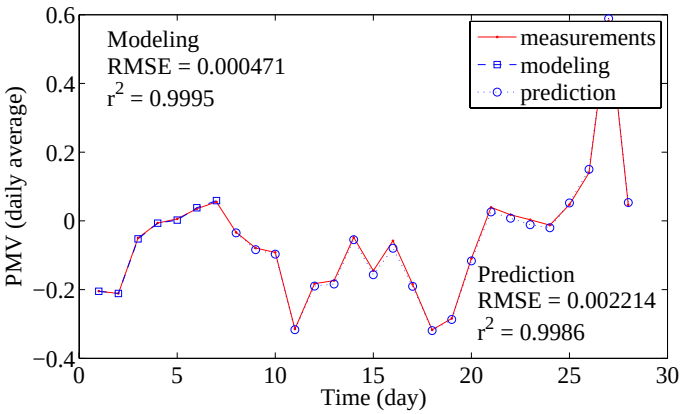

Figure 3.16: Comparison of the measured daily averages of PMV of $V351$ with the modeled and predicted daily averages PMV using the epPMV model. The goodness-of-fit of the prediction is further improved in terms of both root mean square error and coefficient of determination.

3.4.3.2 Second Stage

We can now validate the epPMV model obtained from the two-state regression procedure. The epPMV model contains the coefficients p_{ij}, the architectural parameters and control variables. We use the model to conduct both short-term and long-term predictions of the PMV index with the validation set.

Figure 3.17 shows a short-term prediction of PMV over three days for $V252$. The goodness-of-fit of the prediction is as good as those of VAVs in the training set. The $RMSE$ is 0.0251, and the r^2 is 0.9765. The PMV predictions of other thirteen VAVs in the validation set have similar accuracies, and will not be presented here for the sake of space.

Figure 3.18 shows long-term predictions of PMV for $V254$ over four weeks. The prediction accuracy is good with a $RMSE$ of 0.0111, and a r^2 of 0.9919. Furthermore, the prediction tracks fast fluctuations of PMV or moderate variations in a small range equally well. This indicates that the epPMV model is able to capture the rich dynamics of PMV in the field. Table 3.16 lists the goodness-of-fits of four VAVs in the validation set. The average $RMSE$, MAE, r^2, $MaxAE$ for the long-term predictions are 0.0438, 0.0290, 0.9614, and 0.3550.

Similar to the first-step regression, the prediction of daily average PMV demonstrates improved accuracy. Figure 3.19 shows the prediction of PMV for $V251$ over four weeks.

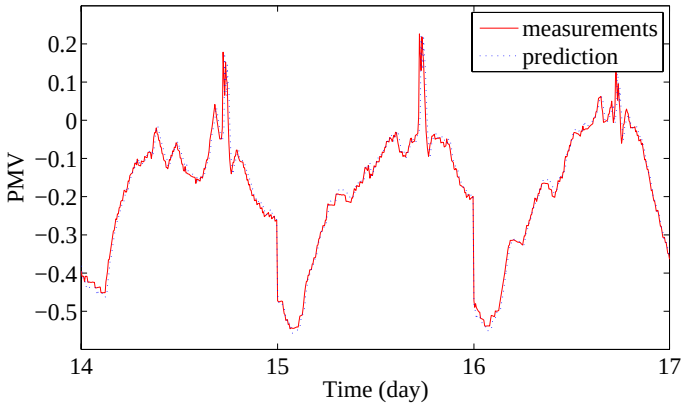

Figure 3.17: Comparison of the measured PMV of $V252$ in the validation set with the predicted PMV using the epPMV model. The goodness-of-fit is almost as excellent as VAVs in the training set.

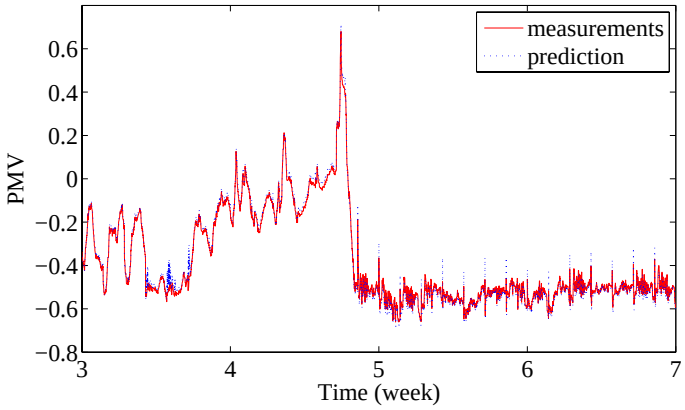

Figure 3.18: Comparison of the measured PMV of $V254$ in the validation set with the predicted PMV using the epPMV model. This four-week prediction is able to track nicely the rich dynamics of the measured PMV.

Table 3.16: The metrics of the goodness-of-fit of four-week predictions of four VAVs in the validation set.

Pr	$RMSE$	MAE	r^2	$MaxAE$
$V251$	0.0640	0.0385	0.9385	0.3996
$V252$	0.0174	0.0196	0.9839	0.2778
$V253$	0.0826	0.0404	0.9315	0.5615
$V254$	0.0111	0.0173	0.9919	0.1812
$Mean$	**0.0438**	**0.0290**	**0.9614**	**0.3550**

As for different finite difference schemes as discussed in Section 2.4.2, the corresponding prediction results demonstrate insignificant differences. Take $V251$ to $V254$ as examples, the average $RMSE$ and r^2 of long-term predictions when forward difference is adopted are 1.2% larger and 0.8% smaller than those when central difference is adopted. Mathematically, central difference produces slightly more accurate prediction results than forward difference for our model.

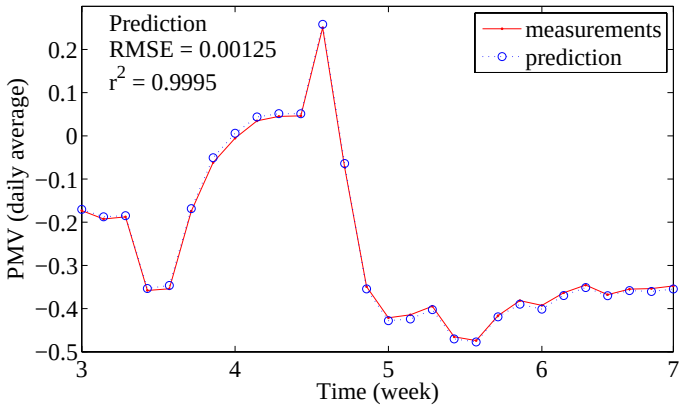

Figure 3.19: Comparison of the measured daily averages of PMV of $V251$ in the validation set with the predicted daily averages PMV using the epPMV model. The goodness-of-fit of the prediction is further improved in terms of both root mean square error and coefficient of determination.

Chapter 4

SENSITIVITY ANALYSIS

4.1 Introduction

In Chapter 2, we have studied the empirical predicted mean vote (epPMV) regression model. This chapter presents a sensitivity analysis of the epPMV model with respect to its parameters, and demonstrates a statistical confirmation of the model.

Sensitivity analysis estimates the rate of change in the output of a model with respect to changes of model parameters. It has been a useful tool in model development and evaluation in many fields [134–136]. In the field of building modeling and simulation, sensitivity analysis has also been applied widely. For example, in one of their pioneering works, Lomas and Eppel investigated and compared three sensitivity analysis techniques for building thermal simulation. They suggested that the differential sensitivity analysis is suitable for obtaining the sensitivities of predictions to individual input parameter uncertainties, and the Monte Carlo analysis is preferable when calculating the total sensitivities in the predictions [137]. Balagangadhar and Roy performed sensitivity analysis to optimize steady fluid-thermal systems, results of which are successfully applied to the heat exchanger fin and duct designs for heating, ventilation and air conditioning (HVAC) systems [138]. Eisenhower and colleagues increased the dimension of sensitivity analysis by studying the influence of about one thousand parameters. They innovated a method to decompose the pathway as uncertainty flows through the dynamics, identified which internal processes transmit the most uncertainty to the final output [139]. By implementing a Monte Carlo approach, Burhenne and colleagues analyzed how a sensitivity analysis could provide a decision support for the optimization of the building operation [140]. Corrado and Mechri performed sensitivity analysis on European Building Energy Rating and discovered that the data uncertainties have only slight impact on the asset energy rating, and the importance of associating an operational energy rating to the asset rating in order to show the influence of user behavior on building energy performance [141]. With an emphasis on quantifying modeling uncertainties, de Wit and Augenbroe demonstrated and justified respectively how uncertainty in building performance assessments can be used in, and is essential to rational design decision [142]. For Building Performance Simulation (BPS), Hopfe and Hensen analyzed uncertainties in physical, scenario, and design parameters, and

demonstrated how to effectively integrate uncertainty analysis in BPS for design information [143]. Mara and Tarantola brought the analysis of variance-based global sensitivity analysis to building thermal modeling with an emphasis on the methodology that is computationally cheap but provides more information [144]. Heiselberg and colleagues implemented one-at-a-time sensitivity analysis to the early design stage of an office building in Denmark. They found that lighting control and the amount of ventilation during winter are the two most important parameters that have the largest effect on the energy use, and concluded that sensitivity analysis can reveal the important parameters during the early stage of design [145]. All of these works demonstrate that sensitivity analysis is essential for evaluating the confidence, examining the robustness, revealing insights, thus assuring the overall quality of statistical modeling [134, 146, 147]. Consequently, it is important and also necessary for us to conduct such an analysis to evaluate the epPMV model.

As discussed in Chapters 2 and 3, the epPMV model is constructed by regression in two stages. In the first-stage regression, the model has eleven independent predictors and nine regression coefficients. In the second-stage regression, the coefficients of the first-stage model are further regressed as functions of architectural parameters. This leads to twenty-seven second-stage regression coefficients. These regression coefficients are the parameters of the epPMV model. The output PMV of the model has variabilities due to uncertainties in all the parameters that are mainly caused by measurement noise, occupancy uncertainty, and environmental changes. The epPMV model is a linear model in both stages. Statistics of the model parameters and the output can be obtained from extensive measurements of the HVAC system over a long period of time.

The objectives of sensitivity analysis in this chapter are to determine: (1) Variability of the PMV predictions when the model parameters contain random uncertainties; (2) Model parameters to which the model output is most and least sensitive; (3) The probabilistic distribution of the model output. We choose the so-called local one-at-a-time (OAT) method [148–150] to compute the parameter sensitivity of the epPMV model. Although the OAT method does not fully explore the input space [134], it is much less computationally demanding, and increases the comparability of results [151]. These advantages suit the structure of the epPMV model and the purpose of our sensitivity analysis.

The rest of this Chapter is organized as follows. In Section 4.2, we present the mathematical background of sensitivity analysis. In Section 4.3, we present the sensitivity analysis of the epPMV model with extensive data from the SE building. Finally, Section 4.4 discusses the findings and Section 4.5 presents the conclusions.

4.2 Mathematical Background

Recall the epPMV model that represents the PMV of a room as

$$PMV_{rm} = \sum_{i=1}^{9} a_i x_i + b_1 x_{10} + b_2 x_{11} \equiv Y(\mathbf{a}), \tag{4.1}$$

where x_i $(i = 1 \dots 11)$ are model predictors that are either sensor measurements or combinations of sensor measurements as listed in Eq. 2.33 and Appendix B, b_1 and b_2 are constants, and a_i $(i = 1 \dots 9)$ are nine model parameters. $\mathbf{a}$ denotes the vector consisting of a_i. Eq. 4.1 is determined with the help of the regression analysis. This is called the first-stage regression in [152]. a_i can be identified as functions of architectural parameters by means of the so-called second-stage regression,

$$a_i = \sum_{j=1}^{3} p_{ij} y_j, \ (i = 1 \dots 9), \tag{4.2}$$

where y_j $(j = 1 \dots 3)$ are grouped architectural parameters [152], and p_{ij} $(i = 1 \dots 9, j = 1 \dots 3)$ are the coefficients of the second-stage regression. In the following, we study the sensitivity of all the regression coefficients in both stages.

We consider two sensitivity functions to assess the quality of the epPMV model. One is the Bode sensitivity [153], the other is a revised Bode sensitivity with a statistical focus.

Let $\bar{Y} = Y(\bar{\mathbf{a}})$ where $\bar{\mathbf{a}}$ is $\mathbf{a}$ when all the parameters a_i take their nominal values $\bar{a}_i$. The Bode sensitivity of Y with respect to a_i for a given index i, i.e. the sensitivity of the first-stage epPMV model, is defined as

$$S_{a_i}^{\bar{Y}} = \lim_{\Delta a_i \to 0} \frac{\Delta Y_i / \bar{Y}}{\Delta a_i / \bar{a}_i} = \frac{\bar{a}_i}{\bar{Y}} \frac{\partial Y}{\partial a_i}\bigg|_{\bar{\mathbf{a}}} = \frac{\bar{a}_i}{\bar{Y}} x_i, \tag{4.3}$$

where ΔY_i is the variation of the model output when $\bar{a}_i$ is subject to a small perturbation Δa_i. The Bode sensitivity is often used in control studies [154]. The sensitivity of $\bar{a}_i$ with respect to p_{ij} can be defined in the same way as

$$S_{p_{ij}}^{\bar{a}_i} = \lim_{\Delta p_{ij} \to 0} \frac{\Delta a_i / \bar{a}_i}{\Delta p_{ij} / \bar{p}_{ij}} = \frac{\bar{p}_{ij}}{\bar{a}_i} \frac{\partial \bar{a}_i}{\partial p_{ij}} = \frac{\bar{p}_{ij}}{\bar{a}_i} y_j. \tag{4.4}$$

By the chain rule of differentiation, we can show that the Bode sensitivity of Y with respect to p_{ij}, is given by

$$S_{p_{ij}}^{\bar{Y}} = S_{a_i}^{\bar{Y}} \cdot S_{p_{ij}}^{\bar{a}_i}. \tag{4.5}$$

In this research, we call $S_{a_i}^{\bar{Y}}$ as the first-stage sensitivity, and $S_{p_{ij}}^{\bar{Y}}$ as the total-stage sensitivity of the epPMV model.

Note that the Bode sensitivity is normalized by the ratio $\bar{Y}/\bar{a}_i$, which does not reveal the statistical relationship between the model parameters and output. To include the statistical information in the sensitivity function, we define a revised Bode sensitivity [135] as

$$Z_{a_i}^{\bar{Y}} = \frac{\sigma_{a_i}}{\Sigma_Y} \left.\frac{\partial Y}{\partial a_i}\right|_{\bar{\mathbf{a}}} = \frac{\sigma_{a_i}}{\Sigma_Y} x_i, \tag{4.6}$$

where σ_Y and σ_{a_i} are the standard deviations of Y and a_i. For the second-stage regression, we have

$$Z_{p_{ij}}^{a_i} = \frac{\sigma_{p_{ij}}}{\Sigma_{a_i}} \left.\frac{\partial a_i}{\partial p_{ij}}\right|_{\bar{\mathbf{p}}} = \frac{\sigma_{p_{ij}}}{\Sigma_{a_i}} y_j. \tag{4.7}$$

Both Σ_Y and Σ_{a_i} are normalization constants to be determined. The revised Bode sensitivity does not obey the chain rule same as in Eq. 4.5. Nevertheless, we choose the values of Σ_Y and Σ_{a_i} such that it still obeys the chain rule in the form. The total-stage revised Bode sensitivity of the output Y with respect to p_{ij} is defined as

$$Z_{p_{ij}}^{\bar{Y}} = Z_{a_i}^{\bar{Y}} \cdot Z_{p_{ij}}^{a_i}. \tag{4.8}$$

Assume that a_i in the epPMV model of Eq. 4.1 are normally distributed, $\mathbf{a} \backsim N(\mathbf{a}_{av}, \mathbf{R}_{aa})$ where $\mathbf{a}_m$ is the mean of $\mathbf{a}$, and $\mathbf{R}_{aa}$ is the correlation matrix of $\mathbf{a}$. Y will also be normally distributed. The squared sum of the revised Bode sensitivity $Z_{a_i}^{\bar{Y}}$ for all i is given by,

$$\sum_i \left(Z_{a_i}^{\bar{Y}}\right)^2 = \frac{1}{\Sigma_Y^2} \sum_i x_i^2 \sigma_{a_i}^2. \tag{4.9}$$

Let

$$\Sigma_Y^2 = \sum_i x_i^2 \sigma_{a_i}^2. \tag{4.10}$$

We have

$$\sum_i \left(Z_{a_i}^{\bar{Y}}\right)^2 = 1. \tag{4.11}$$

Because the squares of all sensitivities add to one, the revised Bode sensitivity $Z_{a_i}^{\bar{Y}}$ with respect to a_i represents its fractional contribution to the output variance of the first-stage model. This property of the revised Bode sensitivity motivates us to define a measure of the fractional contribution to the output variance of the first-stage model. It is called the sensitivity weight W_{a_i} for the model parameter a_i,

$$W_{a_i} = \frac{\left(Z_{a_i}^{\bar{Y}}\right)^2}{\sum_j \left(Z_{a_j}^{\bar{Y}}\right)^2} = \left(Z_{a_i}^{\bar{Y}}\right)^2. \tag{4.12}$$

The above properties also hold for $Z_{p_{ij}}^{a_i}$ of the second-stage model when

$$\Sigma_{a_i}^2 = \sum_j y_j^2 \sigma_{p_{ij}}^2. \tag{4.13}$$

From a control perspective, both the nominal value and variance of the HVAC output are important objectives. The nominal value of the PMV should be maintained at zero so that an average occupant feels neutral in terms of the thermal comfort, while the variance of the output needs to be minimized in order to reduce the fluctuation of human thermal sensation. The Bode sensitivity evaluates the output variation of the nominal model output due to the parameter uncertainty, while the revised Bode sensitivity examines the variance variation of the model output.

Both the sensitivity functions are dependent of the measurement x_i, which can change randomly over time. For this reason, the sensitivities need to be averaged over time. Within the same season, if there is no substantial environmental change, the collected measurements of the HVAC system repeat daily with small variations. Consequently, the HVAC system behaves like a slowly varying periodic dynamic system. Hence, it is reasonable to average the sensitivities over a day or a week in a season.

4.3 Case Study

4.3.1 Data Preprocessing

We use the same measurements collected in Section 3.4. The samples are still divided into a training set and a validation set. The parameter uncertainty is identified with the training set. The parameter sensitivity is evaluated with the validation set and the identified parameter uncertainty. The training set includes all VAVs controlled by AHU 9 and VAVs controlled by AHU 10 on the second floor. The validation set includes VAVs controlled by AHU 10 on the third floor. This yields 44 VAVs for training, and 14 VAVs for validation with 31 and 12 different combinations of second-stage regression predictors y_j ($j = 1 \dots 3$) in each set.

4.3.2 Random Data

The probability distributions of the model parameters a_i and p_{ij} are needed in the sensitivity analysis. Since the model parameters are determined with the training data that contain disturbances and uncertainties, they contain uncertainties as well. In this study, we consider two different schemes to process the measurements separately in order to calculate the mean and standard deviation of the model parameters.

Scheme (a) takes the measurement of one week as a sample. It is intended to reveal the periodic patterns in the measurements, and study the random variations over a long time. Scheme (b) recognizes the slowly varying periodic behavior of the

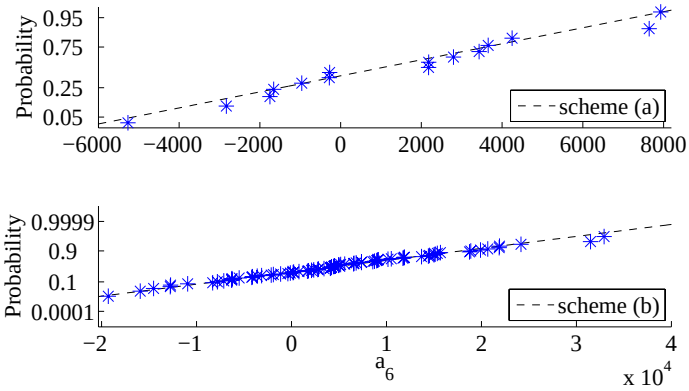

Figure 4.1: Identified first-stage parameter a_6 of $V364$ under both training schemes demonstrate strong normal distributions. While results under scheme (b) are more normally distributed than those obtained under scheme (a) due to more training samples being used.

HVAC system with period equal to one day, and takes the measurement of one day as a sample to study the statistical variations of the HVAC dynamics.

4.3.3 Statistical Analysis

The model parameters a_i $(i = 1 \ldots 9)$ in the first-stage regression are determined with the data prepared with both the schemes. Take parameter a_6 for $V364$, a faculty office, as an example. Figure 4.1 shows the Quantile-Quantile plot results by comparing the identified parameter a_6 under both the schemes to standard normal distributions. To quantitatively verify the parameter normality, we conduct the Sharpiro-Wilk test [155]. With a significance level of 0.05, none of the identified parameters can reject the null hypothesis that they are normally distributed. Table 4.1 lists the mean and standard deviation of the first-stage parameters of all VAVs.

According to the results of the Sharpiro-Wilk test, the average p-values are 0.78 and 0.81 for the model parameters identified with scheme (a) and (b) respectively. This indicates that the model parameters obtained with scheme (a) have a slightly higher probability to reject the hypothesis that they are normal than the parameters obtained with scheme (b). This may be due to the fact that scheme (b) involves more samples than scheme (a).

For each VAV, we have two sets of the first-stage regression parameters a_i with the two schemes. These parameters are taken as the output of the second-stage

Table 4.1: Statistical properties of identified first-stage parameters of VAVs in the training set.

Parameter	μ (scheme a)	σ (scheme a)	μ (scheme b)	σ (scheme b)
a_1	$-2.0495e{+}3$	$4.9504e{+}3$	$1.2968e{+}5$	$2.1817e{+}6$
a_2	0.0010	$5.5515e$-4	0.0011	$6.3658e$-4
a_3	-0.0045	0.0062	0.3126	5.2832
a_4	$2.7902e$-5	$3.3590e$-5	$6.8389e$-5	$2.0299e$-4
a_5	$-1.3271e{+}3$	$2.2193e{+}3$	$-1.3535e{+}5$	$2.1797e{+}6$
a_6	97.1731	$5.0227e{+}3$	$-3.7038e{+}4$	$5.4051e{+}5$
a_7	$-3.1059e{+}5$	$9.4207e{+}4$	$-3.0876e{+}5$	$4.0493e{+}4$
a_8	$2.9550e{+}3$	$2.8894e{+}3$	$3.7478e{+}4$	$5.4042e{+}5$
a_9	$-3.5190e{+}5$	$9.2800e{+}4$	$-3.5345e{+}5$	$3.9609e{+}4$

regression in Eq. 4.2. The second-stage regression identifies the parameters p_{ij}. We now examine their uncertainties.

Take p_{2j} $(j = 1 \ldots 3)$ as examples. Figure 4.2 shows the probability plots of p_{2j} with the two schemes respectively. With a significance level of 0.05, the Sharpiro-Wilk test results show high probabilities, with an average p-value as 0.32, that the hypothesis that these second-stage parameters are normally distributed cannot be rejected. Table 4.2 takes p_{1j}, p_{2j}, p_{3j}, and p_{4j} $(j = 1 \ldots 3)$ as examples and shows the results for all VAVs under consideration and lists their statistical properties.

Under both schemes, all second-stage parameters demonstrate normality. Neither clustered nor segmented distributions have been observed. Results obtained under scheme (b) still fit normal distribution slightly better than those under scheme (a). This is consistent with the observed statistics of the first-stage regression parameters.

4.3.4 Sensitivity of First-Stage Model

4.3.4.1 Bode Sensitivity

Consider the Bode sensitivity function and scheme (a). Figure 4.3 shows the average sensitivities for nine first-stage parameters a_i $(i = 1 \ldots 9)$ over two months for $V359$, as an example. The model output is most sensitive to a_7 and a_9, and least sensitive to a_1 and a_2. Table 4.3 lists the average sensitivities over two months in descending order for all VAVs in the validation set.

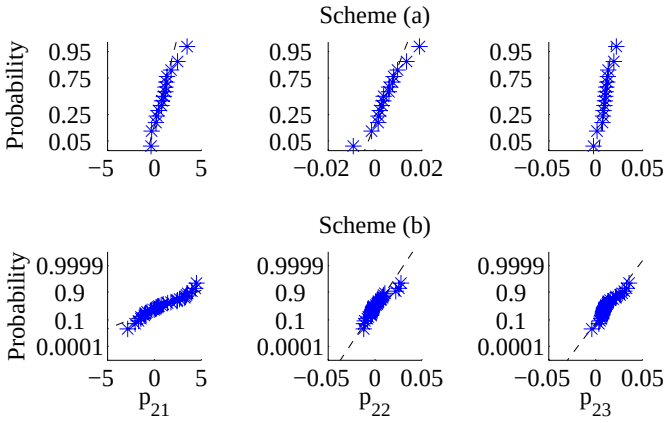

Figure 4.2: All second-stage parameters identified under two training schemes demonstrate strong normal distributions.

Table 4.2: Statistical properties of the second-stage parameters of VAVs in the training set.

Parameter	μ (scheme a)	σ (scheme a)	μ (scheme b)	σ (scheme b)
p_{11}	$-3.2979e{+}5$	$5.3903e{+}6$	$8.3531e{+}8$	$4.4930e{+}9$
p_{12}	$-9.8586e{+}4$	$1.0961e{+}5$	$-8.9591e{+}6$	$1.0143e{+}8$
p_{13}	$9.2234e{+}3$	$4.5954e{+}4$	$9.7172e{+}4$	$5.4315e{+}6$
p_{21}	1.0295	1.0121	0.9003	2.1244
p_{22}	0.0049	0.0065	0.0020	0.0101
p_{23}	0.0103	0.0060	0.0136	0.0096
p_{31}	−15.4584	36.9036	$2.0264e{+}3$	$1.0866e{+}4$
p_{32}	−0.0309	0.1993	−21.4176	245.2422
p_{33}	−0.0132	0.1121	0.0828	13.1325
p_{41}	0.3105	0.5430	0.2132	0.2048
p_{42}	−0.0011	0.0027	$-3.6927e$-4	0.0010
p_{43}	$-6.7366e$-4	0.0014	$-4.0807e$-4	$5.0296e$-4

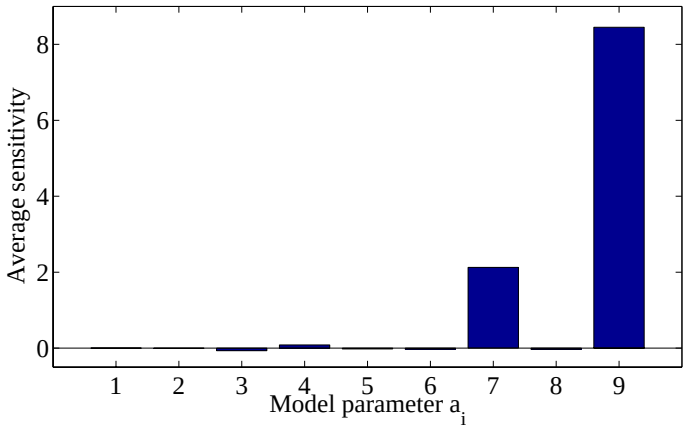

Figure 4.3: By implementing the Bode sensitivity function and training scheme (a), the identified first-stage parameter sensitivities of $V359$ indicates that the model output is most sensitive to two parameters a_7 and a_9 that relate to local control variables.

Table 4.3: Identified first-stage parameter sensitivies of VAVs in the validation set under the Bode sensitivity function and two training schemes (a/b). Three parameters a_9, a_7, and a_5 related to local control variables consistently appear at the top of the list under both training schemes.

Rank	Parameter	$S^Y_{a_i}$	W_{a_i} (%)
1	a_9 / a_9	$+16.6648$ / $+18.5364$	78.97 / 51.16
2	a_7 / a_7	$+8.5959$ / $+15.1589$	21.01 / 41.85
3	a_5 / a_5	-0.2013 / -1.0116	$1.1522e$-2 / 2.79
4	a_8 / a_3	-0.1651 / $+0.7413$	$7.7508e$-3 / 2.05
5	a_4 / a_6	0.0790 / $+0.2557$	$1.7746e$-3 / 0.71
6	a_6 / a_1	-0.0199 / $+0.2302$	$1.1261e$-4 / 0.64
7	a_2 / a_8	-0.0170 / -0.2053	$8.2177e$-5 / 0.57
8	a_3 / a_4	-0.0166 / $+0.0784$	$7.8355e$-5 / 0.22
9	a_1 / a_2	-0.0075 / -0.0107	$1.5995e$-5 / 0.01

It is interesting to point out that the model is less sensitive to parameters a_7 and a_9 for some VAVs with larger windows, e.g., $V351$ and $V352$ in the dean's suite, than other VAVs with smaller windows. Since parameters a_7 and a_9 are closely coupled to local control variables, this indicates that the dominant influence of control variables on the thermal comfort is slightly reduced for these VAVs with larger windows. This observation agrees with the spatial partition found in our previous work [133].

Consider scheme (b) now. Table 4.3 lists the sensitivities for all VAVs over the same period of time. The top three parameters that cover more than 95% total sensitivity weight are identical to those under scheme (a). However, their total sensitivity weight slightly decreases by 4.18%. Meanwhile, the sensitivities with respect to the parameters that are related only to environment conditions increase nearly eight times on average. This indicates that scheme (a) reveals the influence of local control variables better than scheme (b).

4.3.4.2 Revised Bode Sensitivity

Consider now the revised Bode sensitivity in Eq. 4.6. The revised Bode sensitivities are more consistent between the two schemes than the Bode sensitivities.

Figure 4.4 shows daily average sensitivities of the $V351$ model output over the same period of time with respect to the parameters. The magnitude and ranking of the revised Bode sensitivities are almost identical with two schemes. The average sensitivities computed with two schemes for all VAVs in the validation set are presented in Table 4.4. Eight out of nine parameters show similar sensitivity levels. Recall the Bode sensitivities. We conclude that although the first-stage model output has different sensitivities with respect to the parameters near their nominal values, the variance of the model output has a similar sensitivity to most of the parameters. In terms of the revised Bode sensitivity, the model is still most sensitive to parameters a_7 and a_9, and least sensitive to parameter a_2. This pattern is consistent with that in terms of the Bode sensitivity.

4.3.5 Total-Stage Sensitivity

4.3.5.1 Bode Sensitivity

We consider Bode sensitivity and scheme (a) and evaluate total-stage sensitivities in Eq. 4.5 for the model involving all the VAVs in the validation set. Figure 4.5 shows the results for $V359$ as an example.

Similar to the sensitivities of the first-stage model, the second-stage model is found to be most sensitive to some parameters, e.g. p_{9j} and p_{7j} $(j = 1 \dots 3)$, and least sensitive to others, e.g. p_{2j} and p_{1j} $(j = 1 \dots 3)$. According to Eq. 4.2, p_{9j} and p_{7j} $(j = 1 \dots 3)$ are independent predictors of parameters a_7 and a_9, and p_{2j} and

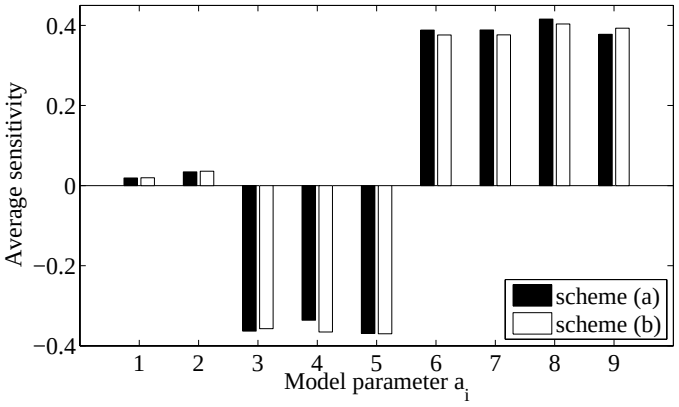

Figure 4.4: By implementing the revised Bode sensitivity function, the majority of first-stage parameters of $V351$ demonstrate similar sensitivities. The results obtained under two schemes are almost identical to each other.

Table 4.4: Identified first-stage parameter sensitivies of VAVs in the validation set under the revised Bode sensitivity function illustrate that their fractional contributions to the variance of model output are similar.

Rank	Parameter	$Z^Y_{a_i}$	W_{a_i} (%)
1	a_6	+0.3794	14.40
2	a_8	+0.3752	14.08
3	a_7	+0.3739	13.98
4	a_9	+0.3680	13.54
5	a_5	−0.3652	13.34
6	a_4	−0.3621	13.11
7	a_3	−0.3556	12.64
8	a_1	−0.2146	4.61
9	a_2	−0.0554	0.30

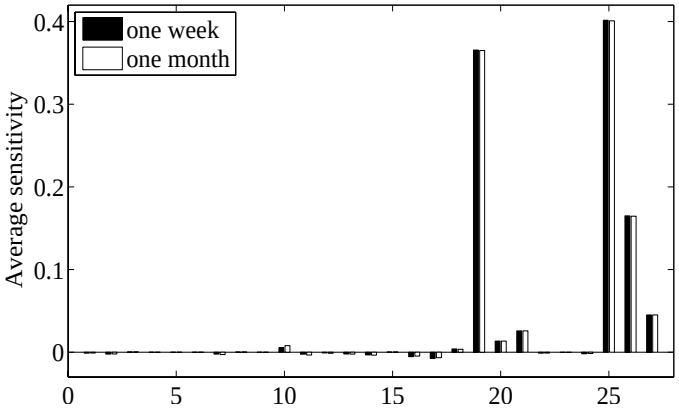

Figure 4.5: By implementing the Bode sensitivity function, the identified second-stage parameter sensitivities (presented in the order of p_{1j}, p_{2j}, ..., p_{9j} $(j = 1 \ldots 3)$) with training scheme (a) are consistent among observations of different lengths in terms of both magnitude and ranking.

p_{1j} $(j = 1 \ldots 3)$ are independent predictors of parameters a_1 and a_2. Therefore, the total-stage sensitivity results are in line with the first-stage results.

The models for all VAVs demonstrate consistent results between the two stages of regression when scheme (b) is adopted. For this reason, we don't discuss scheme (b) further.

4.3.5.2 Revised Bode Sensitivity

With the same VAVs and data sets, the revised Bode sensitivities in Eq. 4.8 of the total model to most second-stage parameters are of similar level, as shown in Figure 4.6. Sensitivities to 77.8% of the second-stage parameter are almost identical with absolute values around 0.2. This is consistent with the results in Section 4.3.4.2 where the revised Bode sensitivities to eight out of nine first-stage parameters are similar. This also suggests that the uncertainties of the second-stage parameters almost evenly contribute to the variance of the model output.

After substituting Eq. 4.2 into 4.1, it becomes clear that the second-stage parameters directly determine the epPMV model. The quality of the PMV predictions depends on the second-stage parameters. We have found that the second-stage parameters are normally distributed. Table 4.2 lists their means and standard deviations.

To evaluate the model output due to variations of the second-stage parameters, we varied one second-stage parameter around its nominal value by two times of its identified standard deviation, while keeping other second-stage parameters fixed

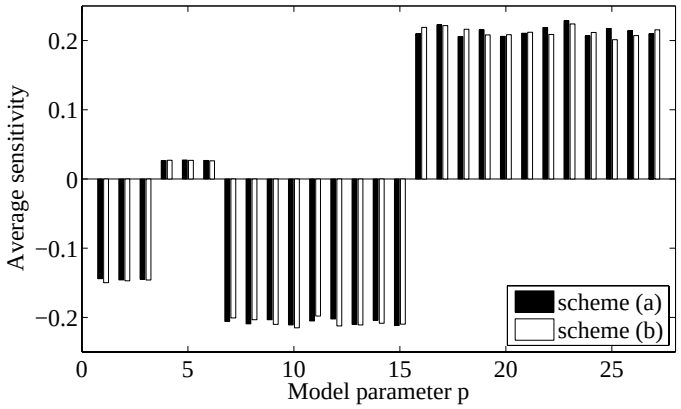

Figure 4.6: Similar total parameter sensitivities (presented in the order of p_{1j}, p_{2j}, ..., p_{9j} $(j = 1 \ldots 3)$) indicate that the second-stage parameters' fractional contributions to the variance of model output are nearly evenly distributed.

at their nominal values. With this setup, we formed the model, calculated and evaluated the model output for each parameter. Table 4.5 lists the standard deviations of the model prediction over one month for VAVs in the validation set. The average standard deviation is 2.63. The largest standard deviations of the model output are most sensitive to perturbations of p_{5j} and p_{3j} $(j = 1 \ldots 3)$, and least sensitive to perturbations of p_{2j} $(j = 1 \ldots 3)$.

4.4 Discussions

1. The probability distributions of the epPMV model parameters depend on many factors. We have found that all model parameters with both schemes are normally distributed.

2. Two different Bode sensitivities lead to slightly different ranking of parameters for the models of the two stages. However, parameters that are related to local control variables consistently are ranked high. Disagreement of the lower ranking is not of practical concern since these parameters have much less influence on the model output [135].

3. The data used in this study were sampled at five-minute interval. The parameter sensitivities of the models are studied with five and thirty minute intervals. We have tested shorter and longer sampling intervals by linearly interpolating

Table 4.5: Standard deviations of the epPMV model output due to variations of second-stage parameters.

Parameter	Output Stdv.	Parameter	Output Stdv.	Parameter	Output Stdv.
p_{11}	2.3700	p_{41}	0.0847	p_{71}	1.8833
p_{12}	0.6845	p_{42}	0.0532	p_{72}	1.1411
p_{13}	1.7576	p_{43}	0.0484	p_{73}	1.3105
p_{21}	0.0008	p_{51}	13.2396	p_{81}	1.4428
p_{22}	0.0007	p_{52}	3.7963	p_{82}	0.4593
p_{23}	0.0008	p_{53}	9.7294	p_{83}	1.1311
p_{31}	10.8216	p_{61}	2.1912	p_{91}	1.7620
p_{32}	3.3777	p_{62}	0.6649	p_{92}	1.2260
p_{33}	7.8832	p_{63}	1.5466	p_{93}	1.2941

the original data or by deleting the extra data points. The resulting parameter sensitivities of the models have less than 3.4% change on average. However, the sensitivity ranking under different sampling intervals are identical.

4. The epPMV model used for sensitivity analysis in this chapter is data-driven and needs to be updated when significant changes occur in the HVAC system including control strategies, major equipment and operation profiles.

4.5 Conclusions

We have investigated the parameter sensitivity of the epPMV model with the help of the Bode sensitivity and revised Bode sensitivity. We first study the sensitivity of the first-stage regression model and have identified the parameters to which the model is most and least sensitive. Top two parameters are discharge air temperature and discharge air relative humidity. These parameters suggest control inputs that have high authority to influence the human comfort. The least sensitive parameter relates to outside environmental conditions, which could be potentially excluded in the epPMV model for model simplification.

We have also found that the parameters to which the second-stage regression model is most and least sensitive are coupled to the parameters to which the first-stage model is most and least sensitive. The parameters of both the first- and second-stage models are found to be nearly normal. Hence, the output of the model is also normal due to its linear structure.

Finally, we have evaluated the effect of different length of the data series used in the regression modeling, and found that the parameters to which the models are most and least sensitive are independent of the different length of the data series.

Chapter 5

FAULT DETECTION AND DIAGNOSIS

5.1 Top-Down Approach with Cross-Level Detection Ability

With the energy flow model and temporal-spatial partition strategy, we develop a system-level FDD structure that can capture abnormal energy consumption irrespective to the nature of faults. By monitoring energy flow and consumption, the FDD strategy can detect faulty HVAC units in a top-down manner. By monitoring the energy flow fluctuations of HVAC units such as bridges and AHUs at higher level, instead of lower level components such as fans and motors, the proposed FDD strategy reduces the computational effort in real-time monitoring of the HVAC system, obtains a system-level view of the HVAC operation, and provides a way to integrate the existing methods for component fault detection when needed.

The proposed FDD method consists of off-line training and real-time application steps as shown in Figure 5.1. The off-line training needs a large number of normal data as well as data with known faults. We propose to use the energy feature extracted from the measurements. The spatial and temporal partitions are created based on architectural, environmental and human factors of the rooms. The off-line training sessions establish the thresholds for normal operation of various units. In the real-time application, the measurement integrity of the sensors are first checked [35–37]. The thresholds calculated from the off-line training are applied to detection algorithm and confirmation of faulty HVAC units. We can further narrow down the faults to the component level. Finally, the real-time application data can be used to update the thresholds identified off-line.

The proposed FDD is reliable and scalable. The top-down strategy can detect and respond to various hardware and software faults, unsatisfactory thermal comfort and abnormal fluctuations of the energy flow, examples of which are discussed in the following Chapter.

5.2 Spatial and Temporal Partition Strategy

HVAC units have certain patterns of flow energy consumption over time domain varying with environmental conditions, architectural factors, human occupancy and control setpoints. We group these units such that the ones in a group share a similar pattern of flow energy consumption over different time intervals. And

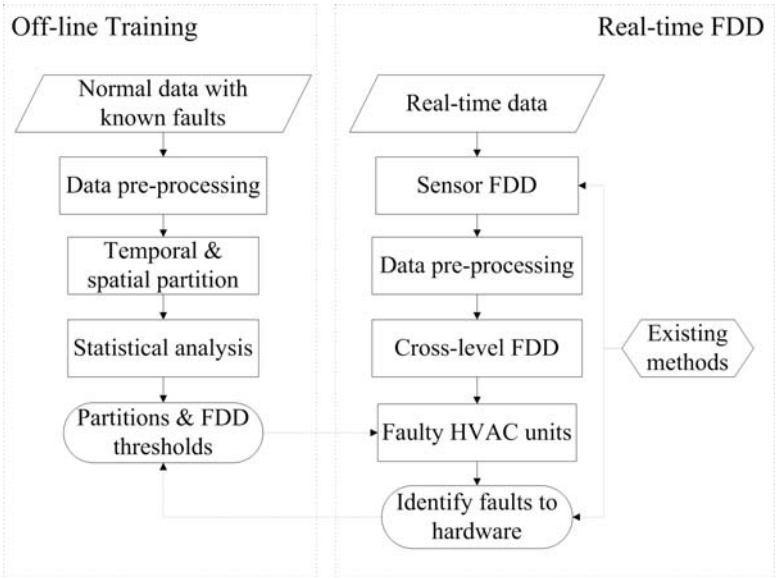

Figure 5.1: Schematics of the proposed top-down FDD.

units in the a group can be monitored with the same and tighter threshold for fault detection.

5.2.1 Spatial Partition

Figure 5.2 shows a floor map of fifteen VAVs ($V351$-$V365$) installed on the third floor of the SE building. We use the sensor data collected from May 1 to May 28 in 2009. As a spatial partition strategy, we investigate the daily correlation between individual VAV's energy performance with the outside temperature, and use the correlation level as the basis to divide these VAVs into different groups.

The sun exposure is one of the factors that influences the energy performance of VAVs, especially in California. Human occupation also plays an important role. Based on these two factors, we can divide these fifteen VAVs into four groups: office suite (OS), faculty office (FO), suite connection (SC) and stairs (S). The office suite has two walls directly exposed to the sun. The stair way has only one large glass wall receiving the sun light. The faculty office also has only one window facing the sun. The suite connection has no direct sun exposure. There are always people walking on the stairs. Students sometimes pass the suite connection. Most of the time, the faculty office is occupied only by one person. The office suite is occupied by the dean with occasional visitors. The correlation analysis seems to suggest that the sun exposure influences the correlation level on average, while the human occupation contributes to fluctuations in the correlation.

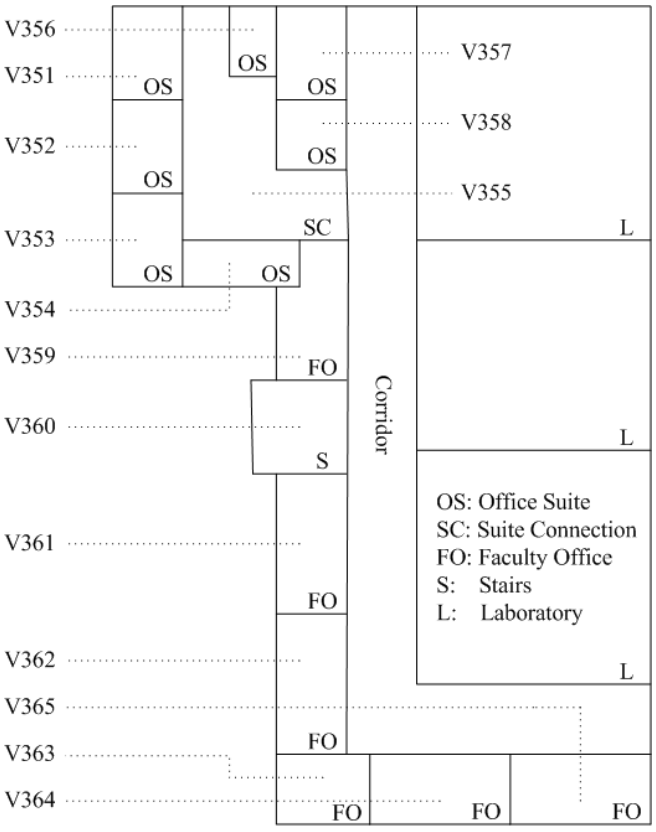

Figure 5.2: VAV distributions on the third floor of Science and Engineering Building on UC Merced campus.

Table 5.1: Groupings of VAVs on the third floor of Science and Engineering Building on UC Merced campus, and the influences of the sun exposure and human occupation.

Group	VAVs Included	Sun Exposure	Occupancy
Office suite	V351-V358	strongest	least complicated
Faculty office	V359, V361-V365	strong	less complicated
Stairs	V360	strong	most complicated
Suite connection	V355	weak	complicated

Table 5.2: Numerical results of correlation analysis and PCA for the groups of VAVs.

Group	Correlation Coefficient (u/σ)	Projection Angles of PCA (u/σ)
Office suite	0.79/0.05	48.8°/5.3°
Faculty office	0.51/0.15	45.4°/7.1°
Stairs	0.31/0.49	20.0°/34.5°
Suite connection	0.15/0.25	-8.7°/27.1°

The architectural design determines the correlation among certain VAVs. Take the dean's suite as an example. VAVs within the suite are thermally coupled because the air flows freely among the rooms in the suite. On the other hand, the faculty offices with $V365$ and $V361$ are more independent because no direct thermal interaction exists between them. Table 5.1 summarizes the influences of the sun exposure and human occupation on the four groups of VAVs.

We compute the flow energy consumption of the VAVs in these four groups and the correlation of flow energy consumption with the outside temperature. Table 5.2 shows the mean (u) and standard deviation (σ) of correlation coefficients for the four groups in May, 2009. We also conduct the PCA of each group of VAVs. The mean and standard deviation of the projection angle between the outside temperature T_O and the flow energy consumption of each group of VAVs are also presented in Table 5.2. The statistical analysis supports the proposed spatial partition scheme.

Table 5.3: Flow energy consumption of AHU 9 and its VAVs during day and night.

Group	Day (u/σ)	Night (u/σ)
AHU9	6.1e+4/3.2e+4	9.2e+3/1.8e+3
VAVs of AHU9	4.4e+5/1.0e+5	2.8e+4/1.6e+4

The VAV group serves as another level between AHU and individual VAVs to improve the efficiency of narrowing down a VAV fault from its upper-level AHU.

5.2.2 Temporal Partition

Temporal conditions, such as day and night, weekday and weekend, and seasons are related to different environmental conditions and human occupancy profiles. The temperature setpoints are different during the day and night. Take the SE building as an example. The difference between the cooling setpoint and heating setpoint during the night is $30°F$, while during the day it is $4°F$. The larger difference leads to less operation of VAVs. Table 5.3 presents the mean and standard deviation of the flow energy consumption of AHU 9 ($A9$) and total flow energy consumption of its VAVs. The data indicate a lower mean and smaller variation during the night (nearly straight line around zero) than the day (fluctuating line above zero).

To demonstrate the effect of temporal partition, we consider thirty-two VAVs of AHU 9 ($A9$) on the third floor of the SE building. The average daily correlation coefficients between the energy consumption of sixteen VAVs and T_{OA} for day and night are shown in Figure 5.3. 16128 samples from 56 days in May and July 2009 are used. The figure shows that all VAVs keep higher mean levels during the day than at night. Variations of energy consumption can also be observed as the season changes as shown in Figure 5.4.

For those HVAC units in certain spatial partitions, we further divide them in different temporal partitions. Specifically, we create summer and spring, day and night partitions with different FDD thresholds for each partition.

The above partition strategies provide two benefits. First, both spatial and temporal partitions increase the FDD sensitivity by providing smaller standard deviation of the correlation coefficients during the normal operation. Table 5.4 compares the mean (u) and standard deviation (σ) of the correlation coefficients in Figures 5.3 and 5.4 with or without temporal partition. Without day and night partition (DTP) or seasonal partition (STP), both standard deviations are 100% bigger than that with temporal partition. Second, the spatial partition makes the top-down strategy

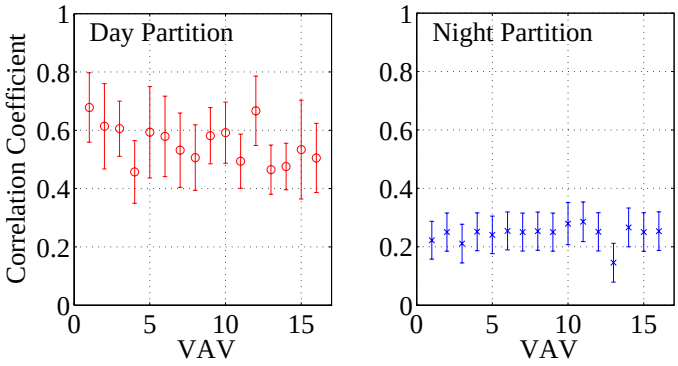

Figure 5.3: The correlation between the energy consumption of VAVs and the outside temperature. The average level of correlation is higher during the day than at night.

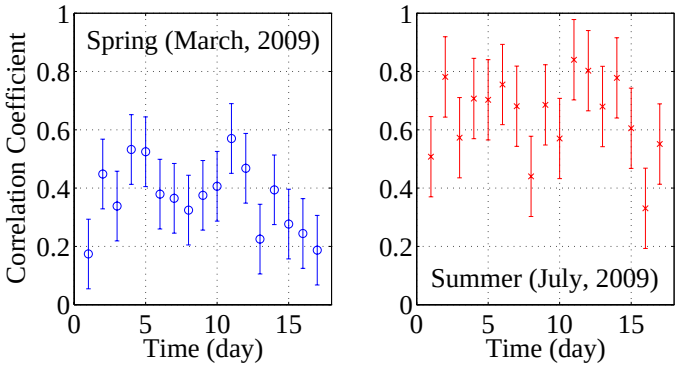

Figure 5.4: The correlation between the energy consumption of $A9$ and the outside temperature. The average level is higher in the summer than in the spring.

Table 5.4: The mean (u) and standard deviation (σ) of the correlation coefficients in Figures 5.3 and 5.4 with or without temporal partition.

	Day	Night	Without DTP	Spring	Summer	Without STP
u	0.50	0.26	0.40	0.36	0.74	0.55
σ	0.13	0.12	0.21	0.11	0.13	0.23

more effective. We shall further demonstrate the advantage of the partition in the following Chapter with examples of detected faults.

5.3 Absolute and Relative References

The difference between the outside temperature and room temperature is one of the driving forces of flow energy consumption. Meanwhile the environmental factors such as solar radiation, internal heat gain, outside humidity in summer etc. also influence the flow energy consumption. Under the influence of these factors, the flow energy consumption of an HVAC unit assumes a certain level of correlation with the outside temperature. The correlation fluctuates in a certain range. An unexpected fluctuation outside a threshold of variations of the correlation is considered abnormal. The units with abnormal flow energy consumption in reference to the outside temperature may be faulty. In this sense, we refer to the outside temperature as an *absolute reference* in a statistical sense.

To confirm the fault, we compare the flow energy consumption of the possibly faulty unit with that of other units at the same level or with a mathematically equivalent measure. This comparison provides a *relative reference*. Specifically, we investigate the correlation between the suspicious unit and other units at the same level or a mathematically equivalent measure. An example of the mathematically equivalent measure is the flow energy consumption of an AHU (E_{AH}) and the sum of flow energy consumption of all its lower-level VAVs (E_{VS}).

We provide a validation example for the relative reference. We apply the energy flow model to sixty-one VAV units and two corresponding AHU units installed in SE building, and calculate the flow energy consumptions of each AHU and its VAVs for 2016 samples in one week. Figure 5.5 shows that E_{AH} provided by $A10$ and E_{VS} track each other well. A further correlation analysis shows that the correlation coefficient between E_{AH} and E_{VS} for $A10$ is 0.94, with an even higher value 0.98 for the AHU 9 ($A9$). This justifies that the E_{VS} can be used as the relative reference to E_{AH}.

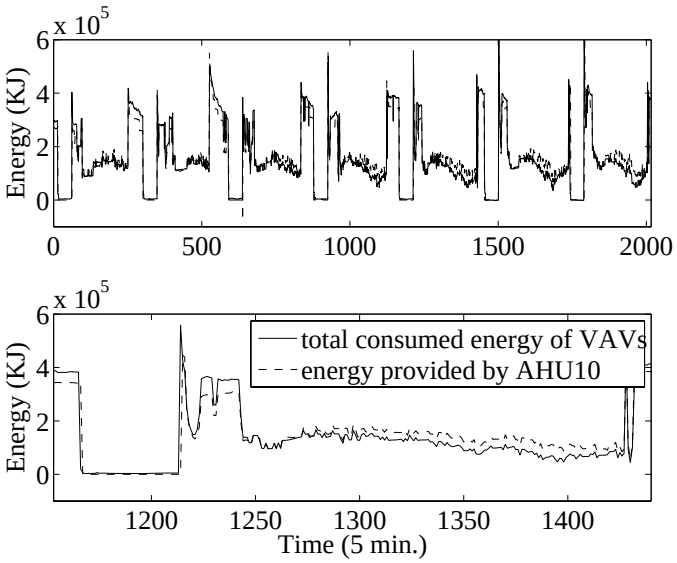

Figure 5.5: Mathematically equivalent measurements show a very high correlation between their flow energy consumption.

5.4 Statistical Analysis

Once spatial and temporal partitions are created, we apply various methods to analyze and recognize the flow energy consumption patterns under normal conditions to obtain corresponding FDD thresholds. In the current work, we select PCA [156–158] and correlation analysis [159] as examples.

In the PCA approach, we project the principal components onto a sub-space spanned by the measurements of interest. Since the principal components are calculated from the whole data set in the higher dimensional space, their projections onto the sub-space contain physical interactions with all other measurements.

As an example, consider a sub-space spanned by two measurements M_1 and M_2. The projection of the principal component into the sub-space is illustrated in Figure 5.6. The measurements are related as

$$M_1 = k \cdot M_2, \tag{5.1}$$

where the slope is $k = \tan\theta$ and θ is the projection angle. When $0 \leq \theta \leq \pi/2$, M_1 and M_2 are linearly and positively correlated within the principal component. When $\pi/2 \leq \theta \leq \pi$, M_1 and M_2 are linearly and negatively correlated. We postulate that the projection angle changes with the conditions of the HVAC system, particularly when a fault occurs.

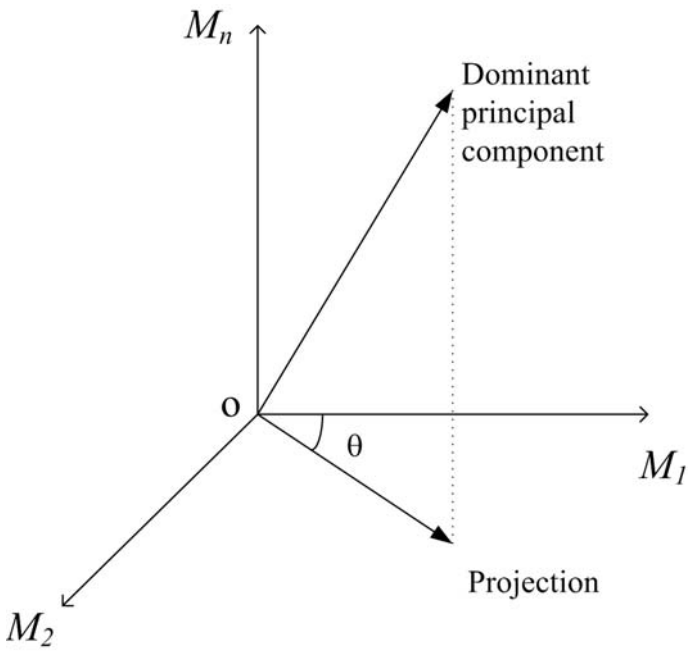

Figure 5.6: The projection of the dominant principal component onto the 2-dimensional plane spanned by two measurements of interest.

Let θ_n denotes the nominal value of the projection angle between the measurements M_1 and M_2 when the system is in normal operational condition. How much deviation of θ from θ_n would signal the existence of a fault? We need a large number of measurement data and various faulty incidents to train an algorithm in order to establish the threshold for θ statistically.

As for the correlation analysis, we propose to use the correlation between the energy performance of VAVs and the outside temperature T_O as the criteria to evaluate whether VAVs are operating at normal condition.

Chapter 6

FDD EXAMPLES

6.1 Detection of Thermostat Setpoint Variation

In summer of 2010, UC Merced participated in the Demand Response program to reduce energy consumption during "critical peak" periods. System-wide thermostat setpoints were raised by four degrees during three peak periods in the afternoon on July 16, August 24 and 25. Since the setpoint related faults are one of the ten typical HVAC faults [26], this demand response program provided us a chance to test our algorithm.

We selected $A10$ and its twenty-nine VAVs, calculated their energy consumption for one hundred and five days from June until the middle of September 2010. A matrix consisting of three measurements: T_{OA}, E_{AH} and E_{VS} is then formed. We computed the dominant principal component of this matrix, and projected it onto 2-dimensional planes formed by two of the three measurements. Temporal partition of day and night is applied to the data set. With 24192 normal samples in day partition, we obtained the mean ($\bar{\theta}$) and standard deviation (σ_θ) of the projection angle (θ). Finally, we considered the 95% probability interval of projection angle $[\bar{\theta} - 1.96\sigma_\theta, \bar{\theta} + 1.96\sigma_\theta]$ as the normal range based on the Gauss distribution assumption, and proposed to use $\pm 1.96\sigma_\theta$ from the mean as the threshold for fault detection. This threshold is adopted for the next numerical example also.

Figure 6.1 (top) presents the projection angle of daily observations on the plane spanned by E_{AH} and T_{OA}. The projection angles of Day 45 (July 16), Day 83 (August 24) and Day 84 (August 25) went beyond the threshold. Furthermore, the projection angles onto the plane spanned by E_{VS}, the mathematically equivalent measure to E_{AH}, and T_{OA} also went beyond the threshold during the same periods, while the correlation between the mathematical equivalents E_{AH} and E_{VS} remains normal. These results indicate that a system-wide abnormal energy performance had occurred during those three periods of time.

On the other hand, if we do not implement the temporal partition of day and night, the entire data set including days and nights has a much larger standard deviation σ_θ, and thus a larger threshold. The same correlation analysis then would not be able to identify the abnormal behavior of the HVAC system during those periods. The result is shown in Figure 6.1 (bottom).

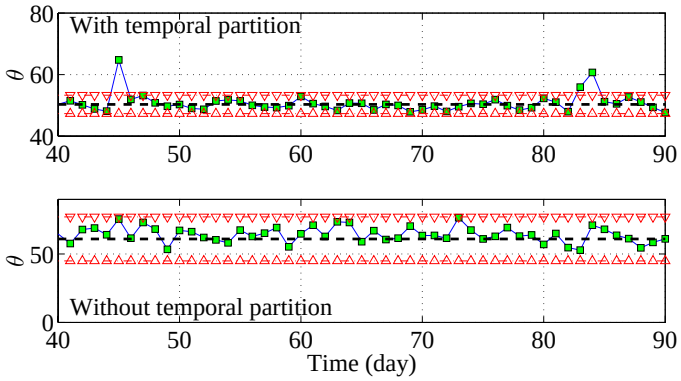

Figure 6.1: The daily average projection angle (solid line with squares) between E_{AH} and T_{OA}. Three abnormal variations of thermostat setpoints are captured beyond the threshold (dash-dot line with triangles) by the proposed FDD with temporal partition (top). Without temporal partition (bottom), the abnormal variations of the thermostat are not detected. The average of daily projection angle is represented as the dash line.

6.2 Frozen Supply Fan of an AHU

$A10$ has twenty-nine lower-level VAVs installed in SE building We use the sensor data collected from May 1 to May 28, and June 15 to July 12 in 2009, apply the energy flow model to calculate the flow energy consumption for all involved units, and use FDD thresholds in the summer day temporal partition introduced in Section 5.2.2.

We construct a matrix consisting of three measurements: T_O, E_{AH} and E_{VS}. We compute the dominant principal component of this matrix, and project it onto 2-dimensional planes formed by two of the three measurements.

We consider the weekly performance of the system. Figures 6.2, 6.3 and 6.4 show the results of projections of the principal component for four weeks (weeks 1-3: magenta dashed line, blue dotted line and green dashdot line. week 4: solid line) in June to July. Figure 6.2 shows projections on 2-dimensional plane spanned by the AHU flow energy consumption E_{AH} and the mathematically equivalent measure, i.e. the total flow energy consumption E_{VS} of its lower-level VAVs. The projection angles of weeks 1, 2 and 3 are close to each other at nearly 45°. The projection angle of week 4 is nearly 160°, far away from that of the other three weeks.

From a physics point of view, every pair of the three measurements T_O, E_{AH} and E_{VS} should be linearly and positively correlated. In weeks 1, 2 and 3, the projection angles are near 45° indicating a linear and positive correlation. The

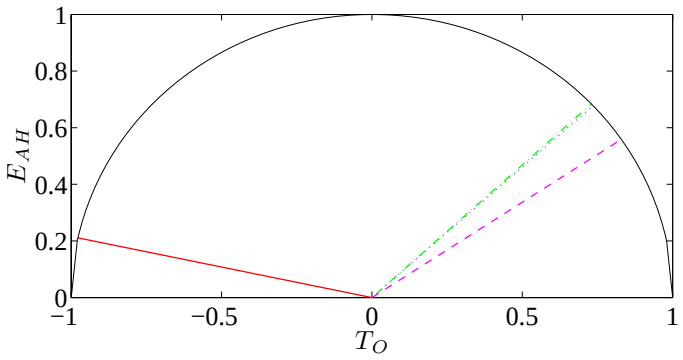

Figure 6.2: The projection angle indicates that the AHU flow energy consumption in week 4 is negatively correlated with the outside temperature, which is physically unreasonable.

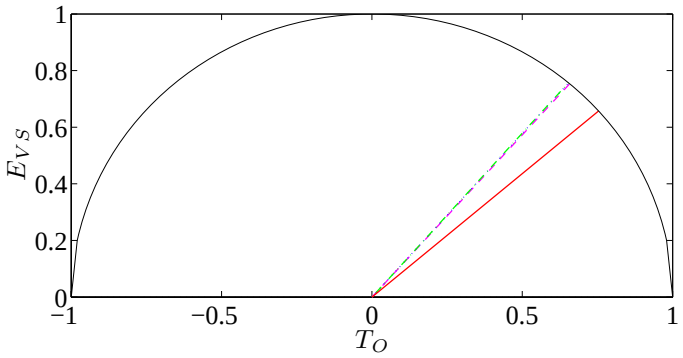

Figure 6.3: All four weekly projections of the dominant principal component are closely clustered in the first quadrant, which suggests that all the VAVs were operating normally in weeks 1-4.

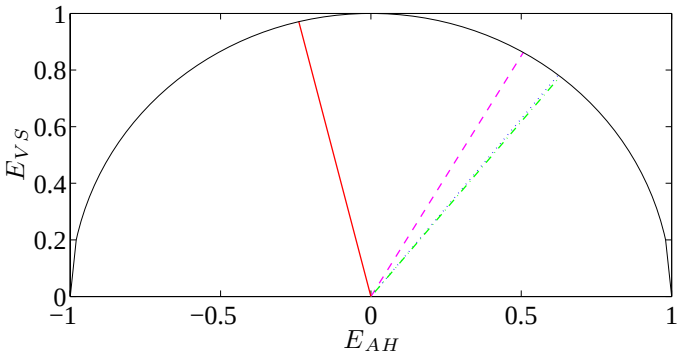

Figure 6.4: AHU flow energy consumption in week 4 is negatively correlated with the mathematically equivalent measure, which confirms the fault.

projection angle of week 4 is 160° indicating a negative correlation. Figure 6.2 suggests that during weeks 1, 2 and 3, the AHU behaved normally as its flow energy consumption is positively correlated with the absolute reference T_O. However, during week 4, a fault may have occurred, which caused abnormal flow energy consumption.

Recall that the mathematically equivalent measure of the flow energy consumption of an AHU is the sum of flow energy consumption of all its lower-level VAVs. Figure 6.3 shows that the projection angles of all four weeks are in the first quadrant around 45°, which verifies that all the VAVs were functioning normally because the sum of their flow energy consumption E_{VS} correlates positively with T_O. Hence, the sum of flow energy consumption of all the VAVs can serve as a relative reference for the AHU.

Figure 6.4 compares the AHU energy E_{AH} with the relative reference E_{VS}. It is clear that the projection angle of week 4 deviates significantly from that of the other three weeks, and this indirectly confirms the fault detected in Figure 6.2.

We have confirmed with the building manager that the supply fan of the AHU was frozen for the entire week in July as shown in Figure 6.5. Meanwhile the return fan of the AHU was working normally. So it was the frozen supply fan that caused the abnormal flow energy consumption for the AHU in week 4 recorded in the June-July data set.

Given the knowledge of the fault, let us now try to determine thresholds of the projection angle. We use four weeks of data in May and the first three weeks in the June-July data set when the system was under normal condition. We compute the standard deviation of the projection angle, and consider the 90% probability interval of the projection angle $[\bar{\theta} - 1.644\sigma_\theta, \bar{\theta} + 1.644\sigma_\theta]$ where $\bar{\theta}$ is the average projection angle and σ_θ is the standard deviation. The probability is computed

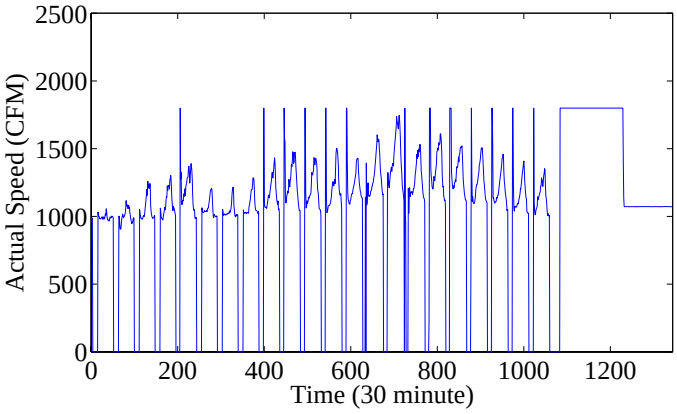

Figure 6.5: The supply fan of $A10$ was frozen for week 4 in June to July.

based on the Gauss distribution assumption. We propose to use $\pm 1.644\sigma_\theta$ from the average as the threshold for fault detection.

Figure 6.6 shows the projection angle of eight weekly observations. It can be seen from the figure that the projection angles of the seven weeks when the system was normal are close to the average value, and the projection angle of week 4 in the June-July data set goes out of the lower boundary of the threshold (the 90% probability thresholds with $\pm 1.644\sigma_\theta$ are marked by red dashdot lines with triangular symbols, the average projection angle by black dashed line, and the blue line with square symbols denotes the actual projection angle of the week).

We present the same analysis with daily observations in Figure 6.7 (legends of Figure 6.7 and following figures are the same as in Figure 6.6). The figure shows a considerable number of samples going beyond the threshold in the last six days, specifically in the first and third subplots where the flow energy consumption of the AHU is compared with the absolute reference T_O and the relative reference, i.e. the mathematically equivalent measure E_{VS}.

6.3 Sensor Fault

We apply the same method as in Section 6.2 to $A9$ in the same building and its thirty-two VAVs from May 31 to July 18 in 2010. During the summer night temporal partition on June 19, shown as Day 20 in Figure 6.8 and Day 12 in Figure 6.9, projection angles on both planes spanned by E_{VS} and T_O, E_{AH} and T_O jump dramatically between $\pm 45°$ as shown in Figure 6.8. The flow energy consumption of HVAC units during the night should be relatively stable with smaller variations. This fluctuation indicates two possible faults. One is that both $A9$ and all thirty-two

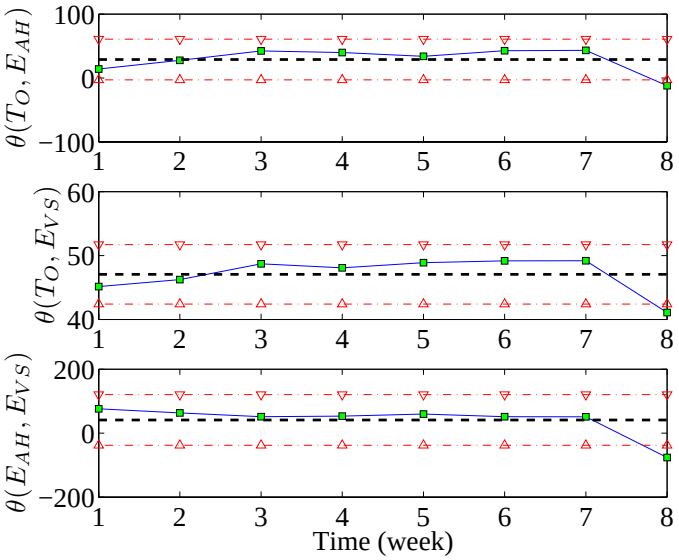

Figure 6.6: The projection angle of week 4 in the June-July data set indicates a possible fault during that time. This fault is verified as the frozen supply fan shown in Figure 6.5.

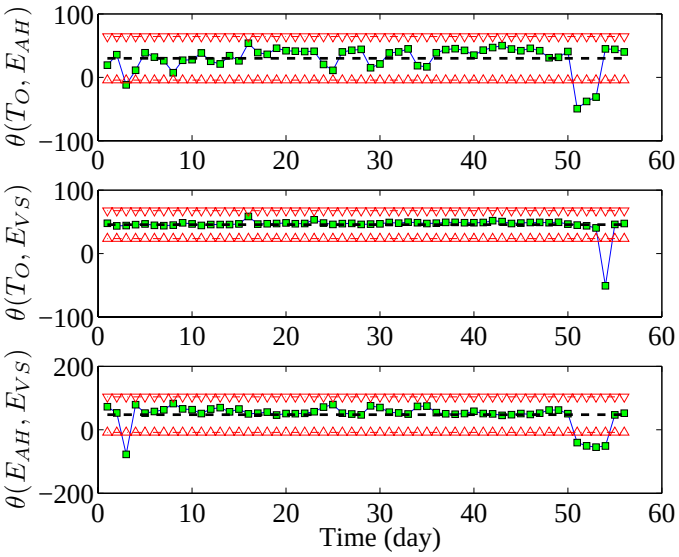

Figure 6.7: Daily observations also show that the projection angles exceed the threshold for three to four days in the last week.

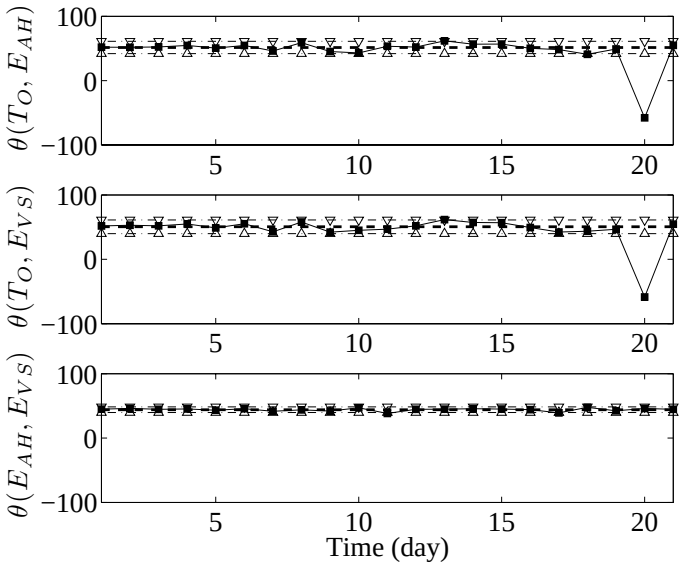

Figure 6.8: Daily projection angles exceed the threshold on June 19 indicating a possible fault during that time.

VAVs encountered some faults simultaneously. The other is that the outside air temperature sensor was broken thus providing inaccurate measurements of the reference T_O. By checking the correlation coefficients between flow energy consumption and outside temperature for all thirty-two VAVs during the same period, we find that 31 out of 32 (97%) VAVs display an out-of-threshold sample during Day 20 as shown in Figure 6.10. This implies that there is a high possibility that the fault is rooted from the inaccurate measurements of outside temperature.

In order to confirm whether the fault is caused by the outside air temperature sensor, we conduct trend analysis on the data recorded by the sensor. Considering that the fluctuation of temperature in the central valley of California is relatively large during the summer time, we use data from six days before June 19 and six days after June 19 as our training data to calculate the threshold under normal condition. We select two properties as the basis. The first one is the temperature difference between the last sample and the first sample in the night partition. We choose this property because during the summer night, temperature has a steady drop from 1 a.m. to 5 a.m. The other one is the slope of the first-order fitting curve of temperature during each night. This property shows the average trend of temperature during the night partition. In Figure 6.9, we notice that both the properties go beyond the thresholds on June 19. Figure 6.11 shows the sensor data in time domain. It is clear that while the rest twenty days demonstrate a steady

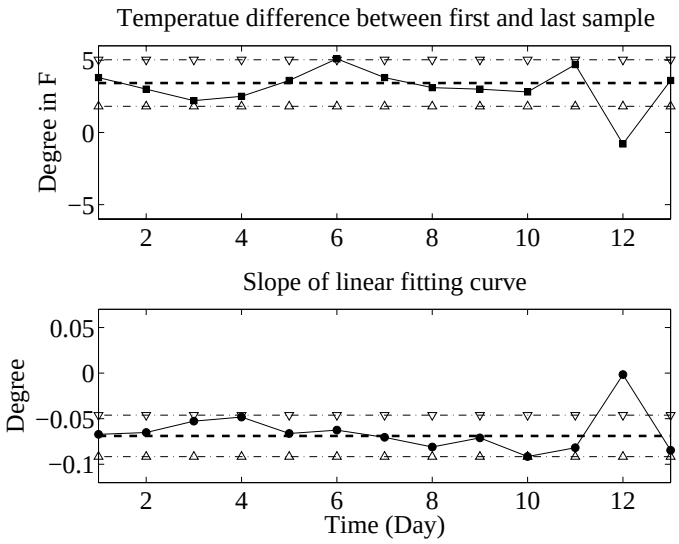

Figure 6.9: Trend analysis of the outside temperature sensor confirms abnormal deviation of the measurements on June 19.

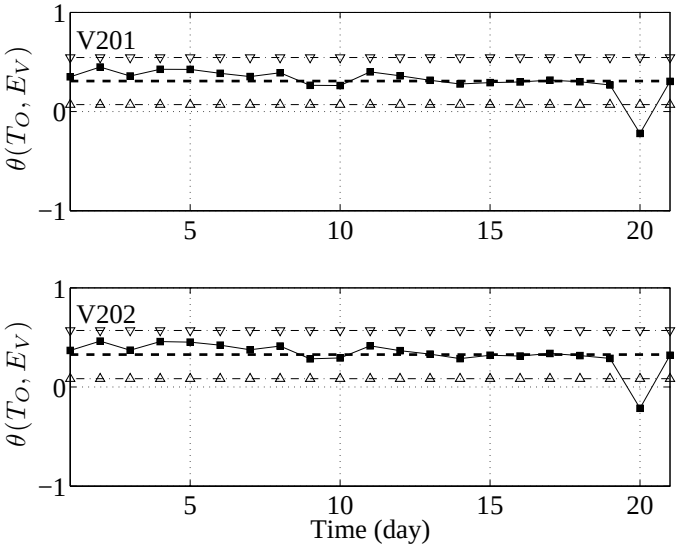

Figure 6.10: Two samples of individual VAVs with the correlation coefficients exceeding the threshold on June 19.

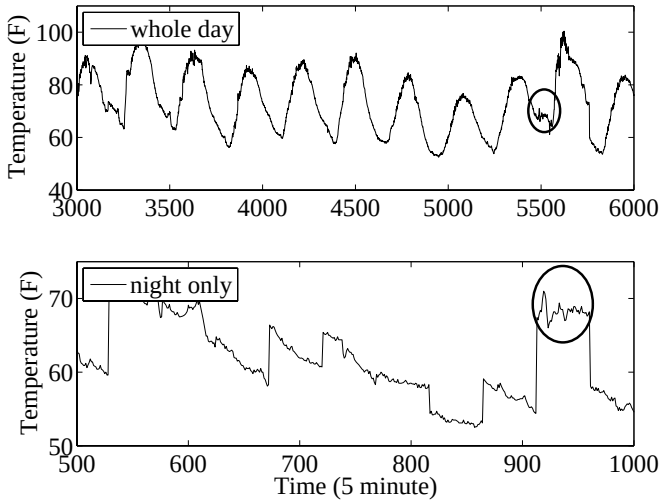

Figure 6.11: Intermittent abnormal fluctuations of the outside temperature sensor during June 19.

temperature drop, temperature reading of June 19 not only shows unreasonable fluctuations, but also displays a flat trend.

6.4 Blockage of Supply Air Inlet

In the same monitored data as in Section 6.3, we apply the spatial partition strategy to create four VAV groups. For the data of the summer day temporal partition, we calculate the total flow energy consumption of all VAVs within each group as our monitoring objects. Then we reserved VAV 201 ($V201$) within VAV group 1 for two weeks from May 31 to June 13 to conduct experiments when there was no human occupancy. In June 14 and 15, we blocked the supply air inlet with air-tight paper as shown in Figure 6.12 to simulate a common VAV hardware fault - a stuck damper.

We monitor the energy performance on the AHU level. In the daily observations of the projection angle on the plane spanned by E_{VS} and T_O shown in Figure 6.13, we notice the samples of day 13 and 14, which are June 14 and 15, go beyond the threshold obtained by training data under normal condition from May 31 to June 12. This indicates during June 14 and June 15, a fault occurred and triggered abnormal total flow energy consumption of all the VAVs. At the same time, projection angles on the plane spanned by E_{AH} and T_O remain within the threshold for all observations. This illustrates the flow energy consumption of $A9$ was normal during this period of time. So we speculate that the fault should come from the VAVs.

Figure 6.12: The supply air inlet of VAV201 was artificially blocked to simulate a stuck damper.

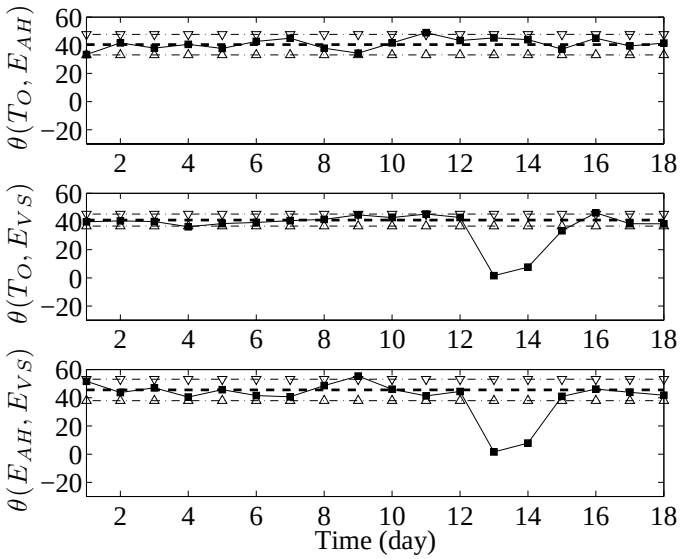

Figure 6.13: Daily observations of projection angles exceed the threshold in June 14 and 15.

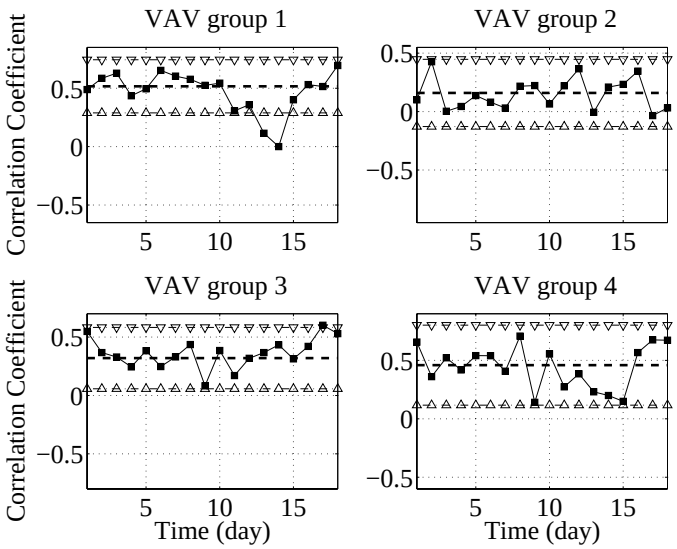

Figure 6.14: Only correlation coefficients of VAV group 1 exceed the threshold in June 14 and 15.

To track down this fault, we check the correlation coefficients between the VAV group's flow energy consumption and the outside temperature. In Figure 6.14, we detect that only VAV group 1 has two samples beyond threshold in June 14 and 15. So we narrow the fault from thirty-two VAVs to eight VAVs within VAV group 1.

Finally, Figure 6.15 shows the results of individual VAVs within VAV group 1. It is clear that only $V201$ manifests two out-of-threshold samples on June 14 and 15. Thus, we successfully detect the fault in a VAV by monitoring its upper-level AHU. Also, since we have only conducted thirteen instances of FDD (one at AHU level, four at VAV group level, and eight at VAV level), the computation is reduced by 60% compared to directly monitoring all 32 VAVs involved. In fact, the more VAVs an AHU has, the more efficient the cross-level FDD is. As more and more skyscrapers being constructed, the cross-level fault detection holds a great potential for practical real-time FDD.

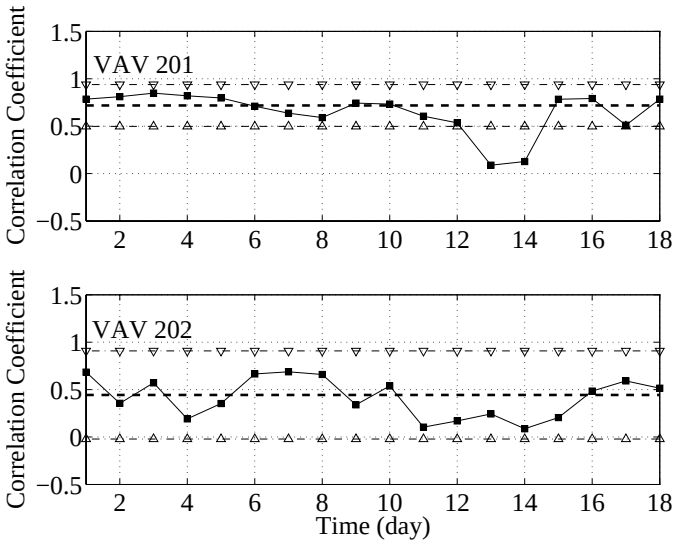

Figure 6.15: Only correlation coefficients of VAV 201 exceed the threshold in June 14 and 15.

Chapter 7

SUMMARY AND FUTURE WORK

7.1 Concluding Remarks

To leverage the untapped capabilities of modern building automation and control systems, we have developed a system-level monitoring and diagnosis tool of HVAC system. Works in this thesis focus on two functionalities, i.e., monitoring and diagnosis. For monitoring, we have combined the strength of both analytical and data-driven approaches, established a parametric modeling approach that incorporates building envelope parameters and control variables in the developed models of room temperature and the PMV index. For diagnosis, we have created a top-down FDD architecture. This FDD architecture utilizes rich performance surveillance of existing sensors and controller hardware in buildings. By employing deductive modeling, statistical analysis, and artificial intelligence, the proposed FDD can automatically detect and diagnose abnormal HVAC system performance. The resulting monitoring and diagnosis tool of HVAC system detect hardware failures, non-optimal operations, and energy inefficiencies, and can be applied on system-wide scales to optimize overall system performance and help achieve the goal of zero energy consumption.

7.1.1 Monitoring

We have developed a physics-based linear parametric room temperature model from the thermodynamic equations. The proposed model has a proper order and much less variables compared to a fully numerical model. Trained with the data over a relatively short time, the proposed model is capable of predicting the room temperature in both short and long terms with a high accuracy. For example, for all sixty-one VAVs involved, the corresponding pbARMAX models trained with the data of optimal length can predict the room temperature over ten weeks with mean squared errors less than 0.01 and coefficients of determination above 0.99 on average. We have also shown that the proposed model provides an analytical foundation for the spatial and temporal partition strategy. With the help of the partition strategy, the proposed temperature model further improves its prediction performance, and could be a basis for intelligent control design of HVAC systems.

Based on the physics-based linear parametric room temperature model, we have further developed a new strategy to incorporate the architectural parameters

into linear parametric regression models of the room temperature. The regression model is developed in two stages. In the first stage, linear parametric regression models of the room temperature based on the thermodynamic equations are developed with the data from VAVs involving different architectural parameters. The collection of the parameters of the regression models are then related to the architectural parameters with the help of another set of linear regression relationships. This is called the second stage regression. The model developed this way has been validated with the data in the validation set. We have found that the two-stage regression models are able to predict the room temperature in both short and long terms with a high accuracy. Furthermore, the proposed models demonstrate improved goodness-of-fit compared to a representative ARMAX model in the literature which has been proved to be very effective in the same application.

To monitor the thermal comfort in a practical approach, we have also developed a strategy to incorporate the architectural parameters and control variables to the empirical PMV model. The epPMV regression model is developed in two stages of regression. The epPMV model has been validated with the data in the validation set. We have found that the two-stage regression models are able to predict the PMV in both short and long terms with high accuracy.

To investigate the statistical reliability of these models developed, we have selected and investigated the parameter sensitivity of the epPMV model with the help of the Bode sensitivity and revised Bode sensitivity. We first study the sensitivity of the first-stage regression model and have identified the parameters to which the model is most and least sensitive. Top two parameters are discharge air temperature and discharge air relative humidity. These parameters suggest control inputs that have high authority to influence the human comfort. The least sensitive parameter relates to outside environmental conditions, which could be potentially excluded in the epPMV model for model simplification. We have also found that the parameters to which the second-stage regression model is most and least sensitive are coupled to the parameters to which the first-stage model is most and least sensitive. The parameters of both the first- and second-stage models are found to be nearly normal. Hence, the output of the model is also normal due to its linear structure. Finally, we have evaluated the effect of different length of the data series used in the regression modeling, and found that the parameters to which the models are most and least sensitive are independent of the different length of the data series.

The proposed models are data-driven. Change of the HVAC control strategies, major equipment, and operation profiles may lead to changes of the system dynamics. In this case, the model needs to be updated with the new measurements. One can even periodically update the model. This leads to an adaptive modeling strategy of HVAC systems. Nevertheless, the mathematical procedure of the modeling remains the same, and could be also extended to other types of buildings equipped with similar HVAC systems.

It should be noted that the proposed model could also be used to predict other parameters of interest in HVAC simulations, e.g., the relative humidity. However, we limit the implementation of the proposed model on the prediction of room temperature only in this thesis due to the unavailability of relative humidity measurements of indoor air. Nevertheless, we believe that the proposed method will be valid when the humidity data becomes available. We are currently trying to manage a way to estimate the relative humidity based on the available measurements at the AHU level. In our future work, we will expand the proposed model to include such important parameters.

7.1.2 Diagnosis

We have innovated and established a system-level FDD architecture on building HVAC systems. Energy description of HVAC units and temporal and spatial partition strategy are two key elements of the method. The energy flow model extract a unique feature shared by HVAC units at different levels. This energy feature reflects abnormal operations caused by faults of different natures, and consequently serves as a foundation for monitoring the HVAC performance in a uniform manner. With the temporal and spatial partition strategy, the proposed FDD monitors the energy performance of HVAC units in groups and at different levels over well-defined time periods. As a result, we have tight thresholds for the HVAC units in the same group, and gain the ability to detect faults by navigating intelligently across different levels of HVAC. We have implemented the PCA in a unique way to extract correlations between signals of interest while preserving interactions among all the related measurements. The concept of absolute and relative references for fault detection has also been introduced, which provides the rationale for logic decisions.

With these core techniques, the resulting FDD is able to detect most typical HVAC faults uniformly. The algorithm can trace down HVAC faults on the lower levels from monitoring the units on the upper levels. This top-down strategy increases the efficiency of fault detection. With the limited real-time data, we have studied the threshold for detecting three typical faults in HVAC systems. Examples of detected hardware failures as well as non-optimal operations from the HVAC system in the SE building have been presented to demonstrate the effectiveness of the proposed FDD. More measurements and extensive studies are needed to establish thresholds for various faults at different levels. It should be noted that the above discussion on the threshold is limited by the availability of the measurements with known faults, and should not be generalized. Also, a too small fault may not be able to trigger enough abnormal performance detectable on the upper level.

7.2 Future Work

7.2.1 Modeling

There are three major operational objectives of building HVAC systems, i.e., occupant comfort, energy efficiency, and low utility bill. Practically, it is desirable that models of these objectives include HVAC control variables, architectural parameters, occupancy, and environmental conditions. With such models, we can optimize the HVAC operation by adjusting control inputs to address various architectural, occupancy, and environmental conditions. However, the mapping between end-use HVAC energy consumption and control inputs, especially from the data driven approach, is still unclear. We shall further our modeling work by exploring this topic and taking advantage of the rich historical data of the entire UC Merced campus. Ideally, the resulting models will reveal not only both local and global relationships between end-use energy consumption and control variables, but also the link between energy consumption and thermal comfort.

7.2.2 Fault Detection and Diagnosis

We shall expand and enhance our FDD techniques by exploring other machine learning techniques. Due to various natures of HVAC faults, an optimal FDD algorithm may exist for a particular type of faults. It is necessary and beneficial to establish some performance metrics to systematically evaluate the effectiveness of applying certain machine learning techniques on detecting certain HVAC faults. Ideally, these evaluation metrics should assist the FDD tool select the optimal machine learning technique for a particular HVAC fault in terms of detection sensitivity, accuracy, and robustness. Besides, machine learning techniques can also be utilized to enhance and provide further support for some key features we have developed. For example, the cluster analysis can be used to provide more insights into the spatial and temporal partition strategy.

We shall expand the fault detection thresholds to include all major HVAC faults that lead to energy consumption inefficiency. This can be achieved by using data involving either identified real faults or artificially simulated faulty environment. For different faults that lead to similar abnormal system performance, a further fault isolation functionality, consisting of more than one machine learning techniques, is also needed in the future.

For some faults that cannot impose detectable abnormal system performance on the high level, a fault detection technique with extraordinary detection sensitivity is critical and indispensable for maintaining and enhancing the cross-level fault detection capability of the proposed FDD in the future.

We shall also complete the entire fault management lifecycle of the proposed FDD. Figure 7.1 shows the workflow of a complete fault management lifecycle of building HVAC systems. Specifically, we need to develop and add three modules to the existing FDD. First, we need to develop an automatic data acquisition and

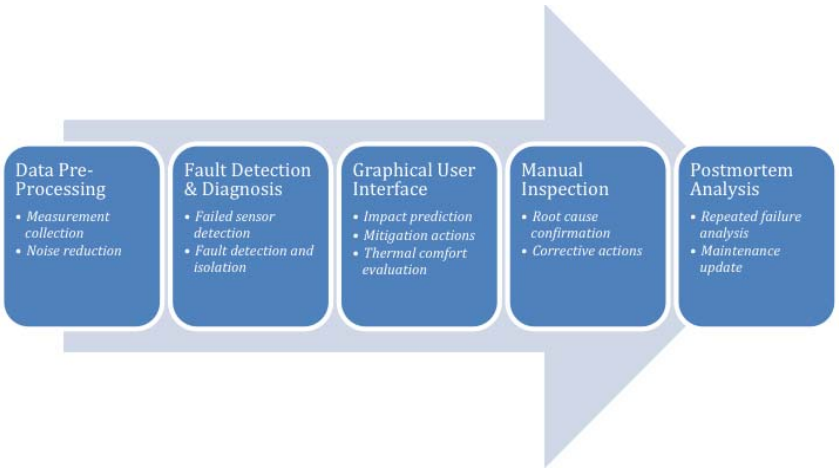

Figure 7.1: The complete workflow of the system-level fault management of building HVAC systems in the future.

pre-processing module. This module will be developed base on widely adopted industrial protocols, such as the Simple Object Access Protocol (SOAP), to automatically collect and store original sensor measurements over the net. The stored measurements will be updated intelligently by considering the past operational records and environmental conditions. Digital filters will be also designed to reduce the measurement noise. Second, we need to develop a graphical user interface (GUI) to present the FDD results in a user-friendly and informative manner. As shown in Figure 7.2, this GUI will describe the distribution of HVAC units, and use different colors to indicate the level of alarm, determined by the FDD algorithm running behind the GUI with the stored measurements. It will also provide other information for fault management including the fault impact prediction, recommended mitigation actions, and thermal comfort evaluation. The information would assist building operators to make better decision in dealing with faults.. Third, we need to develop a postmortem analysis module to identify the failure modes in the HVAC system after some faults have been detected. The postmortem analysis involves a systematic manual confirmation of the underlying physical failure and the root cause. Changes can then be made in the HVAC design, maintenance practice and even thresholds in the FDD algorithm.

We have applied the top-down FDD strategy to investigate faults in one AHU and the groups of VAVs connected to it. We need to scale up the study to handle as many AHUs and groups of VAVs as in a classroom or office building, or even the entire campus. Since the FDD algorithms can be executed in real time on the web, programming efforts will be focused on the development of web-based

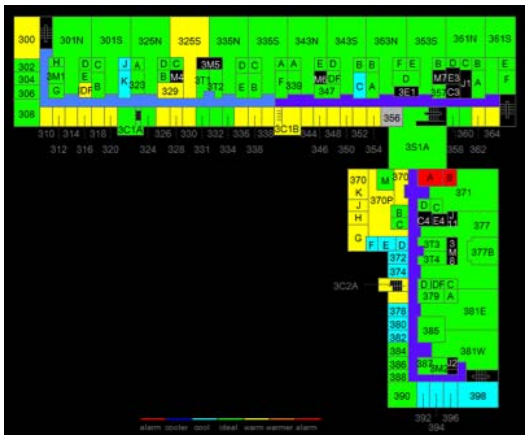

Figure 7.2: A sample GUI of the system-level FDD of building HVAC systems in the future.

software for computing and analyzing the data. Matlab Simulink, MapleSim running Modelica and National Instruments VI offer good options for combining the desktop computing power with the Internet connectivity. This part of the research to scale up the top-down FDD strategy will involve intensive programming effort in one of the three software environments.

7.2.3 Control Design and Optimization

The models we have developed link important indoor thermal properties to HVAC control variables. The simple linear structures of those models pave the way for practical control design and optimization. For example, we can use these models to form objective functions for a multi-objective control design problem, which maximizing the occupant comfort while minimizing energy consumption.

Intelligent lighting and shading plays an important roles in increasing energy efficiency in buildings. We shall explore this area and include its fault detection and diagnosis in our existing FDD. The Social Science and Management (SSM) Building at UC Merced has equipped with intelligent lighting and shading hardware, which provides the access for future research on this topics.

7.2.4 Commercialization

The proposed top-down FDD technology and its software implementation in the office building have a significant potential for commercialization. During the course of its future development, we shall file patents to protect the key technologies and shall license the technologies to private companies.

BIBLIOGRAPHY

[1] http://www.cacx.org/meetings/meetings/2012-08-09/NBI_CCC_Webinar_080912.pdf, Fault detection & diagnostics (FDD) rooftop HVAC units: Technology, code requirements, commissioning.

[2] http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=1.1.1, U.S. residential and commercial buildings total primary energy consumption (quadrillion btu and percent of total).

[3] http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=1.1.3, Buildings share of U.S. primary energy consumption (percent).

[4] http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=1.1.9, Buildings share of U.S. electricity consumption (percent).

[5] J. Schein, S. T. Bushby, N. S. Castro, J. M. House, A rule-based fault detection method for air handling units, Energy and Buildings 38 (12) (2006) 1485–1492.

[6] http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=1.1.8, Shares of U.S. buildings generic quad (percent) (1).

[7] H. Yoshida, S. Kumar, RARX algorithm based model development and application to real time data for on-line fault detection in VAV AHU units, IBPSA Building Simulation 99 (1999) 161–168.

[8] P. O. Fanger, Thermal Comfort, Danish Technical Press., Copenhagen, 1970.

[9] P. O. Fanger, A. K. Melikov, H. Hanzawa, J. Ring, Air turbulence and sensation of draught, Energy and Buildings 12 (1) (1988) 21–39.

[10] American Society of Heating Refrigerating and Air Conditioning Engineers (ASHARE), Thermal Environmental Conditions for Human Occupancy (ASHARE, Standard 55-1992).

[11] International Standards Organization (ISO), Moderate Thermal Environments: Determination of the PMV and PPD Indices and Specification of the Conditions for Thermal Comfort (ISO 7730:1994).

[12] F. Ascione, L. Bellia, A. Capozzoli, F. Minichiello, Energy saving strategies in air-conditioning for museums, Applied Thermal Engineering 29 (4) (2009) 676–686.

[13] C. A. Balaras, E. Dascalaki, A. Gaglia, HVAC and indoor thermal conditions in hospital operating rooms, Energy and Buildings 39 (4) (2007) 454–470.

[14] D. Y. et al., An integrated modeling tool for simultaneous analysis of thermal performance and indoor air quality in buildings, Building and Environment 43 (3) (2008) 287–293.

[15] S. Wei, Y. Sun, M. Li, W. Lin, D. Zhao, Y. Shi, H. Yang, Indoor thermal environment evaluations and parametric analyses in naturally ventilated buildings in dry season using a field survey and pmve-ppde model, Building and Environment 46 (6) (2011) 1275–1283.

[16] F. Engdahl, D. Johansson, Optimal supply air temperature with respect to energy use in a variable air volume system, Energy and Buildings 36 (3) (2004) 205–218.

[17] I. H. Yang, K. W. Kim, Prediction of the time of room air temperature descending for heating systems in buildings, Building and Environment 39 (1) (2004) 19–29.

[18] J. A. Orosa, A new modelling methodology to control HVAC systems, Expert Systems with Applications 38 (4) (2011) 4505–4513.

[19] J. Tanimoto, A. Hagishima, State transition probability for the markov model dealing with on/off cooling schedule in dwellings, Energy and Buildings 37 (3) (2005) 181–187.

[20] D. P. Wyon, Healthy buildings and their impact on productivity, in: Proceedings of Indoor Air, Vol. 6, Helsinki, Finland, 1993, pp. 153–161.

[21] N. Djongyang, R. Tchinda, D. Njomo, Thermal comfort: A review paper, Renewable and Sustainable Energy Reviews 14 (9) (2010) 2626–2640.

[22] http://www.euro.who.int/_data/assets/pdf_file/0008/97091/E89887.pdf, Housing, energy and thermal comfort: A review of 10 countries within the WHO European Region.

[23] http://www.epa.gov/greenbuilding/pubs/gbstats.pdf, Buildings and their impact on the environment: A statistical summary.

[24] D. Veronica, M. Galler, Fault detection and diagnostics for commercial heating, ventilating, and air-conditioning systems, Tech. rep., National Institute of Standards and Technology (2011).

[25] S. Wang, F. Xiao, Detection and diagnosis of AHU sensor faults using principal component analysis method, Energy Conversion and Management 45 (17) (2004) 2667–2686.

[26] J. Qin, S. Wang, A fault detection and diagnosis strategy of VAV air-conditioning systems for improved energy and control performances, Energy and Buildings 37 (10) (2005) 1035–1048.

[27] Y. H. Song, Y. Akashi, J. J. Yee, A development of easy-to-use tool for fault detection and diagnosis in building air-conditioning systems, Energy and Buildings 40 (2) (2008) 71–82.

[28] X. B. Yang, X. Q. Jin, Z. M. Du, Y. H. Zhu, A novel model-based fault detection method for temperature sensor using fractal correlation dimension, Building and Environment 46 (4) (2011) 970–979.

[29] S. H. Cho, H. C. Yang, M. Zaheer-uddin, B. C. Ahn, Transient pattern analysis for fault detection and diagnosis of HVAC systems, Energy Conversion and Management 46 (18-19) (2005) 3103–3116.

[30] Z. Du, X. Jin, X. Yang, A robot fault diagnostic tool for flow rate sensors in air dampers and VAV terminals, Energy and Buildings 41 (3) (2009) 279–286.

[31] Z. Du, X. Jin, Tolerant control for multiple faults of sensors in VAV systems, Energy Conversion and Management 48 (3) (2007) 764–777.

[32] J. Schein, S. T. Bushby, A hierarchical rule-based fault detection and diagnostic method for HVAC systems, HVAC and Research 12 (1) (2006) 111–125.

[33] C. Ghiaus, Fault diagnosis of air conditioning systems based on qualitative bond graph, Energy and Buildings 30 (3) (1999) 221–232.

[34] B. Chen, J. E. Braun, Simple rule-based methods for fault detection and diagnostics applied to packaged air conditioners, ASHRAE Winter Meeting (2001) 975–984.

[35] S. Wang, F. Xiao, AHU sensor fault diagnosis using principal component analysis method, Energy and Buildings 36 (2) (2004) 147–160.

[36] F. Xiao, S. Wang, X. Xu, G. Ge, An isolation enhanced PCA method with expert-based multivariate decoupling for sensor FDD in air-conditioning systems, Applied Thermal Engineering 29 (4) (2009) 712–722.

[37] F. Xiao, S. Wang, Progress and methodologies of lifecycle commissioning of HVAC systems to enhance building sustainability, Renewable and Sustainable Energy Reviews 13 (5) (2009) 1144–1149.

[38] P. Carling, P. Haves, Comparison of three fault detection methods based on field data of an air-handling unit, ASHRAE Transactions 108 PART 1 (2002) 904–921.

[39] L. K. Norford, J. A. Wright, R. A. Buswell, D. Luo, C. J. Klaassen, A. Suby, Demonstration of fault detection and diagnosis methods for air-handling units (ASHRAE 1020-RP), HVAC and R Research 8 (1) (2002) 41–71.

[40] S. M. Namburu, M. S. Azam, J. Luo, K. Choi, K. R. Pattipati, Data-driven modeling, fault diagnosis and optimal sensor selection for HVAC chillers, IEEE Transactions on Automation Science and Engineering 4 (3) (2007) 469–473.

[41] P. Haves, S. K. Khalsa, Model-based performance monitoring: Review of diagnostic methods and chiller case study, Proceedings ACEEE Summer Study on Energy Efficiency in Buildings 3 (2000) 3161–3171.

[42] N. S. Castro, G. Remington, Performance evaluation of a reciprocating chiller using experimental data and model predictions for fault detection and diagnosis, ASHRAE Transactions 108 Part 1 (2002) 889–903.

[43] H. Han, B. Gu, T. Wang, Z. R. Li, Important sensors for chiller fault detection and diagnosis (fdd) from the perspective of feature selection and machine learning, International Journal of Refrigeration 34 (2) (2011) 586–599.

[44] J. Navarro-Esbrl, E. Torrella, R. Cabello, A vapour compression chiller fault detection technique based on adaptative algorithms. application to on-line refrigerant leakage detection, International Journal of Refrigeration 29 (5) (2006) 716–723.

[45] M. B. Bailey, J. F. Kreider, Creating an automated chiller fault detection and diagnostics tool using a data fault library, ISA Transactions 42 (3) (2003) 485–495.

[46] J. U. R. Khan, B. A. Qureshi, S. M. Zubair, A comprehensive design and performance evaluation study of counter flow wet cooling towers, International Journal of Refrigeration 27 (8) (2004) 914–923.

[47] E. Weyer, G. Szederkenyi, K. Hangos, Grey box fault detection of heat exchangers, Control Engineering Practice 8 (2) (2000) 121–131.

[48] S. Wang, Z. Jiang, Valve fault detection and diagnosis based on CMAC neural networks, Energy and Buildings 36 (6) (2004) 599–610.

[49] S. Wang, Y. Chen, Sensor validation and reconstruction for building central chilling systems based on principal component analysis, Energy Conversion and Management 45 (5) (2004) 673–695.

[50] M. Wetter, Modelica-based modelling and simulation to support research and development in building energy and control systems, Journal of Building Performance Simulation 2 (2) (2009) 143–161.

[51] Taylor Engineering LLC, Case study: Automated fault detection and diagnostic software, Tech. rep., California Energy Commission Public Interest Energy Research (PIER) Program (2006).

[52] N. N. Bauman, J. C. Hail, Demonstration of the Whole Building Diagnostician for the federal building U.S. Courthouse at Milwaukee, Wisconsin, and for the University of Wisconsin at Madison, Tech. Rep. PNNL-14515, US DoE (2003).

[53] J. Schein, S. T. Bushby, Fault detection diagnostics for AHUs and VAV boxes, ASHRAE Journal 47 (7) (2005) 58–63.

[54] J. R. Quinlan, Induction of decision trees, Machine Learning 1 (1986) 81–106.

[55] J. A. Clarke, J. Cockroft, S. Conner, J. W. Hand, N. J. Kelly, R. Moore, T. O'Brien, P. Strachan, Simulation-assisted control in building energy management systems, Energy and Buildings 34 (9) (2002) 933–940.

[56] N. Djuric, V. Novakovic, Review of possibilities and necessities for building lifetime commissioning, Renewable and Sustainable Energy Reviews 13 (2) (2009) 486–492.

[57] N. Djuric, V. Novakovic, F. Frydenlund, Heating system performance estimation using optimization tool and bems data, Energy and Buildings 40 (8) (2008) 1367–1376.

[58] L. Song, M. Liu, D. E. Claridge, P. Haves, Study of on-line simulation for whole building level energy consumption fault detection and optimization, in: Proceedings of Architectural Engineering, Building Integration Solutions, Austin, Texas, United States, 2003, pp. 76–83.

[59] A. Pedrini, F. S. Westphal, R. Lamberts, A methodology for building energy modelling and calibration in warm climates, Building and Environment 37 (8-9) (2002) 903–912.

[60] G. Lowry, M. W. Lee, Modelling the passive thermal response of a building using sparse bms data, Applied Energy 78 (1) (2004) 53–62.

[61] H. U. Frausto, J. G. Pieters, J. M. Deltour, Modelling greenhouse temperature by means of auto regressive models, Biosystems Engineering 84 (2) (2003) 147–157.

[62] G. J. Ríos-Moreno, M. Trejo-Perea, R. Castañeda-Miranda, V. M. Hernández-Guzmán, G. Herrera-Ruiz, Modelling temperature in intelligent buildings by means of autoregressive models, Automation in Construction 16 (5) (2007) 713–722.

[63] J. B. Cunha, C. Couto, A. E. Ruano, Real-time parameter estimation of dynamic temperature models for greenhouse environmental control, Control Engineering Practice 5 (10) (1997) 1473–1481.

[64] G. Mustafaraj, J. Chen, G. Lowry, Development of room temperature and relative humidity linear parametric models for an open office using BMS data, Energy and Buildings 42 (3) (2010) 348–356.

[65] D. L. Loveday, C. Craggs, Stochastic modeling of temperatures for a full-scale occupied building zone subject to natural random influences, Applied Energy 45 (4) (1993) 295–312.

[66] J. C. M. Yiu, S. Wang, Multiple ARMAX modeling scheme for forecasting air conditioning system performance, Energy Conversion and Management 48 (8) (2007) 2276–2285.

[67] H. C. Peitsman, V. E. Bakker, Application of black-box models to HVAC systems for fault detection, ASHRAE Transactions 102 (1) (1996) 628–640.

[68] H. Yoshida, S. Kumar, Development of ARX model based off-line FDD technique for energy efficient buildings, Renewable Energy 22 (1-3) (2001) 53–59.

[69] S. Kumar, S. Sinha, T. Kojima, H. Yoshida, Development of parameter based fault detection and diagnosis technique for energy efficient building management system, Energy Conversion and Management 42 (7) (2001) 833–854.

[70] M. J. Jiménez, H. Madsen, K. K. Andersen, Identification of the main thermal characteristics of building components using MATLAB, Building and Environment 43 (2) (2008) 170–180.

[71] C. Ghiaus, A. Chicinas, C. Inard, Grey-box identification of air-handling unit elements, Control Engineering Practice 15 (2007) 421–433.

[72] J. E. Braun, N. Chaturvedi, An inverse gray-box model for transient building load prediction, HVAC and R Research 8 (2002) 73–99.

[73] S. Wang, X. Xu, Simplified building model for transient thermal performance estimation using GA-based parameter identification, International Journal of Thermal Sciences 45 (4) (2006) 419–432.

[74] Q. Zhou, S. Wang, X. Xu, F. Xiao, A grey-box model of next-day building thermal load prediction for energy-efficient control, International Journal of Energy Research 32 (15) (2008) 1418–1431.

[75] K. Zheng, H. Li, Development and validation of a fan performance model for typical packaged HVAC systems, in: Proceedings of 2010 ASHRAE Winter Conference, January 23, 2010 - January 27, 2010, Vol. 116 PART 1, Orlando, Florida, United states, 2010, pp. 323–329.

[76] F. Déqué, F. Ollivier, A. Poblador, Grey boxes used to represent buildings with a minimum number of geometric and thermal parameters, Energy and Buildings 31 (1) (2000) 29–35.

[77] J. Wen, T. F. Smith, Development and validation of online models with parameter estimation for a building zone with VAV system, Energy and Buildings 39 (1) (2007) 13–22.

[78] T. Leephakpreeda, Grey prediction on indoor comfort temperature for HVAC systems, Expert Systems with Applications 34 (4) (2008) 2284–2289.

[79] P. T. Tsilingiris, Wall heat loss from intermittently conditioned spaces–the dynamic influence of structural and operational parameters, Energy and Buildings 38 (8) (2006) 1022–1031.

[80] S. Devgan, A. K. Jain, B. Bhattacharjee, Predetermined overall thermal transfer value coefficients for composite, hot-dry and warm-humid climates, Energy and Buildings 42 (10) (2010) 1841–1861.

[81] N. Aste, A. Angelotti, M. Buzzetti, The influence of the external walls thermal inertia on the energy performance of well insulated buildings, Energy and Buildings 41 (11) (2009) 1181–1187.

[82] R. L. Hwang, S. Y. Shu, Building envelope regulations on thermal comfort in glass facade buildings and energy-saving potential for PMV-based comfort control, Building and Environment 46 (4) (2011) 824–834.

[83] H. Sozer, Improving energy efficiency through the design of the building envelope, Building and Environment 45 (12) (2010) 2581–2593.

[84] I. Korolija, L. Marjanovic-Halburd, Y. Zhang, V. I. Hanby, Influence of building parameters and HVAC systems coupling on building energy performance, Energy and Buildings 43 (6) (2011) 1247–1253.

[85] S. L. Patil, H. J. Tantau, V. M. Salokhe, Modelling of tropical greenhouse temperature by auto regressive and neural network models, Biosystems Engineering 99 (3) (2008) 423–431.

[86] S. Max, A simplified model of thermal comfort, Energy and Buildings 8 (1) (1985) 37–50.

[87] S. Atthajariyakul, T. Leephakpreeda, Neural computing thermal comfort index for HVAC systems, Energy Conversion and Management 46 (15, 16) (2005) 2553–2565.

[88] P. O. Fanger, J. Toftum, Extension of the PMV model to non-air-conditioned buildings in warm climates, Energy and Buildings 34 (6) (2002) 533–536.

[89] M. R. Kulkarni, F. Hong, An experimental technique for thermal comfort comparison in a transient pull down, Building and Environment 39 (2) (2004) 189–193.

[90] G. S. Brager, R. J. de Dear, Thermal adaptation in the built environment: A literature review, Energy and Buildings 27 (1) (1998) 83–96.

[91] R. Z. Homod, K. S. M. Sahari, H. A. F. Almurib, F. H. Nagi, RLF and TS fuzzy model identification of indoor thermal comfort based on PMV/PPD, Building and Environment 49 (2012) 141–153.

[92] M. Hamdi, G. Lachiver, F. Michaud, A new predictive thermal sensation index of human response, Energy and Buildings 29 (2) (1999) 167–178.

[93] Fundamentals: ASHRAE handbook.

[94] S. Atthajariyakul, T. Leephakpreeda, Real-time determination of optimal indoor-air condition for thermal comfort, air quality and efficient energy usage, Energy and Buildings 36 (7) (2004) 720–733.

[95] K. Chen, Y. Jiao, E. S. Lee, Fuzzy adaptive networks in thermal comfort, Applied Mathematics Letters 19 (5) (2006) 420–426.

[96] L. Wang, H. W. Nyuk, S. Li, Facade design optimization for naturally ventilated residential buildings in Singapore, Energy and Buildings 39 (8) (2007) 954–961.

[97] G. M. Stavrakakis, P. L. Zervas, H. Sarimveis, N. C. Markatos, Development of a computational tool to quantify architectural-design effects on thermal comfort in naturally ventilated rural houses, Building and Environment 45 (1) (2010) 65–80.

[98] G. M. Stavrakakis, D. P. Karadimou, P. L. Zervas, H. Sarimveis, N. C. Markatos, Selection of window sizes for optimizing occupational comfort and hygiene based on computational fluid dynamics and neural networks, Building and Environment 46 (2) (2011) 298–314.

[99] G. M. Stavrakakis, P. L. Zervas, H. Sarimveis, N. C. Markatos, Optimization of window-openings design for thermal comfort in naturally ventilated buildings, Applied Mathematical Modelling 36 (1) (2012) 193–211.

[100] J. Yu, C. Yang, L. Tian, D. Liao, Evaluation on energy and thermal performance for residential envelopes in hot summer and cold winter zone of China, Applied Energy 86 (10) (2009) 1970–1985.

[101] N. Ghaddar, K. Ghali, S. Chehaitly, Assessing thermal comfort of active people in transitional spaces in presence of air movement, Energy and Buildings 43 (10) (2011) 2832–2842.

[102] E. Prianto, P. Depecker, Optimization of architectural design elements in tropical humid region with thermal comfort approach, Energy and Buildings 35 (3) (2003) 273–280.

[103] S. H. Ho, L. Rosario, M. M. Rahman, Thermal comfort enhancement by using a ceiling fan, Applied Thermal Engineering 29 (8-9) (2009) 1648–1656.

[104] S. Wu, J. Q. Sun, A top-down strategy with temporal and spatial partition for fault detection and diagnosis of building HVAC systems, Energy and Buildings 43 (9) (2011) 2134–2139.

[105] X. Wang, J. Eisenbrey, M. Zeitz, J. Q. Sun, Multi-stage regression analysis of acoustical properties of polyurethane foams, Sound and Vibration 273 (4-5) (2004) 1109–1117.

[106] X. Wang, J. Q. Sun, On the fatigue analysis of non-gaussian stress processes with asymmetric distribution, Vibration and Acoustics 127 (6) (2005) 556–566.

[107] S. Wu, J. Q. Sun, A physics-based linear parametric model of room temperature in office buildings, Building and Environment 50 (2012) 1–9, DOI: 10.1016/j.buildenv.2011.10.005.

[108] R. B. Bird, W. E. Stewart, E. N. Lightfoot, Transport phenomena.

[109] U. Norlén, Determining the thermal resistance from in situ measurements, Workshop on Application of System Identification in Energy Savings in Building, EUR 15566 EN, Bloem, ed., Joint Research Center, 1993.

[110] J. C. Bakker, G. P. A. Bot, H. Challa, N. J. V. de Braak, Greenhouse climate control. An integrated approach., Wageningen Pers, Wageningen, 1995.

[111] R. Baierlein, Thermal Physics, Cambridge University Press, Cambridge, England, 2003.

[112] P. Perrot, A to Z of Thermodynamics, Oxford University Press, Oxford, England, 1998.

[113] T. Baguley, Beyond ANOVA: From repeated measures to multilevel models, PsyPAG Mathematics Statistics and Computing Workshop.

[114] J. Ahn, Beyond single equation regression analysis: Path analysis and multistage regression analysis, American Journal of Pharmaceutical Education 66 (2002) 37–42.

[115] S. D. Kristjansson, J. C. Kircher, A. K. Webb, Multilevel models for repeated measures research designs in psychophysiology: An introduction to growth curve modeling, Psychophysiology 44 (5) (2007) 728–736.

[116] J. M. Bland, D. G. Altman, Statistics notes: Correlations, regression, and repeated data, British Medical Journal 308 (1994) 896.

[117] K. V. Mardia, J. T. Kent, J. M. Bibby, Multivariate Analysis (Probability and Mathematical Statistics), Academic Press, Waltham, Massachusetts, 1980.

[118] S. E. Maxwell, H. D. Delaney, Designing Experiments and Analyzing Data: A Model Comparison Perspective, Routledge Academic, London, 2003.

[119] C. M. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1995.

[120] C. D. Schunn, D. Wallach, Evaluating goodness-of-fit in comparison of models to data, in: In W. Tack (ed.), Psychologie der Kognition: Reden and vorträge anlässlich der emeritierung von Werner Tack, University of Saarland Press, Saarbrueken, Germany (2005) 115–154.

[121] P. Mazzei, F. Minichiello, D. Palma, HVAC dehumidification systems for thermal comfort: A critical review, Applied Thermal Engineering 25 (5-6) (2005) 677–707.

[122] Fundamentals: ASHRAE Handbook, American Society of Heating Refrigerating and Air Conditioning Engineers, Atlanta, Georgia, 1997.

[123] W. Hyland, A. Wexler, Formulations for the thermodynamic properties of the saturated phases of H2O from 173.15 K to 473.15 K, in: Proceedings of ASHRAE Transactions, Vol. 89, 1983, pp. 500–519.

[124] W. Hyland, A. Wexler, Formulations for the thermodynamic properties of dry air from 173.15 K to 473.15 K, and of saturated moist air from 173.15 K to 372.15 K, in: Proceedings of ASHRAE Transactions, Vol. 89, 1983, pp. 520–535.

[125] D. Sonntag, Important new values of the physical constants of 1986, vapour pressure formulations based on the ITS-90, and psychrometer formulae, Z. Meteorol. 70 (5) (1990) 340–344.

[126] C. Moler, Numerical Computing with MATLAB, The Society for Industrial and Applied Mathematics, Philadelphia, 2004.

[127] T. W. Anderson, The Statistical Analysis of Time Series, Wiley-Interscience, 1994.

[128] D. R. Brillinger, Time Series: Data Analysis and Theory, Society for Industrial and Applied Mathematics, 2001.

[129] M. Norgaard, O. Ravn, N. K. Poulsen, L. K. Hansen, Neural Networks for Modeling and Control of Dynamic Systems, Springer-Verlag, 2000.

[130] U. Norlén, Estimating thermal parameters of outdoor test cells, Building and Environment 25 (1) (1990) 17–24.

[131] http://www.cimis.water.ca.gov/cimis/hourlyReport.do, Merced hourly weather data from california irrigation management information system.

[132] N. C. Harris, C. E. Miller, I. E. Thomas, Solar Energy Systems Design, John Wiley and Sons Inc, New York, 1985.

[133] S. Wu, J. Q. Sun, Cross-level fault detection and diagnosis of building HVAC systems, Building and Environment 46 (8) (2011) 1558–1566.

[134] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global Sensitivity Analysis: The Primer, Wiley, 2008.

[135] D. M. Hamby, A review of techniques for parameter sensitivity analysis of environmental models, Environmental Monitoring and Assessment 32 (2) (1994) 135–154.

[136] J. P. C. Kleijnen, Sensitivity analysis and related analyses: A review of some statistical techniques, Journal of Statistical Computation and Simulation 57 (1997) 111–142.

[137] K. J. Lomas, H. Eppel, Sensitivity analysis techniques for building thermal simulation programs, Energy and Buildings 19 (1) (1992) 21–44.

[138] D. Balagangadhar, S. Roy, Design sensitivity analysis and optimization of steady fluid-thermal systems, Computer Methods in Applied Mechanics and Engineering 190 (42) (2001) 5465–5479.

[139] B. Eisenhower, Z. O. Neill, V. A. Fonoberov, I. Mezić, Uncertainty and sensitivity decomposition of building energy models, Building Performance Simulation (2011) 1–14.

[140] S. Burhenne, M. Elci, D. Jacob, C. Neumann, S. Herkel, Sensitivity analysis with building simulations to support the commissioning process, in: Proceedings of 10th International Conference for Enhanced Building Operations, Kuwait City, 2010.

[141] V. Corrado, H. E. Mechri, Uncertainty and sensitivity analysis for building energy rating, Building Physics 33 (2) (2009) 125–156.

[142] S. de Wit, G. Augenbroe, Analysis of uncertainty in building design evaluations and its implications, Energy and Buildings 34 (9) (2002) 951–958.

[143] C. J. Hopfe, J. L. M. Hensen, Uncertainty analysis in building performance simulation for design support, Energy and Buildings 43 (10) (2011) 2798–2805.

[144] T. Mara, S. Tarantola, Application of global sensitivity analysis of model output to building thermal simulations, Building Simulation 1 (4) (2008) 290–302.

[145] P. Heiselberg, H. Brohus, A. Hesselholt, H. Rasmussen, E. Seinre, S. Thomas, Application of sensitivity analysis in design of sustainable buildings, Renewable Energy 34 (9) (2009) 2030–2036.

[146] D. G. Cacuci, Sensitivity & Uncertainty Analysis, Volume 1: Theory, Chapman and Hall, London, 2003.

[147] B. Ingalls, Sensitivity analysis: From model parameters to system behaviour, Essays in Biochemistry 45 (1) (2008) 177–194.

[148] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, Wiley, 2004.

[149] R. H. Gardner, D. D. Huff, R. V. O'Neill, J. B. Mankin, J. Carney, J. Jones, Application of error analysis to a marsh hydrology model, Water Resources Research 16 (4) (1980) 659–664.

[150] D. J. Downing, R. H. Gardner, F. O. Hoffman, An examination of response-surface methodologies for uncertainty analysis in assessment models, Technometrics 27 (2) (1985) 151–163.

[151] C. Daniel, One-at-a-time plans, Journal of the American Statistical Association 68 (1973) 353–360.

[152] S. Wu, J. Q. Sun, Two-stage regression model of thermal comfort in office buildings, Building and Environment 57 (2012) 88–96.

[153] P. M. Frank, Introduction to System Sensitivity Theory, Academic Press Inc, New York, 1978.

[154] V. Hayward, J. M. Cruz-Hernández, Parameter sensitivity analysis for design and control of force transmission systems, ASME Journal of Dynamics Systems, Meas. and Contr. 120 (2) (1998) 241–249.

[155] S. S. Shapiro, M. B. Wilk, An analysis of variance test for normality (complete samples), Biometrika 52 (3-4) (1965) 591–611.

[156] D. Bau III, L. N. Trefethen, Numerical Linear Algebra, Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[157] G. Strang, Introduction to Linear Algebra 3rd Edition, Wellesley-Cambridge Press, 1998.

[158] R. A. Horn, C. R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, U.K., 1985.

[159] G. U. Yule, An Introduction to the Theory of Statistics, Macmillan Pub Co, 1969.

Appendix A

NOMENCLATURE

$\Delta E_{h,d}$ Internal heat generated by human and devices (W)
α humidity ratio (%)
ρ Air density (kgm^{-3})
ζ Window-to-wall ratio (%)
ϕ Heat gain from solar flux (W)
h Heat convection coefficient $(Wm^{-2}K^{-1})$
k Thermal conductivity coefficient $(Wm^{-1}K^{-1})$
$\dot{m}$ Mass flow rate (kgs^{-1})
n Time sequence of measurements
t Time
flt Pressure drop of moisture filter (Pa)
AHU Air handling unit
C_v Volumetric heat capacity of air $(kJkg^{-1}K^{-1})$
D Damper position (%)
H Specific enthalpy $(kJkg^{-1})$
L Thickness of wall or window (m)
P Parameter group
R Relative humidity (%)
S Surface area exposed to sun (m^2)
T Temperature (K)
V Volume of room (m^3)
X An attribute
VAV Variable air volume unit
Subscripts
i inside
o outside
ea exhaust air
ma mixture air
oa outside air
ra return air
rm room
wa wall

wd window
clg cooling water
htg heating water
fin final
pre preliminary
$surf$ surface
$disch$ discharge
Superscript
$\hat{}$estimated values

Appendix B

MODEL PARAMETERS

B.1 Physics-Based and Multi-Stage Regression Temperature Models

The parameters involved in the pbARMAX and mpbARMAX models are defined as follows.

$$P_1 = \frac{h_{wa_i} L_{wa}}{k_{wa} + h_{wa_i} L_{wa}} \left(1 + \frac{k_{wa}^2}{k_{wa} L_{wa} (h_{wa_i} + h_{wa_o}) + L_{wa}^2 h_{wa_i} h_{wa_o}} \right), \tag{B.1}$$

$$P_2 = \frac{k_{wa} h_{wa_o} L_{wa}}{k_{wa} L_{wa} (h_{wa_i} + h_{wa_o}) + L_{wa}^2 h_{wa_i} h_{wa_o}}, \tag{B.2}$$

$$P_3 = \frac{h_{wd_i} L_{wd}}{k_{wd} + h_{wd_i} L_{wd}} \left(1 + \frac{k_{wd}^2}{k_{wd} L_{wd} (h_{wd_i} + h_{wd_o}) + L_{wd}^2 h_{wd_i} h_{wd_o}} \right), \tag{B.3}$$

$$P_4 = \frac{k_{wd} h_{wd_o} L_{wd}}{k_{wd} L_{wd} (h_{wd_i} + h_{wd_o}) + L_{wd}^2 h_{wd_i} h_{wd_o}}, \tag{B.4}$$

$$P_5 = \frac{1}{\rho_i V C_v}, \tag{B.5}$$

$$P_{rm} = P_5 (S_{wa} (h_{wa_i} (P_1 - 1) + h_{wd_i} \zeta (P_3 - 1)) - \dot{m} C_p), \tag{B.6}$$

$$P_{disch} = P_5 \dot{m} C_p, \tag{B.7}$$

$$P_{oa} = P_5 S_{wa} (h_{wa_i} P_2 + h_{wd_i} \zeta P_4), \tag{B.8}$$

$$\Delta E = P_5 (\phi + \Delta E_{h,d}). \tag{B.9}$$

The polynomials of the transfer functions in the ARMAX_{1113} model are given by

$$A(q) = 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_{n_a} q^{-n_a}, \tag{B.10}$$

$$B_i(q) = 1 + b_{1,i} q^{-1} + b_{2,i} q^{-2} + \cdots + b_{n_b,i} q^{-n_b}, \ i = 1, 2, \cdots, 7 \tag{B.11}$$

$$C(q) = 1 + c_1 q^{-1} + c_2 q^{-2} + \cdots + a_{n_c} q^{-n_c}, \tag{B.12}$$

where n_a, n_b, and n_c are the orders of the polynomials in q.

B.2 Parametric PMV Model

The parameters involved in the epPMV model are defined as follows:

$$P_1 = \frac{h_{wa_i} L_{wa}}{k_{wa} + h_{wa_i} L_{wa}} \left(1 + \frac{k_{wa}^2}{k_{wa} L_{wa}(h_{wa_i} + h_{wa_o}) + L_{wa}^2 h_{wa_i} h_{wa_o}}\right), \tag{B.13}$$

$$P_2 = \frac{k_{wa} h_{wa_o}}{k_{wa}(h_{wa_i} + h_{wa_o}) + L_{wa} h_{wa_i} h_{wa_o}}, \tag{B.14}$$

$$P_3 = \frac{h_{wd_i} L_{wd}}{k_{wd} + h_{wd_i} L_{wd}} \left(1 + \frac{k_{wd}^2}{k_{wd} L_{wd}(h_{wd_i} + h_{wd_o}) + L_{wd}^2 h_{wd_i} h_{wd_o}}\right), \tag{B.15}$$

$$P_4 = \frac{k_{wd} h_{wd_o}}{k_{wd}(h_{wd_i} + h_{wd_o}) + L_{wd} h_{wd_i} h_{wd_o}}, \tag{B.16}$$

$$P_5 = \frac{1}{\rho_i V C_v}, \tag{B.17}$$

$$P_{rm} = P_5\{S_{wa}[h_{wa_i}(P_1 - 1) + h_{wd_i}\zeta(P_3 - 1)] - \dot{m} C_{pa} + 0.62198 C_{pw}\}, \tag{B.18}$$

$$P_{disch} = P_5(\dot{m} C_{pa} - 0.62198 C_{pw}), \tag{B.19}$$

$$P_{oa} = P_5 S_{wa}(h_{wa_i} P_2 + h_{wd_i}\zeta P_4), \tag{B.20}$$

$$P_f = 0.62198 p_a P_5, \tag{B.21}$$

$$F(T, R) = \frac{C_{pw} T + H_{we}}{p_a - p_w}, \tag{B.22}$$

$$F_1(T, R) = \frac{C_{pw}}{\beta_1(p_a - p_w)}, \tag{B.23}$$

$$F_2(T, R) = \frac{C_{pw}\frac{\beta_2}{\beta_1} p_w - H_{we} - \frac{C_{pw}}{\beta_1}\beta_3}{p_a - p_w}, \tag{B.24}$$

$$\Delta E = P_5(\phi + \Delta E_{h,d}), \tag{B.25}$$

$$a = P_{oa}, \tag{B.26}$$

$$b = 0.62198 C_{pw} P_5, \tag{B.27}$$

$$c = P_5 C_{pa}, \tag{B.28}$$

$$d = P_f. \tag{B.29}$$