MACHINE LEARNING FOR HUMAN-CENTERED AND VALUE-SENSITIVE BUILDING ENERGY EFFICIENCY

BY

KADIR AMASYALI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Doctoral Committee:

Associate Professor Nora El-Gohary, Chair
Professor Khaled El-Rayes
Associate Professor Liang Y. Liu
Associate Professor Mani Golparvar-Fard
Professor Chimay J. Anumba, University of Florida

ABSTRACT

Enhancing building energy efficiency is one of the best strategies to reduce energy consumption and associated CO_2 emissions. Recent studies emphasized the importance of occupant behavior as a key means of enhancing building energy efficiency. However, it is also critical that while we strive to enhance the energy efficiency of buildings through improving occupant behavior, we still pay enough attention to occupant comfort and satisfaction.

Towards this goal, this research proposes a data-driven machine-learning-based approach to behavioral building energy efficiency, which could help better understand and predict the impact of occupant behavior on building energy consumption and occupant comfort; and help optimize occupant behavior for both energy saving and occupant comfort. Three types of models were developed and tested – simulation-data-driven, real-data-driven, and hybrid.

Accordingly, the research included five primary research tasks. First, the importance levels of energy-related human values (e.g., thermal comfort) to building occupants and their current satisfaction levels with these values were identified, in order to better understand the factors that are associated with higher/lower importance and/or satisfaction levels and identify the potential factors that could help predict occupant comfort. Second, a data sensing and occupant feedback collection plan was developed, in order to capture and monitor the indoor environmental conditions, energy consumption, energy-related occupant behavior, and occupant comfort in real buildings. Third, a set of buildings were simulated, in order to model the energy consumption of different buildings in different contexts – in terms of occupant behavior, building sizes, weather conditions, etc.; and a simulation-data-driven occupant-behavior-sensitive machine learning-based model, which learns from simulation data, was developed for predicting hourly cooling energy consumption. Fourth, a set of real-data-driven occupant-behavior-sensitive machine learning-

based models, which learn from real data (data collected from real buildings and real occupants), were developed for predicting hourly cooling and lighting energy consumption and thermal and visual occupant comfort; and a genetic algorithm-based optimization model for determining the optimal occupant behavior that can simultaneously reduce energy consumption and improve occupant comfort was developed. Compared to the simulation-data-driven approach, the real-data-driven approach aims to better capture and model the real-life behavior and comfort of occupants and the real-life energy-consumption patterns of buildings. Although successful in this regard, the resulting models may not generalize well outside of their training range. Fifth, a hybrid, occupant-behavior-sensitive machine learning-based model, which learns from both simulation data and real data, was developed for predicting hourly cooling and lighting energy consumption. The hybrid approach aims to overcome the limitations of both simulation-data-driven and real-data-driven approaches – especially the limited ability to capture occupant behavior and real-life consumption patterns in simulation-data-driven approaches and the limited generalizability of real-data-driven approaches to different cases – by learning from both types of data simultaneously.

The experimental results show the potential of the proposed approach. The energy consumption prediction models achieved high prediction performance, and the thermal and visual comfort models were able to accurately represent the individual and group comfort levels. The optimization results showed potential behavioral energy savings in the range of 11% and 22%, with significant improvement in occupant comfort.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest gratitude to my advisor, Prof. Nora El-Gohary, for her continuous guidance, encouragement, and support during my Ph.D. study. I would also like to deeply thank my Doctoral Committee Members – Prof. Chimay Anumba, Prof. Khaled El-Rayes, Prof. Liang Y. Liu, and Prof. Mani Golparvar-Fard – for their insightful comments and advice.

I thank my research groupmates, Jiansong Zhang, Lu Zhang, Xuan Lv, Peng Zhou, Marwan Ammar, Kaijian Liu, Lufan Wang, Ruichuan Zhang, Peter Liu, Nidia Bucarelli, and Xiyu Wang, for our stimulating discussions and their inspiring research work, and for all the fun we had. I also thank all the rest of my friends and colleagues at the University of Illinois.

Many thanks to my dear friends, Alperen Gunay, Ebru Toprak, Emir Ruzgar, Gizem Tabak, Nuri Ersahin, Okan Ilhan, Ozgun Alp Numanoglu, Sevket Sureyya Taskan, and Yusuf Akemoglu for all the great time we spent together. A big smile grows on my face when I think about all the memories that we collected together.

I would like to thank my parents, Esra and Ugur, my brother, Kerem, and the rest of family for their love, encouragement, and support. Without their support, I could not have been standing where I am now.

I would also like to thank the Philadelphia Business and Technology Center (PBTC) and the Penn State Consortium for Building Energy Innovation (CBEI) for providing access to their building energy data; Prof. Chimay Anumba, Dean and Professor, University of Florida and Prof. Yewande Abraham, Assistant Professor, Rochester Institute of Technology for their collaboration in the data collection; and the occupants of the PBTC for the occupant feedback they provided. Finally, I gratefully acknowledge the financial support from the Qatar National Research Fund (a member of Qatar Foundation), under NPRP Grant #6-1370-2-552.

TABLE OF CONTENTS

CHAPTER 1 – INTRODUCTION

1.1 Introduction and Motivation

Globally, the building sector is responsible for 40% of the total delivered energy consumption and one-third of the total CO_2 emissions (Yang et al. 2014), with a major role in global warming (Ascione et al. 2017a). The recent growth in energy demand comes at a time of increasing global concern over carbon emissions and resulting global climate change. The production and consumption of non-renewable energy, including oil and natural gas, pose adverse environmental impacts on the ecosystem in terms of air pollution and global warming. With its high energy consumption and growing trend, the building sector presents the most significant energy cost saving and CO_2 emission reduction opportunity. Improving building energy efficiency is one of the best strategies to reduce the environmental and economic impacts of energy consumption (Becerik-Gerber et al. 2014). For example, the International Energy Agency (IEA) calls on all countries to pay more attention to building energy efficiency in order to achieve the target of low carbon emissions by 2050 (IEA 2013). IEA estimates a 50% increase in the energy demand caused primarily by buildings by 2050, and highlights that this increase can be capped to 10% without any sacrifice in the comfort of building occupants, if necessary improvements in energy efficiency can be achieved (IEA 2013).

Recent studies emphasized the importance of occupant behavior as key means of enhancing building energy efficiency. Occupant behavior is one of the most significant factors that affect building energy consumption. A significant amount of energy can be saved through improving occupant behavior. Many studies demonstrated the impact of occupant behavior on building energy consumption. For example, Parker et al. (2012) showed that the energy consumed by identical buildings can vary by as much as three-fold and emphasized the role of occupant behavior in such differences. Hong and Lin (2014) simulated three identical offices with different occupant behaviors and demonstrated that an austerity workstyle consumed up to 50% less energy while a wasteful workstyle consumed up to 90% more energy, compared to the base model. Bonte et al. (2014) investigated the impact of some occupant behavior (e.g., blinds, lights, windows, temperature setpoints, fans, and clothing) on building energy consumption and occupant thermal sensations through simulating eight buildings that represent combinations of different climates (Oceanic and Mediterranean), thermal inertia levels (high and low), and air conditioning states (on and off). The results showed that conventional design strategies undervalue building energy consumption and overvalue occupant thermal sensation because of neglecting occupant behavior.

Human values are an important element that affect occupant behavior. “Values influence behavior because people emulate the conduct they hold valuable” (Boundless 2015). On the other hand, people spend the majority of their time in residential and commercial buildings, and therefore it is essential that while we aim to reduce building energy consumption that we also satisfy their values; it is important that we maintain comfortable, healthy, and productive indoor environmental conditions to the occupants of the building (Frontczak and Wargocki 2011) and it also important to consider other values such as cost saving. It is, thus, critical that while we strive to improve the energy efficiency of buildings through the understanding of energy-use behavior that we also understand the values [such as thermal comfort, indoor air quality (IAQ), productivity] of building occupants, how these values may impact energy-use behavior, and how we can improve energy efficiency without negatively impacting these values (e.g., while maintaining the satisfaction levels with these values).

Despite the significance of occupant behavior in enhancing the energy efficiency of buildings, major knowledge gaps exist in this area. First, there is a lack of understanding of how important these human values are to residential and office building occupants, what their satisfaction levels with these values are, and how these importance and satisfaction levels vary across different types of occupants, different states, and different energy-related factors. Second, there is a lack of building energy consumption prediction and simulation models that take energy-use behavior into account. Third, there is a lack of studies on capturing real energy-use behavior, real energy-consumption patterns, and real occupant-comfort information to better understand the impact of occupant behavior on building energy consumption and occupant comfort; and help optimize occupant behavior for both energy saving and occupant comfort. Fourth, there is a lack of studies on comparing simulation-data-driven and real-data-driven energy consumption prediction approaches and studying how to leverage their strengths and reduce their limitations. On one hand, real-data-driven approaches are suitable for capturing nonlinearity and patterns in complex systems, which makes them good candidates for capturing real energy consumption patterns that could be highly variable and complex due to occupant behavior. But, they require a large amount of data to train a model well and the resulting model often does not generalize well outside of the training range (Foucquier et al. 2013). On the other hand, simulation-based models are still limited in representing the complexity and stochastic nature of occupant behavior and comfort.

To address these gaps, this thesis proposes a human-centered, data-driven, and machine-learning-based approach to behavioral building energy efficiency. The proposed thesis work involves six primary research tasks: (1) conducting a comprehensive literature review; (2) identifying the importance levels of the human values (that are related to energy-use behavior and energy consumption) to residential and office building occupants and their current satisfaction levels with these values, the factors that are associated with higher/lower importance and/or satisfaction levels, and the potential factors that could help predict occupant satisfaction levels; (3) developing a data sensing and occupant feedback collection plan for capturing and monitoring indoor environmental conditions, energy consumption (heating/cooling, lighting, and plug loads), energy-related occupant behavior, and occupant comfort in real buildings; and conducting the data collection; (4) modeling and simulating a set of buildings with different sizes and occupant behaviors in different weather conditions; and developing an occupant-behavior-sensitive machine learning-based model, which learns from simulation data, for predicting hourly cooling energy consumption; (5) developing occupant-behavior-sensitive machine learning-based models, which learn from real data (data collected from real buildings and real occupants), for predicting hourly cooling and lighting energy consumption and thermal and visual occupant comfort; and developing a genetic algorithm-based optimization model for determining the optimal occupant behavior that can simultaneously reduce energy consumption and improve occupant comfort; and (6) developing a hybrid, occupant-behavior-sensitive machine learning-based model, which learns from both simulation data (data created through simulating a set of reference buildings) and real data (data collected from real buildings and real occupants), for predicting hourly cooling energy consumption. The hybrid approach aims to overcome the limitations of both simulation-data-driven and real-data-driven approaches – especially the limited ability to capture occupant behavior and real-life consumption patterns in simulation-data-driven approaches and the limited generalizability of real-data-driven approaches to different cases – by learning from both types of data simultaneously.

1.2 State of the Art and Knowledge Gaps

Four main knowledge gaps were identified: (1) knowledge gaps in understanding energy-related human values, including importance and satisfaction levels of these values and the associated factors, (2) knowledge gaps in occupant behavior-sensitive building energy consumption prediction, (3) knowledge gaps in real-data-driven energy consumption prediction, occupant comfort prediction, and occupant behavior optimization, and (4) knowledge gaps in hybrid approaches for building energy consumption prediction.

1.2.1 State of the Art and Knowledge Gaps in Understanding Energy-Related Human Values

A body of research efforts has been undertaken towards enhancing building energy efficiency. Despite the importance of these efforts, two primary knowledge gaps in the area of energy-related human values are identified. First, there is a lack of empirical knowledge on discovering the values that are related to energy-use behavior and energy consumption of residential and office building occupants. On one hand, a significant amount of existing research has examined energy efficiency in buildings and supported the importance of occupant behavior as key means of enhancing building energy efficiency. For example, Peschiera et al. (2010) showed that occupant behavior plays an important role in energy consumption and that there is an energy-saving potential by only improving occupant behavior; Azar and Menassa (2012b) emphasized that occupancy behavioral parameters have a significant impact on energy consumption and developed an agent-based energy consumption forecasting model to account for different occupant behaviors; Petersen et al. (2007) showed that supplying electricity consumption data to building occupants resulted in 32% reduction in energy consumption; and a simulation study by Klein et al. (2012) showed that utilizing occupancy information, occupant preferences, and a meeting relocation agent resulted in 12% reduction in building energy consumption. These efforts provided important contributions to the field of occupant behavior-based energy efficiency. However, there is still a lack of studies on how the values (health, comfort, productivity, etc.) of occupants impact their occupant behavior and how to reduce energy consumption while maintaining the satisfaction levels with these values. On the other hand, there are number of studies (e.g., Lai and Yik 2009, Zalejska-Jonsson and

Wilhelmsson 2013; Frontczak et al. 2012; Bluyssen et al. 2011) that focused on occupant comfort with indoor environmental quality (IEQ). For example, Lai and Yik (2009) studied the importance of thermal comfort, air cleanliness, odor, and noise to high-rise building occupants in Hong Kong. These studies provide an important contribution towards understanding the values that affect occupant comfort, but they are not focused on studying the values to understand how they are related to energy-use behavior and energy consumption.

Second, there is a lack of understanding of how the importance levels of values and the satisfaction levels with these values vary across different types of occupants, different states, and different potential energy-related factors (PEFs). A number of research studies have conducted questionnaire surveys or instrumental measurements to understand how occupant satisfaction with some values [including thermal comfort, visual (lighting) comfort, and IAQ] varies across different PEFs. For example, Frontczak et al. (2012) investigated the effects of 11 indoor environmental and building parameters on occupant satisfaction through post occupancy survey data collected by the Center for Built Environment (CBE). A literature survey by Frontczak and Wargocki (2011) discovered occupant, building, and outdoor climate characteristics that affect the comfort of occupants in buildings. Despite the importance of these studies, these efforts did not analyze how satisfaction levels with values vary across different types of occupants and states and across a number of important PEFs such as energy efficiency building features and occupant behavior.

1.2.2 State of the Art and Knowledge Gaps in Occupant-Behavior-Sensitive Building Energy Consumption Prediction

A significant number of recent studies have focused on building energy consumption prediction. Despite the importance of these studies, three primary research gaps are identified. First, there is a lack of energy consumption prediction and simulation approaches that take occupant behavior into account. Existing machine learning-based building energy consumption prediction efforts have used many features that affect the consumption, such as outdoor weather conditions (e.g., dry-bulb temperature, solar radiation), indoor environmental conditions (e.g., room temperature, room relative humidity), building characteristics (e.g., geometry, orientation), time [e.g., type of day (e.g., weekday, weekend, holiday), type of hour (e.g., daytime, nighttime)], operation characteristics (e.g., building use schedule, number of occupants). For example, Dong et al. (2005) predicted building energy consumption using mean outdoor dry-bulb temperature, relative humidity, and global solar radiation. Hong et al. (2014) predicted annual heating and electrical energy consumption using 23 building and climate-related features. Ascione et al. (2017b) predicted energy consumption for space heating and cooling using 9 geometry, 30 envelope, 6 building operation, and 3 HVAC-related features. Zhang et al. (2015) predicted hot water energy rate using only dry-bulb temperature. Rastogi et al. (2017) predicted heating loads using window-to-floor ratio (WFR), window-to-wall ratio (WWR), form factor (volume / wall area), annual sum of internal heat gain, and average sunlit percentage of envelope and predicted cooling loads using median dry-bulb temperature, inter-quartile range of dew-point temperature, WFR, and WWR. Paudel et al. (2014) predicted heating demand using outside temperature, solar radiation, day type, occupancy profiles, operational power level characteristics, and time-dependent attributes of operational power-level characteristics. Deb et al. (2016) predicted diurnal cooling load using only previous loads. Jain et al. (2014a) predicted electrical consumption using previous electrical consumption values, temperature, day type, sine of the current hour, and cosine of the current hour. Platon et al. (2015) predicted electricity consumption using outdoor air temperature, relative humidity, indoor air temperature, and some HVAC-related variables. Yu et al. (2010) classified building energy use intensity (EUI) levels using climatic conditions (e.g., annual average air temperature), building characteristics (e.g., house type, construction type), household characteristics (e.g., number of occupants), and household appliance energy sources (e.g., space heating source, hot water supply source). These studies were able to predict building energy consumption within acceptable accuracy ranges and provided important contributions to the field. However, despite their importance, none of them focused on taking the impact of occupant behavior into account. In developing machine learning-based prediction models, it is important to take the behavior of the building occupants into account, because it is one of the major factors that affect the energy consumption of the building and is one of the most significant contributors to the uncertainty in the prediction of this consumption (Hong et al. 2016a).

Second, there is a lack of energy consumption prediction studies that explored a deep learning approach. Deep learning approaches have been proven to outperform other machine learning algorithms in many fields (Amasyali and El-Gohary 2018). In the area of building energy consumption prediction, a significant number of studies have used Artificial Neural Networks (ANN) with shallower architectures (i.e., ANN with a single hidden layer). For example, Li et al. (2015) proposed a hybrid improved Particle Swarm Optimization algorithm (iPSO) – ANN model to predict hourly building electricity consumption, in which the iPSO algorithm was applied to adjust the weights and threshold values of the ANN structure. An et al. (2013) proposed a multi-output feedforward neural network (FFNN)-based approach, which combines FFNN with empirical mode decomposition (EMD)-based signal filtering and seasonal adjustment, to predict the electricity demand of a week in 30-min intervals. Ekici and Aksoy (2009) proposed an ANN-based model to predict annual heating energy consumption. ANN with deeper architectures, DNN [i.e., ANN with more than a single hidden layer], on the other hand, are less explored in the field of building energy consumption prediction. DNN was used in only a few studies such as Fan et al.

(2017). Additional studies are, thus, needed to better understand the applicability and limitations of DNN in the building energy consumption domain – for example to understand how the number of hidden layers and the sizes of the datasets impact prediction accuracy and computational efficiency.

Third, there is a lack of energy consumption prediction studies that explored an ensemble learning approach. Ensemble models consist of a number of models and are therefore likely to provide more accurate predictions than single models due to their stability (Wang et al. 2018a). A few studies have utilized ensemble techniques in the area of building energy consumption prediction. For example, Wang et al. (2018a) developed an ensemble bagging tree (EBT) model to predict the electricity demand of an institutional building. Wang et al. (2018b) developed a random forest (RF) model for hourly building energy consumption prediction. Tsanas and Xifara (2012) developed an RF model to predict cooling and heating loads of residential buildings. Lahouar and Slama (2015) proposed an RF model for short-term load forecasting. Jovanovic et al. (2015) compared the performance of three ANN models – including FFNN, radial basis function network (RBFNN), and adaptive neuro-fuzzy interference system (ANFIS) – to the ensemble of these three models and showed that the ensemble model achieved the most accurate prediction results. Such studies successfully applied ensemble techniques to building energy consumption prediction problems and showed that ensemble models can achieve more accurate predictions. More studies are needed to better understand the applicability and limitations of ensemble models in this domain (e.g., understand how the size of the training dataset affects prediction accuracy and computational efficiency), and their relative performance in comparison to deep learning algorithms such as DNN.

1.2.3 State of the Art and Knowledge Gaps in Real-Data-Driven Energy Consumption Prediction, Occupant Comfort Prediction, and Occupant Behavior Optimization

Building energy consumption prediction models are one of the essential tools to assess different energy saving strategies and therefore are used in conjunction with optimization algorithms. There are two main approaches for prediction: physical modeling approach and data-driven approach (Amasyali and El-Gohary 2018). Physical models (also known as engineering methods or white-box models) predict building energy consumption based on some predefined physical rules and require significant amount of building-related data. Building performance simulation (BPS) software, such as EnergyPlus, IDA Indoor Climate and Energy (IDA ICE), and IES Virtual Environment (IESVE), are examples of physical models. For example, Ascione et al. (2015) presented an EnergyPlus and MATLAB coupling-based optimization methodology to optimize the thermal design of the building envelope for reduced primary energy demand for the annual space conditioning and reduced thermal discomfort hours. Grygierek and Ferdyn-Grygierek (2018) presented an EnergyPlus and Non-dominated Sorting Genetic Algorithm II (NSGA-II) coupling-based optimization to optimize the selected design parameters (e.g., types of windows and building orientation) in a single-family building in temperate climate conditions for life cycle costs and thermal comfort.

BPS software, however, require a significant amount of time for accurate results (Chari and Christodoulou 2017). Therefore, the use of physical models is not practical in some cases due to the high number of iterations required by the optimization algorithms to reach a convergence. To overcome this, machine learning-based surrogate models, which can still accurately predict energy consumption but require much less time, were developed using simulation results (Reynolds et al. 2018). For example, Magnier and Haghighat (2010) developed a multilayer feed-forward ANN-based energy consumption prediction model. Then, the developed model was used in combination with a multi-objective genetic algorithm to optimize 20 building envelope-related, and HVAC system-related design variables for minimum energy consumption and maximum thermal comfort. Asadi et al. (2014) developed a three-layer feed-forward ANN-based model for predicting sanitary hot water, space heating and cooling energy consumption, and total percentage of discomfort hours. Then the developed model was used in combination with a multi-objective genetic algorithm to determine external wall insulation material, roof insulation material, windows, installation of solar collector and different HVAC system variables for optimal energy consumption, retrofit cost, and total percentage of discomfort hours.

Such efforts provided important contributions to the field of building energy optimization. However, despite the importance of all these efforts, there is still a lack of real-data-driven approaches for optimizing occupant behavior. Existing research efforts on the optimization of occupant behavior [e.g., Reynolds et al. (2018); Magalhães et al. (2017); Ahmad et al. (2016)] are based on simulation data. While there has been significant increase in the number of occupant behavior modeling in BPS [e.g., Gaetani et al. (2016)], simulation data are still limited in representing the complexity and stochastic nature of occupant behavior (Amasyali and El-Gohary 2019). The insufficient understanding of real occupant behavior in building design, operation, and retrofit leads to incorrect simplifications (Hong et al. 2016b) and incorrect comfort assumptions in simulations.

1.2.4 State of the Art and Knowledge Gaps in Hybrid Approaches for Building Energy Consumption Prediction

There are two main traditional approaches to building energy consumption prediction: data-driven approach and physical modeling approach. The data-driven approach uses machine learning algorithms to model building energy consumption based on previously recorded time-series data such as past energy consumption data, outdoor weather conditions, and occupancy schedule. With the advancement of data analytics, a significant number of data-driven prediction models, with various intended uses, have been developed. The physical modeling approach, on the other hand, relies on thermodynamic rules to model building energy consumption. Many whole building energy simulation programs – such as EnergyPlus, IDA Indoor Climate and Energy (IDA ICE), ESP-r, the IES Virtual Environment (IESVE), TRNSYS, and eQUEST – take a physical modeling approach for prediction.

Each approach has its own strengths and limitations. On one hand, data-driven approaches are suitable for capturing nonlinearity and patterns in complex systems, which makes them good candidates for capturing real energy consumption patterns that could be highly variable and complex due to occupant behavior. But, they require a large amount of data to train a model well and the resulting model often does not generalize well outside of the training range (Foucquier et al. 2013). For example, a model that was trained by learning from a limited dataset (e.g., data collected from a small set of buildings) may not perform well outside of the training data (e.g., different types of buildings in terms of physical properties, operation strategies, weather conditions, and occupant behavior) (Amasyali and El-Gohary 2018). Nevertheless, in many cases (e.g., Neto and Fiorelli 2008), data-driven prediction models were shown to be more accurate than physical models. On the other hand, physical models minimize the efforts of data sensing to collect the data for model training but require a significant amount of calibration effort to be able to accurately predict actual consumption (Lam et al. 2014). The calibration efforts involve determining and inputting the most accurate values for all the relevant parameters and variables – such as weather conditions, occupant and equipment schedules, and system parameters – for given energy consumption levels (IPMVP 2003). Also, the modeling of occupant behavior is an essential part of the model calibration (Gaetani et al. 2016) but in many cases physical models are still limited in representing the complexity and stochastic nature of occupant behavior.

In recent years, there has been growing interest in taking a hybrid modeling approach to building energy consumption prediction due to their unique capability to leverage the strengths and eliminate the limitations of the traditional data-driven and physical modeling approaches, by coupling them (Wang and Srinivasan 2017). The hybrid approach, often, requires relatively less training data and only a rough description of the building geometry (Foucquier et al. 2013) and still outperform the two traditional approaches (Amasyali and El-Gohary 2018). However, despite the importance of these studies, there is a lack of hybrid approaches for occupant-behavior-sensitive energy consumption prediction, which could leverage the strengths and reduce the limitations of the traditional data-driven and physical modeling.

1.3 Problem Statement

Recent studies emphasized the importance of occupant behavior as key means of enhancing building energy efficiency. Occupant behavior is one of the most significant factors that affect building energy consumption. A significant amount of energy can be saved through improving occupant behavior. Despite the significance of occupant behavior in enhancing the energy efficiency of buildings, major knowledge gaps exist in this area: (1) there is a lack of understanding of how important energy-related human values are to residential and office building occupants, what their satisfaction levels with these values are, and how these importance and satisfaction levels vary across different types of occupants, different states, and different energy-related factors; (2) there is a lack of building energy consumption prediction and simulation models that take energy-use behavior into account; (3) there is a lack of studies on capturing real energy-use behavior, real energy-consumption patterns, and real occupant-comfort information to better understand the impact of occupant behavior on building energy consumption and occupant comfort; and help optimize occupant behavior for both energy saving and occupant comfort; and (4) there is a lack of studies on comparing the simulation-data-driven and the real-data-driven energy consumption prediction approaches, and a lack of hybrid models that aim to leverage the strengths and reduce the limitations of the two approaches.

1.4 Research Objectives and Questions

The aim of this thesis is to propose and test a data-driven machine-learning-based approach to behavioral building energy efficiency, which could help better understand and predict the impact of occupant behavior on building energy consumption and occupant comfort; and help optimize occupant behavior for both energy saving and occupant comfort. Three types of approaches were investigated, in this regard: simulation-data-driven, real-data-driven, and hybrid.

Accordingly, five specific objectives are defined, as follows.

(1) <u>**Objective #1**</u>: Identify the importance levels of the human values (that are related to energy-use behavior and energy consumption) to residential and office building occupants and their current satisfaction levels with these values, the factors that are associated with higher/lower importance and/or satisfaction levels, and the potential factors that could help predict occupant satisfaction levels.

<u>Research Questions</u>: What are the potential occupant values that could be related to energy-use behavior and energy consumption in residential and office buildings? What are the importance levels of these values to residential and office building occupants? What are the satisfaction levels of occupants with these values? What are the factors (e.g., occupant

characteristics, level of occupant building control, building energy efficiency features, energy-use behavior) that are associated with higher/lower importance and/or satisfaction levels?

Outcomes: (1) Identifying the importance levels of the values (that are related to energy-use behavior and energy consumption) to residential and office building occupants; (2) Identifying their current satisfaction levels with these values; and (3) Determining the factors (e.g., occupant characteristics, level of occupant building control, building energy efficiency features, energy-use behavior) that are associated with higher/lower importance and/or satisfaction levels.

(2) **Objective #2**: Develop a data sensing and occupant feedback collection plan for capturing and monitoring indoor environmental conditions, energy consumption (heating/cooling, lighting, and plug loads), energy-related occupant behavior, and occupant comfort in real buildings; and conducting the data collection.

Research Questions: What are the indoor environmental condition data that need to be collected? What are the temporal and spatial granularities (including submetering level) of energy consumption data that need to be collected? What are the energy-use behavior data that need to be captured, how, and how frequent? What are the occupant comfort data that need to be captured, how, and how frequent? What are the sensors, meters, and user-interactive devices that are needed to capture these data?

Outcomes: (1) A data sensing and occupant feedback collection plan for capturing and monitoring for capturing and monitoring indoor environmental conditions, energy consumption (heating/cooling, lighting, and plug loads), energy-related occupant behavior, and occupant comfort in real buildings; and (2) Datasets for training and testing the machine-learning based models.

(3) **Objective #3:** Model and simulate a set of buildings with different sizes and occupant behaviors in different weather conditions; and develop an occupant-behavior-sensitive machine learning-based model, which learns from simulation data, for predicting hourly cooling energy consumption.

Research Questions: What are the buildings that need to be modeled and simulated? How to represent occupant behavior in building energy simulations? Which features to use for predicting building cooling energy consumption? Which machine learning algorithms will perform better in terms of prediction accuracy, computational efficiency (training time), and sensitivity to variations in sample sizes?

Outcome: An occupant behavior-sensitive machine learning-based model for predicting hourly cooling energy consumption based on simulation data.

(4) **Objective #4**: Develop occupant-behavior-sensitive machine learning-based models, which learn from real data (data that were collected from real buildings and real occupants), for predicting hourly cooling and lighting energy consumption and thermal and visual occupant comfort; and develop a genetic algorithm-based optimization model for determining the optimal occupant behavior that can simultaneously reduce energy consumption and improve occupant comfort.

Research Questions: Are the occupant-behavior features discriminating? Which will perform better, the models with or without occupant-behavior features? Which will perform better, the models with the past one- or two-hour outdoor weather condition features, or with no past-hour features? Which machine learning algorithms will perform better in terms of prediction accuracy? How to optimize occupant behavior for simultaneously reducing energy consumption and improving occupant comfort? Can energy savings and occupant comfort be achieved simultaneously?

Outcomes: (1) A set of occupant-behavior-sensitive machine learning-based models for predicting hourly cooling and lighting energy consumption and thermal and visual occupant comfort based on real data; and (2) A genetic algorithm-based optimization model for determining the optimal occupant behavior.

(5) **Objective #5**: Develop a hybrid, occupant-behavior-sensitive machine learning-based model, which learns from both simulation data (data created through simulating a set of reference buildings) and real data (data collected from real buildings and real occupants), for predicting hourly cooling energy consumption.

Research Questions: How to learn from both simulation and real data for improving prediction performance? How to represent the impact of outdoor weather conditions on energy consumption? How to reveal the impact of occupant behavior on energy consumption? Which features to use for predicting building cooling energy consumption? Which machine learning algorithms will perform better in terms of prediction accuracy?

Outcome: A hybrid, occupant-behavior-sensitive machine learning-based model for predicting hourly cooling energy consumption based on both simulation and real data.

1.5 Research Tasks and Methodology

The research methodology includes six primary research tasks, as summarized in Figure 1.1. A more detailed explanation of the methodology of each task is presented in the following subsections.

Figure 1.1 – Summary of Research Tasks

1.5.1 Task 1: Literature Review

A comprehensive literature review was conducted in seven primary domains: building energy efficiency, energy-related occupant values, occupant behavior, machine learning algorithms and machine learning-based prediction models, hybrid modeling approach, time-series clustering, and weather normalization. The following points provide a summary of the literature review in each of these domains:

- Building energy efficiency: the literature review covered existing studies that have been conducted in the area of building energy efficiency, with especial emphasis on studies that focused on occupant behavior.
- Energy-related occupant values: the literature review focused on existing literature on occupant values that are related to energy-use behavior of occupants in residential and office buildings.
- Occupant behavior: the literature review focused on studies that demonstrated the significant impact of occupant behavior on building energy consumption.
- Machine learning algorithms and machine learning-based prediction models: the literature review focused on existing machine learning-based models and studies for predicting building energy consumption. The literature review covered the scope of the existing prediction models, the data collection and data preprocessing techniques of the existing models, the machine learning algorithms used for training the existing models, and the performance of the existing models. The literature review also focused on understanding the state-of-the-art machine learning algorithms, such as support vector machine (SVM).
- Hybrid modeling approach: the literature review covered the state-of-the-art studies for coupling data-driven and physical approaches in the area of building energy consumption prediction.

- Time-series clustering: the literature review focused on existing time-series clustering methods and studies for clustering building energy consumption data.
- Weather normalization: the literature review focused on existing weather normalization methods.

1.5.2 Task 2: Identification of Value Importance and Satisfaction Levels and Associated Factors through Occupant Surveys

This task aimed to identify the importance levels of the human values (that are related to energy-use behavior and energy consumption) to residential and office building occupants and their current satisfaction levels with these values, the factors that are associated with higher/lower importance and/or satisfaction levels, and the potential factors that could help predict occupant satisfaction levels. Two questionnaire surveys were conducted to solicit the input of a randomly selected set of residential and office building occupants in Arizona (AZ), Illinois (IL), and Pennsylvania (PA), on (1) the importance levels of occupant values and (2) the current satisfaction levels with these values. According to the Köppen-Geiger climate classification, IL and PA have a hot summer continental climate (Dfa), whereas a significant part of AZ has a hot desert climate (Bwh) (Peel et al. 2007). This research task is composed of four primary research subtasks: (1) questionnaire design, (2) validation of questionnaire design, (3) respondent recruitment and survey implementation, and (4) analysis of survey results.

1.5.2.1 Subtask #2.1 – Questionnaire Design

Both questionnaires were composed of four sections. Section 1 included two filtering questions that were asked to verify eligibility of participation in terms of occupancy type and residency state (e.g., for the office survey, occupancy of an office building and residency in AZ, IL, or PA). Responses that failed to pass Section 1 were disregarded. In Section 2, respondents were asked to rate the importance levels of occupant values to them on a 6-point Likert scale (very unimportant, unimportant, moderately unimportant, moderately important, important, very important). Section 3 aimed to solicit the satisfaction levels with the values. Section 4 aimed to collect data about PEFs, in order to explore the potential differences in the importance levels of values and satisfaction levels with these values across different PEFs, including occupant characteristics, health symptoms, primary building characteristics, level of occupant building control, energy efficiency building features, energy cost and consumption feedback (for residential survey only), energy-use behavior to control indoor environmental conditions, workspace characteristics (for office survey only), and job characteristics (for office survey only). Due to the variability in occupancy and building characteristics across residential and office buildings, the questions in this section varied across both questionnaires.

1.5.2.2 Subtask #2.2 – Validation of Questionnaire Design

Prior to launching the survey, a pilot study was conducted to test the effectiveness and clarity of the questionnaire. Participants were requested to complete the residential or office building survey and, then, to provide feedback on the format and content of the questionnaire. Feedback was solicited on different aspects of the questionnaire, such as question wording, response options and evaluation scale, instructions to respondents, visual appearance, and clarity of value concepts. The questionnaire was revised based on the feedback, if/as needed.

1.5.2.3 Subtask #2.3 – Respondent Recruitment and Survey Implementation

Potential respondents were recruited by Qualtrics, a provider of online panels (potential respondents). Panels were generated using samples from various databases and were verified to prevent any fraudulent or duplicate respondents (Qualtrics 2014). Qualtrics hosted the survey and sent emails to potential respondents inviting them to complete the survey, for research purposes, in return for incentives. Two response quality filters were used: (1) an attention filter question, and (2) a minimum survey completion time of two minutes. Responses that failed to pass these two filters were disregarded.

1.5.2.4 Subtask #2.4 – Analysis of Survey Results

The analysis of the survey results focused on answering the following research questions: What are the ratings and the rankings of the importance levels of values by residential and office building occupants in AZ, IL, and PA? What are the ratings and the rankings of the satisfaction levels of residential and office building occupants with the values in AZ, IL, and PA? What are the differences in the importance levels and satisfaction levels of/with the values across different types of occupants (residential and office), different states (AZ, IL, and PA), and PEFs?

Five statistical analysis methods were utilized to address the above research questions: (1) mean indexing, (2) Spearman's rank correlation, (3) Kendall’s coefficient of concordance, (4) Mann-Whitney U test, and (5) Kruskal-Wallis H Test. Mean indexing was used to determine the mean ratings of values. Spearman's rank correlation was used to assess the general similarity between occupants of residential and office buildings. Kendall’s coefficient of concordance was computed to examine whether there are significant agreements among (1) occupants of residential buildings across the three states, and (2) occupants of office buildings across the three states. When there were two groups to compare, the Mann-Whitney U test was used to identify whether specific values are rated differently (e.g., across residential and office building occupants and across male and female office occupants). When there were more than two groups to compare, the Kruskal-Wallis H test was conducted to identify whether specific values were rated differently (e.g., across the three states). The Statistical Package for Social Sciences (SPSS) version 20.0 was used to conduct these statistical analyses.

1.5.3 Task 3: Data Sensing and Occupant Feedback Collection Plan Development and Data Collection

This task aimed to develop a data sensing and occupant feedback collection plan for capturing and monitoring indoor environmental conditions, energy consumption (heating/cooling, lighting, and plug loads), energy-related occupant behavior, and occupant comfort in real buildings; and conducting the data collection. This research task is composed of two primary subtasks: (1) data sensing and occupant feedback collection plan development, and (2) and data collection.

1.5.3.1 Subtask #3.1 – Data Sensing and Occupant Feedback Collection Plan Development

A data collection plan was developed in collaboration with the research collaborators in this effort: Prof. Chimay Anumba, Dean and Professor, University of Florida; Prof. Yewande Abraham, Assistant Professor, Rochester Institute of Technology; and the Penn State Consortium for Building Energy Innovation (CBEI). The plan defines the data that need to be collected, the method of collection (e.g., sensor versus mobile occupant feedback system), and the frequency of the data collection. Based on the data collection plan, for indoor environmental condition and energy consumption data, a data sensing and instrumentation plan was developed by the aforementioned research collaborators. The plan shows the details of the sensing and instrumentation, including specific types of sensors and meters that are needed, the locations of installation in the buildings, etc. The sensors/meters were installed based on this plan. A data interface software was used to display and download the data. The software provided access to review/download historical and real-time data. For occupant behavior and comfort data, an occupant feedback system was utilized to collect occupant feedback. The occupant feedback system was developed on Qualtrics Mobile by the aforementioned research collaborators.

1.5.3.2 Subtask #3.2 – Data Collection

The aforementioned data were collected from the Philadelphia Business and Technology Center (PBTC) building between October 05, 2015 and December 31, 2018. The building was selected because Pennsylvania is the home of one of the biggest national research efforts for enhancing building energy efficiency, the Consortium for Building Energy Innovation (CBEI), and the research team agreed to use one of their partially instrumented buildings (and build on existing instrumentation) to save on financial resources.

1.5.4 Task 4: Simulation-Data-Driven Occupant-Behavior-Sensitive Machine Learning-based Model Development

This task aimed to model and simulate a set of buildings with different sizes and occupant behaviors and develop an occupant-behavior-sensitive machine learning-based model, which learns from simulation data, for predicting hourly cooling energy consumption. This research task is composed of five primary subtasks: (1) building and occupant behavior modeling, (2) energy simulations, (3) data preprocessing, (4) machine learning model development, and (5) performance evaluation.

1.5.4.1 Subtask #4.1 – Building and Occupant Behavior Modeling

A number of buildings were modeled to represent different occupant behaviors and building sizes. To capture the impact of different occupant behaviors on consumption, a set of cases that represent different behaviors were modelled. To model these cases, a set of proxy variables that could represent behavior differences were identified and modeled in a parametric way: cooling setpoint, window status, lighting power density, occupancy density, and electric equipment power density. To create a dataset that represents different office buildings in the U.S., different building sizes were modeled.

1.5.4.2 Subtask #4.2 – Energy Simulations

The building models were simulated in five cities, using EnergyPlus, to represent the five main climate zones in the United Sates. The typical meteorological year 3 (TMY3) weather data of the five locations were used. In order to have an undisturbed consumption pattern throughout the simulation period, the holiday schedules in EnergyPlus were removed.

1.5.4.3 Subtask #4.3 – Data Preprocessing

The EnergyPlus data were preprocessed in preparation for the machine learning. This included three primary steps: feature generation, feature selection, and data sampling.

1.5.4.4 Subtask #4.4 – Machine Learning Model Development

A set of hourly cooling energy consumption prediction models were developed to test and compare different machine learning algorithms in terms of prediction accuracy, computational efficiency (training time), and sensitivity to variations in sample sizes. Four machine learning algorithms were tested: Classification and Regression Tree (CART), ANN, EBT, and DNN. CART and ANN are among the most popular machine learning algorithms in the field of building energy consumption prediction, whereas EBT and DNN are potentially superior but relatively less explored in this field. To assess the effect of ensembling on the prediction, the performances of the CART (single model) and the EBT (ensemble model) models were compared. To understand the effect of the depth of the neural network models on the prediction, four different neural network models with different number of hidden layers were compared.

1.5.4.5 Subtask #4.5 – Performance Evaluation

Three performance metrics were used to evaluate the prediction performance of the models: coefficient of variation (CV), root mean square error (RMSE), and coefficient of determination (R^2). The trained models were used to predict the hourly cooling energy consumptions of the instances in the testing dataset. The predicted values were compared to the actual (or simulated) values and the CV, RMSE, and R^2 were calculated, as per Eq. (1.1) to (1.3).

$$\text{CV}\ (\%) = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_{predict,i} - y_{data,i})^2}{n}}}{\overline{y}_{data}} \times 100 \tag{1.1}$$

$$\text{RMSE(kWh)} = \sqrt{\frac{\sum_{i=1}^{n}(y_{predict,i} - y_{data,i})^2}{n}} \tag{1.2}$$

$$R^2(\%) = \frac{\sum_{i=1}^{n}(y_{predict,i} - \overline{y}_{data})^2}{\sum_{i=1}^{n}(y_{data,i} - \overline{y}_{data})^2} \times 100 \tag{1.3}$$

where $y_{predict,i}$ is the predicted energy consumption at hour i, $y_{data,i}$ is the actual (or simulated) energy consumption at hour i, n is the number of hours in the dataset, and $\overline{y}_{data}$ is the average energy consumption.

1.5.5 Task 5: Real-Data-Driven Occupant-Behavior-Sensitive Machine Learning-based Model Development and Optimization

This task aimed to develop occupant-behavior-sensitive machine learning-based models, which learn from real data (data collected from real buildings and real occupants), for predicting hourly cooling and lighting energy consumption and thermal and visual occupant comfort; and develop a genetic algorithm-based optimization model for determining the optimal occupant behavior that can simultaneously reduce energy consumption and improve occupant comfort. This research task is composed of five primary subtasks: (1) data preprocessing, (2) data analysis, (3) machine learning model development, (4) prediction performance evaluation, and (5) optimization.

1.5.5.1 Subtask #5.1 – Data Preprocessing

Data preprocessing included four primary steps: data cleaning and outlier filtering, data aggregation, data integration, and data normalization.

1.5.5.2 Subtask #5.2 – Data Analysis

Data analysis included three primary steps: data transformation, clustering of energy-use modes, and statistical analysis.

1.5.5.3 Subtask #5.3 – Machine Learning Model Development

A set of machine learning-based occupant-behavior-sensitive prediction models for real-data-driven prediction of cooling and lighting energy consumption and thermal and visual occupant comfort were developed. For the comfort prediction models, two types of models were developed: group and individual. To verify that occupant-behavior features are discriminating, and hence that the prediction models can be used to predict the impact of the behavior, the performance of the models were compared to others without occupant-behavior features. To consider the delayed effects of outdoor weather, prediction models with the past one- and two-hour outdoor weather condition features and with no past-hour features were tested and compared. A set of machine learning algorithms were also tested to determine the most accurate, among the tested ones, in energy consumption and comfort prediction.

1.5.5.4 Subtask #5.4 – Prediction Performance Evaluation

A 10-cross fold validation was utilized to assess the performance, because it minimizes the bias due to the randomness in choosing the testing data (Chou and Bui 2014). The following performance metrics were used to evaluate the performance of the prediction models, which were calculated using Eq. (1.1) to (1.3): CV, RMSE, and R^2.

1.5.5.5 Subtask #5.5 – Optimization

A genetic algorithm-based optimization model for optimizing occupant behavior was developed. The main purpose of the optimization is to minimize energy consumption while maximizing thermal and visual comfort. The decision variables, objective functions, and optimization computations of the proposed optimization were formulated. The NSGA-II (Deb 2001) was used for conducting the optimization due to the algorithm's capabilities of fast non-dominated sorting approach, fast crowded distance estimation procedure, and simple crowded comparison operator (Yusoff et al. 2011). In conducting the optimization, four types of optimal solutions were considered: energy-priority, thermal-comfort-priority, and visual-comfort-priority, and balanced solutions.

1.5.6 Task 6: Hybrid Machine Learning-based Model Development for Coupling Data-Driven and Physical Approaches

This task aimed to develop a hybrid, occupant-behavior-sensitive machine learning-based model, which learns from both simulation data (data created through simulating a set of reference buildings) and real data (data collected from real buildings and real occupants), for predicting hourly cooling energy consumption. This research task is composed of four primary subtasks: (1) weather factor prediction model development, (2) occupant-behavior factor prediction model development, (3) ensemble model development, and (4) performance evaluation.

1.5.6.1 Subtask #6.1 – Weather Factor Prediction Model Development

The development of the weather-factor prediction model included three primary steps: energy simulations, time-series clustering, and factor prediction model development. The simulation-generated data were created by simulating a number of reference models of small, medium, and large office and midrise apartment buildings. Time-series clustering was conducted to group similar consumption patterns. It included four main steps: data organization, data normalization, time-series clustering, and cluster validation. For each cluster, a machine learning model that predicts the hourly weather factor based on outdoor weather conditions was developed.

1.5.6.2 Subtask #6.2 – Occupant-Behavior Factor Prediction Model Development

The development of the occupant-behavior factor prediction model included three primary steps: data preprocessing, weather normalization, and factor prediction model development. Data preprocessing included five main steps: data aggregation, data integration, data cleaning and outlier filtering, data normalization, and data splitting. Weather normalization was performed to remove the effect of weather conditions and better reveal the impact of occupant behavior. The daily weather normalization method proposed by Hydro One (Hydro One 2006) was adapted so that it can be used for hourly normalization. The adapted method includes three main steps: model development, expected energy calculation, and occupant-behavior factor calculation. A machine learning model that predicts the hourly occupant-behavior factor based on occupant behavior was developed.

1.5.6.3 Subtask #6.3 – Ensembler Model Development

An ensembler model was developed to predict hourly cooling energy consumption. The ensembler model takes the weather- and occupant-behavior-factors as features.

1.5.6.4 Subtask #6.4 – Performance Evaluation

The performance of the three constituent models and therefore the whole hybrid model was evaluated. The performance of the weather-factor prediction model was evaluated on the simulation-generated data using 10-fold cross validation. The performance of the occupant-behavior-factor prediction model was evaluated on the real training data using 10-fold cross validation. The performance of the ensembler model, and therefore the whole hybrid model, was evaluated using the real testing data. The following metrics were used to evaluate the performance of the prediction models, which were calculated using Eq. (1.1) to (1.3): CV, RMSE, and R^2.

1.6 Contribution to the Body of Knowledge

The proposed research contributes to the body of knowledge in four primary ways. First, it advances the theoretical and empirical knowledge in the area of energy-related human values by identifying the importance levels of the human values to residential and office building occupants and their current satisfaction levels with these values, the factors that are associated with higher/lower importance and/or satisfaction levels, and the potential factors that could help predict occupant satisfaction levels. Second, it offers a simulation-data-driven, machine-learning approach for predicting building energy consumption in an occupant behavior-sensitive manner. The proposed approach could help better understand the impact of occupant behavior on building energy consumption, as well as identify opportunities for behavioral energy-saving measures and efficient building-operation strategies. Third, it offers a real-data-driven approach for predicting building energy consumption in an occupant behavior-sensitive manner and incorporating occupant behavior into building energy optimization. The use of real-life data helps better represent and understand the complex and stochastic nature of occupant behavior and its impact on energy consumption and comfort. Also, combining the prediction models with optimization offers a powerful tool for finding the right energy-use behavioral changes that can achieve, both, energy saving and comfort improvement. Fourth, it offers a novel hybrid modeling approach for energy consumption prediction. The proposed approach offers a direction towards leveraging the strengths and reducing the limitations of the traditional data-driven and physical modeling approaches by learning from both simulation and real data. Learning from both types of data aims to overcome two main limitations: the limited generalizability of data-driven approaches to different cases and the limited ability of physical modeling approaches to capture occupant behavior and real-life consumption patterns.

1.7 Publications

The thesis contains material published in the following journal and conference papers:

- **Amasyali, K.**, & El-Gohary, N. (2015). Discovering the values of residential building occupants for value-sensitive improvement of building energy efficiency. *Proceedings of the Canadian Society for Civil Engineering's 5th International/11th Construction Specialty Conference (ICSC15),* Vancouver, BC, Canada, June 7-10, 2015.
- **Amasyali, K**., & El-Gohary, N. M. (2016). Energy-related values and satisfaction levels of residential and office building occupants. *Building and Environment, 95*, 251-263.
- **Amasyali, K.**, & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews, 81*, 1192-1205.
- **Amasyali, K**., & El-Gohary, N. (2018). Machine learning-based occupant energy use behavior optimization. *Proceedings of the 2018 Construction Research Congress (CRC 2018),* New Orleans, LA, April 2- 5, 2018.
- **Amasyali, K**., & El-Gohary, N. (2019). Predicting energy consumption of office buildings: a hybrid machine learning-based approach. *Proceedings of the 35th CIB W78 2018 Conference: IT in Design, Construction, and Management*, Chicago, IL, October 1- 3, 2018.

CHAPTER 2 – SUMMARY OF LITERATURE REVIEW

This chapter presents a summary of literature review that focuses on the following research domains: building energy efficiency, energy-related occupant values, occupant behavior, machine learning algorithms and machine learning-based prediction models, hybrid modeling approach, time-series clustering, and weather normalization.

2.1 Building Energy Efficiency

Buildings are an important cause of energy consumption and CO_2 emissions. The IEA "Technology Roadmap for Energy Efficient Buildings: Heating and Cooling Equipment" report highlights that globally, buildings represent 33% of the energy consumption and CO_2 emissions (2011). In the developed world, the building share of the total energy consumption is higher than the other primary sectors such as transportation and industrial (Perez-Lombard et al. 2008). For example, in the U.S. and Europe, buildings account for 40% of the energy consumption (DOE, 2012; Lewis et al., 2013). In addition, in the U.S. and Europe, an average person spends 90% or more of their time indoors (EPA 2009; European Commission 2003). Economic booms, population growth, and increase in the building comfort demand of people will increase the time spent indoors and, therefore, the energy consumption of buildings will increase further in parallel with the increase in economies, populations, and comfort demands of countries (Perez-Lombard et al. 2008). IEA estimates a 50% increase in the energy demand caused primarily by buildings by 2050, and highlights that this increase can be capped to 10% without any sacrifice in the comfort of building occupants, if necessary improvements in energy efficiency can be achieved (IEA 2013).

For these reasons, with its high energy consumption and growing trend, the building sector presents the most significant energy cost saving and CO_2 emission reduction opportunity. Enhancing building energy efficiency is one of the most effective ways of reducing both energy consumption and CO_2 emissions. Adapting existing technologies (e.g., human-sensitive automated building control, real-time energy cost and consumption feedback) to the field of building energy efficiency is one of the best and affordable ways for reducing energy consumption without any sacrifice from user features. van der Hoeven, executive director of IEA, calls on all countries to pay more attention to building energy efficiency in order to achieve the target of low carbon emissions by 2050 (IEA 2013). Investment in energy efficiency should not be considered as an extra cost because it can pay back its cost in a very short period through reduced energy cost and lower CO_2 emissions, if it is correctly applied (Kneifel 2010).

There are seven factors which affect building energy consumption and efficiency: (1) climate, (2) building-related characteristics, (3) occupant-related characteristics, (4) building services systems and operation (5) building occupants' behavior and actions, (6) social and economic factors, and (7) IEQ required. The role of occupants in enhancing energy efficiency in buildings is high. Using technology-only-oriented approaches, such as using more energy efficient equipment, is not enough. "The behavior of occupants in a building can have as much impact on energy consumption as the efficiency of equipment" (WBCSD 2007). Occupants have the most important impact on energy consumption since all the technological features are dependent on occupant action (Masoso and Grabler, 2010).

2.2 Energy-Related Occupant Values

In the context of building energy efficiency, a comprehensive literature review was conducted to identify all human values that could be related to energy-use behavior and energy consumption, with focus on studies presenting importance levels of occupant values and satisfaction levels with these values.

Based on this literature review, seven primary values were identified and classified into three categories: (1) values that may impact energy-use behavior and energy consumption levels: thermal comfort, visual comfort, and IAQ, (2) values that may be impacted by the set of values in the first category: health and personal productivity, and (3) values that may motivate enhanced energy-use behavior towards reduced energy consumption: environmental protection and energy cost saving.

Thermal comfort is "that condition of mind that expresses satisfaction with the thermal environment" (ASHRAE 2010). There are six primary factors that affect thermal comfort: metabolic rate, clothing insulation, air temperature, radiant temperature, air speed, and humidity (ASHRAE 2010). Among these factors, metabolic rate depends on a number of sub-factors such as activity level, age, gender, height, weight, and health conditions (Maiti 2014; Boston 2013). Clothing insulation varies by occupant clothing type (ASHRAE 2010). Air temperature, radiant temperature, air speed, and humidity, on the other hand, are highly dependent on the settings and parameters of the HVAC system or other heating and cooling sources of buildings, which in turn may affect energy consumption.

Visual comfort is defined as "a subjective condition of visual well-being induced by the visual environment" (EN 2011). Visual comfort or discomfort is impacted by luminance distribution, illuminance and its uniformity, glare, color of light, color rendering, flicker rate, and amount of daylight (EN 2011). Illuminance is the factor which associates visual comfort with energy consumption.

IAQ is "a term referring to the air quality within and around buildings and structures" (EPA 2015). The amounts of indoor pollutants and of ventilation are the major factors that affect IAQ. Air pollutants entering from outdoor space, building materials, combustion sources (wood, coal, oil etc.), cleaning products, and tobacco, are main causes of indoor pollutants (EPA 2015). On the other hand, the amount of ventilation is determined by the amount of air that enters the building. Poor IAQ is seen as the primary environmental health risk (EPA 2015). In order to maintain good IAQ to building occupants, the amounts of pollutants should be controlled and proper amount of ventilation should be provided (EPA 2015). While controlling the amount of pollutants can be achieved by improving energy-use behavior and eliminating the causes of pollutants, the amount of ventilation is highly dependent on the building ventilation system which may consume energy.

Health and personal productivity are the values that may be impacted by the set of values in the first category. With the majority of people spending about 90% of their time indoors, the impact of thermal comfort, visual comfort, and IAQ on occupant health and productivity has been emphasized in recent years. Good thermal comfort, visual comfort, and IAQ are linked to decreased number of illnesses and sick building syndrome symptoms and enhanced productivity. Recent studies have further shown that maintaining good thermal conditions increases productivity (Lan et al. 2012) and that an amount of $17 billion to $26 billion can be saved annually, as a result of decreased sick building syndrome symptoms and increased productivity, by only enhancing thermal comfort and IAQ in offices (Fisk et al. 2011). Health and personal productivity are also highly impacted by occupant characteristics such as age (Skirbekk 2004), weight (Finkelstein et al. 2010), fitness level (Sharifzadeh 2013), and smoking habits (Halpern et al. 2001).

Environmental protection and energy cost saving are values that may motivate enhanced energy-use behavior towards reduced energy consumption. Energy consumption is associated with both environmental impacts and cost. For example, residential buildings account for 20.8% of the US total CO_2 emissions (EPA 2009) and residential building occupants spent 2.7% of their household income for home energy bills in 2012 (EIA 2013). The role of energy-use behavior in reducing energy consumption, and in turn in environmental protection and energy cost saving, is vital. The International Energy Agency (IEA) estimates a 50% increase in energy demand caused primarily by buildings by 2050, and highlights that this increase can be capped to 10% without any sacrifice in the comfort of building occupants, if necessary improvements in energy-use behavior and energy efficiency can be achieved (2013). A number of measures may help motivate enhanced energy-use behavior such as providing energy cost feedback, energy consumption feedback, and real-time pricing (Darby 2006; Faruqui et al. 2010; Fischer 2008; Faruqui and Sergici 2010). Several other factors such as building characteristics and energy efficiency features may also affect energy consumption, and in turn environmental protection and energy cost saving (Yu et al. 2011).

Zalejska-Jonsson and Wilhelmsson (2013) analyzed the impact of values on overall occupant satisfaction through ordinal logistic regression. The results showed that IAQ has the highest impact on overall satisfaction. Lai et al. (2009) showed that thermal and visual (aural) quality are the most important contributors of IEQ; while IAQ is the least. Cao et al. (2012) introduced an overall satisfaction calculation equation for public buildings. Based on the equation, thermal environment is the most important contributor of overall satisfaction. Lai and Yik (2009) discovered importance of occupant values through analytical hierarchical process. Thermal comfort was identified as the most important value. Results of a study by Frontczak et al. (2012) suggested that acceptability of thermal conditions, visual conditions, acoustic conditions, and air quality are equally important for occupants. Lai and Yik (2009) found that thermal comfort is the least satisfied and least important value in commercial buildings.

Today, most of the buildings are being operated based on the standards that assess occupant satisfaction (with values) based on predefined rules and equations. For example, predicted mean vote (PMV) is a widely-used thermal comfort index provided by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). The PMV index predicts the occupant satisfaction level with thermal comfort based on several assumptions about clothing levels, activity levels, and metabolic rate, but these assumptions, are, often, inaccurate (Jazizadeh et al. 2014). Indoor carbon dioxide (CO_2) concentration and comfort, satisfaction, and performance (CSP) index, on the other hand, are used for assessing the satisfaction levels with IAQ and visual comfort, respectively (Lai et al. 2009). However, studies (e.g., Barlow and Fiala 2007) have shown that the occupant satisfaction levels that are calculated by the standards are different from the actual occupant satisfaction levels and, therefore, operating the buildings based on the standards causes occupant dissatisfaction. In this regard, developing a personalized method for predicting satisfaction levels of occupants is essential (Daum et al. 2011).

2.3 Occupant Behavior

Occupant behavior is the actions and decisions taken by building occupants that affect building energy consumption and/or occupant satisfaction (Klein et al. 2012). It includes interactions of occupants with operable windows, lights, blinds, thermostats, and plug-in appliances (Yan et al. 2015). Occupant behavior is a major factor affecting building energy consumption and contributing to the uncertainty associated with building energy consumption prediction (Hong et al. 2016a). The understanding of occupant behavior in building design, operation, and retrofit is limited, which causes inaccurate modeling and analysis (Hong et al. 2015).

A body of research efforts has been undertaken to demonstrate the significant impact of occupant behavior on building energy consumption. For example, Jian et al. (2015) compared 44 identical apartments in Beijing and showed that there are significant differences in household electricity consumption among the monitored apartments due to the significant impact of occupant behavior. Clevenger et al. (2014) showed that occupant behavior can impact annual energy consumption as much as 75% for residential buildings and 150% for commercial buildings. Guerra-Santin et al. (2018) highlighted the uncertainties related to energy savings due to occupant behavior, and proposed an approach to minimize such uncertainties. Azar and Menassa (2014) showed that the current energy consumption level of the office building stock in the U.S. can be reduced by 21% due to improved human actions and more efficient operation of building systems.

2.4 Machine Learning Algorithms and Machine Learning-Based Prediction Models

Predicting energy consumption is a challenging task, because it is related to a variety of factors. For example, Kwok and Lee (2011) summarized the factors that can play a role in building energy consumption including physical properties, installed equipment, outdoor weather, and occupant behavior. In this regard, physical models (engineering methods) were developed to take several factors that play a role in building energy consumption into account. Physical models, however, require many input parameters that are usually not available to users (Zhao and Magoules 2012). Some building energy software tools, such as EnergyPlus, eQuest, and Ecotect, are examples of such physical models. A review on the use of physical models in energy consumption prediction is provided in Crawley et al. (2008).

To address the limitations of physical models, data-driven models have been introduced to the field of building energy consumption prediction. Data-driven models, as opposed to physical models, do not require many input parameters. Instead, historical input data (e.g., outdoor weather conditions and electricity consumption) are utilized in prediction.

2.4.1 Machine Learning Algorithms

In any data-driven approach, developing a prediction model consists of four steps: data collection, data preprocessing, model training, and model testing. In the field of building energy consumption prediction, data collection involves collecting historical input data for model training such as outdoor weather conditions and electricity consumption. Data preprocessing includes data cleaning, data integration, data transformation, and data reduction. Model training is the training of the prediction model using the historical data. SVM, ANN, decision tree (DT), and other statistical algorithms are the most commonly-used supervised learning algorithms for model training. SVM is a kernel-based machine learning (ML) algorithm and can be used for both regression and classification (Wu et al. 2008). The goal of this algorithm is to find a function f(x) that has at most epsilon (ε) deviation from the actually obtained target y_i for all the training data and at the same time is as flat as possible (Vapnik 1995). The algorithm can solve non-linear problems even with small amount of training data (Zhao and Magoules 2012). SVM is one of the most robust and accurate algorithms and has been listed in the top ten most influential data mining algorithms in the research community by the IEEE International Conference on Data Mining (Wu et al. 2008). It was found to outperform other ML algorithms in numerous applications. In order to increase the computational efficiency of SVM, least squares support vector machines (LS-SVM) (e.g., Edwards et al. 2012) and parallel support vector machines (e.g., Zhao and Magoules 2010) were also implemented in the field of building energy consumption prediction. ANN is a non-linear computational model, inspired by the human brain. A typical ANN includes three layers: the input layer, the hidden layer, and the output layer. Each layer has a number of interconnected neurons which has an activation function. Three types of parameters are typically used to define ANNs: the interconnection pattern between the neurons of the different layers; the learning process of updating the weights of the interconnections; and the activation function that converts a neuron's weighted input to its output activation (Wang and Srinivasan 2015). ANN is the most popular algorithm used in building energy consumption prediction (Ahmad et al. 2014). Based on their activation functions, ANN can be classified as back propagation neural network (BPNN), RBFNN, general regression neural network (GRNN), or feed forward neural network (FFNN). Hierarchical mixture of experts (HME), fuzzy c-means (FCC), and multilayer perceptron (MLP) are other methods that can be used in conjunction with ANN. DT algorithms use a tree to map instances into predictions. In a DT model, each non-leaf node represents one feature, each branch of the tree represents a different value for a feature, and each leave node represents a class of prediction. DT is a flexible algorithm that could grow with increased amount of training data (Domingos 2012). CART, chi-squared automatic interaction detector (CHAID), RF, and boosting tree (BT) are the most widely used DT methods in the area of building energy consumption prediction. Other statistical algorithms include multiple linear regression (MLR), general linear regression (GLR), ordinary least squares regression (OLS), autoregressive (AR), autoregressive integrated moving average (ARIMA), Bayesian regression, polynomial regression (poly), exponential regression, multivariate adaptive regression splines (MARS), ANFIS, case-based reasoning (CBR), and k-nearest neighbors (KNN). Model testing, which is the last step of developing a model, is the evaluation of the prediction model using some standard evaluation measures.

2.4.2 Data-Driven Energy Consumption Prediction Models

2.4.2.1 Scope of Prediction

The scope of the studies was classified in terms of type of building, temporal granularity, and type of energy consumption predicted. Two types of buildings (residential and non-residential), five types of temporal granularities (sub-hourly, hourly, daily, monthly, and yearly), and four types of energy consumption (heating, cooling, lighting, and overall energy consumption) were defined.

Existing models covered residential and/or non-residential buildings, with different temporal granularities and for different types of energy consumption. Figure 2.1 shows the distribution of the reviewed models according to type of building, temporal granularity, and type of energy consumption. Only 19% of these models focused on residential buildings, with the remaining models focusing on non-residential buildings including commercial and educational buildings. The majority of these models, 57%, were developed for predicting hourly energy consumption, while 12%, 15%, 4%, and 12% of the models focused on sub-hourly, daily, monthly, and yearly consumption, respectively. Overall, 47% of the models focused on predicting overall energy consumption, with 31% and 20% focusing on cooling and heating energy consumption, respectively, and only 2% focusing on lighting energy consumption prediction. The scope of each reviewed model is summarized in Table 2.1, in terms of building type, temporal granularity, type of energy consumption, and purpose of prediction.

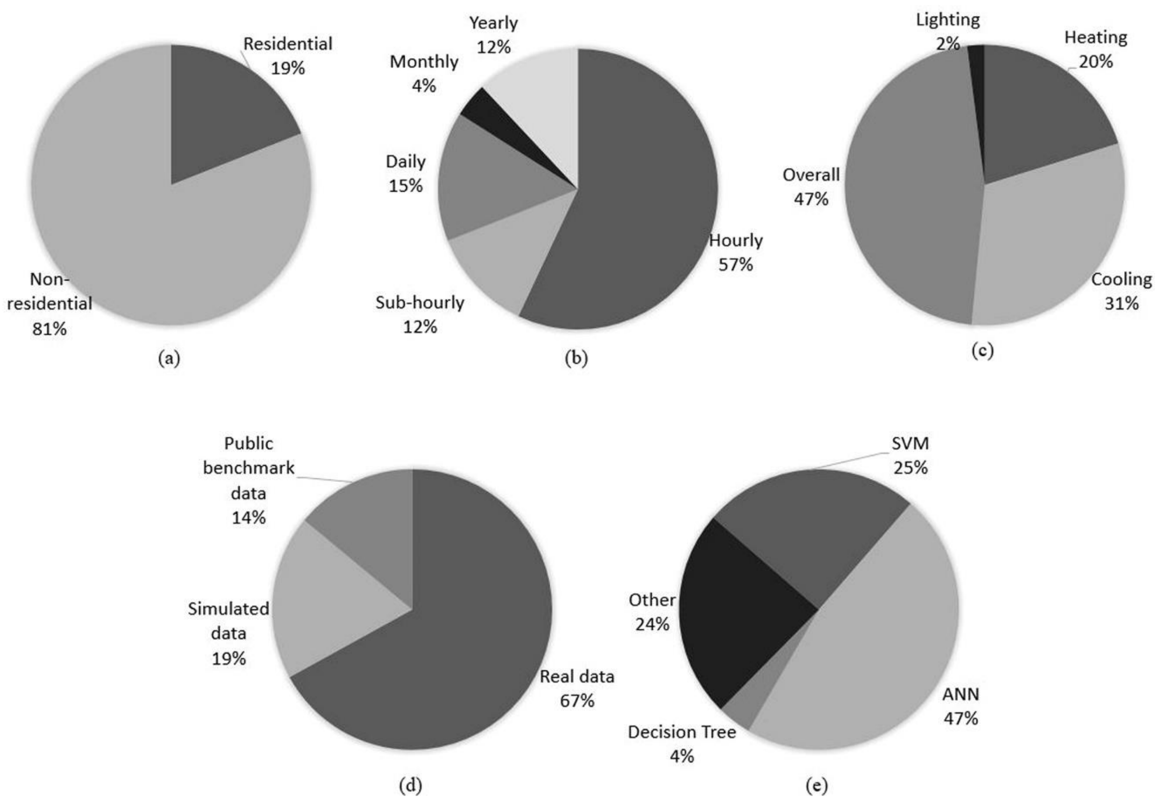


Figure 2.1 – Distribution of the Reviewed Models According to (a) Type of Building, (b) Temporal Granularity, (c) Type of Energy Consumption, (d) Type of Data, (e) Machine Learning Algorithm

2.4.2.2 Data Properties and Data Preprocessing

2.4.2.2.1 Types of Data: Real, Simulated, or Benchmark

Data were classified into three types: (1) real data, (2) simulated data, and (3) public benchmark data (e.g., datasets provided for energy consumption prediction competitions). Figure 2.1 shows the distribution of the reviewed studies by type of data used for training and testing. The majority (67%) of these studies used real data to train and test their models, while 19% and 14% of the studies used simulated and public benchmark data, respectively. Table 2.1 shows the types of data used in the reviewed studies.

Real data cover data collected through smart energy meters, sensors, building management systems, and weather stations; in addition to utility bills, energy consumption surveys, and energy consumption statistics and reports Jain et al (2014a). Sensor-based approaches have several advantages and disadvantages. On one hand, sensor-based approaches provide actual indoor environmental condition data and energy consumption levels. On the other hand, installing sensors brings an additional cost and effort not only to install the required sensors, but also to test and ensure the quality of the data collected (Edwards et al. 2012). Otherwise, sensor data may include noise, missing values, and/or outliers, which would affect the performance of the prediction models adversely.

Simulation-based studies, on the other hand, model an existing or unexisting building in a building energy simulation software tool – such as EnergyPlus, DeST, DOE2, or Ecotect – and obtain the needed data through running the simulations. By nature of modeling, a model cannot fully represent its prototype or exactly behave same as it does. For example, Li et al. (2015a) showed that current building energy software tools are, in some cases, limited in evaluating the performance of energy conservation measures. Simulation data are, however, useful in cases where real data are limited (e.g., when instrumenting a building is difficult due to technical difficulties and/or economic reasons).

Other studies [e.g., (Edwards et al. 2012, Karatasou et al. (2006), González and Zamarreño 2005] utilized publicly-available benchmark datasets such as the ASHRAE's Great Building Energy Predictor Shootout and EUNITE dataset. This type of datasets provides benchmark data that can be used to compare the performance of different models.

2.4.2.2.2 Types of Features

A machine learning model predicts energy consumption based on a set of features. These features can be related to outdoor weather conditions, indoor environmental conditions, building characteristics, time, occupancy and occupant energy-use behavior, and/or historical energy consumption. Outdoor weather condition features include dry-bulb temperature, dew point temperature, relative humidity, global solar radiation, wind speed, wind direction, degree of cloudiness, pressure, rainfall amount, and evaporation. Indoor environmental condition features include room temperature, room relative humidity, and indoor lighting level. Building characteristic features include relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, glazing area distribution, mean heat transfer coefficient of building walls, mean thermal inert index of building walls, roof heat transfer coefficient, building size coefficient, absorption coefficient for solar radiation of exterior walls, eastern WWR, western WWR, southern WWR, northern WWR, mean WWR, shading coefficient (SC) of eastern window, SC of western window, SC of southern window, SC of northern window, and integrated SC. Time features include the type of day (e.g., weekday, weekend, holiday) and the type of hour (e.g., daytime, nighttime). Occupant energy-use behavior and occupancy features include building use schedule, heat gain through lights and people, water temperature, and number of occupants.

For all these types of features, some studies used data considering various past time steps (e.g., past hour) in history. For example, Li et al. (2009) used current outdoor dry-bulb temperature, outdoor dry-bulb temperature of an hour ago, outdoor dry-bulb temperature of two hours ago, current relative humidity, current solar radiation, and solar radiation of an hour ago to predict building cooling load. Jain et al. (2014) used electricity consumption of the previous two time steps, current temperature, current solar flux, a denote for weekend/holiday or weekday, sine of current hour, and cosine of current hour to predict the electricity consumption of a multi-family residential building. Table 2.1 summarizes the features used in the reviewed models.

2.4.2.2.3 Data Sizes

The sizes of datasets varied from 2-week [e.g., (Liu and Chen 2013)] to 4-year energy consumption data [e.g. Dagnely et al. 2015, Dong et al. (2005)]. A small dataset may not be able to capture a representative sample of data, whereas a large dataset requires a lot of computational effort to process. The majority (56%) of the reviewed studies utilized one-month to one-year long datasets; 9% utilized datasets shorter than one-month; and 31% utilized datasets longer than one-year. Table 2.1 shows the dataset sizes used in the reviewed studies.

2.4.2.2.4 Data Preprocessing

Data preprocessing is essential for any data-driven approach, because any incorrect or inconsistent data can cause errors in the analysis (Hellerstein 2008). Data preprocessing may include data cleaning, data integration, data transformation, and/or data reduction. Data cleaning is the process of detecting and correcting (completing, modifying, replacing, and/or removing) the incomplete, incorrect, inaccurate, irrelevant, and/or noisy parts of the data. For example, data collected through sensors are usually noisy and often incomplete (Pattipati 2008). Data integration is the process of combining multiple data from different sources. For example, outdoor weather condition data and hourly electricity consumption data come from different sources, but are combined in a single dataset for training and testing. Data transformation is the process of transforming the data into the format that is required by the learning algorithm. Data transformation may include normalization, smoothing, aggregation/disaggregation, and/or generalization of the data. Data reduction is the process of reducing the dimensionality of the dataset, which is not only computationally more efficient but may also enhance the performance of the machine learning algorithm by removing non-discriminative features. There are different techniques for data reduction including principal component analysis (PCA) and kernel PCA (KPCA). For example, Xuemei et al. (2010b) applied PCA and KPCA for reducing the dimensionality of the data and compared the performances of SVM with PCA, SVM with KPCA, and SVM without any data reduction techniques. They also applied C-mean clustering to ensure that the training samples were chosen based on the similarity degree of the input samples and compared the performances of fuzzy C-means (FCM) fuzzy SVM, FCM-SVM, and SVM without any clustering (Xuemei et al. 2010a).

2.4.2.3 Machine Learning Algorithms

A machine learning algorithm is needed to train an energy consumption prediction model. Previous studies in data-driven building energy consumption prediction have utilized SVM, ANN, decision trees, and/or other statistical algorithms. Figure 2.1 shows the distribution of the studies by type of machine learning algorithm. Overall, 47% and 25% of the studies utilized ANN and SVM, respectively, to train their models. Only 4% of the studies utilized decision trees. On the other hand, 24% of the studies utilized other statistical algorithms such as MLR, OLS, and ARIMA.

Some studies also compared the effectiveness of different algorithms in energy consumption prediction. For example, Li et al. (2009a) compared SVM and BPNN; Borges et al. (2013a) compared SVM and AR; Xuemei et al. (2009) compared LS-SVM and BPNN; Liu and Chen

(2013) compared SVM and ANN; Penya et al. (2011b) compared poly, exponential, mixed, AR, ANN, SVM, and Bayesian Network; Platon et al. (2015) compared ANN and CBR; Jain et al. (2014b) compared SVM and MLR; Hou et al. (2006) compared ARIMA and ANN; Penya et al. (2011a) compared AR, ARIMA, ANN, and Bayesian Network; Fan et al. (2014) compared MLR, ARIMA, SVM, RF, MLP, BT, MARS, and KNN; Chou and Bui (2014) compared ANN, SVM, CART, CHAID, and GLR; Edwards et al. (2012) compared MLR, FFNN, SVM, LS-SVM, HME-FFNN, and FCM-FFNN; Li et al. (2009b) and Li et al. (2010) compared SVM, BPNN, RBFNN, and GRNN; Dagnely et al. (2015) compared OLS and SVM; Massana et al. (2015) compared MLR, MLP, and SVM; and Fernandez et al. (2011) compared AR, poly, ANN, and SVM.

2.4.2.4 Performance Evaluation

Model testing is the evaluation of the prediction model using some standard evaluation measures. The most commonly-used evaluation measures of energy consumption prediction models are CV, RMSE and R^2. These measures can be calculated using Eq. (1.1) to (1.3). Other measures used for evaluating energy consumption prediction include mean absolute percentage error (MAPE) [also referred to as absolute proportional error (APE), mean relative error (MRE), and absolute relative error (ARE)], mean absolute error (MAE), mean bias error (MBE), and mean squared error (MSE). These measures can be calculated using Eq. (2.1) to (2.4). Precision and recall are also used occasionally for performance evaluation.

$$\text{Mean Absolute Percentage Error (MAPE)(\%)} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{predict,i}-y_{data,i}}{y_{data,i}}\right| \times 100 \tag{2.1}$$

$$\text{Mean Absolute Error (MAE)} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{predict,i} - y_{data,i}\right| \tag{2.2}$$

$$\text{Mean Bias Error (MBE)} = \frac{\frac{\sum_{i=1}^{n}\left(y_{predict,i} - y_{data,i}\right)}{n}}{\bar{y}_{data}} \times 100 \tag{2.3}$$

$$\text{Mean Squared Error (MSE)} = \frac{1}{n}\sum_{i=1}^{n}(y_{predict,i} - y_{data,i})^2 \tag{2.4}$$

where $y_{predict,i}$ is the predicted energy consumption at time point i, $y_{data,i}$ is the actual (or simulated) energy consumption at time point i, $\bar{y}_{data}$ is the average energy consumption, and n is the total number of data points in the dataset.

Table 2.1 Scope, Data Properties, Algorithms, and Performance of the Energy Consumption Prediction Models

Reference	Learning algorithm	Spatial scale (building type)	Temporal granularity	Type of energy consumption predicted	Type of dataset (simulation tool)	Types of input data	Amount of data	Performance (metric)
Xuemei et al. (2010a)	SVM	Building (non-residential)	Hourly	Cooling	Simulated (N/A)	Date, daily average temperature, daily lowest temperature, daily highest temperature, cooling load	620 instances	0.17 (RMSE)
	PCA-SVM							0.04 (RMSE)
	KPCA-SVR							0.02 (RMSE)
Jinhu et al. (2010)	SVM	Building (non-residential)	Hourly	Cooling	Simulated (N/A)	Date, daily average temperature, daily lowest temperature, daily highest temperature, cooling load	620 instances	0.17 (RMSE)
	PCA-SVM							0.04 (RMSE)
	PCA-WSVM							0.03 (RMSE)
Li et al. (2009a)	SVM	Building (non-residential)	Hourly	Cooling	Simulated (DeST)	Dry-bulb temperature, relative humidity, solar radiation	5 months	1.15% – 1.18% (RMSE)
	BPNN							2.22% - 2.36% (RMSE)
Li et al. (2009b)	SVM	Building (non-residential)	Hourly	Cooling	Simulated (DeST)	Dry-bulb temperature, relative humidity, solar radiation	5 months	1.15% – 1.18% (RMSE)
	BPNN							2.22% - 2.36% (RMSE)
	RBFNN							1.43% - 1.51% (RMSE)
	GRNN							1.19% – 1.20% (RMSE)
Yu et al. (2014)	SVM	Building (non-residential)	Hourly	Cooling	Real (N/A)	Outside temperature, date, humidity, discomfort-index, wind velocity, solar irradiation index, degree of cloudiness	~12 months	93.7 – 96.6 (recall) 93.8 – 95.6 (precision)
Xuemei et al. (2009)	LS-SVM	Building (non-residential)	Hourly	Cooling	Simulated (DeST)	Dry-bulb temperature, relative humidity, solar radiation	4 months	5.56% (RMSE)
	BPNN							11.84% (RMSE)
Xuemei et al. (2010b)	SVM	Building (non-residential)	Hourly	Cooling	Real (N/A)	N/A	6 months	3.85% (RMSE)
	FCM-SVM							2.68% (RMSE)
	FCM-FSVM							1.24% (RMSE)
Solomon et al. (2011)	SVM	Building (non-residential)	Hourly	Overall	Real (N/A)	Temperature, dew point temperature, pressure, wind direction, wind speed, humidity, precipitation	~27.5 months	0.71 – 0.95 (R^2)

Table 2.1 (cont.)

Reference	Learning algorithm	Spatial scale (building type)	Temporal granularity	Type of energy consumption predicted	Type of dataset (simulation tool)	Types of input data	Amount of data	Performance (metric)
Chou and Bui (2014)	ANN	Building (residential)	Hourly	Cooling	Simulated (Ecotect)	Relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, glazing area distribution	N/A	1.68 (RMSE)
	SVM							1.65 (RMSE)
	CART							1.84 (RMSE)
	CHAID							1.86 (RMSE)
	GLR							1.74 (RMSE)
	ANN			Heating				0.61 (RMSE)
	SVM							0.35 (RMSE)
	CART							0.80 (RMSE)
	CHAID							0.91 (RMSE)
	GLR							1.04 (RMSE)
Edwards et al. (2012)	MLR	Building (residential)	Hourly	Overall	Real (N/A)	140 different sensor data	A year	26.27% - 38.53% (CV)
	FFNN							24.32% - 37.15% (CV)
	SVM							21.32% - 31.88% (CV)
	LS-SVM							20.05% – 30.66% (CV)
	HME-REG							26.14% - 38.22% (CV)
	HME-FFNN							20.15% - 32.98% (CV)
	FCM-FFNN							20.53% - 32.92% (CV)
	MLR	N/A			ASHRAE dataset	Temperature, solar flux, date, sin of the current hour, cosine of the current hour	6 months	4.07% (CV)
	FFNN							2.93% (CV)
	SVM							3.97% (CV)
	LS-SVM							6.35% (CV)
	HME-REG							4.05% (CV)
	HME-FFNN							2.75% (CV)
	FCM-FFNN							2.71% (CV)
Dagnely et al. (2015)	OLS	Building (non-residential)	Hourly	Overall	Real (N/A)	Occupancy, recency, temperature, irradiance, time	4 years	2.05 (MAE)
	SVM							1.94 (MAE)
Massana et al. (2015)	MLR	Building (non-residential)	Hourly	Overall	Real (N/A)	Temperature, relative humidity, solar radiation, indoor temperature, indoor relative humidity, indoor light level, occupancy, date	~A year	4.68% (MAPE)
	MLP							0.45% (MAPE)
	SVM							0.06% (MAPE)

Table 2.1 (cont.)

Reference	Learning algorithm	Spatial scale (building type)	Temporal granularity	Type of energy consumption predicted	Type of dataset (simulation tool)	Types of input data	Amount of data	Performance (metric)
Penya et al. (2011a)	Poly	Building (non-residential)	Hourly	Overall	Real (N/A)	Day of the week, type of day, season, wind direction, humidity, precipitation, sigma direction, sigma speed, air temperature, average speed	N/A	7.43% - 13.86% (MAPE)
	Exponential							18% (MAPE)
	Mixed							7.59% - 23.00% (MAPE)
	AR							5.30% - 8.72% (MAPE)
	ANN							7.89% - 12.55% (MAPE)
	SVM							5.79% - 9.28% (MAPE)
	Bayesian Network							5.92% - 11.31% (MAPE)
Zhao and Magoules (2009)	SVM	Building (non-residential)	Hourly	Overall	Simulated (EnergyPlus)	day type indications if the current day is holiday or not, weather conditions, zone mean air temperatures, infiltration volume, heat gain through each window, heat gain through lights and people, zone internal total heat gain	100 instances	0.00 (MSE)
	Parallel SVM							0.00 (MSE)
Paudel et al. (2015)	SVM	Building (non-residential)	Hourly	Overall	Real (N/A)	Outside air temperature	7 months	50 - 51 (RMSE)
Liu and Chen (2013)	SVM	Building (non-residential)	Hourly	Lighting	Real (N/A)	Number of people in building, solar radiation	168 instances	0.66 (MSE)
	ANN							3.14 (MSE)
Jain et al. (2014a)	SVM	N/A	Hourly	Overall	ASHRAE dataset	Temperature, solar flux, date, sine of current hour, cosine of current hour	6 months	3.30% (CV)
		Building (residential)	Sub-hourly,		Real (N/A)	Temperature, date, sine of current hour, cosine of current hour	~3.5 months	10.47% - 133.24% (CV)
			Hourly					2.16% - 11.30% (CV)
			Daily					5.52% - 11.39% (CV)
Jain et al. (2014b)	SVM	Building (residential)	Sub-hourly	Overall	Real (N/A)	Outdoor temperature, data, sine of current hour, cosine of current hour	84 days	10.66 – 202.29 (CV)
	MLR							14.88 – 86.18 (CV)
	SVM		Hourly					11-31 – 88.71 (CV)
	MLR							12.03 – 97.39 (CV)

Table 2.1 (cont.)

Reference	Learning algorithm	Spatial scale (building type)	Temporal granularity	Type of energy consumption predicted	Type of dataset (simulation tool)	Types of input data	Amount of data	Performance (metric)
Mena et al. (2014)	ANN	Building (non-residential)	Hourly	Overall	Real (N/A)	Date, outdoor temperature, outdoor humidity, solar radiation, outdoor wind speed, outdoor wind direction, state of the pumps, state of the boilers, state of the absorption machine, state of the cooling tower, state of the heat pump	18 months	0.81% - 1.73% (MAPE)
Kwok and Lee (2011)	ANN	Building (non-residential)	Hourly	Cooling	Real (N/A)	Outdoor temperature, relative humidity, rainfall, wind speed, bright sunshine duration, solar radiation, occupancy area, occupancy rate	1053 instances	11.41% - 17.17% (RMSPE)
Yezioro et al. (2008)	ANN	Building (non-residential)	Monthly, yearly	Cooling, heating	Simulated	Outdoor temperature, relative humidity, setpoint temperature, occupancy schedule	159 instances	N/A
Ben-Nakhi and Mahmoud (2004)	ANN	Building (non-residential)	Hourly	Cooling	Simulated (ESP-r)	Temperature	4 years	0.91 – 0.96 (R^2)
Hou et al. (2006)	MRAN	N/A	Hourly	Cooling	Real (N/A)	Parameters of 11 ahus	288 hours	3.65% (MRE)
	ARIMA							9.17% (MRE)
Cardenas et al. (n.d.)	ANN	Building (non-residential)	Hourly, daily	Overall	Real (N/A)	Week day, time in minutes, continuous working day	10 months	0.05 (RMSE)
Penya et al. (2011b)	AR	Building (non-residential)	Hourly	Overall	Real (N/A)	Day of the week, type of day, season, wind direction, humidity, precipitation, sigma direction, sigma speed, air temperature, average speed, temperature humidity index, wind chill index	N/A	4.26% – 8.14% (MAPE)
	ARIMA							13.54% – 19.13% (MAPE)
	ANN							3.46% - 4.11% (MAPE)
	Bayesian Network							6.87% - 22.75% (MAPE)
Yokoyama et al. (2009)	ANN	Building (non-residential)	Hourly	Cooling	Real (N/A)	Air temperature, relative humidity	45 days	7.89% (RMSE)

Table 2.1 (cont.)

Reference	Learning algorithm	Spatial scale (building type)	Temporal granularity	Type of energy consumption predicted	Type of dataset (simulation tool)	Types of input data	Amount of data	Performance (metric)
Platon et al. (2015)	ANN	Building (non-residential)	Hourly	Overall	Real (N/A)	Outside air temperature, outside air relative temperature, boiler outlet water temperature, boiler outlet water flowrate, chiller outlet water temperature, chiller outlet water flowrate, supply air temperatures - hot duct for ahus, supply air temperatures - cold duct for ahus, supply air fan VFD control settings for ahus, return air fan VFD control settings for ahus, indoor air temperatures of different zones	A year	8.38% - 8.48% (CV-RMSE)
	CBR							13.37% - 14.32% (CV-RMSE)
Karatasou et al. (2006)	ANN	Building (non-residential)	Hourly	Overall	ASHRAE dataset	Temperature, solar flux, humidity, wind speed, date, sine and cosine of the hour of the day, sin and cosine of the day of the week, sin and cosine of the day of the year	6 months	2.44 (CV)
					Real (N/A)	Temperature, humidity, date, sine and cosine of the hour of the day, sin and cosine of the day of the week, sin and cosine of the day of the year	A year	2.95 (CV)
Gonzalez and Zamarreno (2005)	ANN	Building (non-residential)	Hourly	Overall	ASHRAE dataset	Current and forecasted temperature, date	N/A	1.50 (CV)
					Proben dataset		N/A	2.55 (CV)
Yang et al. (2005)	ANN	Building (non-residential)	Hourly	Cooling	Simulated (N/A)	On/off status of compressors, temperature of the water entering the ice tank, temperature of the water entering the evaporator, temperature of the water leaving the evaporator, outdoor relative humidity, outdoor temperature, is the chilled water prepared in the ice tanks, percentage of chilled water prepared in the ice tanks, holiday indicator, date	9 months	0.04 (CV)
					Real (N/A)			0.23 (CV)

Table 2.1 (cont.)

Reference	Learning algorithm	Spatial scale (building type)	Temporal granularity	Type of energy consumption predicted	Type of dataset (simulation tool)	Types of input data	Amount of data	Performance (metric)
Kwok et al. (2011)	ANN	Building (non-residential)	Hourly	Cooling	Real (N/A)	Outdoor temperature, relative humidity, rainfall, wind speed, bright sunshine duration, solar radiation, occupancy area, occupancy rate	1053	12.12% - 16.36% (RMSPE)
Fernandez et al. (2011)	AR	Building (non-residential)	Hourly	Overall	Real (N/A)	Time	18 months	7.34% - 13.78% (MAPE)
					ASHRAE dataset		6 months	5.74% (MAPE)
					EUNITE dataset		24 months	6.69% (MAPE)
	SVM				Real (N/A)		18 months	7.92 - 14.25% (MAPE)
					ASHRAE dataset		6 months	5.88% (MAPE)
					EUNITE dataset		24 months	7.34% (MAPE)
	Poly				Real (N/A)		18 months	11.91 - 19.78% (MAPE)
					ASHRAE dataset		6 months	6.94% (MAPE)
					EUNITE dataset		24 months	7.36% (MAPE)
	ANN				Real (N/A)		18 months	13.46 – 17.64% (MAPE)
					ASHRAE dataset		6 months	6.63% (MAPE)
					EUNITE dataset		24 months	7.78% (MAPE)
Georgescu et al. (2014)	SVM	Building (mixed)	Sub-hourly	Overall	Real (N/A)	Outdoor air temperature, relative humidity, solar radiation, wind speed, wind direction	A year	N/A
Paudel et al. (2014)	ANN	Building (non-residential)	Sub-hourly	Heating	Real (N/A)	Outside temperature, solar radiation, work/off day, occupancy profiles, operational power level characteristics, transitional characteristics	27 days	0.15 (MSE)
Escrivá-Escrivá et al. (2011)	ANN	Building (non-residential)	Sub-hourly	Overall	Real (N/A)	External temperature	N/A	3.16 (MAPE)
Kamaev et al. (2012)	ANN	Building (non-residential)	Sub-hourly	Overall	Real (N/A)	Current temperature of external environment, status, building usage profile,	N/A	14.78 kwh (RMSE)
Bagnasco et al. (2015)	ANN	Building (non-residential)	Sub-hourly	Overall	Real (N/A)	Date, 24-hour-ahead average load, day-ahead load, 7 days-ahead load, day-ahead temperature	A year	5.64% - 8.97% (MAPE)

Table 2.1 (cont.)

Reference	Learning algorithm	Spatial scale (building type)	Temporal granularity	Type of energy consumption predicted	Type of dataset (simulation tool)	Types of input data	Amount of data	Performance (metric)
Borges et al. (2013a)	SVM	Building (non-residential)	Hourly	Overall	Real (N/A)	Time	~18 months	6.38% - 13.29% (MAPE)
					ASHRAE dataset		6 months	4.62% (MAPE)
	AR				Real (N/A)		~18 months	6.03% - 12.86% (MAPE)
					ASHRAE dataset		6 months	4.63% (MAPE)
Fan et al. 2014	MLR	Building (non-residential)	Daily	Overall	Real (N/A)	Maximum dry-bulb temperature, average dry-bulb temperature, minimum dry-bulb temperature, average dew point temperature, average relative humidity, average pressure, average amount of cloud, total rainfall, number of hours of reduced visibility, solar radiation, total evaporation, average wind speed	A year	4.23% (MAPE)
	ARIMA							5.45% (MAPE)
	SVM							3.11% (MAPE)
	RF							3.17% (MAPE)
	MLP							4.75% (MAPE)
	BT							4.07% (MAPE)
	MARS							3.97% (MAPE)
	KNN							4.01% (MAPE)
Lai et al. (2008)	SVM	Building (residential)	Daily	Overall	Real (N/A)	Date, outdoor temperature, bedroom temperature, living temperature, living room humidity, bedroom humidity, outdoor humidity, water temperature	15 months	0.88 (Pearson coefficient)
Iwafune et al. (2014)	MLR	Building (residential)	Daily	Overall	Real (N/A)	Outside temperature, date	3 years	2.39 kwh/day (RMSE)
Wong et al. (2010)	ANN	Building (non-residential)	Daily	Cooling	Simulated	Daily average dry-bulb temperature, daily average wet-bulb temperature, daily global solar radiation, daily average clearness index, solar aperture, daylight aperture, overhang, side-fins projections, date	8760 instances	725 - 1410 kWh (RMSE)
				Heating				607 - 785 kWh (RMSE)
				Lighting				224 - 396 kWh (RMSE)
				Overall				2118 – 2904 kWh (RMSE)

Table 2.1 (cont.)

Reference	Learning algorithm	Spatial scale (building type)	Temporal granularity	Type of energy consumption predicted	Type of dataset (simulation tool)	Types of input data	Amount of data	Performance (metric)
Lam et al. (2010)	MLP	Building (non-residential)	Daily	Cooling	Simulated	Temperature, humidity, solar radiation	29 years	1.4 – 1.5 MWh (RMSE)
				Overall				1.2 – 1.3 MWh (RMSE)
Dong et al. (2005)	SVM	Building (non-residential)	Monthly	Overall	Real (N/A)	Dry-bulb temperature, relative humidity, solar radiation	4 years	0.99% – 2.69% (CV)
Popescu and Ungureanu (2013)	ANN	Building (residential)	Monthly	Heating	Real (N/A)	Monthly average external temperature, heat transfer rate through the envelope, heat transfer rate through the wall next to the staircase, heat flow rate due to infiltration/natural ventilation, solar gain through transparent elements, internal gains, income level, occupant per room	5 years	0.83 (R)
Ferlito et al. (2015)	ANN	Building (non-residential)	Monthly	Overall	Real (N/A)	N/A	3 years	15.70% - 17.97% (RMSPE)
Li et al. (2010)	SVM	Building (residential)	Yearly	Overall	Real (N/A)	Average heat transfer coefficient of building walls, mean thermal inert index of building walls, roof heat transfer coefficient, building size coefficient, absorption coefficient for solar radiation of exterior walls, eastern window-wall ratio, western window-wall ratio, southern window-wall ratio, northern window-wall ratio, mean window-wall ratio, shading coefficient of eastern window, shading coefficient of western window, shading coefficient of southern window, shading coefficient of northern window and integrated shading coefficient	59 instances	2.40% (RMSE)
	BPNN							14.46% (RMSE)
	RBFNN							12.44% (RMSE)
	GRNN							5.24% (RMSE)

Table 2.1 (cont.)

Reference	Learning algorithm	Spatial scale (building type)	Temporal granularity	Type of energy consumption predicted	Type of dataset (simulation tool)	Types of input data	Amount of data	Performance (metric)
Turhan et al. (2014)	ANN	Building (residential)	Yearly	Overall	Simulated	Width/length, wall overall heat transfer coefficient, area/volume, total external surface area, total window area/total external surface area	148 instances	5.06 (MAPE)
Hawkins (2012)	ANN	Building (mixed)	Yearly	Overall except heating	Real (N/A)	Building activity, building environment, heating fuel, age, primary material, geometry data, adjacency shading data, adjacency sheltering factor, orientation, glazing, weather data	1872 instances	34.00% (MAPE)
				Heating				25.10% (MAPE)
Ekici and Aksoy (2009)	ANN	N/A	Yearly	Overall	Simulated (N/A)	Transparency ratio, orientation, insulation thickness	45 instances	94.83% - 98.51% (accuracy)
Yu et al. (2010)	DT	Building (residential)	N/A	Energy use intensity	Simulated (N/A)	Annual average air temperature, house type, construction type, floor area, heat loss coefficient, equivalent leakage area, number of occupants, space heating, hot water supply, kitchen	80 instances	95.16 (recall) 95.17 (precision)
Yi and Ying (2010)	LS-SVM	Country (N/A)	Yearly	Overall	Real (N/A)	GDP, population, import, export, the ratio of the tertiary industry to GDP, government expenditure	19 years	1.57% - 5.01% (relative error)
Jain and Satish (2009)	SVM	N/A	Sub-hourly	Overall	Real (N/A)	Average temperature, forecasted day's average temperature, day of the week	2 years	0.58% - 2.05% (relative error)
Xing-Ping and Rui (2007)	SVM	Country (N/A)	Yearly	Overall	Real (N/A)	GDP per capita, heavy industry share, efficiency improvement	21 years	1.94 (APE)
Borges et al. (2013b)	Poly	State (N/A)	Hourly	Overall	Real (N/A)	Time	3 years	7.55%
		City (N/A)					10 years	6.79%
	SVM	State (N/A)					3 years	7.45%
		City (N/A)					10 years	6.38%
	AR	State (N/A)					3 years	7.22%
		City (N/A)					10 years	6.28%
Zhang and Qi (2009)	SVM	City (N/A)	Hourly	Heating	Real (N/A)	Temperature difference to multiply the flux of first thermo network	120 days	0.00 – 0.06 (relative error)
Chen et al. (2004)	SVM	N/A	Daily	Overall	EUNITE dataset	Average daily temperature, dates of holidays	24 months	1.95% - 2.86% (MAPE)

Table 2.1 (cont.)

Reference	Learning algorithm	Spatial scale (building type)	Temporal granularity	Type of energy consumption predicted	Type of dataset (simulation tool)	Types of input data	Amount of data	Performance (metric)
Farzana et al. (2014)	ANN	Residential buildings of a city (N/A)	Yearly	Overall	Real (N/A)	Locale (i.e., urban and rural), total population in urban areas, average number of people per household, electrification rate, penetration of device or appliance, types of lighting bulb, number of lighting bulb of type per household, power of bulb of type, hours of use of bulb of type, fuel type, lighting energy use of fuel, cooking and water heating energy use of fuel per household per year, space heating and cooling energy use of fuel, other end use devices	12 years	0.09% (MRPE)
Junfang and Dongxiao (2009)	LS-SVM	Province (N/A)	Daily	Overall	Real (N/A)	Temperature-related values, calendar information	A year	N/A
Mohandes (2002)	SVM	Province (N/A)	Hourly	Overall	Real (N/A)	Time	6 years	0.02 (RMSE)
Xu et al. (2015)	ARMA	Province (N/A)	Yearly	Overall	Real (N/A)	Time	23 years	7.17% (ARE)

2.4.3 Analysis of the Reviewed Studies

2.4.3.1 Temporal Granularities of Prediction Models

Both short-term (e.g., sub-hourly, hourly, or daily) and long-term (e.g., yearly) energy consumption prediction are essential for building and grid design and operation. For example, “HVAC operations including adjusting the starting time of cooling to meet start-up loads, minimizing or limiting the electric on-peak demand, optimizing the costs and energy utilization in cool storage systems, and related energy and cost needs in other HVAC systems” all benefit from short-term energy consumption prediction (Xuemei et al. 2010b). Short-term energy consumption prediction models are also utilized for maintaining economic and secure operation of power grids and for providing energy consumption data to building occupants to better negotiate energy prices with energy retailers (Fernandez et al. 2011). Among the reviewed literature, 84% of the studies focused on short-term energy consumption prediction because of its direct relation to the day-to-day operations of buildings (Fan et al. 2014).

Only 12% of the studies focused on long-term (yearly) energy consumption prediction. This might be caused by several reasons. First, to achieve good performance, long-term energy consumption prediction requires a relatively higher amount of data that covers a long time span (Zhao 2011). For example, prediction errors of annual energy consumption prediction models, which were developed based on 1-day, 1-week, and 3-month measurements, were 100%, 30%, and 6%, respectively (Cho et al 2004). Second, nonlinearity in long-term data is usually more prominent compared to short-term data (Li et al. 2015b). Third, uncertainties in long-term energy consumption prediction are usually higher because of the many changes that may occur in the supply and demand over a long time span. Long-term energy consumption prediction, thus, requires specific long-term prediction models due to the non-homogeneity and significant changes that may occur on the long-run (Azadeh et al. 2008). Despite their challenges, long-term energy consumption prediction models are essential; they are required when studying decisions of long-term implications such as capacity expansion, energy supply strategy, and capital investment (Ekonomou 2010).

2.4.3.2 <u>Building Types</u>

About 81% of the reviewed research efforts focused on developing energy consumption prediction models for commercial and/or educational buildings, with only 19% focusing on residential buildings. The relative lack of studies on residential buildings could be due to a number of reasons. First, the lack of data – specifically sensor-based data – could be a main reason. The majority, 73%, of non-residential building energy consumption prediction models rely on sensor data for algorithm training. Such data are much harder to obtain for residential buildings because the majority of buildings are not sufficiently metered in a way that allows for sensing at high granularity (Wang and Srinivasan 2017). Another reason could be the complexity of predicting energy consumption in residential contexts because of the relatively higher variability of occupant behavior compared to the commercial context (Jain et al. 2014a). Occupant behavior is the greatest uncertainty in building energy consumption prediction (O'Brien and Gunay 2015); ignoring, misunderstanding, and/or underestimating the role of occupant behavior in affecting energy consumption is one of the main causes for the deviations between the predicted and the actual consumption levels (Azar and Menassa 2012a).

Despite their challenges, residential building energy consumption predictions are needed because of the high energy consumption share of this sector and the potential high gain that can be achieved if successful energy reducing strategies are implemented. Residential buildings represent 21% of the total energy consumption in the US, which is greater than the share of commercial buildings (EIA 2019). Further studies are, thus, needed on the residential sector. For example, experimental studies could be conducted to see if/how existing data-driven commercial building energy consumption prediction models could be extended to the residential context.

2.4.3.3 Energy Consumption Types

As discussed in Section 2.4.2, 46%, 31%, 20%, and 2% of the reviewed research efforts focused on predicting overall, cooling, heating, and lighting energy consumption, respectively. This shows a relative lack of studies on predicting lighting loads. This might be caused by the predominant impact of occupant behavior on lighting energy consumption. Lighting use is directly impacted by building occupancy and occupant behavior patterns (Yun et al. 2012). For example, 500 lx is the recommended illuminance level for office buildings (CIE 2001). Theoretically, people who have access to natural lighting, when the outdoor illumination is sufficient, are expected to use artificial lightings less (Zhou et al. 2015). However, Yun et al. (2012) showed that there are no statistically significant relationships between outdoor illuminance and artificial lighting use patterns.

Despite these reasons, lighting energy consumption prediction is essential for building energy efficiency and for efficient supply-side management. Lighting represents almost 20% of the global electricity consumption (IEA 2015). Since it is a major heat source, lighting is not only a significant piece of building energy consumption by itself, but it also impacts the cooling energy demand (Chae et al. 2016). In general, one-third of the cooling energy consumption can be saved if a good balance between natural light and solar heat can be achieved (Wong et al. 2010). In addition, different building design features – in terms of building envelope, architectural features, and building materials – may have different impacts on lighting energy consumption (Cheng et al. 2015). Lighting energy consumption prediction models, thus, require more attention to better understand lighting energy consumption trends and conservation opportunities, the interaction between cooling load and lighting, and the impacts of various design features on consumption levels.

2.5 Hybrid Modeling Approach

In recent years, there has been growing interest in taking a hybrid modeling approach to building energy consumption prediction due to their unique capability to leverage the strengths and eliminate the limitations of the traditional data-driven and physical modeling approaches, by coupling them (Wang and Srinivasan 2017). The hybrid approach, often, requires relatively less training data and only a rough description of the building geometry (Foucquier et al. 2013) and still outperforms the two traditional approaches (Amasyali and El-Gohary 2018). There are two major strategies for coupling data-driven and physical approaches in the area of building energy consumption prediction: (1) using machine learning to estimate parameters of the physical model and (2) using simulation-generated data to train machine learning models (Foucquier et al. 2013). The first strategy uses an algorithm to calibrate the physical model for more accurate predictions. For example, Ramos Ruiz et al. (2016) used the NSGA-II for building envelope calibration. New et al. (2012) proposed a novel machine-learning-based tool, “Autotune”, to calibrate EnergyPlus models in an automated fashion. The second strategy uses a physical model to generate a dataset for training a machine learning-based prediction model. For example, Amasyali and El-Gohary (2017) developed a deep neural network (DNN) model to predict cooling energy consumption of a building using a dataset generated by EnergyPlus simulations in five locations. And, Li and Huang (2013) developed short-term load prediction models using the data generated by TRNSYS simulations.

2.6 Time-Series Clustering

Clustering aims to organize unlabeled data into homogenous groups where the within-group-object similarity is minimized, and the between-group-object dissimilarity is maximized (Liao 2005). Due to the advances in data storages and processors in recent years, data in many fields (e.g., energy, stock market, and weather) are now available in time-series format (Aghabozorgi et al. 2015). Time-series clustering – a type of clustering to handle time-series data – is usually conducted by modifying the existing clustering methods for static data in a way that time-series data can be handled, or by converting time-series data into static data so that the existing clustering methods for stationary data (e.g., partitional, hierarchical, and density based) can directly be used. The former approach is called raw-data-based approach because it does not require any modification in the data. Instead, the distance/similarity measure used for the static data is replaced with another measure which is suitable for time-series data. The latter approach includes feature-based and model-based approaches. In the feature-based approach, a set of features are extracted from the time series, and then a conventional clustering method is applied to the extracted features. The model-based approach assumes that each time series is generated by some kind of model or by a mixture of underlying probability distributions and applies a conventional clustering method to the parameters of the assumed model (Liao 2005).

In time-series clustering, the choice of the distance measure is more important than the choice of the clustering method (Roelofsen 2018). The distance measure is used to assess the similarity between two time series. There are three popular distance measures: Manhattan, Euclidean, and dynamic time warping (DTW) distances. The Manhattan (l_1) and Euclidean (l_2) distances are the most commonly used due to high their computational efficiencies. However, these measures can only be used for time-series with equal lengths and are sensitive to noise, scale, and time shifts. To overcome these limitations, DTW was introduced (Ratanamahatana and Keogh 2004) to find an optimal alignment of a pair of time-series data via nonlinear mapping so that the two time-series match each other to the best extent (Matsumoto et al. 2018). DTW has been extensively used for time-series clustering because it can be used for series with different lengths. DTW, however, is relatively computationally expensive. For this reason, a number of modifications to the DTW were proposed. Such modifications were classified into three main categories adding constraints, abstracting the data, and indexing. For the details of these categories, the readers can refer to Roelofsen (2018).

Load profiling is one of the many application areas of time-series clustering in the field of building energy efficiency. Load profiling is a procedure to classify temporal subsequences of energy data for characterizing customer behavior (Miller et al. 2018). For example, Park et al. (2019) identified three fundamental load-shape profiles using a k-means algorithm and discussed the potential use of the identified profiles for portfolio management. Shahzadeh et al. (2015) applied a k-means clustering-based load profiling to enhance the performance of load forecasting. Pan et al. (2015) proposed a kernel Principal Component Analysis (PCA)-based nonparametric clustering method to better understand how to provide more efficient demand-oriented services. Panapakidis et al. (2015) proposed a new clustering algorithm, tested the performance of the new algorithm against the existing algorithms (e.g., k-means), and discussed the potential implementation of the cluster for demand side management.

2.7 Weather Normalization

Weather normalization aims to remove the contribution of weather conditions to building energy consumption. It is a procedure to adjust energy consumption to a hypothetical common scale (Wang et al. 2017). Thus, it provides a more equitable comparison of energy performance under different weather conditions. For example, weather normalization enables the comparison of the energy performance of a building in a colder climate, such as in Chicago, with a building in a warmer climate, such as in Phoenix, by removing the contribution of weather conditions. Such a normalization would also allow the assessment of the impact of building energy efficiency measures over time. For example, it would help understand whether a drop in energy consumption is due to retrofitting or weather changes. Depending on the purpose of comparison and benchmarking, energy consumption of a building can be also normalized per visitor, employee, and/or floor area (Schmidt and Ahlund 2018).

There are three main weather normalization methods: degree-days method, modified utilization factor (MUF) method, and climate severity index (CSI) method. The degree-day method is one of the most popular weather normalization methods. It utilizes the heating and cooling degree-days

(HDD and CDD) as a factor to quantify how cold or how hot the temperature was for a given period. For example, cooling energy consumption of a month can be adjusted by a CDD factor (Wang et al. 2017). The MUF method computes the actual and normalized energy consumption by adjusting indoor temperature to the set-point temperature. It is used to normalize space heating loads only (Wang et al. 2017), because normalizing space cooling loads using the MUF method is complicated (Beheshti et al. 2019). The CSI method uses temperature, solar radiation, and wind speed to normalize energy consumption. It utilizes an index to adjust building energy consumption for different regions (Wang et al. 2017).

CHAPTER 3 – ENERGY-RELATED VALUES AND SATISFACTION LEVELS OF RESIDENTIAL AND OFFICE BUILDING OCCUPANTS

3.1 Research Methodology

Two questionnaire surveys were conducted to solicit the input of a randomly selected set of residential and office building occupants in AZ, IL, and PA on (1) the importance levels of occupant values and (2) the current satisfaction levels with these values. The scope of the energy studies are focused on IL and PA. AZ was additionally selected to capture potential variability in responses as a result of a different climate, which provides an opportunity of investigating the impact of climate on occupant values and satisfaction level with the values. According to the Köppen-Geiger climate classification (Peel et al. 2007), IL and PA have a humid continental (warm summer) climate (Dfa), whereas AZ has a dessert climate (Bwh). The research methodology was composed of four primary research subtasks: (1) questionnaire design, (2) validation of questionnaire design, (3) respondent recruitment and survey implementation, and (4) survey results analysis.

3.1.1 Questionnaire Design

Two different questionnaires were used, one for residential and one for office building occupants. Both questionnaires were composed of four sections. Section 1 included two filtering questions that were asked to verify eligibility of participation in terms of occupancy type and residency state (e.g., for the office survey, occupancy of an office building and residency in AZ, IL, or PA). Responses which failed to pass Section 1 were disregarded. In Section 2, respondents were asked to rate the importance levels of occupant values to them on a 6-point Likert scale (very unimportant, unimportant, moderately unimportant, moderately important, important, very important). Section 3 was composed of four questions, all which aimed at soliciting the satisfaction levels with the values. Question 1 directly asked respondents to rate their satisfaction levels with the following values on a scale of 1 to 6 (very dissatisfied, dissatisfied, moderately dissatisfied, moderately satisfied, satisfied, very satisfied): thermal comfort in winter, thermal comfort in summer, visual comfort, IAQ, energy cost saving, and environmental protection. Because both productivity and health are values which may be impacted by the values in first category (i.e., thermal comfort, visual comfort, IAQ), Question 2 and 3 asked respondents to rate how they think their perceived personal productivity and health, respectively, are decreased or increased by the current indoor environmental conditions (temperature, lighting, IAQ) at home (or work) using a 3-point scale (decrease, no effect, increase).

Section 4 aimed to collect data about PEFs, in order to explore the potential differences in the importance levels of values and satisfaction levels with these values across different PEFs, including (1) occupant characteristics: gender, age, years spent in this state, region of origin, weight, height, physical exercise, education, occupation, gross household annual income, chronic respiratory disease, and smoking (2) health symptoms: fatigue, dry skin, nose irritation, headaches, sleepiness, sore throat, concentration lapses, eye irritation, trouble focusing eyes, and dizziness, (3) primary building characteristics: building type and building age, (4) level of occupant building control: level of heating control, level of cooling control, level of ventilation control, and level of lighting control, (5) energy efficiency building features: Energy Star certification, LEED certification, energy efficient wall and roof assemblies, high efficiency thermal insulation, energy efficient windows and doors, energy efficient air barriers/vapor diffusion retarders, energy efficient weatherstripping and caulking, controlled ventilation, high efficiency cooling systems, high efficiency heating systems, Energy Star home electronics (or Energy Star office equipment for office survey), Energy Star appliances, Energy Star lighting bulbs and fixtures, and Energy Star water heaters, (6) energy cost and consumption feedback (for residential survey only):

monthly energy cost paid, energy consumption data provided, utility real-time pricing, (7) energy-use behavior to control indoor environmental conditions: turn on/off light switches, adjust window blinds or shades, open/close windows, adjust thermostat, use/adjust portable heater, adjust permanent heater, use/adjust portable humidifier, use/adjust room air conditioning unit, use/adjust portable fan, use/adjust ceiling fan, adjust air ventilation in wall or ceiling, adjust floor air ventilation, open/close internal doors, open/close external doors, (8) workspace characteristics (for office survey only): type of personal workspace, years spent in personal workspace, average number of hours spent in workspace, and (9) job characteristics (for office survey only): overall job satisfaction, employee benefits, psychosocial atmosphere at work. Only those PEFs that can be solicited through questionnaire surveys were included in Section 4. For example, PEFs that affect thermal comfort such as activity level, age, gender, height, weight, and health conditions can be solicited through questionnaire surveys and were included; but metabolic rate, clothing insulation, air temperature, radiant temperature, air speed, and humidity are difficult to capture through questionnaire surveys and were thus excluded. Due to the variability in occupancy and building characteristics across residential and office buildings, the questions in this section varied across both questionnaires.

3.1.2 Validation of Questionnaire Design

Prior to launching the survey, a pilot study on fifteen building occupants was conducted to test the effectiveness and clarity of the questionnaire. Participants were requested to complete the residential or office building survey and, then, to provide feedback on the format and content of the questionnaire. Feedback was solicited on different aspects of the questionnaire, such as question wording, response options and evaluation scale, instructions to respondents, visual appearance, and clarity of value concepts. The questionnaire was revised based on the feedback. For example, the wording of the scales in some questions were modified to enhance clarity.

3.1.3 Respondent Recruitment and Survey Implementation

The surveys were conducted from October to November 2014. Potential respondents were recruited by Qualtrics, a provider of online panels (potential respondents). Panels were generated using samples from various databases and were verified to prevent any fraudulent or duplicate respondents (Qualtrics 2014). Qualtrics hosted the survey and sent emails to potential respondents inviting them to complete the survey, for research purposes, in return for incentives. Two response quality filters were used: (1) an attention filter question and (2) a minimum survey completion time of two minutes. Responses that failed to pass these two filters were disregarded.

3.1.4 Survey Results and Analysis

The analysis of the survey results aimed at answering the following research questions: What are the ratings and the rankings of the importance levels of values by residential and office building occupants in AZ, IL, and PA? What are the ratings and the rankings of the satisfaction levels of residential and office building occupants with the values in AZ, IL, and PA? What are the differences in the importance levels and satisfaction levels of/with the values across different types of occupants (residential and office), different states (AZ, IL, and PA), and PEFs?

Five statistical analysis methods were utilized to address the above research questions: (1) mean indexing, (2) Spearman's rank correlation, (3) Kendall's coefficient of concordance, (4) Mann-Whitney U test and, and (5) Kruskal-Wallis H Test. Mean indexing was used to determine the mean ratings of values. Spearman's rank correlation was used to assess the general similarity between occupants of residential and office buildings. Kendall's coefficient of concordance was computed to examine whether there was a significant agreement among (1) occupants of residential buildings across the three states and (2) occupants of office buildings across the three states. When there were two groups to compare, Mann-Whitney U test was used to identify whether specific values were rated differently (e.g., across residential and office building occupants and across male and female office occupants). When there were more than two groups to compare, Kruskal-Wallis H test was conducted to identify whether specific values were rated differently (e.g., across the three states). The Statistical Package for Social Sciences (SPSS) version 20.0 was used to conduct these statistical analyses.

3.2 Survey Results and Analysis

3.2.1 Classification of Responses

A total of 310 and 308 valid responses from occupants of residential and office buildings, respectively, were collected. Qualtrics identified approximately 9,600 potential respondents and invited them via email. A total of 729 responses (including invalid responses) were received, representing a response rate of 7%. This is consistent with the reported response rates for online panels (Neslin 2009). This sample size is statistically significant with 95% confidence level and 10 confidence interval. Responses were classified into six subgroups in terms of combinations of (1) occupant type: residential and office building occupants and (2) state: AZ, IL, and PA. The descriptive statistics of the six subgroups are shown in Table 3.1.

Table 3.1 Descriptive Statistics of Responses

Type of occupants	Arizona		Illinois		Pennsylvania		Total			
	Number of participants	Number of valid responses	Number of participants	Number of valid responses	Number of participants	Number of valid responses	Number of people contacted	Number of participants	Response rate	Number of valid responses
Residential building occupants	123	104	119	102	129	104	N/A	371	N/A	310
Office building occupants	110	104	111	102	110	102	N/A	331	N/A	308
Total	233	208	230	204	239	206	~9600	702	7%	618

3.2.2 Reliability of Values

In order to validate the internal consistency of the data, prior to data analysis, a Cronbach's alpha reliability analysis was conducted. Internal consistency indicates the extent to which all the items in a test measure the same concept. Alpha values greater than 0.7 indicate adequacy of internal consistency (Tavakol and Dennick 2011). The overall Cronbach's alpha values for the residential and office building occupant surveys are 0.883 and 0.891, respectively, which indicates a high level of reliability.

3.2.3 Importance Levels of Values

The importance rating frequencies of the values by residential and office building occupants are shown in Figure 3.1. Overall, 85% or more and 86% or more of the residential and office building occupants, respectively, rated the values as "moderately important" or higher, which indicates that all the values were at least moderately important to the majority of residential and office building occupants.

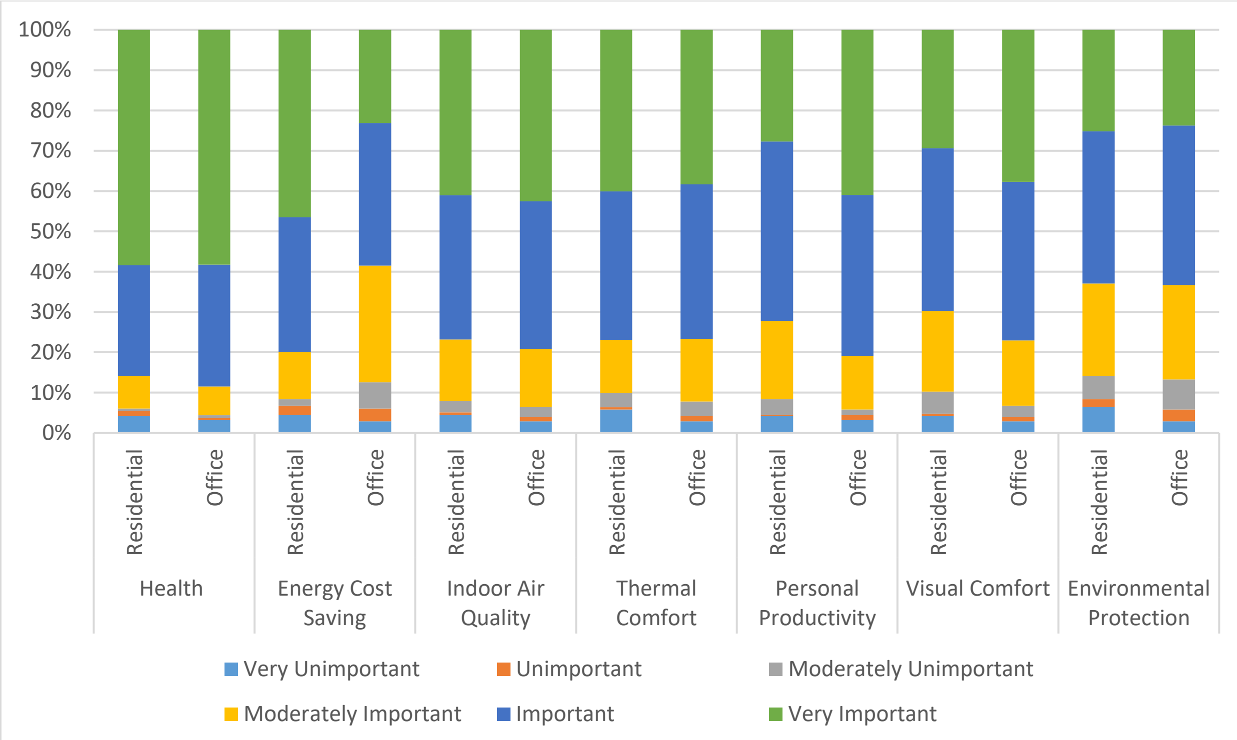


Figure 3.1 – Importance Ratings of the Values by Residential and Office Building Occupants

3.2.3.1 <u>Comparison of Importance Levels across Residential and Office Building Occupants</u>

The values were ranked based on their mean importance rating scores. The higher the mean importance rating score, the higher the rank, and vice versa. Table 3.2 shows the mean importance ratings and importance rankings of the values of residential and office building occupants overall, in AZ, in IL, and in PA. As shown in the table, all mean scores are higher than 4.00, which indicates that on average all the values were moderately important or higher to residential and office building occupants. On average, health was ranked the highest among the values – across both occupant types, and across the three states, which indicates that both occupants of residential and office buildings, across AZ, IL, and PA, valued health the most among the seven values. Other than the highest ranked value, the ranking differed across residential and office building occupants. For example, energy cost saving was ranked as the second most important value by residential building occupants, but ranked as the least important value (seventh) by office building occupants.

Table 3.2 Mean Importance Ratings and Ranks of the Values across Different Occupant Types and States

Values	Residential general		Office general		Residential AZ		Residential IL		Residential PA		Office AZ		Office IL		Office PA	
	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
Health	5.28	1	5.35	1	5.18	1	5.37	1	5.30	1	5.29	1	5.42	1	5.33	1
Energy cost saving	5.07	2	4.60	7	5.07	2	5.08	2	5.06	2	4.58	7	4.58	7	4.64	7
Indoor air quality	5.00	3	5.08	2	4.96	3	5.08	2	4.96	4	5.06	2	5.13	2	5.07	3
Thermal comfort	4.95	4	5.00	5	4.84	4	5.02	4	4.98	3	5.06	2	4.96	5	4.98	5
Personal productivity	4.83	5	5.08	2	4.77	5	4.88	6	4.84	5	5.05	4	5.06	3	5.14	2
Visual comfort	4.80	6	5.01	4	4.73	6	4.90	5	4.76	6	4.98	5	5.06	3	4.99	4
Environmental protection	4.59	7	4.65	6	4.53	7	4.66	7	4.59	7	4.68	6	4.60	6	4.67	6

Spearman's rank correlation analysis was conducted to further assess the general similarity between the rankings of both groups. The value of Spearman's rank correlation coefficient (r_s) ranges from -1 (perfect negative correlation in ranking) through 0 (no correlation in ranking) to 1 (perfect positive correlation in ranking). The correlation is considered significant when the significance level of the correlation coefficient is less than 0.05 (i.e., $p<0.05$). The r_s value between the rankings of both groups is 0.33 (with $p>0.05$), which indicates that the correlation between the importance rankings of both groups is weak and not significant. The null hypothesis that the importance rankings of residential and office building occupants are different cannot, thus, be rejected. The rankings across both groups are, thus, different.

Mann-Whitney U test was conducted to further identify whether specific values were rated differently across the two types of occupants. The results were interpreted based on the probability value (p-value). If the p-value is less than 0.05, there is a significant difference across the two groups. The results showed that the importance ratings of visual comfort, personal productivity, and energy cost saving were significantly different between residential and office building occupants. Energy cost saving was more important for residential building occupants, whereas visual comfort and personal productivity were more important for office building occupants. These findings are supported by common sense, statistics, and previous studies. The difference in the rating of energy cost saving across residential and office building occupants is likely related to "who pays the bill"; for office buildings, the energy bill is paid the employer not the occupants (employees). An energy survey report by DeCicco et al. (2015) shows that 13%, 5%, and 4% of energy consumer in the bottom, middle, and top tercile of income levels, respectively, think their home energy cost is unaffordable. The higher ranking of productivity by office occupants is likely related to work context; productivity is typically more relevant/important in workplace settings than in others (Meyer 2003). Previous studies [e.g., (Haynes 2007)] have emphasized the importance of office indoor environments in enabling productivity. The importance of visual comfort in office buildings may be attributed to its relation to health and productivity, which were both rated highly by occupants. A previous study has shown a 5% to 15% increase in productivity and 15% reduction in absenteeism as a result of improvements in office lighting conditions (Edwards and Torcellini 2002). Overall, these results show that residential and office building occupants have different value priorities. Such differences in value priorities should be taken into consideration when aiming to understand and improve energy-use behavior. For example, since office building occupants attach less importance to energy cost saving than residential building occupants, energy cost may have a lower (or no) impact on the energy-use behavior of office building occupants. There is, thus, a need to find more innovative ways to incentivize office building occupants to improve their energy-use behavior.

3.2.3.2 Comparison of Importance Levels across Different States

Figure 3.2 shows the importance rating frequencies of the values by residential and office building occupants across AZ, IL, and PA. Kendall's coefficients of concordance (Kendall's W) were computed to examine whether there were significant agreement on the importance rankings. The results of the test were interpreted based on the W value and the significance level of the test. If Kendall's W is 1 there is complete agreement and if W is 0 there is no agreement at all, with the result being significant if the significance level is less than 0.05 (Kendall and Gibbons 1990). Both Kendall's W values are greater than 0.9 and their significance levels are less than 0.05, which indicates that there were significant high levels of agreements on the importance rankings of values among (1) residential building occupants across the three states [$W=0.960$ ($p<0.05$)] and (2) office building occupants across the three states [$W=0.904$ ($p<0.05$)]. The difference in the ratings of importance levels across the three states was also examined using Kruskal-Wallis H test, which is the non-parametric version of one way analysis of variance. The results of the test were interpreted based on the significance level of the test. If the significance level is less than 0.05, then there are significant differences across the groups. No significant differences in the ratings of importance levels were shown among (1) residential building occupants across the three states and (2) office building occupants across the three states. Overall, these results indicate that occupants across the three states have similar value priorities.

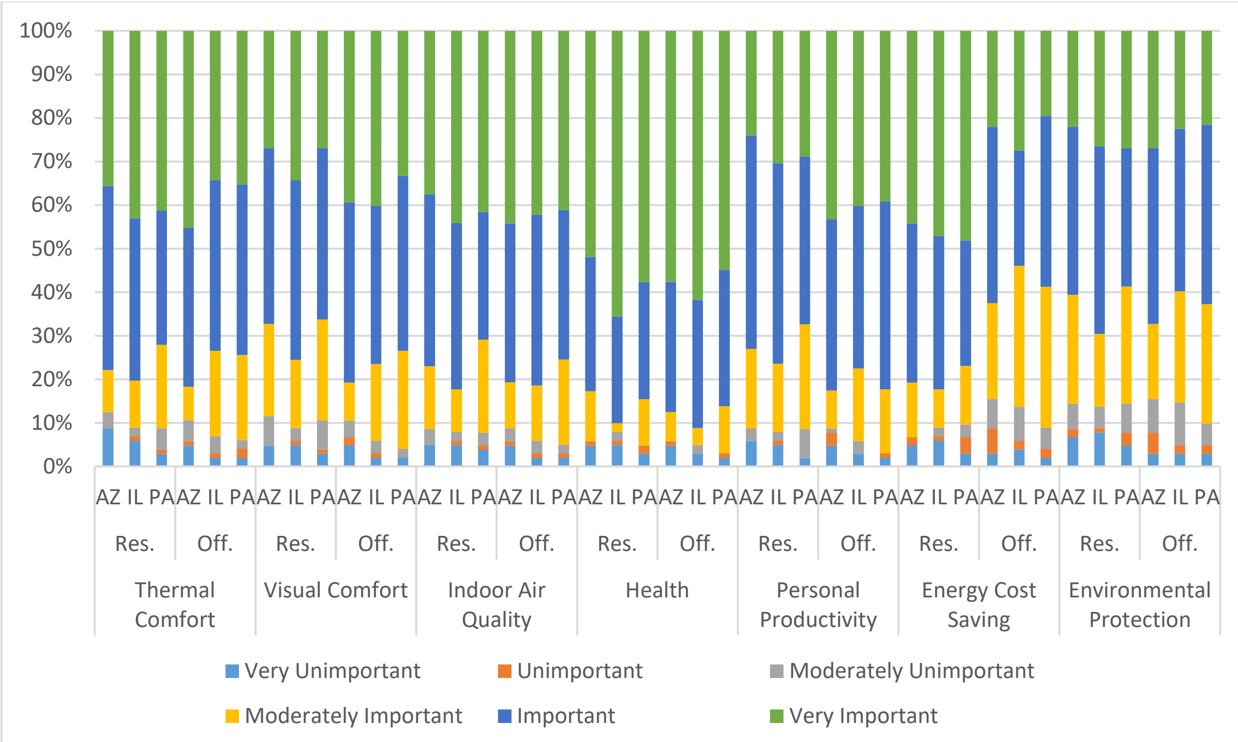


Figure 3.2 – Importance Ratings of the Values by Residential and Office Building Occupants across AZ, IL, and PA

3.2.4 Satisfaction Levels with Values

Figure 3.3 and Figure 3.4 show the rating frequencies of residential and office building occupants of their satisfaction levels with the values and the impact of indoor environmental conditions on their perceived personal productivity and health, respectively. Overall, on average, both residential and office building occupants rated their satisfaction levels with energy cost saving, IAQ, thermal comfort (in summer and winter), visual comfort, and environmental protection as "moderately satisfied" or higher, but a considerable percentage of occupants from both groups indicated that their satisfactions with these values are "moderately unsatisfied" or lower. For example, 19.7% of the residential building occupants were "moderately unsatisfied" or lower with energy cost saving, which was the least satisfied value for residential building occupants. Environmental protection, thermal comfort in summer, thermal comfort in winter, IAQ, and visual comfort were rated by

15.8%, 15.5%, 13.9%, 11.6%, and 11.2% of the residential building occupants as "moderately unsatisfied" or lower, respectively. For office building occupants, environmental protection was the least satisfied value with 21.4% "moderately unsatisfied" or lower ratings. Energy cost saving, thermal comfort in summer, thermal comfort in winter, IAQ, and visual comfort were rated by 20.4%, 18.8%, 18.1%, 17.5%, and 12.0% of the office building occupants as "moderately unsatisfied" or lower, respectively. For health and personal productivity, 27.8% and 24.8% of the residential and 34.4% and 38.6% of the office building occupants believed that the current indoor environmental conditions have a negative effect on their perceived health and personal productivity, respectively.

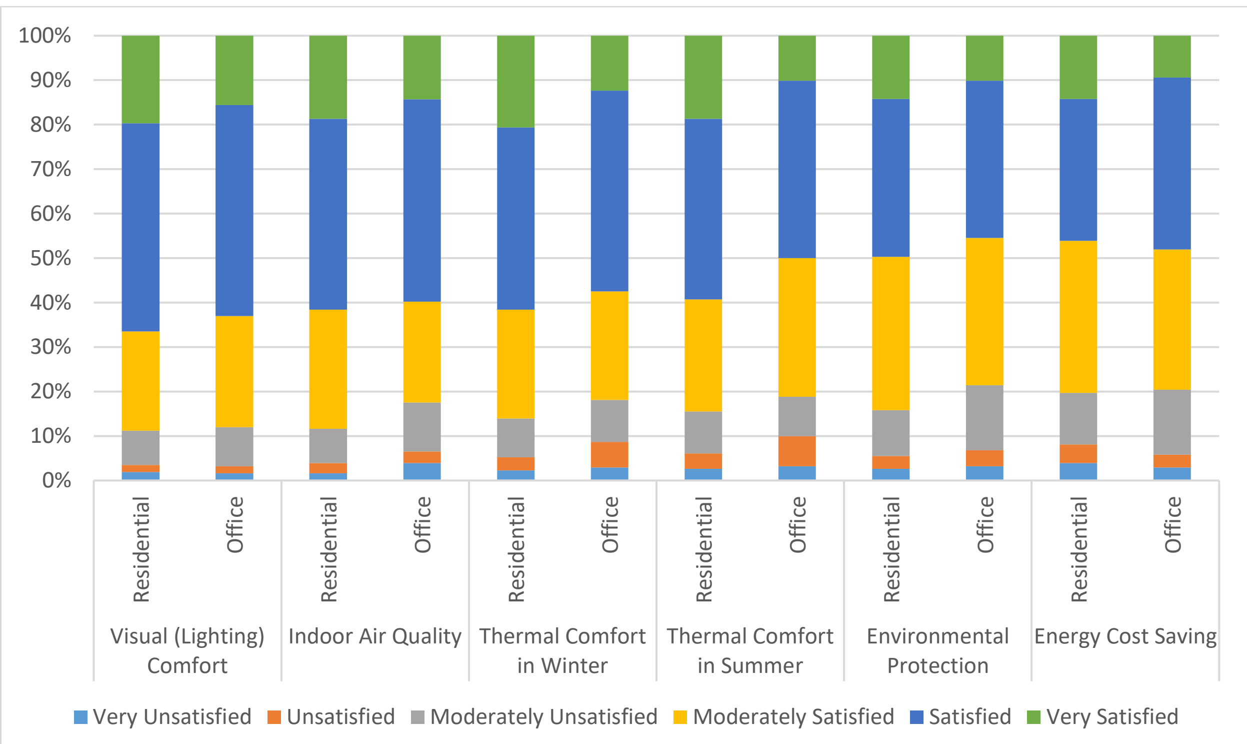


Figure 3.3 – Satisfaction Levels of Residential and Office Building Occupants with the Values

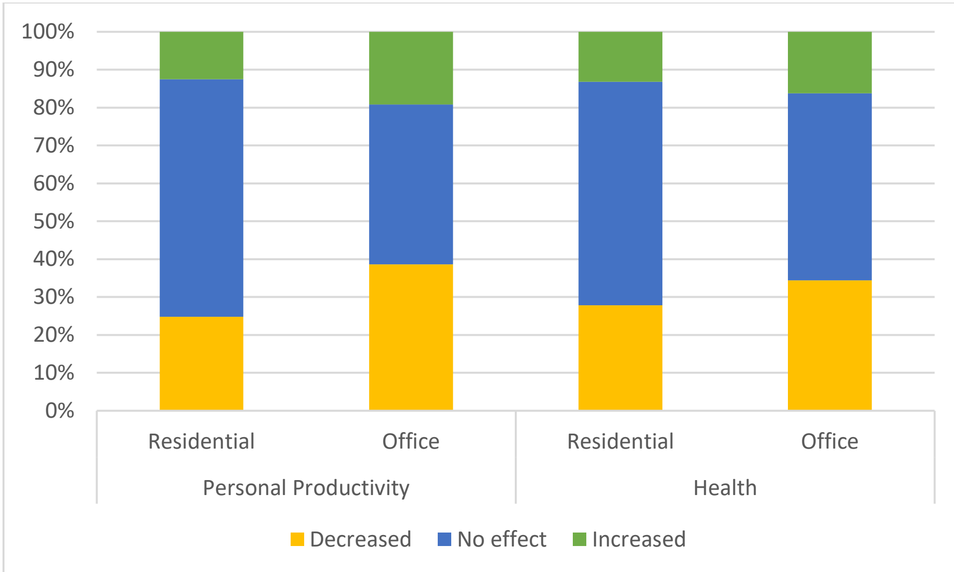


Figure 3.4 – Reported Impact of Indoor Environmental Conditions on Perceived Health and Personal Productivity of Residential and Office Building Occupants

3.2.4.1 Comparison of Satisfaction Levels across Residential and Office Building Occupants

The values were ranked based on their mean satisfaction rating scores. Table 3.3 shows the mean satisfaction ratings and satisfaction rankings of residential and office building occupants with the values overall, in AZ, in IL, and in PA. On average, visual comfort was ranked the highest among the values in satisfaction – overall and across both occupant types. Overall, there was a similarity in the satisfaction ranking by residential and office building occupants, except for environmental protection and energy cost saving. For example, energy cost saving was ranked as the least (sixth) satisfied value by residential building occupants, but ranked as the fourth most satisfied value by office building occupants.

Table 3.3 Mean Satisfaction Ratings and Ranks with the Values across Different Occupant Types and States

Values	**Residential general**		**Office general**		**Residential AZ**		**Residential IL**		**Residential PA**		**Office AZ**		**Office IL**		**Office PA**	
	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
Visual comfort	4.70	1	4.62	1	4.70	2	4.73	1	4.67	1	4.57	1	4.57	1	4.71	1
Indoor air quality	4.63	2	4.46	2	4.61	3	4.62	2	4.67	1	4.41	3	4.44	4	4.53	2
Thermal comfort in winter	4.61	3	4.40	3	4.82	1	4.62	2	4.40	4	4.51	2	4.47	2	4.22	3
Thermal comfort in summer	4.54	4	4.28	4	4.51	4	4.56	4	4.54	3	4.22	6	4.46	3	4.16	6
Environmental protection	4.40	5	4.24	6	4.48	5	4.44	5	4.29	6	4.26	4	4.25	6	4.21	5
Energy cost saving	4.29	6	4.28	4	4.29	6	4.27	6	4.31	5	4.26	4	4.38	5	4.22	3

Spearman’s rank correlation coefficient was calculated to assess the general similarity between the satisfaction rankings of both groups. The r_s value between the rankings of both groups is 0.90 (with $p<0.05$), which indicates that there is a significant strong correlation between the satisfaction rankings of both occupant groups. The null hypothesis that the satisfaction rankings of residential and office building occupants are different was, thus, rejected.

In order to find whether satisfaction levels with some specific values are rated differently across residential and office building occupants, Mann-Whitney U test was conducted. Residential building occupants rated their satisfaction levels with thermal comfort (in summer and winter) higher than office building occupants. The variance across both occupant types could be attributed to different perceptions of people in different environmental contexts (Tablada et al. 2009). Nicol and Humphreys (2002) stated that building type is one of the factors which influence occupant satisfaction with thermal environment.

3.2.4.2 Comparison of Satisfaction Levels across Different States

Figure 3.5 and Figure 3.6 show the rating frequencies of residential and office building occupants of their satisfaction levels with the values and the impact of indoor environments on their perceived personal productivity and health, respectively, across AZ, IL, and PA. Kendall's W and p values indicate significant high levels of agreements on the satisfaction rankings of values among (1) residential building occupants across the three states [W=0.841 ($p<0.05$)] and (2) office building occupants across the three states [W=0.770 ($p<0.05$)]. Kruskal-Wallis H test showed significant differences across the three states, in the satisfaction levels of residential occupants with thermal comfort in winter. In order to identify where the differences between the groups lie, a post-hoc pairwise comparison test was conducted. The test showed that residential occupants in AZ are more satisfied with thermal comfort in winter, than those in PA. This difference could be explained by the different weather characteristics of the two states. According to the National Climatic Data Center, the average temperatures of AZ and PA in winter months from 1915 to 2015 are 42.3^{o} F and 27.3^{o} F, respectively (2015). Other than thermal comfort in winter, there were no significant differences in the ratings of satisfaction levels among (1) residential building occupants across the three states and (2) office building occupants across the three states. Similarly, for each occupant group, there were no significant differences across the three states in the ratings of impact of indoor environmental conditions on their perceived health and personal productivity.

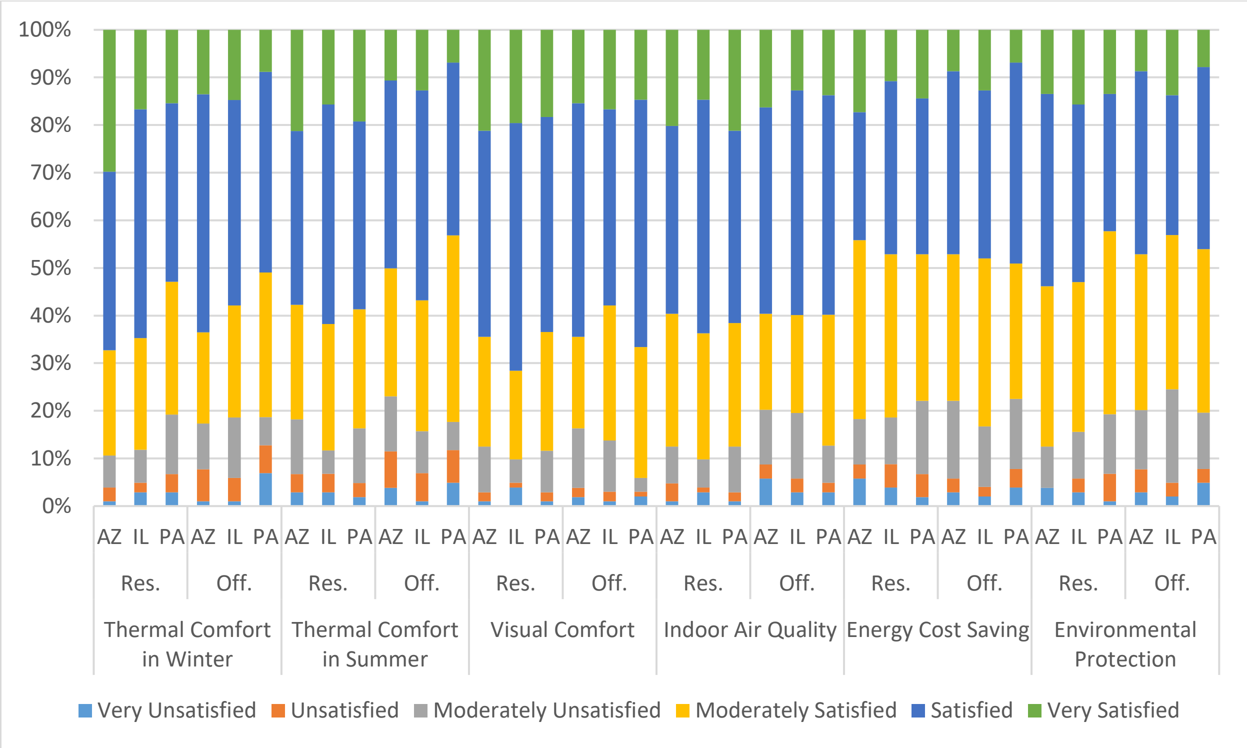


Figure 3.5 – Satisfaction Levels of Residential and Office Building Occupants with the Values across AZ, IL, and PA

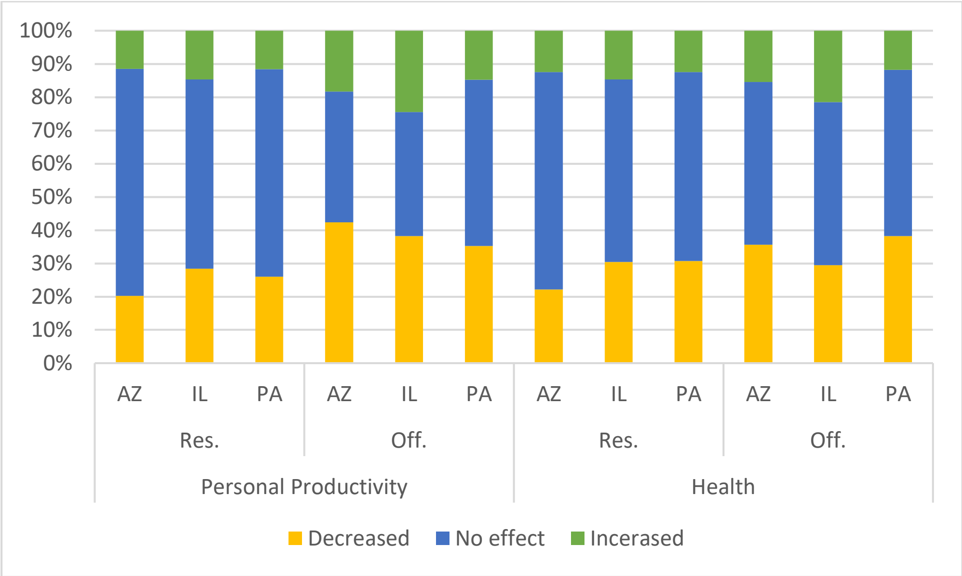


Figure 3.6 – Reported Impact of Indoor Environmental Conditions on Perceived Health and Personal Productivity of Residential and Office Building Occupants across AZ, IL, and PA

3.2.5 Differences of Importance and Satisfaction Levels across Different PEFs

3.2.5.1 Occupant Characteristics

As shown in Table 3.4, for residential building occupants, there were significant differences in the importance ratings of values across males and females. Female occupants gave more importance to all seven values. The existence of gender differences has been shown in previous studies in the area of building environment. For example, a recent study (Kim et al. 2013) has shown gender differences in the perception of occupants about various factors of IEQ. The fact that these differences were only seen among residential building occupants could be attributed to different gender roles at home and in the workplace (Easterlin 1987). For office buildings, there were significant differences in the satisfaction ratings of all values between male and female occupants. Female occupants were less satisfied with all the values than male occupants. This supports the existence of gender differences in satisfaction levels. For example, previous studies (Kim et al. 2013; Karjalainen 2012; Bakke et al. 2007; Aries et al. 2010) have shown that females were more dissatisfied with their overall comfort. The fact that these differences were only seen among office building occupants could be attributed to variations in control levels of females over indoor environmental conditions across home and office. For example, a survey showed that females felt that they have less control over room temperatures at office than males in winter and summer but have more control over room temperatures at home in summer (Karjalainen 2007).

A significant difference in the importance rating of IAQ was also shown between residential building occupants with respiratory disease and with no respiratory disease; occupants with respiratory disease attach more importance to IAQ. This makes sense when the potential risks of poor IAQ are considered; indoor air pollutants are becoming a serious threat to the respiratory health of residential and office building occupants (American Lung Association 2015).

Overall, the above results show that a number of occupant characteristics are associated with higher/lower importance and/or satisfaction levels. Such differences in importance and/or satisfaction levels can cause variations in the patterns of energy-use behavior across different occupants. It can also help determine which means would be more effective for improving that behavior while maintaining satisfaction levels. It may also indicate that the process of behavioral change could become quite complicated. For example, on one hand, it might be easier to motivate female occupants than male occupants to improve their energy-use behavior because they attach higher importance levels to values such as environmental protection and energy cost saving; but, on the other hand, because of their lower satisfaction levels, female occupants might be more sensitive to the impact of such behavioral changes on their satisfaction with values such as thermal comfort.

Table 3.4 Associations between PEFs and Importance and Satisfaction Ratings of/with the Values by Residential and Office Building Occupants

PEF type	Potential energy-related factor (PEF)	Occupant value						
		Health	Energy cost saving	Indoor air quality	Thermal comfort	Personal productivity	Visual comfort	Environmental protection
State	Residency state				SR^{W}			
Occupant characteristics	Gender	IR SO	IR SO	IR SO	IR SO	IR SO	IR SO	IR SO
	Age							
	Years spent in this state							
	Region of origin							
	Weight							
	Height							
	Physical exercise							
	Education							
	Occupation							
	Gross household annual income							
	Chronic respiratory disease			IR				
	Smoking							
Health symptoms	Fatigue					SR		
	Dry skin							
	Nose irritation							
	Headaches			SO			SO	
	Sleepiness				SO^{S}	SO		
	Sore throat	SR		SR		SR		
	Concentration lapses	SR			SO	SO		
	Eye irritation	SO				SO	SO	
	Trouble focusing eyes				SO^{W}	SO	SO	
	Dizziness							
Primary building characteristics	Building type							
	Building age							
Level of occupant building control	Level of heating control							
	Level of cooling control							
	Level of ventilation control							
	Level of lighting control							
Energy efficiency building or space features	Energy Star certification		SR SO					SR SO
	LEED certification							
	Energy efficient wall and roof assemblies		SO					SR SO
	High efficiency thermal insulation		SR SO					SR SO
	Energy efficient windows and doors		SR SO					SR SO
	Energy efficient air barriers/vapor diffusion retarders		SR SO					SR SO
	Energy efficient weatherstripping and caulking		SR SO					SR SO
	Controlled ventilation		SR SO					SR SO
	High efficiency cooling systems		SR SO					SR SO
	High efficiency heating systems		SO					SR SO
	Energy Star home electronics*							
	Energy Star office equipment**		SO					SO
	Energy Star appliances		SO					SR SO
	Energy Star lighting bulbs and fixtures		SR SO					SO
	Energy Star water heaters		SO					SR SO
Energy cost & consumption feedback	Energy cost*							
	Energy consumption data provided*							
	Utility real-time pricing*							
Occupant behavior (to control indoor environmental conditions)	Adjust lighting					SR		
	Adjust window blinds or shades					SO	SO	
	Open/close windows			SO		SR SO		
	Adjust thermostat				SO	SO		
	Adjust portable heater				SR^{w}			
	Adjust permanent heater				SO^{w}			
	Adjust humidifier	IR		IR SO		IR		
	Adjust room air conditioning unit				SO			
	Adjust portable fan							
	Adjust ceiling fan				SO			
	Adjust air ventilation in wall or ceiling							
	Adjustable floor air ventilation							
	Open/close internal doors			SO	SO^{w}			
	Open/close external doors			SR				
Workspace characteristics	Type of personal workspace**		IO		SO		SO	SO
	Years Spent in personal workspace**							
	Average number of hours spent in workspace**							
Job characteristics	Overall job satisfaction**							
	Employee benefits**	SO				SO		
	Psychosocial atmosphere at work**							

*Residential Only
**Office only
IR Importance of the values to residential building occupants
IO Importance of the values to office building occupants
SR Satisfaction of residential building occupants with values
SO Satisfaction of office building occupants with the values
^{S} Summer only
^{W} Winter only

3.2.5.2 Occupant Health Symptoms

As shown in Table 3.4, occupants who experience some of the health symptoms showed lower satisfaction levels with some values. Significant differences were shown in the satisfaction ratings of residential building occupants with some values across different frequencies of experiencing fatigue, sore throat, and concentration lapses. Occupants who experience higher frequencies of fatigue were less satisfied with personal productivity. Occupants who experience higher frequencies of sore throat showed lower satisfaction levels with IAQ, health, and personal productivity. Occupants who experience higher frequencies of concentration lapses showed lower satisfaction levels with health.

For office building occupants, more health symptoms were shown to be associated with lower satisfaction levels. Occupants who experience higher frequencies of headache indicated lower satisfaction levels with visual comfort and IAQ. Occupants who experience higher frequencies of sleepiness rated their satisfaction levels with thermal comfort in summer and personal productivity lower. Occupants who experience higher frequencies of concentration lapses showed lower satisfaction levels with thermal comfort (in winter and summer) and personal productivity. Occupants who experience higher frequencies of eye irritation rated their satisfaction levels with visual comfort, health, and personal productivity lower. Occupants who experience higher frequencies of trouble focusing eyes indicated lower satisfaction with thermal comfort in winter, visual comfort, and personal productivity. These findings could be explained by the existence of sick building syndrome (SBS) at office buildings; according to the Environmental Protection Agency, SBS causes dissatisfaction with comfort in buildings (EPA 1991).

Overall, occupants who experience higher frequencies of health symptoms expressed lower satisfaction levels. A number of measures for improving indoor environmental conditions, such as doubling ventilation rates, were proposed in the literature to reduce such health symptoms and their associated occupant dissatisfaction (Fisk 2000). However, such measures are usually not implemented because of their additional energy needs (Fisk 2000). More studies are thus needed to identify measures for improving indoor environmental conditions that are energy efficient (e.g., outside air economizer, heat recovery from exhaust ventilation air, nighttime precooling using outdoor air) and understand the impact of implementing such measures on both occupant satisfaction and building energy consumption.

3.2.5.3 Building Characteristics, Occupant Building Control, and Energy Efficiency Features

Energy Star aims to enhance energy cost saving of buildings (Energy Star 2015a). The survey results showed that occupants of Energy Star certified buildings, both residential and office buildings, were more satisfied with environmental protection and energy cost saving than occupants of non-Energy Star buildings. Recent reports indicate that Energy Star-certified new residential buildings consume 15-30% less energy compared to an average new residential building (Energy Star 2015b). No significant differences in the satisfaction levels of occupants of LEED-certified and non-LEED buildings were shown. Similar results were shown in the literature. For example, a study has shown that LEED and non-LEED building occupants have equal satisfaction with IEQ parameters (Schiavon and Altomonte 2014). The fact that there were significant differences in satisfaction levels across Energy Star and non-Energy Star building occupants but not across LEED and non-LEED building occupants can be explained by the broad scope of LEED certification. LEED has many credit categories and stakeholders can receive certification in many ways, not only through energy efficiency. Energy Star certification could thus be an effective means to both reduce energy consumption and enhance occupant satisfaction.

No significant differences in satisfaction levels were shown across different types (single-family house, apartment building, etc.) and ages of buildings. However, significant differences in satisfaction levels were shown across buildings with and without energy efficiency features. Occupants of buildings with energy efficiency features indicated higher satisfaction with environmental protection and energy cost saving. This may indicate that cost saving and environmental protection expectations of occupants can be fulfilled, regardless of building type and age, through the use of energy efficiency measures (Frey et al. 2012).

3.2.5.4 Energy Cost and Consumption Feedback

For residential buildings, no significant differences in importance or satisfaction levels of/with values were shown across: (1) occupants who pay different levels of monthly energy cost, (2) occupants who receive energy consumption feedback and those who do not, and (3) occupants who participate in a utility real time pricing program and those who do not. This may suggest that although energy consumption feedback systems have had some success in encouraging energy conservation, the outcomes of these systems are either limited or not fully perceived by building occupants. This is supported by previous studies that reported limited effectiveness of energy consumption feedback systems in energy conservation on the long-run (Faruqui et al. 2010; Alahmad et al. 2012; van Dam et al. 2010). More studies are thus needed to explore how to sustain energy conservation benefits brought by energy consumption feedback systems on the long-term (Pierce et al. 2010).

3.2.5.5 Energy-Use Behavior

A number of energy-use behavior (actions) were associated with higher/lower importance levels and/or satisfaction levels of/with values. For importance levels, residential building occupants who use/adjust humidifiers indicated higher importance levels of IAQ, health, and personal productivity. For satisfaction levels of residential building occupants, occupants who use/adjust portable heater were less satisfied with thermal comfort in winter, occupants who open/close windows and occupants who adjust lighting were less satisfied with personal productivity but occupants who open/close external doors were more satisfied with IAQ. For office building occupants, as shown in Table 3.4, adjusting window blinds or shades, opening/closing windows, and adjusting thermostat were associated with higher satisfaction levels with personal productivity; adjusting thermostat, adjusting permanent heaters, using/adjusting room air conditioning units, using/adjusting ceiling fans, and opening/closing internal doors were associated with higher satisfaction levels with thermal comfort in winter; adjusting thermostat, using/adjusting room air conditioning units, and using/adjusting ceiling fans were associated with higher satisfaction levels with thermal comfort in summer; opening/closing windows, using/adjusting humidifier, opening/closing internal doors were associated with higher satisfaction levels with IAQ; and adjusting window blinds or shades was associated with higher satisfaction levels with visual comfort.

These significant differences in satisfaction levels across different energy-use behavior/actions could be attributed to energy-use behavior aiming at controlling indoor environmental conditions to reach high/higher satisfaction levels. On one hand, control over environmental conditions may result in higher satisfaction levels (Haldi 2010; Lehrer 2006). For example, access to thermostat and having operable windows were associated with higher satisfaction levels with thermal comfort (Huizenga et al. 2006). On the other hand, having some personal devices to compensate deficiencies of indoor environmental conditions may indicate lower satisfaction levels (Lehrer 2006). For example, in a previous study, the use of portable fans and portable heaters were associated with lower satisfaction with thermal comfort (Huizenga et al. 2006). In a similar study, it was shown that 19% of the occupants who used portable fans indicated that air motion is too low, whereas only 8% of the occupants who did not use portable fans indicated that air motion is too low (Zhang et al. 2007).

The differences across residential and office building occupants could be attributed to the different context of buildings as well as the complexity of energy-use behavior. Energy-use behavior is affected by various parameters such as household characteristics, lifestyle, motivation, and interaction between the occupant and the building (Santin 2013) and it is not fully understood (Hong 2014). Better understanding, through experimental studies, of energy-use behavior and how it is associated with energy consumption and occupant satisfaction is needed.

3.2.5.6 Workspace and Job Characteristics

For office building occupants, there was a significant difference in the importance ratings of energy cost saving across different types of personal workspaces. Occupants with private workspaces showed higher importance ratings for energy cost saving than those with shared workspaces. This could be attributed to private workspaces being mostly occupied by managerial-level employees (or employers), who are typically concerned about business expenditures. Also, occupants with shared workspaces may believe that they have limited opportunities for energy conservation because they often share office appliances with multiple employees (Carrico and Riemer 2011).

Significant differences in satisfaction levels were shown across occupants with different types of personal workspaces. Occupants of private workspaces showed higher satisfaction levels with thermal comfort (in winter and summer), visual comfort, and environmental protection. This finding supports the reported results that private office occupants are highly comfortable (Kim and de Dear 2013).

Significant differences in satisfaction levels were also shown across different levels of employee benefits. Occupants who are more satisfied with the employee benefits were more satisfied with their health and personal productivity. This finding coincides with that of Newsham et al. (2009) who showed a positive relationship between job satisfaction and overall satisfaction.

CHAPTER 4 – DATA SENSING, OCCUPANT FEEDBACK, AND DATA COLLECTION

4.1 Data Sensing and Occupant Feedback Collection Plan

The data collection plan is summarized in Table 4.1. The following types of data were collected: (1) occupant characteristics: clothing level, activity level, health symptoms, and personal characteristics (age, gender, education, etc.), (2) importance levels of values to occupants: importance levels of thermal comfort, IAQ, visual comfort, health, personal productivity, energy cost saving, and environmental protection, (3) occupant satisfaction levels with the values: satisfaction levels with thermal comfort, IAQ, visual comfort, health, personal productivity, energy cost saving, and environmental protection, (4) energy-use behavior: occupancy, turn on/off light, open/close window blinds or shades, open/close door, open/close window, turn on/off/adjust portable heater, turn on/off/adjust portable humidifier, turn on/off/adjust portable fan, turn on/off/adjust ceiling fan, turn on/off/adjust room air conditioning unit, adjust floor/ceiling/wall air ventilation, and adjust thermostat, (5) indoor environmental conditions: indoor temperature (F), relative humidity (%), CO_2 level (ppm), and illumination level (lux), (6) outdoor weather conditions: ambient temperature (F), ambient relative humidity (%), global horizontal solar radiation (W/m^2), wind speed (mph), and (7) building energy consumption levels: heating/cooling energy demand (kW), lighting energy demand (kW), and plug loads (kW).

Table 4.1 Data Collection Plan

Data Type	Data	Unit/scale	Collection method	Frequency of collection
Occupant characteristics	Clothing level	Multiple choice	Occupant feedback system	Twice a day
	Activity level			
	Health symptoms			
	Personal characteristics (age, gender, education, etc.)	Multiple choice	Questionnaire survey	Once only
Importance levels of values to occupants	Thermal comfort	Likert	Questionnaire survey	Once only
	IAQ			
	Visual comfort			
	Health			
	Personal productivity			
	Energy cost saving			
	Environmental protection			
Occupant satisfaction levels with the values	Thermal comfort	Likert	Occupant feedback system	Per change
	IAQ			
	Visual comfort			
	Health			
	Personal productivity			
	Energy cost saving			
	Environmental protection			
Energy-use behavior	Occupancy	Yes/no	Occupancy	Per change
	Turn on/off light	On/off	Occupant feedback system	
	Open/close window blinds or shades	Open/close		
	Open/close door	Open/close		
	Open/close window	Open/close		
	Turn on/off/adjust portable heater	On/off/adjust		
	Turn on/off/adjust portable humidifier	On/off/adjust		
	Turn on/off/adjust portable fan	On/off/adjust		
	Turn on/off/adjust ceiling fan	On/off/adjust		
	Turn on/off/adjust room air conditioning	F	Data logger	
	Adjust floor/ceiling/wall air ventilation	cfm		
	Adjust thermostat	F		
Indoor environmental conditions	Indoor temperature	F	Temperature	15min
	Relative humidity	%	Humidity sensor	15min
	CO_2 level	ppm	CO_2 sensor	15min
	Illumination level	lux	Luminance meter	15min
Outdoor weather conditions	Ambient temperature	F	Weather stations	1h
	Ambient relative humidity	%		1h
	Global horizontal solar radiation	W/m2		1h
	Wind speed	mph		1h
Building energy consumption level	Heating/cooling energy demand	kW	Power meter	15min
	Lighting energy demand	kW		15min
	Plug loads	kW		15min

4.2 Data Collection

Data were collected from the Philadelphia Business and Technology Center (PBTC) building between June 1, 2016 and August 31, 2016. The PBTC is a 6-story masonry office building with an estimated total floor area of 272,000 ft^2. The windows of the building are not operable. The west wing of the 4th floor was already instrumented for empirical data collection as part of an earlier research project by the CBEI (CBEI 2017). For this study, the sensors were tested and were fixed/replaced if/as needed. The instrumented area is 10,000 ft^2, which consists of 12 offices and two thermal zones (Figure 4.1). The instrumented area is occupied on weekdays from 8AM to 5PM, mostly. The data collection methods are summarized in Table 4.2. The building uses electricity for cooling and artificial lighting.

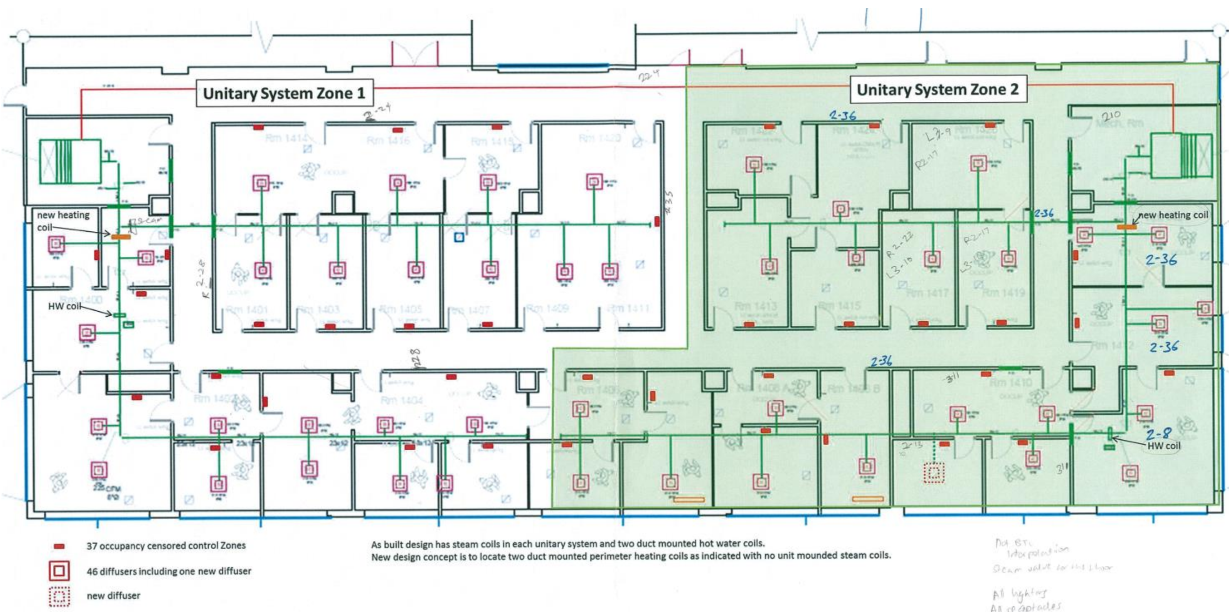


Figure 4.1 – Floor Layout

Table 4.2 Data Collection Methods

Data collection	**Cooling energy consumption**	**Outdoor weather conditions**	**Occupant behavior**	**Occupant comfort**
Method	Power meter (WattNode WNC-3Y-208-BN)	Local weather station and NSRDB[1]	Preference monitoring application	Preference monitoring application
Granularity	15 min	1 hour	As reported	As reported

[1]NSRDB: National Solar Radiation Database

Cooling and lighting energy consumption was metered in 15-min intervals using power meters installed on the air handling units (AHUs) and panelboards, and monitored using the PI CoreSight web-based application. Outdoor weather condition data were gathered from a weather station at the Philadelphia International Airport (Pennsylvania State Climatologist 2017) and the National Renewable Energy Laboratory's (NREL) National Solar Radiation Database (NSRDB) (Sengupta et al. 2018) in hourly intervals. The outdoor weather data included temperature, dewpoint temperature, relative humidity, wind speed, direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), and global horizontal irradiance (GHI) data.

Occupant behavior and occupant comfort data were collected from 12 occupants and captured through a preference monitoring application (PMA) (Abraham et al. 2017). The PMA was developed using an online survey tool to capture the energy-use behavior actions taken by the occupants and solicit feedback on their thermal and visual comfort. The behavior actions included adjusting thermostat, turning on/off a portable heater, opening/closing a door, opening/closing a shading device, and turning on/off a light. For behavior, feedback was required whenever they have taken an action. For thermal and visual comfort, a 6-point Likert satisfaction scale (1=very unsatisfied, 2=unsatisfied, 3=moderately unsatisfied, 4=moderately satisfied, 5=satisfied, 6=very satisfied) was used and feedback was required at least daily, or as frequent as occupants wish to report (e.g., whenever they feel not satisfied). The use of the PMA was completely voluntary. However, prior to data collection, a meeting with the occupants was held and the importance of their feedback on this research was communicated. Also, a reward program was launched to encourage building occupants to give frequent feedback. Each month, a $50 gift card was given to the most frequent feedback-provider of the month.

CHAPTER 5 – SIMULATION-DATA-DRIVEN OCCUPANT-BEHAVIOR-SENSITIVE MACHINE LEARNING-BASED ENERGY CONSUMPTION PREDICTION

5.1 Methodology

The proposed occupant-behavior-sensitive energy consumption prediction approach included five primary steps: (1) modeling a set of buildings with different sizes and occupant behaviors, (2) conducting energy simulations in several locations using EnergyPlus, (3) preprocessing the simulation-generated data, (4) developing a set of machine learning-based prediction models that learn from these data, and (5) evaluating the performance of the developed models. EnergyPlus was selected to conduct the building energy simulations due to its capability to simulate a building close to its real situation (Yousefi et al. 2017).

5.1.1 Building and Occupant Behavior Modeling

A total of 1,152 buildings were modeled to represent different occupant behaviors and building sizes. Other building characteristics, such as envelope thermal properties, were held constant to reduce the number of variables and therefore the simulation time. These properties were determined according to the industry standards [e.g., ASHRAE Standard 189.1-2013 (ASHRAE 2013b)]. The limitations of the scope and how they can be addressed in future work are further discussed in Sections 8.2.2 and 8.4.2, respectively.

To capture the impact of different occupant behaviors on consumption, a set of cases that represent different behaviors were modeled. To model these cases, a set of five proxy variables that could represent behavior differences were identified and modeled in a parametric way: cooling setpoint, window status, lighting power density, occupancy density, and electric equipment power density. Three cooling setpoint cases, within the temperature range recommended by the ASHRAE Standard 55-2013, were considered. For each of the lighting power, occupancy, and electric equipment power densities, two densities were considered: ±10% of the density for open offices in the ASHRAE Standard 189.1-2013. Two window operation cases were considered: non-operable windows and open during working hours. The natural ventilation rate through the windows was assumed as 5 air changes per hour. The window operation strategies were adopted from Sun and Hong (2017a) and Wang and Greenberg (2015). The building operational characteristics (e.g., cooling setpoint) were determined as per the ASHRAE Standard 189.1-2013. Table 5.1 summarizes the variables and their values. It is important to note that the aforementioned proxy variables represent some types of occupant behavior, only to a certain extent. Simulated data may not reflect the reality as accurate as the actual data collected during building operation (Naganathan et al. 2016). Compared to the simulation-data-driven approach, the real-data-driven approach, in Chapter 6, aims to better capture and model the real-life behavior and comfort of occupants and the real-life energy-consumption patterns of buildings. Compared to both approaches, the hybrid approach, in Chapter 7, aims to overcome the limitations of both simulation-data-driven and real-data-driven approaches – by learning from both types of data simultaneously. The limitations of this section are further discussed in Section 8.2.2.

Table 5.1 Occupant Behavior Variables

Variable	**Values**
Cooling setpoint (occupied)	{22.8ºC, 24.0ºC, 25.2ºC}
Lighting power density	{9.59 W/m^2, 11.72 W/m^2}
Occupancy density	{0.05 people/m^2, 0.07 people/m^2)
Electric equipment power density	{6.88 W/m^2, 8.41 W/m^2}
Window operation	{Not operable, Open during working hours}

To create a dataset that represents different office buildings in the U.S., different building sizes were modeled. The geometric properties of the buildings were determined based on the categories given in the 2012 Commercial Buildings Energy Consumption Survey (CBECS) (EIA 2015). For the building floor areas, a value from each size category was used: 232.26 m^2 (2500 ft^2), 696.77

m^2 (7500 ft^2), 1858.06 m^2 (20000 ft^2), 3716.12 m^2 (40000 ft^2), 7432.24 m^2 (80000 ft^2), 11612.88 m^2 (125000 ft^2), 23225.75 m^2 (250000 ft^2), and 46451.50 m^2 (500000 ft^2). For the number of floors, the most common three categories were considered: one-, two-, and three-story.

For the other building characteristics, the envelope thermal properties (e.g., wall and slab materials) were determined based on the ASHRAE Standard 189.1-2013. As such, the floor-to-floor height of the buildings was 3.66 m (12 ft). The perimeter zones of the buildings had operable windows for natural ventilation and cooling, and the WWR of each external surface of the buildings was 36%. The buildings were equipped with two packaged rooftop units, which use direct expansion (DX) cooling coils to supply cooling, with a rated cooling coefficient of performance (COP) of 3. The perimeter and core zones of the buildings had individual packaged rooftop units for cooling. The entire spaces of the buildings were designed as an open office. Table 5.2 shows the detailed envelope thermal properties of the buildings.

Table 5.2 Envelope Thermal Properties of the Buildings

Surface	Materials (from outside to inside of the building)	U Value (W/m^2K)
Exterior wall	25.3 mm stucco + 203.3 mm heavyweight concrete block + 45.2 mm wall insulation + 12.7 mm gypsum	0.78
Interior wall	19 mm gypsum board + air gap + 19 mm gypsum board	0.14
Ground slab	101.6 mm concrete + carpet	0.10
Roof	9.5 mm roof membrane + 210.5 mm roof insulation + 1.5 mm metal decking	0.23
Window	3 mm theoretical glass	13.83

For implementation, the different occupant behavior cases were simulated using the direct input approach due to its straightforwardness and accuracy (Hong et al. 2018). The building geometries were modeled in SketchUp Make 2017. The operational characteristics, envelope thermal properties, and HVAC systems of the buildings were defined in OpenStudio 2.4.0. The combinations of 8 building floor areas, 3 numbers of floors, 3 cooling setpoints, 2 lighting power densities, 2 electric equipment power densities, 2 occupancy densities, and 2 window operation strategies resulted in a total of 1,152 cases. Due to the repetitive nature of modeling such a relatively large number of cases, a model producer script was written to create the EnergyPlus input files that represent the 1,152 cases.

5.1.2 Energy Simulations

The 1,152 building models were simulated in EnergyPlus. The models were simulated in five cities, which represent the five main climate zones in the United Sates, resulting in a total of 5,760 model instances. According to the ASHRAE 169-2013 standard, the continental United States has five main climate zones (ASHRAE 2013a). Table 5.3 shows the selected cities and their climate properties. The simulations were conducted from June 1 to August 31, with hourly time steps, resulting in a total of 2,208 hours. The TMY3 weather data of the five locations were used. Prior to the energy simulations, the HVAC system of each model was autosized by EnergyPlus based on the corresponding city's summer design days. In order to have an undisturbed consumption pattern throughout the simulation period, the holiday schedules in EnergyPlus were removed. The simulations were conducted on a four-core personal computer, in parallel on all 4 cores, using EnergyPlus 8.8.0.

Table 5.3 Model Locations and Climate Properties

Location	Climate type	CDDs[1]
Chicago	Cool - Humid	468
San Jose	Warm - Marine	398
Phoenix	Very Hot - Dry	2532
Houston	Hot - Humid	1667
New York	Mixed - Humid	672

[1] CDDs = cooling degree days

5.1.3 Data Preprocessing

The EnergyPlus data were preprocessed in preparation for the machine learning. This included three primary steps: feature generation, feature selection, and data sampling.

A total of 36 initial features were generated: five occupant-behavior features (see Section 5.1.1), two building-size features (building floor area and number of floors), and 29 outdoor weather-condition features (extracted from the TMY3 weather data). The weekend and nonworking-weekday hours, during which the buildings are unoccupied and cooling energy consumption is zero, were removed from the dataset. One-hot encoding was used to convert the categorical variables to numerical. The ordinal and continuous variables were normalized using their means and standard deviations to avoid the dominating effect of the features with high values. The resulting dataset included over 10 million hourly data instances.

For feature selection, first, features that are obviously irrelevant to energy consumption prediction (e.g., source of weather data) were removed, based on engineering judgement. Then, further feature selection was carried out, using Neighborhood Component Analysis (NCA), to select the discriminating and non-redundant features. The NCA is a non-parametric and embedded feature selection method, which learns a feature weighting vector by minimizing an objective function that measures the average leave-one-out regression loss with a regularization term (Yang et al. 2012).

Subsequently, the data were sampled to reduce the computational cost of training the machine learning models using such a large dataset and to evaluate the performance of the machine learning algorithms with different sample sizes. The conditional Latin hypercube sampling (cLHS) method was used for the sampling, which is a stratified random procedure for sampling existing ancillary data (Minasny and McBratney 2006). Datasets with 1,000, 2,000, 5,000, 10,000, 20,000, 50,000, 100,000, 200,000, 500,000, and 1,000,000 data instances were sampled. Then, each dataset was randomly split into training, validation, and testing datasets with proportions of 65%, 10%, and 25%, respectively.

5.1.4 Machine Learning Model Development

A set of hourly cooling energy consumption prediction models were developed to test and compare different machine learning algorithms in terms of prediction accuracy, computational efficiency (training time), and sensitivity to variations in sample sizes. Four machine learning algorithms were tested: CART, ANN, EBT, and DNN. CART and ANN are among the most popular machine learning algorithms in the field of building energy consumption prediction, whereas EBT and DNN are potentially superior but relatively less explored in this field.

To assess the effect of ensembling on the prediction, the performances of the CART (single model) and the EBT (ensemble model) models were compared. The CART algorithm was trained with a minimum of four-leaf-node observations. The training of the EBT algorithm consisted of three main steps. First, 50 data subsets were generated by randomly sampling the training dataset with replacement. Second, a regression-tree-based weak learner was trained for each of the generated data subsets. Third, the weak learners were ensembled using the bagging algorithm, where the EBT's prediction is the average of the predictions made by the weak learners.

To understand the effect of the depth of the neural network models on the prediction, four different neural network models with different number of hidden layers were compared: an ANN algorithm, with 22 input neurons and 15 neurons in a single hidden layer; and three DNN algorithms, also with 22 input neurons, but with 15 neurons in two, three, and four hidden layers. The models were trained using the training dataset, the MATLAB's neural network training tool, and the statistical and machine learning toolbox. The parameters of the machine algorithms were tuned through parameter grid search using the validation datasets, for each model to maximize the prediction performance. The final parameters are shown in Table 5.4.

Table 5.4 Parameter Tuning Results

Algorithm	Parameter	Value
CART	Minimum leaf size	4
EBT	Minimum leaf size	2
	Number of learners	50
	Fraction of training set to resample	1.0
ANN (FFNN)	Number of neurons in the hidden layer	15
	Training function	Bayesian regularization
	Activation function	Tan-Sigmoid
DNN	Number of hidden layers	2, 3, 4
	Number of neurons in each layer	15
	Training function	Bayesian regularization
	Activation function	Tan-sigmoid

5.1.5 Performance Evaluation

Three performance metrics were used to evaluate the prediction performance of the models: CV, RMSE, and R^2. The trained models were used to predict the hourly cooling energy consumptions of the instances in the testing dataset. The predicted values were compared to the actual (simulated) values and the CV, RMSE, and R^2 were calculated, as per Eq. (1.1) to (1.3). CV is a measure to assess the variability between the predicted and the actual energy consumptions. RMSE is the standard deviation of the residuals between the predicted and the actual energy consumptions. R^2 is a measure to assess how much of the variance in the actual energy consumption levels are explained by the model (Wang et al. 2016). The lower the CV and RMSE and the higher the R^2, the more similar dispersions are between the predicted and the actual consumptions. CV was utilized as the primary performance metric, while RMSE and R^2 were only utilized as tie breakers when the CV did not show a significant difference between the models.

5.2 Results and Discussion

5.2.1 EnergyPlus Simulation Results

The EnergyPlus simulations generated a dataset with a range of occupant behaviors, building sizes, outdoor weather conditions, and resulting energy consumption values. Figure 5.1 shows a sample of these simulation results. Figure 5.2 aggregates the simulated hourly cooling energy consumption to annual energy consumption for all models. As shown in Figure 5.2, the range of energy consumption levels is wide. For example, the highest energy-consuming model used 3432.50 times more energy than the lowest-consuming model, which demonstrates the high impact of the variables considered in this study on building energy consumption. When only comparing same-size models, the highest consumer consumed 20.48 times more than the lowest consumer. This shows the combined impact of occupant behavior and outdoor weather conditions on energy consumption. Comparing models with the same building characteristics and in the same location, the highest energy consumer used 7.36 times more energy than the lowest consumer. This shows that occupant behavior, alone, has a major impact on building energy consumption.

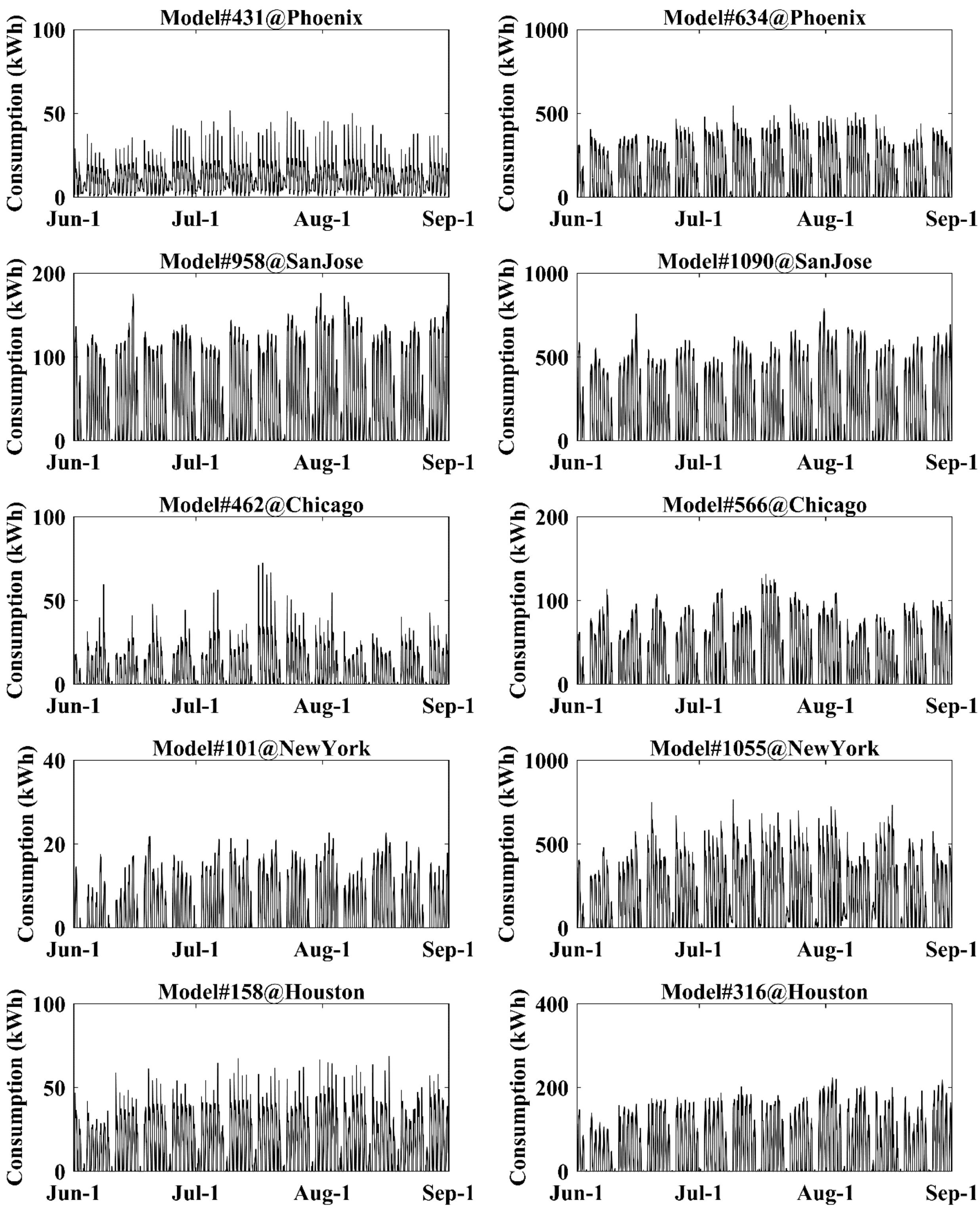


Figure 5.1 – Hourly Cooling Energy Consumption Levels of Randomly Selected Models

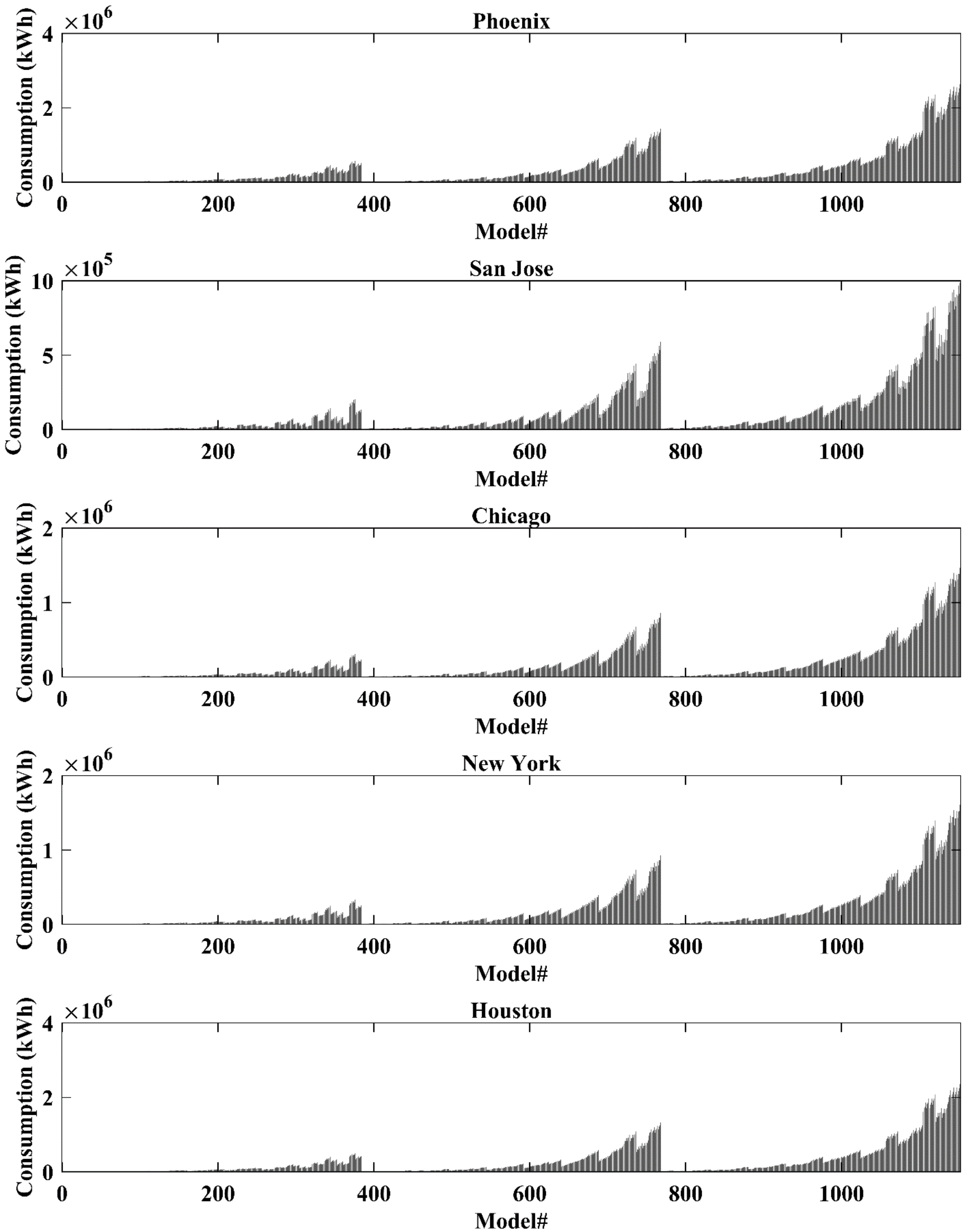


Figure 5.2 – Aggregated Hourly Cooling Energy Consumption Levels for All Models

5.2.2 Feature Selection and Data Sampling

The feature selection and data sampling steps reduced the dimensionality and sample size of the dataset. The feature size was reduced from 36 to 12. The final, discriminating features that were used for the prediction included: (1) two building-property features: building size and number of floors, (2) eight outdoor weather-condition features: dry bulb temperature, atmospheric pressure, extraterrestrial horizontal radiation, extraterrestrial direct normal radiation, horizontal infrared radiation intensity from sky, wind direction, wind speed, and precipitable water, and (3) two occupant-behavior features: cooling setpoint and window operation. Previous hours' values of these features were not utilized, because such approach makes prediction models more complicated and computationally expensive (Fan et al. 2017). Looking at the energy simulation results, these discriminating features had the highest impact on energy consumption. For example, a model consumed as much as 3.86 times more energy than another model with the same building characteristics, in the same location, and with the same occupant behavior except for only cooling setpoint. When comparing the models with differences in only window operation, the difference was as much as 2.69 times more. On the other hand, such high differences were not observed for the remaining, nondiscriminating features. For example, when comparing models with differences in only lighting power density, the maximum energy consumption difference was only 1.20 times more.

For data sampling, the sampled datasets (see Section 5.1.3) represent the characteristics of the original dataset. Figure 5.3 shows the histogram of the cooling energy consumption of the original dataset. Figure 5.4 shows the histograms of the cooling energy consumption of the sampled datasets. The sampled datasets preserve the characteristics of the original dataset, which are summarized in Table 5.5 and Figure 5.3. For example, the frequency of energy consumption levels within the range of 0-250 kWh is 68.14% for the original dataset; while the same frequency is 67.20%, 68.80%, 68.84%, 68.15%, 67.59%, 68.06%, 68.11%, 68.19%, 68.16%, and 68.14% for the sampled datasets with 1,000, 2,000, 5,000, 10,000, 20,000, 50,000, 100,000, 200,000, 500,000, and 1,000,000 hourly data instances, respectively.

Table 5.5 Statistics of the Original Dataset

Feature	Min	Mean	Median	Max
Building size	2500	128125	60000	500000
Number of floors	1	2	2	3
Cooling setpoint	22.8	24.0	24.0	25.2
Dry bulb temperature	6.7	25.5	25.0	44.4
Atmospheric pressure	96200.0	100038.8	101000.0	102400.0
Extraterrestrial horizontal radiation	0	459.3	243.5	1311.0
Extraterrestrial direct normal radiation	0	781.6	1321.0	1342.0
Horizontal infrared radiation intensity from sky	273.0	395.5	396.0	531.0
Wind direction	0	183.2	190.0	360.0
Wind speed	0	3.6	3.6	16.0
Precipitable water	50.0	286.9	279.0	569.0

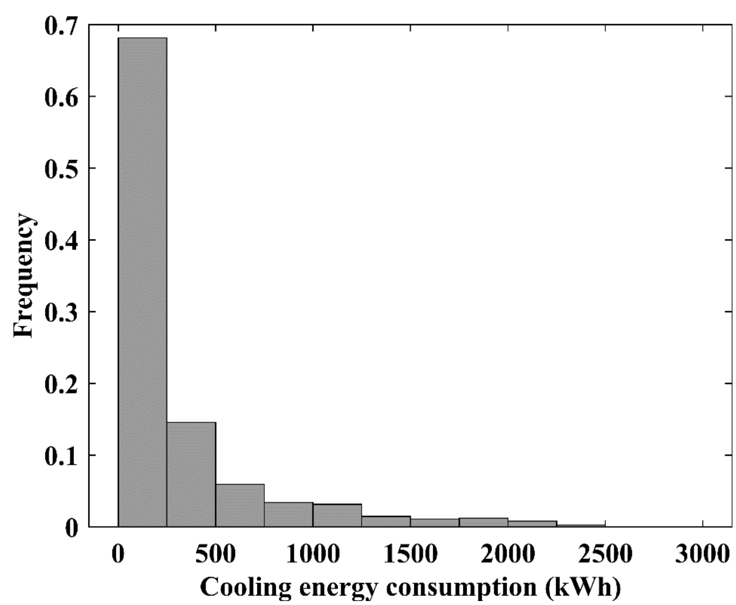


Figure 5.3 – Hourly Cooling Energy Consumption Levels of the Original Dataset

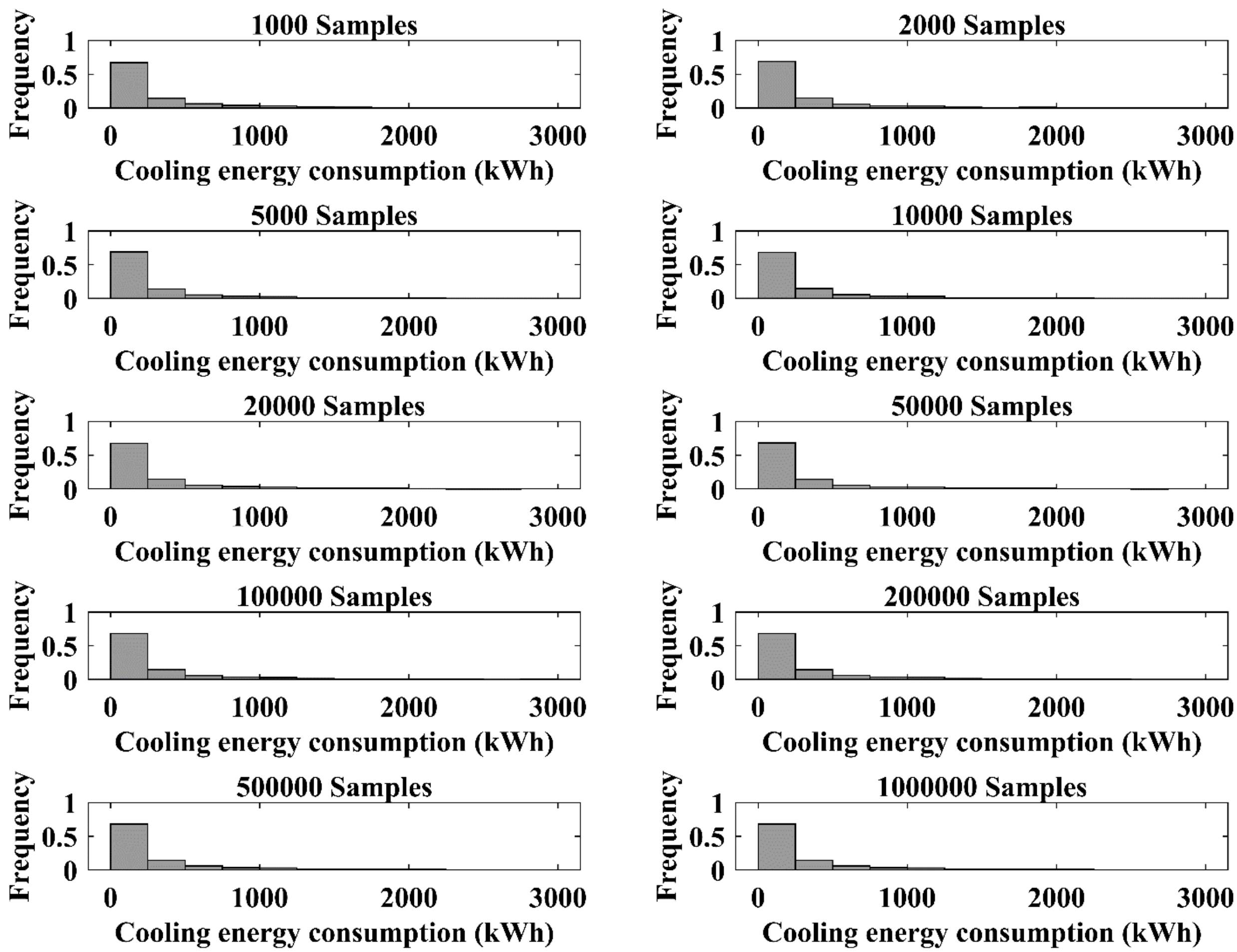


Figure 5.4 – Hourly Cooling Energy Consumption Levels of the Sampled Datasets

5.2.3 Prediction Performance

5.2.3.1 Prediction Performance

Table 5.6 summarizes the prediction performance results of the four algorithms: CART, EBT, ANN, and DNN. When sufficient amounts of data were used for training, all four algorithms were able to achieve accurate predictions within reasonable training times. The ANN and DNN models may become computationally expensive to train in some sample sizes but can achieve very high prediction accuracies compared to the CART and EBT models. A number of findings can be drawn from the results of this study. First, the neural network models with the optimal number of hidden layers outperformed both the CART and EBT models in terms of prediction accuracy, for all 10 sample sizes. The highest prediction accuracy, 2.97%, was achieved by the DNN model with four hidden layers (trained using the 1,000,000 dataset in about 10 hours). Second, the CART and EBT algorithms required more training data than the ANN and DNN algorithms to be considered calibrated. According to the ASHRAE Guideline 14, an hourly prediction model is considered calibrated if its hourly CV values fall below 30%. For example, the ANN model required a minimum sample size of 1,000, compared to 5,000 for the EBT model. Third, for all sample sizes, the training times of the CART and EBT algorithms were less than those for the ANN and DNN algorithms. For example, for a sample size of 100,000, the DNN model with four hidden layers required 4,122 seconds to train, while the CART model required only 1 second. Also, for a given training time period, the CART and EBT models achieved the highest prediction accuracies. For example, the EBT model could achieve 6.35% CV in 169 seconds (for a sample size of 1,000,000). But, in about the same time, the ANN model with a single hidden layer (for a sample size of 50,000) and the DNN model with two hidden layers (for a sample size of 10,000) could achieve 6.97% and 6.48%, respectively. Fourth, the DNN models were able to achieve a CV of less than 5% for sample sizes larger than 50,000, but the other algorithms were never able to achieve such low CV. However, the training time of the DNN to achieve such a high prediction accuracy was significantly higher than the other algorithms (e.g., 10 hours vs. 7 seconds). The choice of the algorithm to use, therefore, depends on the application requirements (e.g., needed accuracy) and constraints (e.g., data availability).

Table 5.6 Performance of the Cooling Energy Consumption Prediction Models

Sample size	Algorithm	CV	RMSE (kWh)	R^2	Training time (sec)
1,000	ANN	5.55%	49.83	98.97%	7
	DNN {2}	8.38%	75.25	97.00%	67
	DNN {3}	168.73%	1515.69	-966.24%	1
	DNN {4}	220.72%	1982.64	-2051.82%	2
	CART	99.63%	894.96	74.72%	<1
	EBT	89.52%	804.11	79.59%	<1
2,000	ANN	11.11%	30.13	99.49%	11
	DNN {2}	12.96%	35.13	99.18%	87
	DNN {3}	20.02%	54.27	97.91%	217
	DNN {4}	15.19%	41.19	98.83%	385
	CART	34.41%	93.28	93.90%	<1
	EBT	30.51%	82.72	95.20%	<1
5,000	ANN	8.22%	22.95	99.72%	16
	DNN {2}	10.00%	27.93	99.59%	127
	DNN {3}	11.36%	31.72	99.38%	293
	DNN {4}	15.00%	41.90	99.14%	510
	CART	30.72%	85.81	96.39%	<1
	EBT	25.63%	71.60	97.49%	<1
10,000	ANN	7.65%	21.66	99.75%	15
	DNN {2}	6.48%	18.35	99.80%	177
	DNN {3}	8.46%	23.98	99.67%	390
	DNN {4}	9.88%	28.00	99.59%	739
	CART	22.04%	62.44	97.94%	<1
	EBT	18.14%	51.39	98.60%	1.55
20,000	ANN	7.04%	20.31	99.77%	43
	DNN {2}	5.50%	15.85	99.86%	267
	DNN {3}	5.93%	17.10	99.84%	591
	DNN {4}	7.40%	21.35	99.74%	997
	CART	17.36%	50.06	98.59%	<1
	EBT	13.77%	39.71	99.11%	2.1
50,000	ANN	6.97%	19.93	99.78%	149
	DNN {2}	4.88%	13.93	99.89%	548
	DNN {3}	4.43%	12.66	99.91%	1313
	DNN {4}	3.27%	9.33	99.95%	2172
	CART	14.59%	41.69	99.04%	<1
	EBT	10.82%	30.91	99.47%	6
100,000	ANN	7.48%	21.34	99.75%	323
	DNN {2}	4.65%	13.26	99.90%	1065
	DNN {3}	3.72%	10.62	99.94%	2496
	DNN {4}	3.21%	9.14	99.95%	4122
	CART	12.55%	35.80	99.27%	1
	EBT	9.00%	25.67	99.62%	11
200,000	ANN	7.20%	20.57	99.76%	481
	DNN {2}	4.77%	13.64	99.90%	1751
	DNN {3}	3.68%	10.51	99.94%	4503
	DNN {4}	3.19%	9.11	99.96%	7454
	CART	11.03%	31.50	99.46%	1.5
	EBT	7.94%	22.68	99.72%	22
500,000	ANN	7.40%	21.15	99.75%	1323
	DNN {2}	4.71%	13.45	99.90%	4800
	DNN {3}	3.88%	11.08	99.93%	11920
	DNN {4}	3.54%	10.10	99.94%	17534
	CART	9.27%	26.48	99.61%	3
	EBT	6.79%	19.40	99.79%	57
1,000,000	ANN	7.23%	20.67	99.77%	2559
	DNN {2}	4.88%	13.96	99.89%	9138
	DNN {3}	3.57%	10.22	99.94%	21819
	DNN {4}	2.97%	8.50	99.96%	35154
	CART	8.48%	24.27	99.67%	7
	EBT	6.35%	18.16	99.82%	169

DNN {x} x: number of hidden layers

5.2.3.2 Comparison of CART and EBT

In comparison to the CART models, the EBT models achieved higher accuracies but took longer times to train. These accuracy differences could have partially resulted from the characteristics of the original dataset. In general, the accuracy margin between the two types of models is more apparent in sparse datasets, because CART models suffer more from instability in sparse datasets (Wang et al. 2018). In this study, the datasets were relatively diverse and sparse (because they represent different building sizes, weather conditions, and occupant behaviors), resulting in these significant accuracy differences. Such differences, however, may or may not be observed if other training datasets are used.

For all sample sizes, the CART models required less times to train than the EBT models. For example, for a sample size of 1,000,000, training the EBT model took about 25 times more than training the CART model. This is probably due to two reasons. First, EBT models require extra time due to data sampling. Second, for an EBT model, multiple models are trained on the same dataset, in comparison to only a single model for a CART model. Nevertheless, for all sample sizes, the training times of both types of models remained in acceptable ranges.

The most important advantage of the CART models is their transparent and explicit structures. The EBT models, however, are not as explicit as the CART models. But, in terms of prediction accuracy, the EBT models always outperformed the CART models. Since accuracy is highly important in many applications and contexts, EBT models can be preferred to CART models for building energy consumption prediction (Wang et al. 2018). In general, the results show that the CART and EBT models are computationally inexpensive to train and can achieve good prediction accuracies. Thus, bagging should be one of the ensembling methods to consider when developing a building energy consumption prediction model.

5.2.3.3 Comparison of ANN and DNN

Shallower architectures outperformed deeper architectures for smaller datasets but fell below deeper architectures for larger datasets. For example, for sample sizes of 1,000, 2,000, and 5,000, the typical ANN model with a single hidden layer outperformed the others in terms of prediction accuracy. For sample sizes of 10,000 and 20,000, the DNN model with two hidden layers outperformed. For all other larger datasets, the DNN models with four hidden layers outperformed as well. These results indicate that the increase in the number of hidden layers does not always guarantee an increase in the model performance. But, the increase in the number of hidden layers increased the model complexity and therefore caused longer training times. With larger datasets, on the other hand, deeper architectures always outperformed shallower architectures. This is consistent with the conclusions drawn by Fan et al. (2017): "The increase in the number of hidden layers leads to a dramatic increase in the number of model coefficients. To develop robust and reliable estimations of these model coefficients, a huge amount of data is needed".

In most sample sizes, the shallower architectures required much less time to train than the deeper architectures. For example, for a sample size of 1,000,000, training the DNN model with four hidden layers took about 14 times more than training the typical ANN model with a single hidden layer. One interesting result is the very low prediction accuracy and the very short training time of the DNN models with three and four hidden layers for the sample size of 1,000. A possible explanation for this is overfitting. The use of larger number of hidden layers, more than the needed number to represent the complexity of the prediction function, leads to overfitting problems (Karsoliya 2012). In addition, the training of the DNN model with four hidden layers using the sample size of 1,000,000 took about 10 hours, which brings some practicality concerns despite the very high accuracy (2.97% CV) of the resulting model.

CHAPTER 6 – REAL-DATA-DRIVEN OCCUPANT-BEHAVIOR-SENSITIVE MACHINE LEARNING-BASED ENERGY CONSUMPTION PREDICTION AND BEHAVIOR OPTIMIZATION

6.1 Methodology

A data-driven approach to determine the optimal occupant behavior that can simultaneously reduce energy consumption and improve comfort was proposed. The proposed approach consists of two components: (1) a set of machine learning-based occupant-behavior-sensitive models for real-data-driven prediction of hourly cooling and lighting energy consumption and thermal and visual occupant comfort, and (2) a genetic algorithm-based optimization model for optimizing occupant behavior settings for each hour using the prediction models. To test and evaluate the proposed approach, the PBTC building was instrumented for real data collection (see Chapter 4). Cooling and lighting energy consumption, indoor environmental condition, outdoor weather condition, and occupant behavior data were collected for about three months. The collected data were then preprocessed – cleaned, aggregated, integrated, and normalized. Figure 6.1 shows an overview of the research methodology.

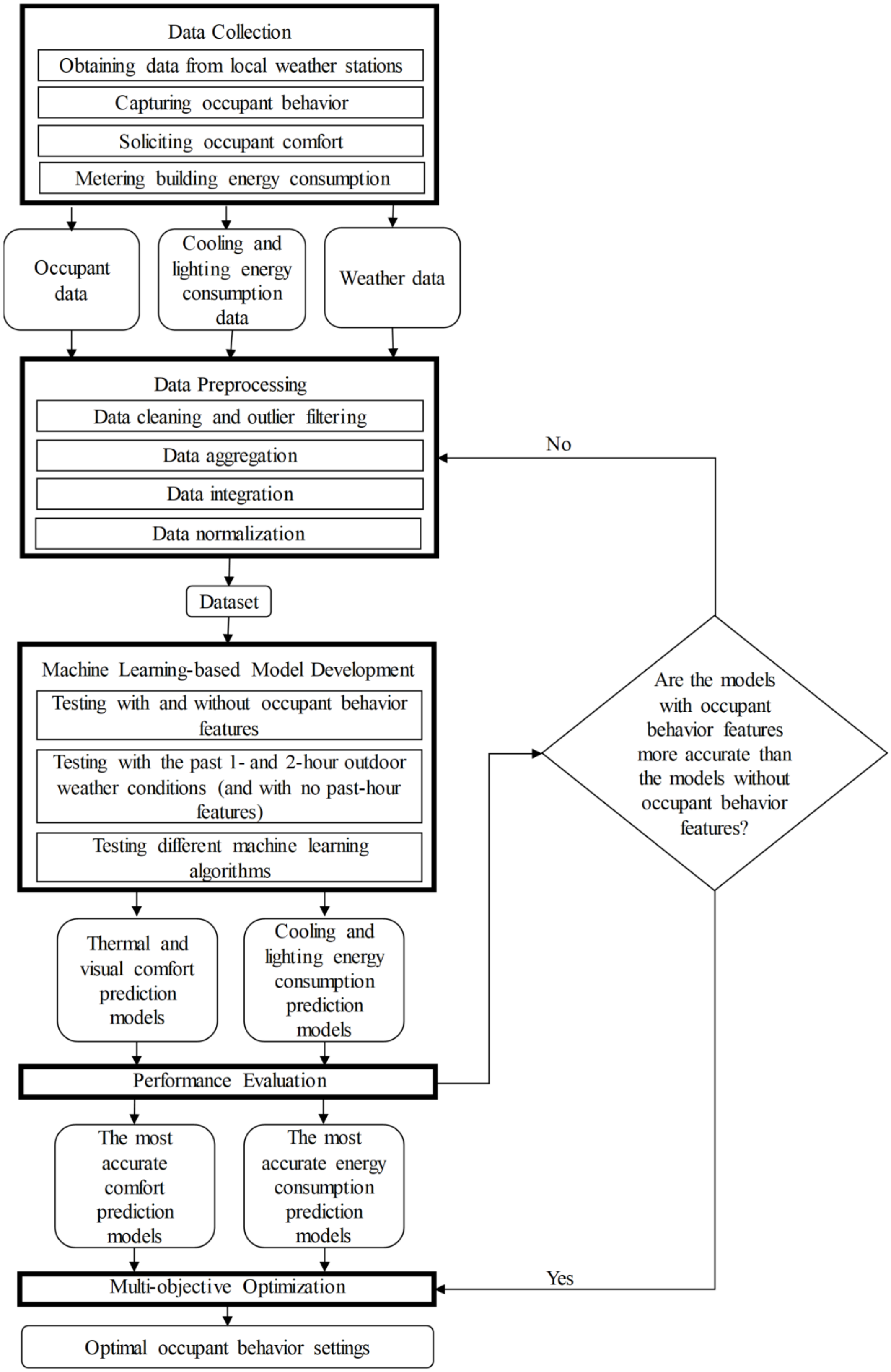


Figure 6.1 – An Overview of the Research Methodology

6.1.1 Data Preprocessing

Data preprocessing included four primary steps: data cleaning and outlier filtering, data aggregation, data integration, and data normalization. First, the data instances that have missing and/or outlier values were removed from the dataset. Cook's distance was used to compute and identify outliers. The instances with outliers represent 0.9% of the entire dataset. Second, 15-min data intervals of cooling and lighting energy consumption were aggregated into hourly consumption values using summation. Third, occupant behavior data from multiple occupants were aggregated using a majority voting strategy. The behavior of an occupant was assumed unchanged until another behavior is reported or until the end of the day. Data from multiple sources including energy consumption data, indoor environmental conditions data, outdoor weather conditions data, and occupant behavior data, were integrated using their date and time. The date and time variables were then replaced with a working-hour indicator. Fourth, each variable in the dataset was normalized between 0 and 1 to avoid overflowing of an individual variable.

6.1.2 Data Analysis

The data analysis methodology was composed of three primary steps: (1) data transformation: the collected data were transformed in order to represent building operation efficiencies, (2) clustering of energy-use modes: the transformed data were clustered into different energy-use modes based on operation efficiency, and (3) statistical analysis: the correlation between building operation efficiency and thermal comfort was assessed, and thermal comfort levels under the three energy-use modes were compared.

A principal component analysis (PCA) was conducted to transform the collected data into a more meaningful form that can better indicate energy efficiency. PCA is a technique that is commonly used to transform data into a more meaningful representation to obtain strong patterns. The temperature difference and energy consumption data, which are the original variables, were transformed into a new variable (PC2) that represents building operation efficiency.

The building operation efficiencies (PC2s) were clustered into different groups that represent different energy-use modes, using the k-means clustering algorithm, which is an unsupervised algorithm to cluster data into a given number of groups based on a similarity measure (e.g., distance functions). For this study, the data were clustered into three energy-use modes: saver, balanced, and spender. After the clustering, PC2 was transformed back to the original variables for better visualization.

A Spearman's rank correlation analysis and Kruskal-Wallis H test were used for statistical analysis. A Spearman's rank correlation analysis was conducted to assess the correlation between building operation efficiency and thermal comfort. The thermal comfort levels under the three energy-use modes were then compared using the Kruskal-Wallis H test, which is a rank-based nonparametric test to determine whether two or more groups are significantly different. The results of the Kruskal–Wallis H test were interpreted based on the significance level (p) of the test. If the significance level is less than 0.05, then there is a significant difference across the modes.

6.1.3 Machine Learning Model Development

A set of machine learning-based occupant-behavior-sensitive prediction models for real-data-driven prediction of cooling and lighting energy consumption and thermal and visual occupant comfort were developed. The prediction of cooling and lighting energy consumption is a regression problem, because cooling and lighting energy consumption data are continuous. The prediction of thermal and visual comfort, on the other hand, is a classification problem, because thermal and visual comfort data are categorical. For cooling energy consumption and thermal comfort prediction, the following features were used: working-time indicator, hourly average temperature, hourly average dewpoint temperature, hourly average relative humidity, hourly average wind speed, DNI, DHI, GHI, thermostat setpoint, portable heater status, door status, and window shade status. For lighting energy consumption and visual comfort prediction, the following features were used: working-time indicator, DNI, DHI, GHI, window shade status, and light status. For the comfort prediction models, two types of models were developed: group and individual. The group comfort models were used to find optimal group solutions. The individual comfort models are personalized and were used to predict the impact of the optimal group solutions on individual comfort.

To verify that the behavior features are discriminating, and hence that the prediction models can be used to predict the impact of the behavior, the performance of the models were compared to others without occupant-behavior features. This included all aforementioned features except thermostat setpoints, portable heater status, door status, window shade status, and light status. To consider the delayed effects of outdoor weather, prediction models with the past one- and two-hour outdoor weather condition features and with no past-hour features were tested and compared. A feature selection was also conducted, using the least absolute shrinkage and selection operator (LASSO), to verify that all the aforementioned behavior and outdoor weather condition features are discriminating.

A set of machine learning algorithms for regression and classification were also tested to determine the most accurate in energy consumption and comfort prediction. For energy consumption prediction, the following regression algorithms were tested: Support Vector Regression (SVR), ANN, CART, and MLR. For comfort prediction, the following classification algorithms were tested: SVM, ANN, DT, and KNN. These algorithms were selected because they are among the most popular and accurate machine learning algorithms.

For all comparisons, the Welch's t-test was performed to assess the statistical significance of the differences in the prediction results by the different models. The difference was considered significant when the resulting *p*-value was less than 0.05 (i.e., $p<0.05$).

For all prediction models, the parameters of the machine algorithms were individually tuned using grid search for each model to maximize the prediction performance. The models were trained using the MATLAB's neural network training tool, and the statistical and machine learning toolbox. In training the models, outdoor weather condition and thermostat setpoint data were represented as continuous variables, whereas working-time indicator and occupant-behavior (except thermostat setpoint) data were represented as binary variables.

6.1.4 Prediction Performance Evaluation

A 10-cross fold validation was utilized to assess the performance, because it minimizes the bias due to the randomness in choosing the testing data (Chou and Bui 2014). The following performance metrics were utilized for consumption prediction, which were calculated using Eq. (1.1) to (1.3): CV, RMSE, and R^2. RMSE is the standard deviation of the residuals between the predicted and the actual energy consumption values. CV is a measure to assess the variability between the predicted and the actual energy consumption values. R^2 is a measure to assess how much of the variance in the actual energy consumption values are explained by the model. The lower the RMSE and CV and the higher the R^2, the more similar dispersions are between the predicted and the actual consumptions. The RMSE was utilized as the primary performance metric and CV and R^2 were only utilized as tie breakers when the RMSE did not show a significant difference between the models.

The following performance metrics were utilized for comfort prediction, which were calculated using Eq. (6.1) and Eq. (6.2): MAE and MSE.

$$\text{MAE} = \frac{\sum_{i=1}^{n} \left| y_{predict,i} - y_{data,i} \right|}{n} \tag{6.1}$$

$$\text{MSE} = \frac{\sum_{i=1}^{n} (y_{predict,i} - y_{data,i})^2}{n} \tag{6.2}$$

where $y_{predict,i}$ is the predicted thermal (or visual) comfort level at hour i, $y_{data,i}$ is the actual thermal (or visual) comfort level at hour i, and n is the number of hours in the dataset.

6.1.5 Optimization

A genetic algorithm-based optimization model for optimizing occupant behavior was developed. The main purpose of the optimization is to minimize energy consumption while maximizing thermal and visual comfort. The decision variables, objective functions, and optimization computations of the proposed optimization were formulated as follows. The decision variables included thermostat setpoint, portable heater status, door status, window shade status, and light status for each hour. Three objective functions were formulated: (1) minimizing hourly energy consumption (sum of cooling and lighting energy consumptions), (2) maximizing hourly group thermal comfort, and (3) maximizing hourly group visual comfort. The hourly cooling and lighting energy consumptions, and thermal and visual comfort were calculated using the prediction models (see Section 6.1.4). The NSGA-II (Deb 2001) was used for conducting the optimization due to the algorithm's capabilities of fast non-dominated sorting approach, fast crowded distance estimation procedure, and simple crowded comparison operator (Yusoff et al. 2011). Similar to any genetic algorithm, the NSGA-II is based on the evolution of a population of individuals (or chromosomes), each representing a solution for the optimization problem (Gou et al. 2017). The optimization was conducted using the MATLAB's optimization toolbox. A population size of 50 individuals, crossover probability of 90%, mutation probability of 10%, tolerance of 0.0001 and termination criteria of 700 generations were selected as the parameters of the algorithm.

In conducting the optimization, four types of optimal solutions were considered: energy-priority, thermal-comfort-priority, and visual-comfort-priority, and balanced solutions. An energy-priority solution places priority on minimizing energy consumption, while a comfort-priority solution places priority on maximizing comfort. These three types of solutions are useful for demonstrating the extreme savings or comfort improvements that are possible, but do not improve all three objectives equally. In a balanced solution, an equal importance is given to all three objectives using the weighted sum method, which turns the three objectives into a single objective by adding each objective pre-multiplied by an equal weight.

6.2 Results and Discussion

6.2.1 Analysis of Cooling Energy Consumption and Thermal Comfort

Figure 6.2 – (a) and Figure 6.3 – (a) show the thermal comfort levels and corresponding cooling energy consumption levels and temperature differences for thermal zone 1 and thermal zone 2, respectively. The data is scattered along PC1; there is more variation on PC1 than PC2. This can be explained by the various outdoor and indoor temperatures and their corresponding energy consumption levels. On the other hand, PC2 shows the variation in cooling energy consumption for the unit temperature difference. When PC2 increases, the energy consumed for a unit temperature difference decreases. In Figure 6.2 – (b) and Figure 6.3 – (b) the data were projected on PC2 as it represents building operation efficiency. The higher the value of PC2 is, the less efficient the building operation is.

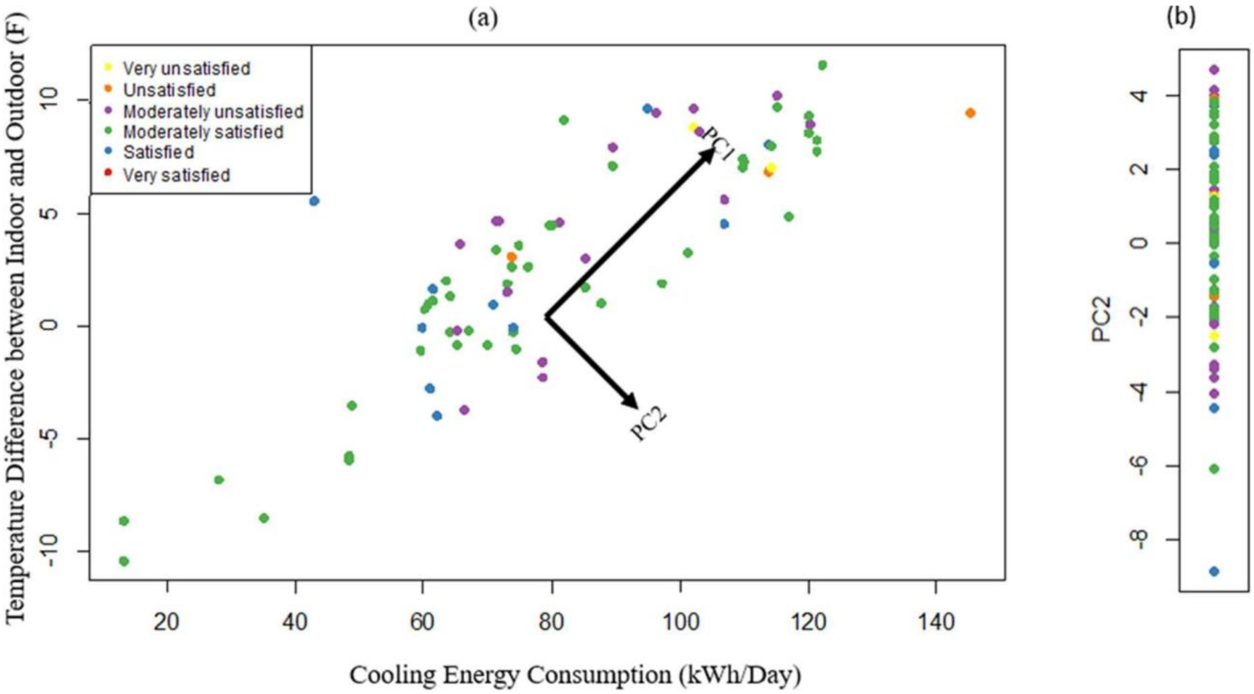


Figure 6.2 – (a) Thermal Zone 1 – Cooling Energy Consumption, Temperature Difference, and Occupant Thermal Comfort (b) Data Projection on Principal Component 2

Figure 6.4 shows the clustered data based on PC2 and their corresponding energy-use modes for thermal zone 1 and thermal zone 2. There is a balanced distribution across the three energy-use modes. There are 26 spender, 32 balanced, and 34 saver instances for thermal zone 1; and 10 spender, 14 balanced, and 10 saver instances for thermal zone 2. For example, for thermal zone 1, an instance which belongs to the saver mode consumed 42.9 kWh/day for 5.5^{o} F temperature difference and provided a “satisfied” thermal comfort. On the other hand, another instance which belongs to the spender mode consumed 78.6 kWh/day for -2.3^{o} F temperature difference and provided a “moderately unsatisfied” thermal comfort. This shows that a saver operation can provide more satisfaction with thermal comfort. However, this case is only one example and does not necessarily represent a statistical difference. The results of the Kruskal-Wallis H test are, thus, necessary to test whether the differences in the thermal comfort levels under the different energy use-modes are statistically significant or not.

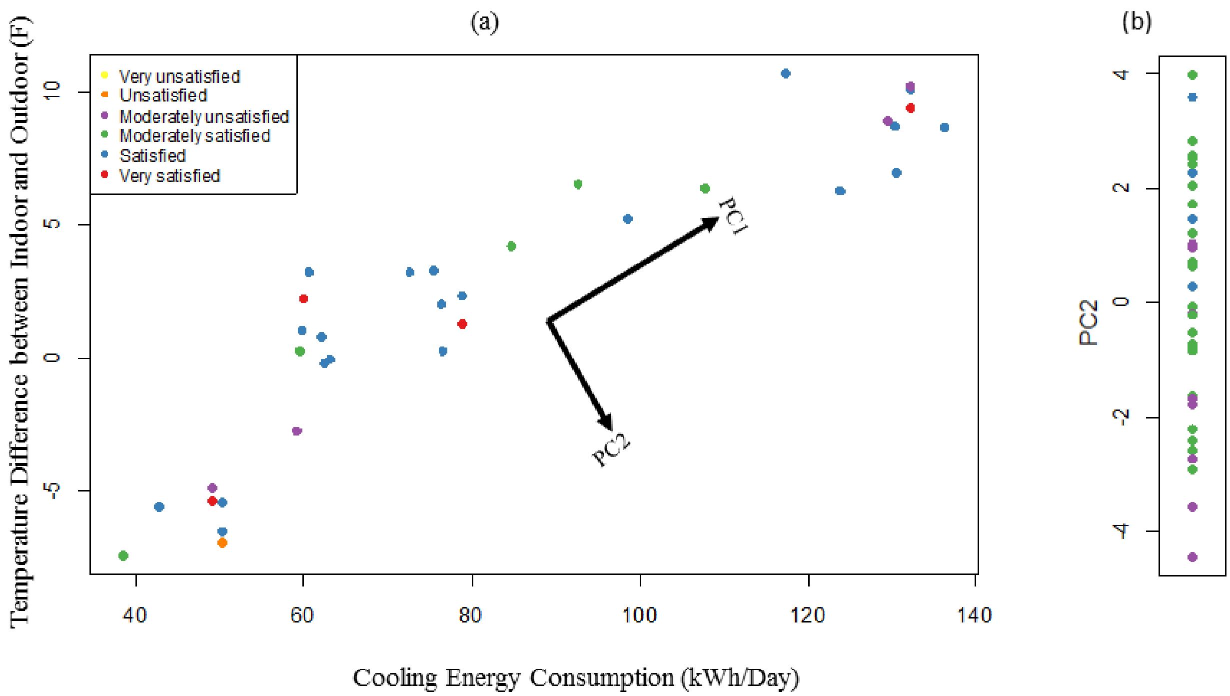


Figure 6.3 – (a) Thermal Zone 2 – Cooling Energy Consumption, Temperature Difference, and Occupant Thermal Comfort (b) Data Projection on Principal Component 2

The values of the correlation coefficient between PC2 and thermal comfort are 0.10 and -0.14 for thermal zone 1 and 2, respectively. PC2, which is an indicator of building operation efficiency, has a very weak correlation with thermal comfort. This indicates that building operation strategies that reduce building energy consumption without sacrifice in thermal comfort can be found.

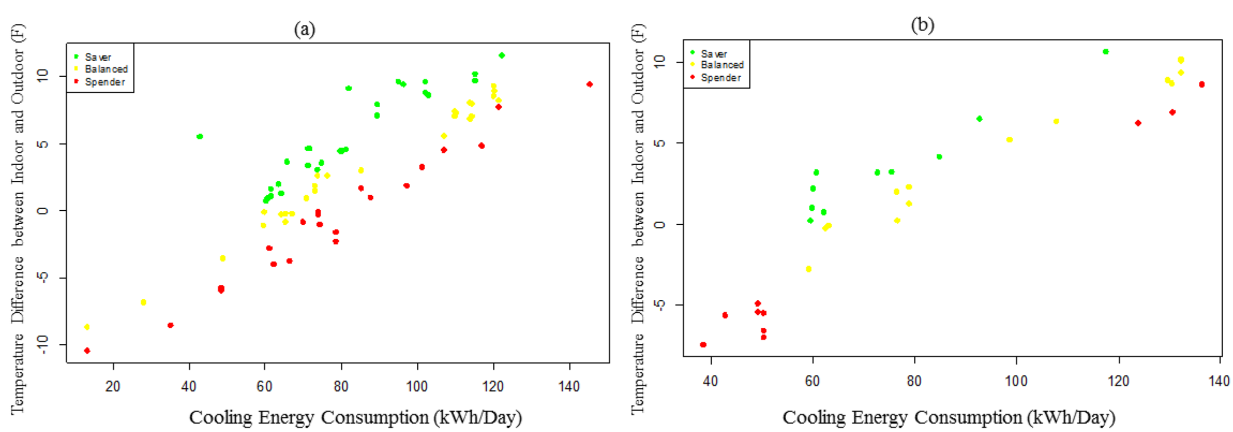


Figure 6.4 – Energy-Use Modes for (a) Thermal Zone 1 and (b) Thermal Zone 2

Figure 6.5 – (a) shows the satisfaction levels of occupants with thermal comfort across the three energy-use modes for thermal zone 1. On average, occupants are in between moderately

unsatisfied and moderately satisfied with their thermal comfort under all three energy-use modes. Overall, 35%, 25%, and 16% of the occupants under the saver, balanced, and spender energy-use modes, respectively, are moderately unsatisfied or lower. However, based on the Kruskal–Wallis H test, no significant differences in the thermal comfort levels of the occupants across the different energy-use modes were found (p=0.40). Figure 6.5 – (b) shows the satisfaction levels of occupants with thermal comfort across the three energy-use modes for thermal zone 2. On average, occupants are in between moderately satisfied and satisfied with their thermal comfort under all three energy-use modes. Overall, 0%, 21%, and 20% of the occupants under the saver, balanced, and spender energy-use modes, respectively, are moderately unsatisfied or lower. Similarly, no significant differences in the thermal comfort levels of the occupants across the different building energy-use modes were found (p=0.95).

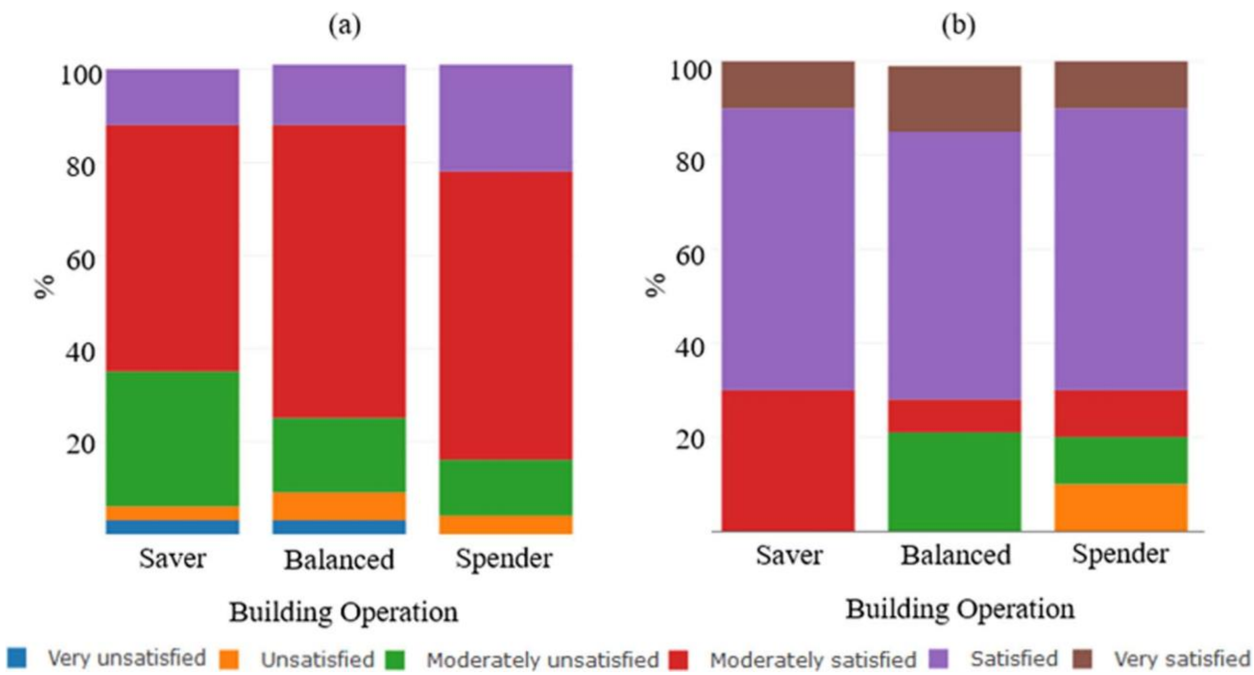


Figure 6.5 – Thermal Comfort Levels for (a) Thermal Zone 1 and (b) Thermal Zone 2

6.2.2 Cooling and Lighting Energy Consumption Prediction

6.2.2.1 Prediction Performance

Table 6.1 summarizes the prediction performance results of the cooling and lighting energy consumption prediction models. For both cooling and lighting energy consumption prediction, SVR models (Model #5 and Model #29) that predict energy consumption using occupant-behavior features, past two-hour outdoor weather condition features, and working-time indicator achieved the best performance: 0.98 kWh and 0.26 kWh RMSE, respectively. Figure 6.6 shows the linear regression between the actual and predicted energy consumption values, for Model #5 (cooling) and Model #29 (lighting). The predicted energy consumption values show a good agreement with the actual levels. Model #5 and Model #29, therefore, were chosen to be used for the optimization.

Table 6.1 Results of the Cooling and Lighting Energy Consumption Prediction for all Models

Model #[1]	Occupant -behavior features	Past hour weather features	Machine learning algorithm[2]	Cooling consumption[3]			Lighting consumption[3]		
				RMSE (kWh)	CV	R^2	RMSE (kWh)	CV	R^2
1 (25)	✓	0	SVR	1.09	17.81%	98.13%	0.32	14.88%	96.97%
2 (26)	✗	0	SVR	1.25	30.21%	96.56%	0.38	19.25%	96.15%
3 (27)	✓	1	SVR	1.02	16.80%	98.18%	0.31	14.54%	97.61%
4 (28)	✗	1	SVR	1.31	31.58%	96.52%	0.38	19.09%	96.35%
5 (29)	✓	2	SVR	**0.98**	**16.11%**	**98.19%**	**0.26**	**13.08%**	**98.42%**
6 (30)	✗	2	SVR	1.30	31.49%	96.17%	0.33	15.01%	97.11%
7 (31)	✓	0	ANN	1.35	22.15%	96.97%	0.41	20.07%	95.71%
8 (32)	✗	0	ANN	1.47	35.55%	95.54%	0.48	22.17%	94.36%
9 (33)	✓	1	ANN	1.31	21.48%	97.03%	0.40	19.98%	96.20%
10 (34)	✗	1	ANN	1.45	35.06%	95.59%	0.47	21.98%	95.18%
11 (35)	✓	2	ANN	1.28	21.13%	97.05%	0.38	19.19%	96.27%
12 (36)	✗	2	ANN	1.44	34.80%	95.85%	0.44	21.31%	95.69%
13 (37)	✓	0	CART	1.43	23.50%	96.53%	0.33	15.34%	97.47%
14 (38)	✗	0	CART	1.47	35.60%	95.65%	0.40	18.14%	95.27%
15 (39)	✓	1	CART	1.42	23.33%	96.66%	0.32	15.12%	98.18%
16 (40)	✗	1	CART	1.42	34.40%	95.57%	0.38	17.49%	96.15%
17 (41)	✓	2	CART	1.27	20.96%	97.33%	0.27	14.90%	98.28%
18 (42)	✗	2	CART	1.40	33.87%	95.70%	0.35	16.23%	96.57%
19 (43)	✓	0	MLR	2.35	38.60%	90.78%	0.75	32.33%	89.48%
20 (44)	✗	0	MLR	2.14	51.80%	89.80%	0.78	35.26%	88.34%
21 (45)	✓	1	MLR	2.35	38.61%	90.45%	0.75	31.07%	90.21%
22 (46)	✗	1	MLR	2.14	51.64%	89.84%	0.77	33.81%	89.15%
23 (47)	✓	2	MLR	2.36	38.96%	90.68%	0.71	29.27%	91.34%
24 (48)	✗	2	MLR	2.13	51.41%	89.36%	0.75	32.11%	89.55%

[1]Numbers outside of the parentheses indicate the model numbers for the cooling energy consumption prediction models, and in parentheses indicate those for the lighting energy consumption prediction models.
[2]SVR: Support Vector Regression; ANN: Artificial Neural Networks; CART: Classification and Regression Tree; MLR: Multiple Linear Regression.
[3]RMSE: Root Mean Square Error; CV: Coefficient of Variation.
Best results are shown in **bold font**.

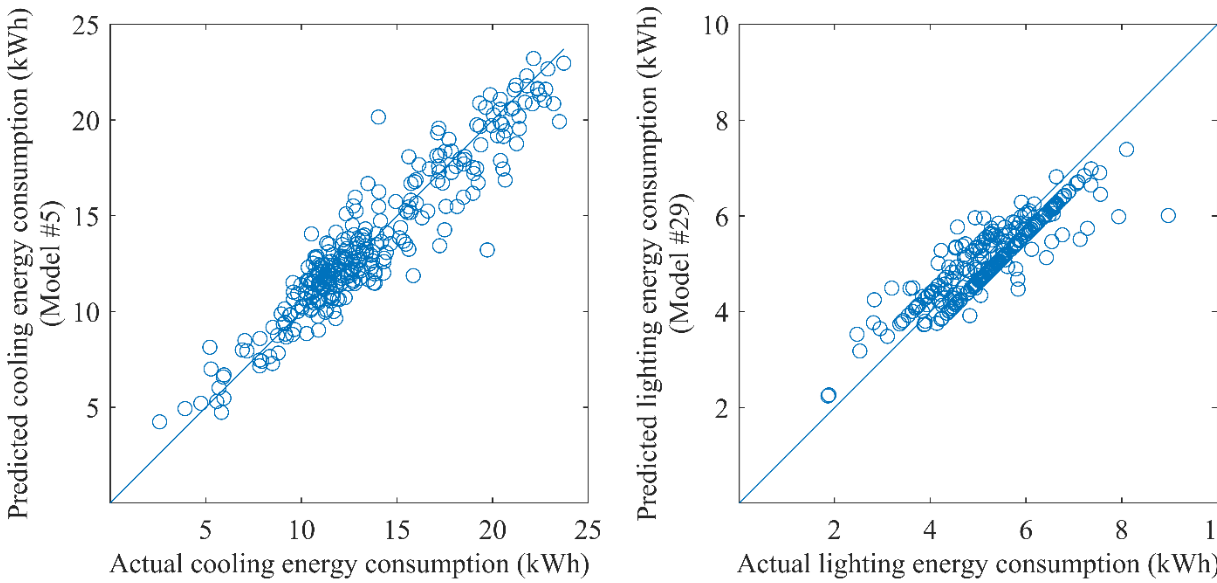
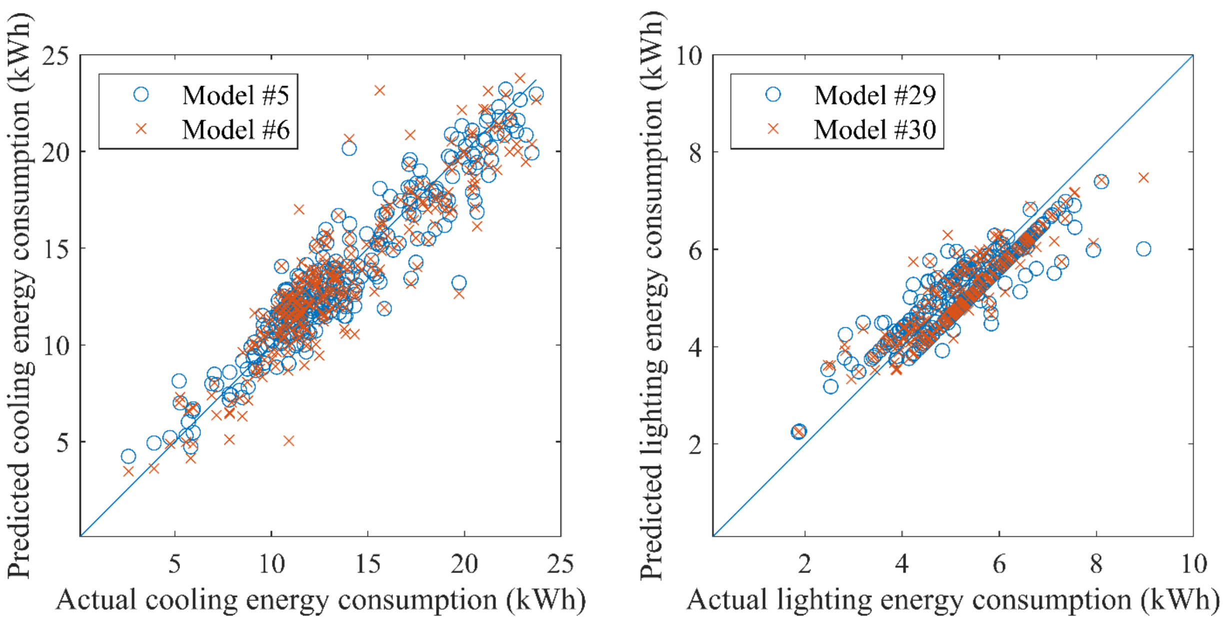


Figure 6.6 – Regression between the Actual and Predicted Cooling and Lighting Energy Consumption Values

6.2.2.2 Occupant Behavior Feature Analysis

The SVR, ANN, and CART models with occupant-behavior features (i.e., all behavior features mentioned in Section 6.1.3) achieved more accurate prediction performance than the models without behavior features. For example, for cooling energy consumption prediction, Model #5 achieved 0.98 kWh RMSE, which is the most accurate; while Model #6, which utilized the same features as Model #5 except for behavior features, achieved 1.30 kWh RMSE. For lighting energy consumption prediction, Model #29 achieved 0.26 kWh RMSE, which is the most accurate; while Model #30, which utilized the same features as Model #29 except for behavior features, achieved 0.33 kWh RMSE. The RMSE improvement due to the inclusion of behavior features ranged from 0.1% to 24.6%. Figure 6.7 shows the regression between the actual energy consumption values and the predictions by these models, which similarly shows Models #5 and #29 provided more accurate predictions than Models #6 and #30. The t-test p-values for the comparisons between the predicted energy consumption values by the models with and without occupant-behavior features were less than 0.05, which indicates that these differences in the predictions are significant. Such performance improvement was not seen for the MLR models, which is probably because linear approaches, such as MLR, cannot capture the nonlinear relationships and higher-order interactions between the features and the energy consumption. Overall, these results indicate that the behavior features are discriminating, and hence that the prediction models can be used to predict the impact of the behavior.



Model #5: SVR with occupant-behavior and past two-hour outdoor weather condition features
Model #6: SVR with no occupant-behavior and past two-hour outdoor weather condition features
Model #29: SVR with occupant-behavior and past two-hour outdoor weather condition features
Model #30: SVR with no occupant-behavior and past two-hour outdoor weather condition features

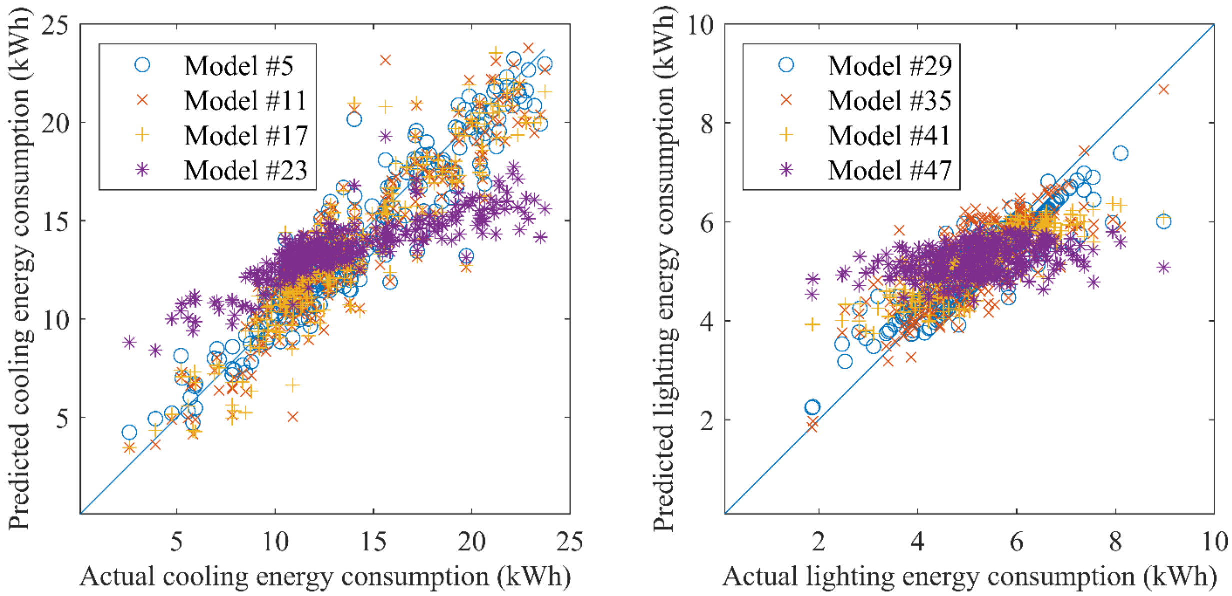
Figure 6.7 – Comparison of Actual and Predicted Cooling Energy Consumption Values Predicted by Models with and without Occupant-Behavior Features

6.2.2.3 Past-Timestep Feature Analysis

The use of additional past-timestep features led to some marginal performance improvement. For example, for cooling energy consumption prediction, Model #5, which utilized the past two-hour weather data (i.e., all weather features mentioned in Section 6.1.3), occupant-behavior data, and the working-time indicator as the features, achieved 0.98 kWh RMSE compared to 1.02kWh RMSE for Model #3 (which utilized the past one-hour data) and 1.09 kWh RMSE for Model #1 (which utilized no past-hour data). Similarly, for lighting energy consumption prediction, Model #29, which utilized the past two-hour data, achieved 0.26 kWh RMSE compared to 0.31 kWh RMSE for Model #27 (which utilized the past one-hour data) and 0.32 kWh RMSE for Model #25 (which utilized no past-hour data). The t-test p-values for the comparisons between the energy consumption values predicted by the models with different past-timestep features were less than 0.05, which shows that although performance improvement is marginal, the differences in the values are significant. The impact of including more than two past hour data was, however, not tested, because the performance differences between including two, one, or no past-hour outdoor weather condition features was marginal.

6.2.2.4 Machine Learning Algorithms

As per Table 6.1, the best prediction performance was achieved by SVR models (Model #5 and Model #29), with no clear outperformer across the ANN and CART models, and with the MLR models achieving the worst performance. For example, for cooling energy consumption prediction, Model #5, which is an SVR model utilizing the past two-hour weather data, occupant-behavior data, and the working-time indicator as the features, achieved 0.98 RMSE. However, Models #11 and #17, which are ANN and CART models utilizing the same set of features as Model #5, achieved 1.28 RMSE and 1.27 RMSE, respectively. On the other hand, Model #23, an MLR model with the same set of features as others, achieved 2.36 RMSE. Similarly, Figure 6.8 illustrates the performance gap between the MLR models and the other models. It shows the cooling and lighting energy consumption prediction levels by four models, which use the same set of features (occupant behavior, past two-hour outdoor weather, and the working-time indicator) but different algorithms. These results could be attributed to the fact that SVR is good at solving nonlinear problems even with a relatively small amount of training data (Amasyali and El-Gohary 2018), whereas MLR is limited in modeling complex and nonlinear relationships (Fan et al. 2017).



Model #5: SVR with occupant-behavior and past two-hour outdoor weather condition features
Model #11: ANN with occupant-behavior and past two-hour outdoor weather condition features
Model #17: CART with occupant-behavior and past two-hour outdoor weather condition features
Model #23: MLR with occupant-behavior and past two-hour outdoor weather condition features
Model #29: SVR with occupant-behavior and past two-hour outdoor weather condition features
Model #35: ANN with occupant-behavior and past two-hour outdoor weather condition features
Model #41: CART with occupant-behavior and past two-hour outdoor weather condition features
Model #47: MLR with occupant-behavior and past two-hour outdoor weather condition features

Figure 6.8 – Comparison of Actual and Predicted Cooling Energy Consumption Values Predicted by Models with different Algorithms

6.2.3 Thermal and Visual Comfort Prediction

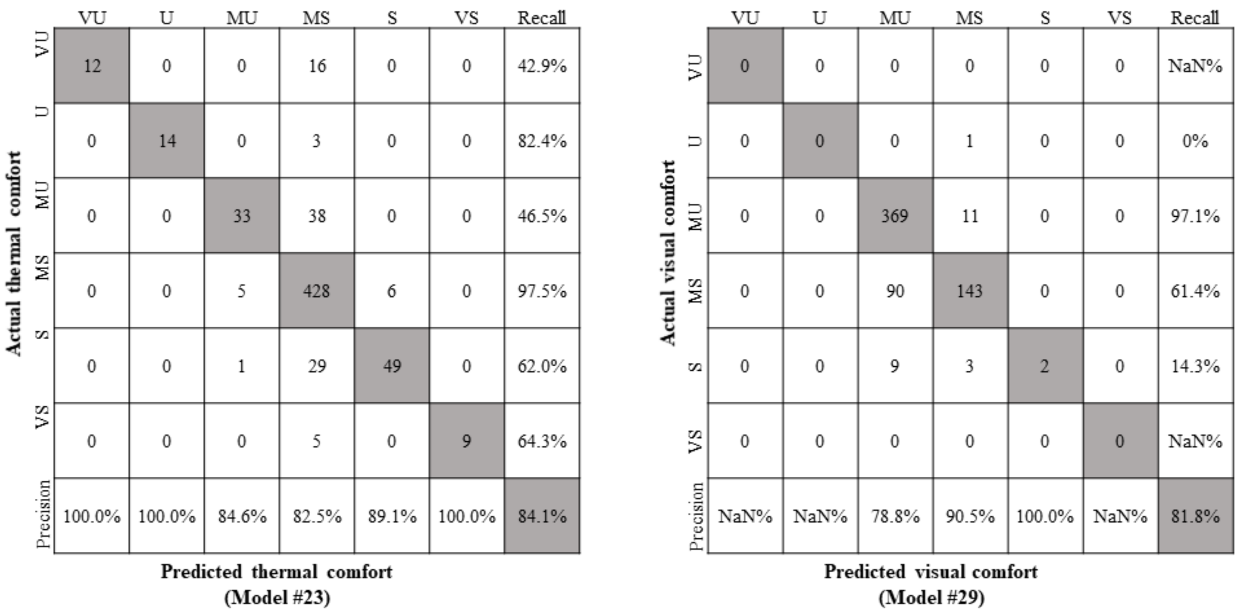
6.2.3.1 Group Comfort Prediction

Table 6.2 summarizes the prediction performance results of the thermal and visual comfort prediction models. For thermal comfort prediction, Model #23 – a KNN model that predicts thermal comfort using occupant-behavior features, past two-hour outdoor weather conditions, and working-time indicator – achieved the best performance, 0.20 MAE. For visual comfort prediction, Model #29 – an SVM model that predicts visual comfort using occupant-behavior features, past two-hour outdoor weather conditions, and working-time indicator – achieved the best performance, 0.40 MAE. Figure 6.9 shows the confusion matrices of thermal and visual comfort prediction by Models #23 and #29. The predicted levels show a good agreement with the actual comfort levels. Models #23 and #29 were, therefore, chosen to be used for the optimization.

Table 6.2 Results of the Thermal and Visual Comfort Prediction for all Models

Model #[1]	Occupant -behavior features	Past hour weather features	Machine learning algorithm[2]	Thermal comfort[3]		Visual comfort[3]	
				MAE	MSE	MAE	MSE
1 (25)	✓	0	SVM	0.27	0.46	0.41	0.45
2 (26)	✗	0	SVM	0.44	0.79	0.46	0.50
3 (27)	✓	1	SVM	0.25	0.43	0.42	0.46
4 (28)	✗	1	SVM	0.42	0.78	0.49	0.52
5 (29)	✓	2	SVM	0.24	0.41	**0.40**	**0.44**
6 (30)	✗	2	SVM	0.41	0.75	0.49	0.52
7 (31)	✓	0	ANN	0.26	0.45	0.42	0.46
8 (32)	✗	0	ANN	0.44	0.78	0.46	0.50
9 (33)	✓	1	ANN	0.25	0.43	0.49	0.54
10 (34)	✗	1	ANN	0.43	0.76	0.50	0.53
11 (35)	✓	2	ANN	0.23	0.41	0.44	0.49
12 (36)	✗	2	ANN	0.41	0.74	0.50	0.53
13 (37)	✓	0	DT	0.23	0.41	0.41	0.45
14 (38)	✗	0	DT	0.41	0.75	0.46	0.50
15 (39)	✓	1	DT	0.23	0.41	0.44	0.48
16 (40)	✗	1	DT	0.41	0.75	0.49	0.53
17 (41)	✓	2	DT	0.22	0.40	0.42	0.46
18 (42)	✗	2	DT	0.40	0.72	0.50	0.53
19 (43)	✓	0	KNN	0.26	0.45	0.41	0.45
20 (44)	✗	0	KNN	0.44	0.78	0.46	0.50
21 (45)	✓	1	KNN	0.25	0.43	0.43	0.46
22 (46)	✗	1	KNN	0.42	0.77	0.49	0.53
23 (47)	✓	2	KNN	**0.20**	**0.35**	0.42	0.45
24 (48)	✗	2	KNN	0.41	0.76	0.49	0.53

[1]Numbers outside of the parentheses indicate the model numbers for the thermal comfort prediction models, and in parentheses indicate those for the visual comfort prediction models.
[2]SVM: Support Vector Machine; ANN: Artificial Neural Networks; DT: Decision Tree; KNN: k-Nearest Neighbors
[3]MAE: Mean Absolute Error; MSE: Mean Squared Error.
Best results are shown in **bold font**.



VU = Very unsatisfied; U = Unsatisfied; MU = Moderately unsatisfied; MS = Moderately satisfied; S = Satisfied; VS = Very satisfied

Figure 6.9 – Confusion Matrices for Thermal Comfort Prediction by Model #23 and Visual Comfort Prediction by Model #29

All comfort prediction models with occupant-behavior features always produced more accurate results than the models without behavior features. For example, Model #23 achieved 0.20 MAE, which is the most accurate; while Model #24, which utilized the same features as Model #23 except for behavior features, achieved 0.41 MAE. The MAE improvement due to the inclusion of behavior features ranged from 38.6% to 45.0%. The t-test p-values for the comparisons between the predicted comfort levels by the models with and without occupant-behavior features were less than 0.05, which indicates that these performance differences are significant. The results indicate that the behavior features are discriminating, and hence that the prediction models can be used to predict the impact of the behavior.

Same as energy consumption prediction, the use of additional past-timestep features led to some marginal improvement for comfort prediction. For example, for thermal comfort prediction, Model #23, which utilized the past two-hour weather data, occupant-behavior data, and the working-time indicator as the features, achieved 0.20 MAE compared to 0.25 MAE for Model #21 (which utilized the past one-hour data), and 0.26 MAE for Model #19 (which utilized no past hour data). Similarly, for visual comfort prediction, Model #29, which utilized the past two-hour data, achieved 0.40 MAE compared to 0.42 MAE for Model #27 (which utilized the past one-hour data) and 0.41 MAE for Model #25 (which utilized no past hour data). The t-test p-values for the comparisons were less than 0.05, which indicates that the differences in the comfort levels predicted by these models are significant.

For the machine learning algorithms, as per Table 6.2, although, the best prediction performance was achieved by Model #23 (a KNN model) for thermal comfort and Model #29 (an SVM model) for visual comfort, there was no clear outperformer across the SVM, ANN, DT, and KNN models.

6.2.3.2 Individual Comfort Prediction

A total of 24 individual models were developed, two for each occupant. The same features and algorithms of the group comfort models were used for individual comfort prediction. For the remainder of this chapter, the results of three example occupants are shown for illustrative purposes. For example, for thermal comfort prediction, KNN models with occupant-behavior features, past two-hour outdoor weather conditions, and working-time indicator achieved 82.2%, 82.6%, and 79.3% accuracies for Occupants #1, #2, and #3, respectively. For visual comfort prediction, SVM models with occupant-behavior features, past two-hour outdoor weather conditions, and working-time indicator achieved 81.3%, 77.4%, and 79.3% accuracies for Occupants #1, #2, and #3, respectively.

6.2.4 Occupant Behavior Optimization

6.2.4.1 Group Comfort Results

The following prediction models were used for the multi-objective occupant behavior optimization: Model #5 for cooling energy consumption prediction, Model #29 for lighting energy consumption prediction, Model #23 for thermal comfort prediction, and Model #29 for visual comfort prediction. The optimal occupant behavior settings were determined, for each hour throughout the three months during which the data were collected. For example, Figure 6.10 shows seven optimal solutions at hour 1301, which included energy consumption levels varying from 9.76 kWh to 16.61 kWh, thermal comfort levels varying from very unsatisfied to very satisfied, and visual comfort levels ranging from moderately satisfied to satisfied. The 9.76 kWh solution is the energy-priority-solution, which provides the minimum energy consumption. The 16.61 kWh solution is the balanced solution, which offers the highest weighed sum of the objectives, resulting in satisfied thermal and comfort levels.

Figure 6.11 shows the cumulative energy consumption of the three extreme and the balanced solutions as well as the actual energy consumption. The energy-priority, thermal-comfort-priority, visual-comfort-priority, and balanced solutions consumed 8,838 kWh, 10,077 kWh, 9,696 kWh, and 9,715 kWh respectively, in comparison to 11,299 kWh actual consumption. Overall, the energy-priority, thermal-comfort-priority, visual-comfort-priority, and balanced solutions achieved 21.8%, 10.8%, 14.2%, and 14.0% energy savings, respectively. All solutions consumed less energy relative to the actual energy consumption levels.

Figure 6.12 shows thermal and visual comfort distributions of the solutions as well as the actual comfort distributions. Overall, before optimization, 17.7% of the time the actual thermal comfort level was moderately unsatisfied or lower. After the optimization, only 3.4%, 0.0%, 1.1%, and

0.1% of the time the thermal comfort level was moderately unsatisfied or lower for the energy-priority, thermal-comfort-priority, visual-comfort-priority, and balanced solutions, respectively. Before optimization, the actual 3-month average thermal comfort level was 3.9 (i.e., moderately unsatisfied to moderately satisfied). After optimization, the average comfort level increased to 5.1 (i.e., satisfied to very satisfied) for the thermal-comfort-priority solution, 4.9 (i.e., moderately satisfied to satisfied) for the visual-comfort-priority and balanced solutions, and 4.4 (i.e., moderately satisfied to satisfied) for the energy-priority solution.

For visual comfort, the time when the comfort was moderately unsatisfied or lower was reduced from 60.7% to 17.5%, 18.9%, 11.3%, and 15.3% for the energy-priority, thermal-comfort-priority, visual-comfort-priority, and balanced solutions, respectively. The 3-month average visual comfort level was increased from 3.4 (i.e., moderately unsatisfied to moderately satisfied) to 4.4 (i.e., moderately satisfied to satisfied) for the visual-comfort-priority solution, 4.3 (i.e., moderately satisfied to satisfied) for the balanced solution, and 4.2 (i.e., moderately satisfied to satisfied) for the thermal-comfort-priority and energy-priority solutions.

The optimal occupant-behavior settings differed for the four optimal solutions. For example, Figure 6.13 shows the settings for the balanced solution: 69.8 F average thermostat setpoint, portable heaters on 16.2% of the time, doors open 77.3% of the time, shading devices open 6.1% of the time, and lights on 74.2% of the time. In general, the average thermostat setpoint (70.0 F) for the energy-priority solution was higher than the setpoints for other solutions (68.9 F – 69.8 F). For all solutions, more than 80% of the time, the thermostat setpoints were in the range of recommended indoor temperature levels (67^{o} – 82^{o}) by the ASHRAE. As expected, higher thermostat setpoints resulted in less cooling energy consumption. Turning on portable heaters was optimal only less than 20% of the time for all solutions. The thermal-comfort-priority solution had the least use of personal heaters due to avoiding over-cooling (Derrible and Reeder 2015). Opening doors was optimal 82.4% of the time for the energy-priority solution; while it was only optimal in the range of 70.8% to 77.3% of the time for other solutions. Rupp and Ghisi (2014) showed that less air-conditioning use hours is needed when the internal doors are open. For the visual-comfort-priority solution, the lights were on 77.9% of the time and the shading devices were only open 4.2% of the time. However, for all other solutions, compared to the visual-comfort-priority solution, the lights were on for shorter durations and the shading devices were open for longer durations. For example, for the energy-priority solution, the shading devices were open double the time of that for the thermal-comfort and visual-comfort priority solutions. This shows that opening shading devices can help save energy, although it cannot always fully substitute turning on lights because natural light may not provide sufficient visual comfort. On one hand, opening shading device lets more natural light in and reduces lighting energy consumption. On the other hand, when there is less artificial lighting, the lighting devices produce less heating and therefore the demand for cooling is also reduced. Lighting is, thus, not only a significant piece of building energy consumption by itself, but it also impacts cooling energy demand. One-third of the cooling energy consumption can be saved if a good balance between natural light and solar heat can be achieved (Wong et al. 2010).

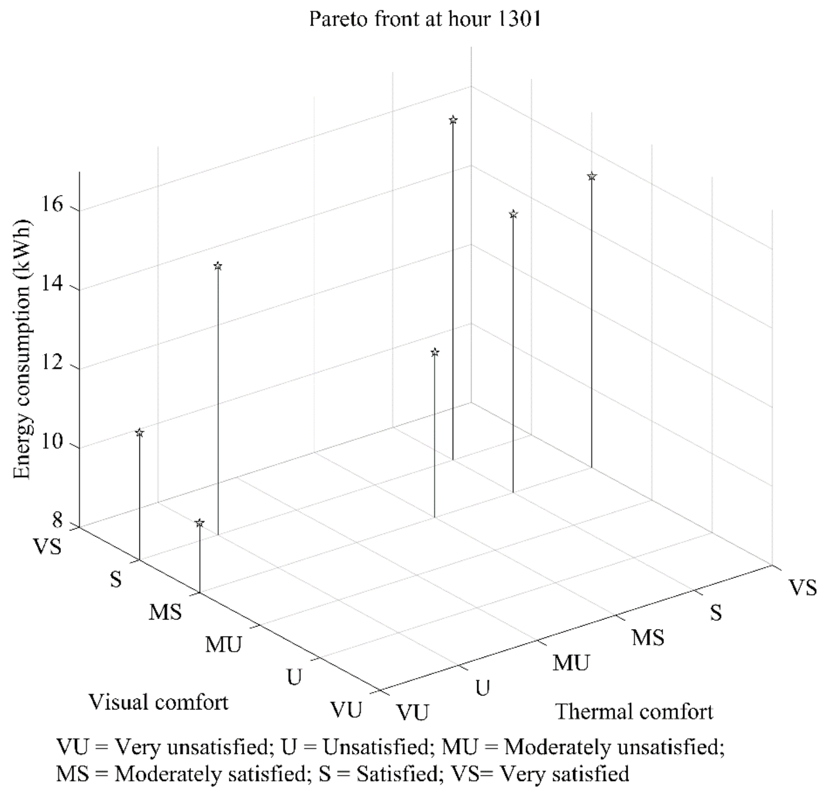


Figure 6.10 – Pareto Front Solutions at Hour 1301

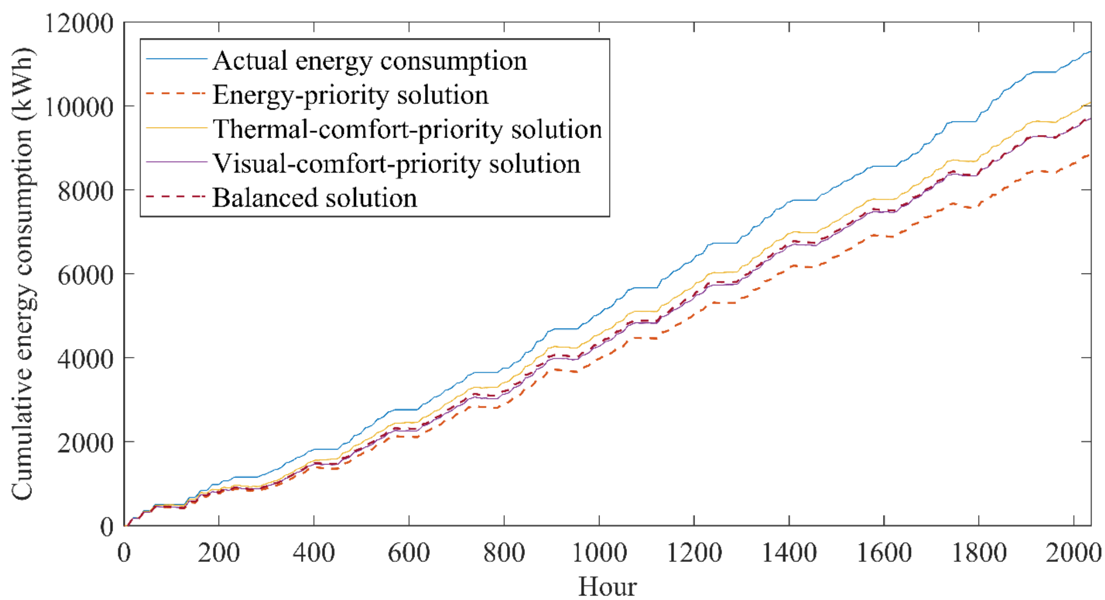
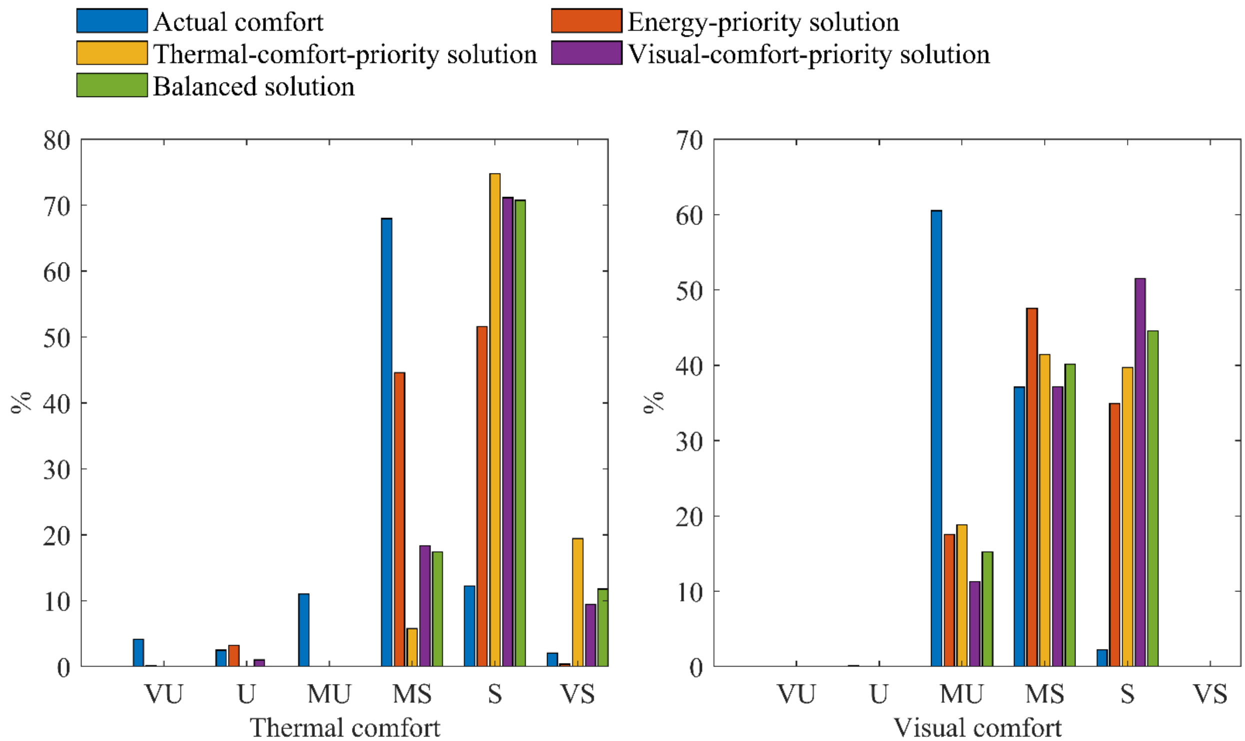


Figure 6.11 – Energy Consumption of Optimal Solutions



VU = Very unsatisfied; U = Unsatisfied; MU = Moderately unsatisfied; MS = Moderately satisfied; S = Satisfied; VS= Very satisfied

Figure 6.12 – Thermal and Visual Comfort Distributions

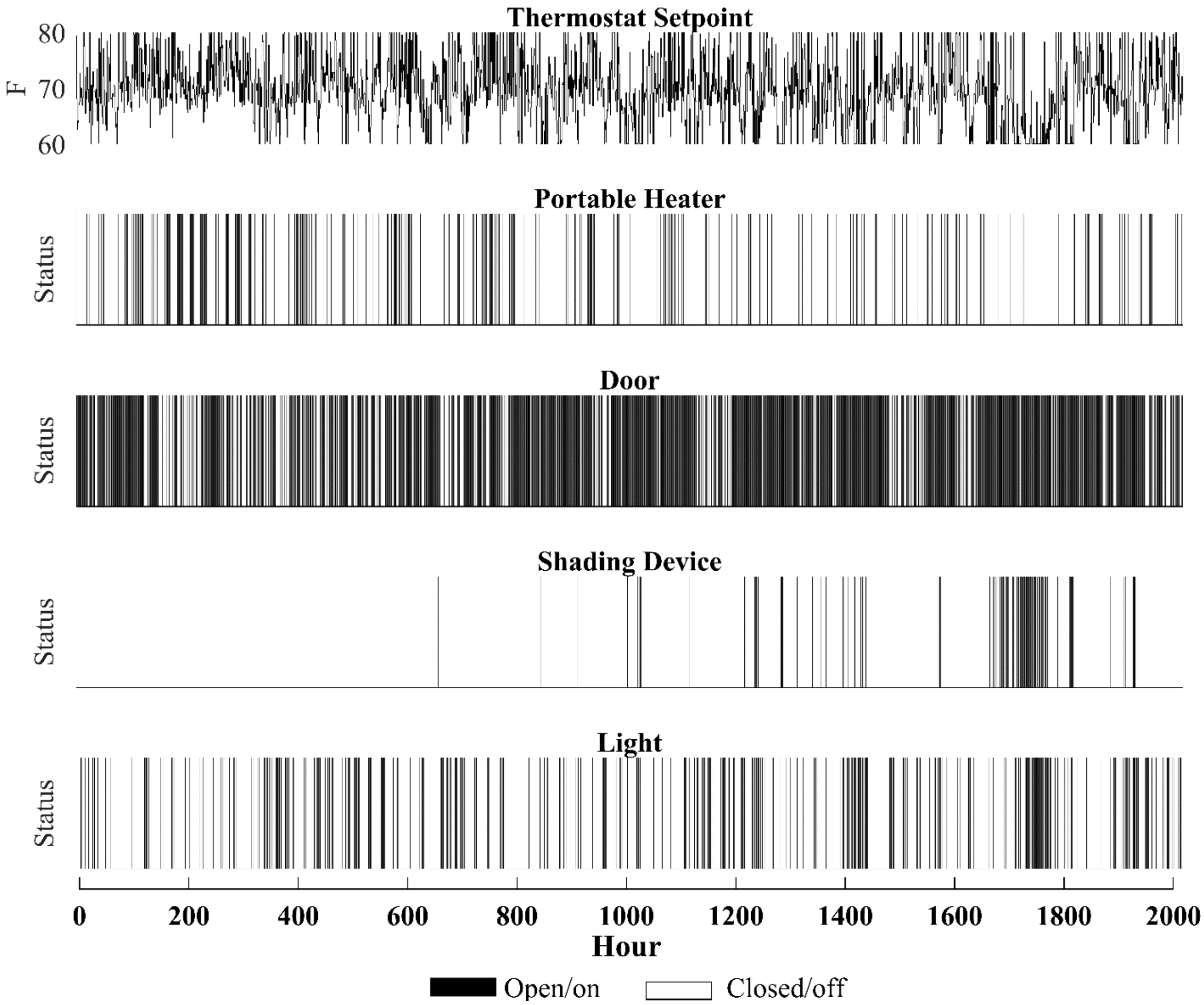


Figure 6.13 – Optimal Occupant Behavior Variables for the Balanced Solution

6.2.4.2 Individual Comfort Results

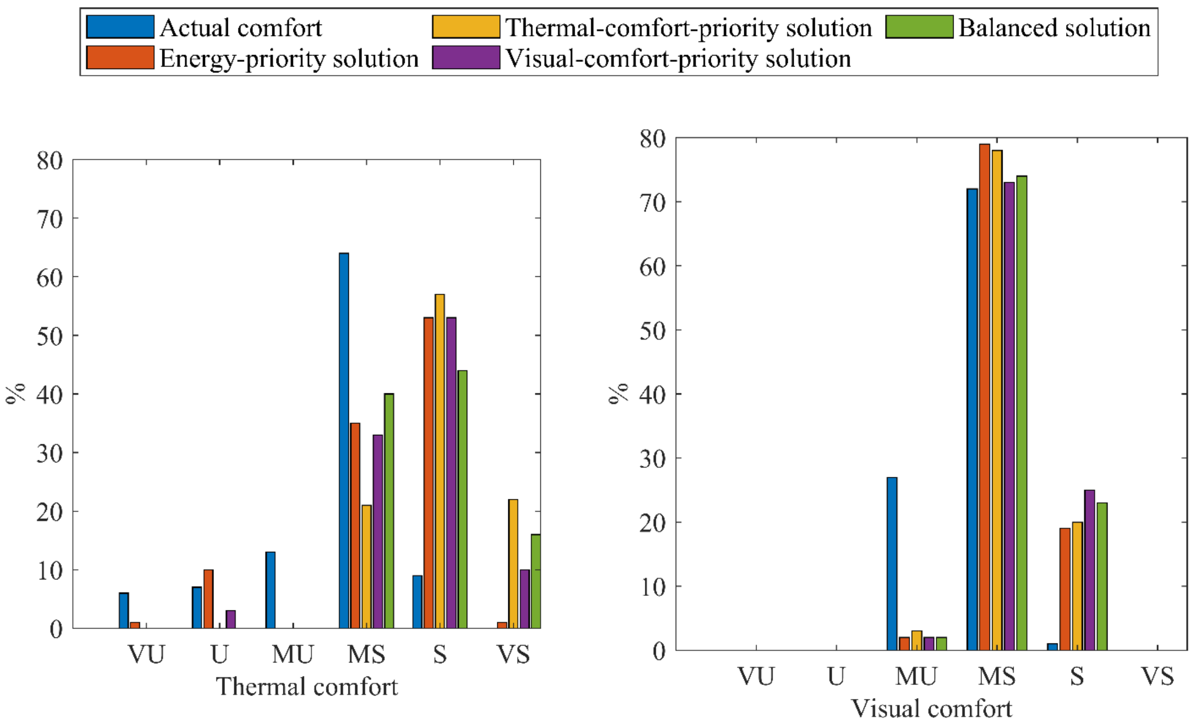
Figure 6.14 to Figure 6.16 show the predicted thermal and visual comfort levels of the three occupants for the optimal solutions. Table 6.3 summarizes the actual (reported) comfort levels. These occupants reported different thermal and visual comfort levels when they were exposed to similar indoor environmental conditions, which indicates that these three occupants have different thermal and visual preferences. Overall, Occupant #2 was the most comfortable, while Occupant #3 was the least comfortable. For example, for thermal comfort, Occupant #2 was satisfied 60.9% of the time, moderately satisfied 19.6% of the time, and moderately unsatisfied or lower 19.5% of the time. Occupant #1 was satisfied 9.1% of the time, moderately satisfied 64.9% of the time, and moderately unsatisfied or lower 26.4% of the time. And, Occupant #3 was satisfied 14.0% of the time, moderately satisfied 41.3% of the time, and moderately unsatisfied or lower 44.6% of the time.

For all optimal solutions, the optimal thermal and visual comfort levels of the three occupants were higher than their actual comfort levels. For example, before the optimization, the actual 3-month average thermal comfort levels of Occupants #1, #2, and #3 were 3.6 (i.e., moderately unsatisfied to moderately satisfied), 4.2 (i.e., moderately satisfied to satisfied), and 3.5 (i.e., moderately unsatisfied to moderately satisfied), respectively. After the optimization, the average comfort levels of these three occupants increased to 5.0 (i.e., satisfied), 5.3 (i.e., satisfied to very satisfied), and 5.0 (i.e., satisfied), respectively for the thermal-comfort-priority solution. For the energy-priority, visual-comfort-priority, and balanced solutions, the average comfort levels of the three occupants increased to the ranges of 4.3 – 4.7 (i.e., moderately satisfied to satisfied), 4.6 – 5.0 (i.e., moderately satisfied to satisfied), and 4.1 – 4.6 (i.e., moderately satisfied to satisfied), respectively.

Similarly, before optimization, the actual 3-month average visual comfort levels of the three occupants were 3.7 (i.e., moderately unsatisfied to moderately satisfied), 4.1 (i.e., moderately satisfied to satisfied), and 2.5 (i.e., unsatisfied to moderately unsatisfied), respectively. These levels increased to 4.2 (i.e., moderately satisfied to satisfied), 4.8 (i.e., moderately satisfied to satisfied), and 3.1 (i.e., moderately unsatisfied to moderately satisfied), respectively, for the visual-comfort-priority solution. For other solutions, these levels increased to the ranges of 4.1 – 4.2 (i.e., moderately satisfied to satisfied), 4.7 – 4.8 (i.e., moderately satisfied to satisfied), and 3.0 – 3.1 (i.e., moderately unsatisfied to moderately satisfied), respectively.

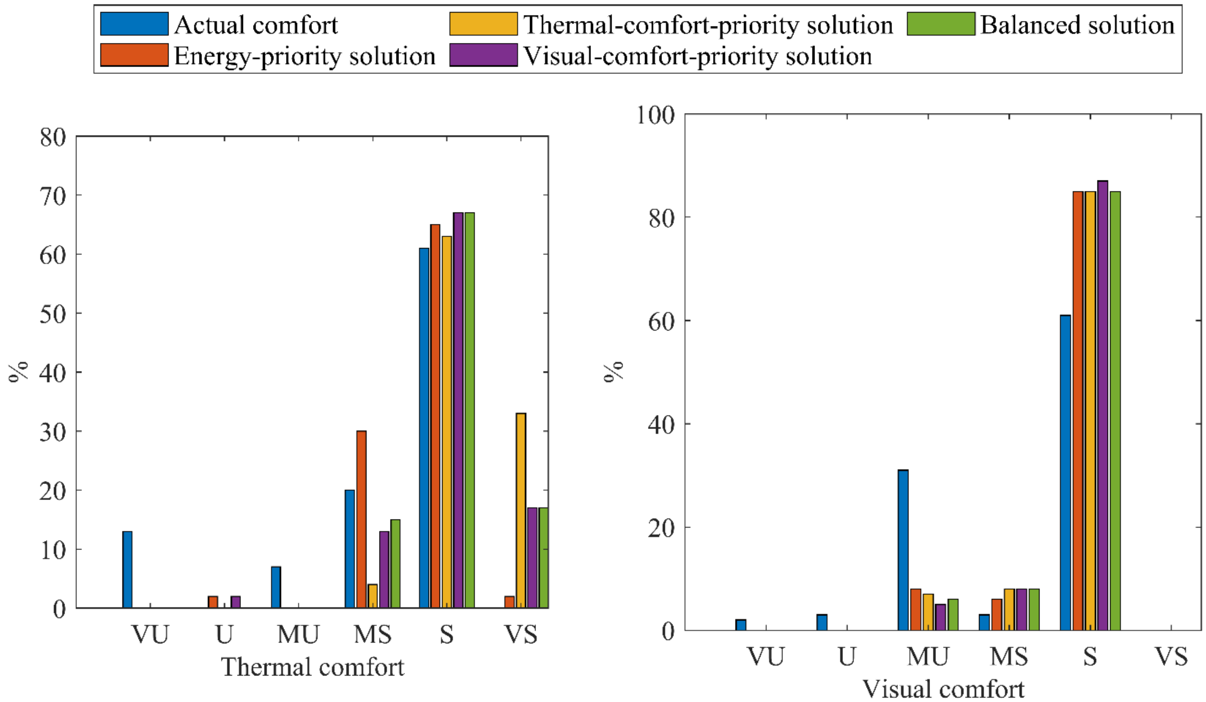
Table 6.3 Reported Thermal and Visual Comfort Levels (Percentage of Time)

Comfort level	Thermal comfort			Visual comfort		
	Occupant #					
	1	2	3	1	2	3
Very unsatisfied	5.9%	13.0%	3.3%	0.0%	1.6%	3.3%
Unsatisfied	7.3%	0.0%	7.3%	0.0%	3.2%	47.3%
Moderately unsatisfied	13.2%	6.5%	34.0%	26.8%	30.7%	47.3%
Moderately satisfied	64.9%	19.6%	41.3%	72.3%	3.2%	0.7%
Satisfied	9.1%	60.9%	14.0%	1.0%	61.3%	0.7%
Very satisfied	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%



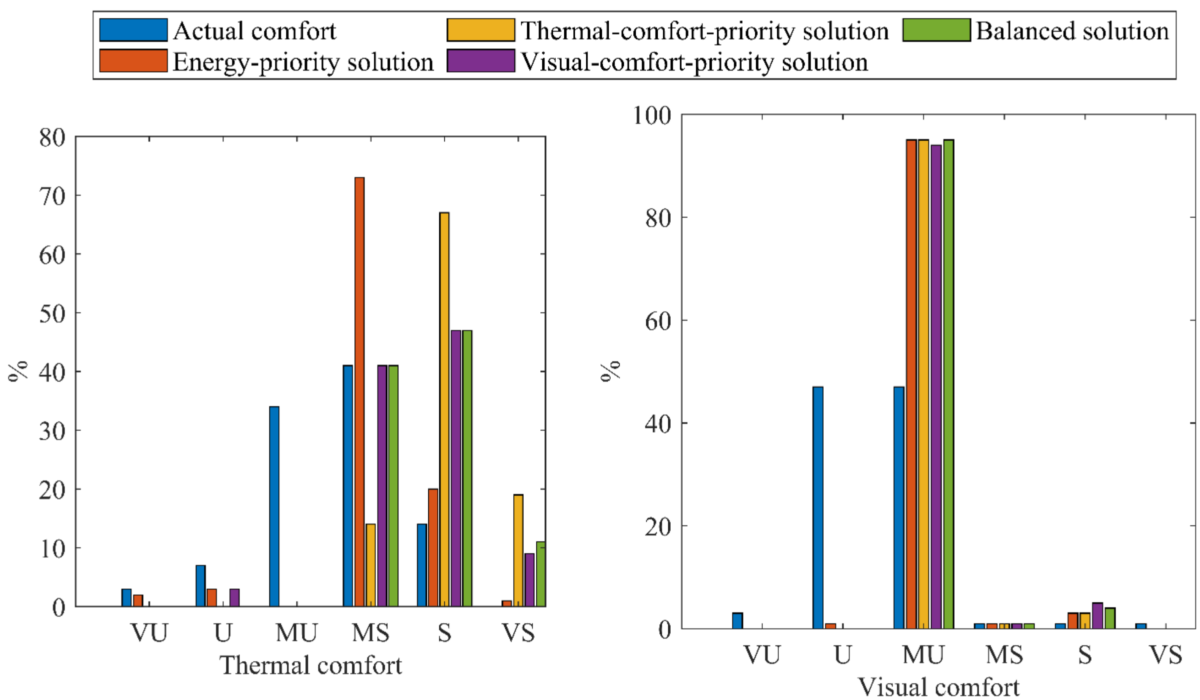
VU = Very unsatisfied; U = Unsatisfied; MU = Moderately unsatisfied; MS = Moderately satisfied; S = Satisfied; VS= Very satisfied

Figure 6.14 – Thermal and Visual Comfort Levels of Occupant #1



VU = Very unsatisfied; U = Unsatisfied; MU = Moderately unsatisfied; MS = Moderately satisfied; S = Satisfied; VS= Very satisfied

Figure 6.15 – Thermal and Visual Comfort Levels of Occupant #2



VU = Very unsatisfied; U = Unsatisfied; MU = Moderately unsatisfied; MS = Moderately satisfied; S = Satisfied; VS= Very satisfied

Figure 6.16 – Thermal and Visual Comfort Levels of Occupant #3

CHAPTER 7 – HYBRID MACHINE LEARNING-BASED ENERGY CONSUMPTION PREDICTION: COUPLING DATA-DRIVEN AND PHYSICAL APPROACHES

7.1 Hybrid Machine Learning-Based Prediction Approach

A hybrid machine learning model, which learns both from simulation-generated data and real data, was developed. The hybrid model is composed of three constituent models: (1) a machine learning model that predicts the hourly values of a weather factor: the weather factor represents the impact of outdoor weather conditions on cooling energy consumption, at a specific hour. The model was trained on simulation-generated data, because a simulation environment enables the generation of datasets in which energy consumption differences are due to outdoor weather conditions only; (2) a machine learning model that predicts the hourly values of an occupant-behavior factor: the occupant-behavior factor represents the impact of occupant behavior on cooling energy consumption, at a specific hour. The model was trained on real data, because the stochastic and complex nature of occupant behavior can be better captured in a real-world setting; and (3) an ensembler model that predicts the hourly cooling energy consumption values based on the predicted values of the two factors.

7.1.1 Weather-Factor Prediction Model Development

The development of the weather-factor prediction model included three primary steps: energy simulations, time-series clustering, and factor prediction model development.

7.1.1.1 Energy Simulations

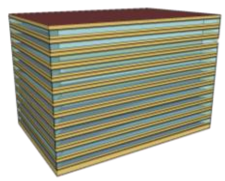
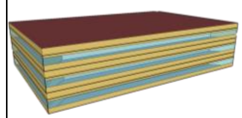
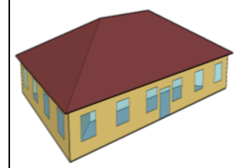
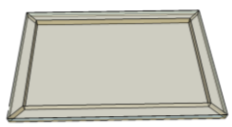
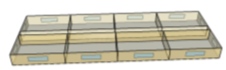
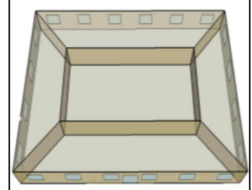
The simulation-generated data were created by simulating the reference models of the small, medium, and large office and midrise apartment buildings, provided by the U.S. Department of Energy (DOE) (Deru et al. 2011), in EnergyPlus. These models were selected because the reference models are modeled in a very detailed and precise way for benchmarking purposes and therefore, are expected to result in more accurate results than some other simulation efforts (e.g., Li et al. 2015a) which showed that simulations are not very accurate. The models were simulated in 16 cities, which represent all U.S. climate zones, resulting in a total of 64 cases. The small office has 5 conditioned zones, including core and four perimeter zones. The medium office has 15 conditioned zones, including core and four perimeter zones on each floor. The large office has 16 conditioned zones, including core zone with four perimeter zones on each floor and a basement. The midrise apartment has 24 conditioned zones, including apartments and an office. Table 7.1 shows the selected cities and their climate properties. Table 7.2 shows the properties of the simulated buildings. The simulations were conducted using the TMY3 weather data of the selected cities. The simulations were conducted with hourly time steps on a four-core personal computer, in parallel on all 4 cores, using EnergyPlus 9.0.1.

Table 7.1 Model Locations and Climate Properties

Location #	**City**	**Climate zone**	**Climate type**	**CDD**
1	Miami, Florida	1A	Very Hot – Humid	2521
2	Houston, Texas	2A	Hot – Humid	1699
3	Phoenix, Arizona	2B	Hot – Dry	2570
4	Atlanta, Georgia	3A	Warm – Humid	1052
5	Los Angeles, California	3B-Coast	Warm – Dry	323
6	Las Vegas, Nevada	3B	Warm – Dry	1937
7	San Francisco, California	3C	Warm – Marine	80
8	Baltimore, Maryland	4A	Mixed – Humid	701
9	Albuquerque, New Mexico	4B	Mixed – Dry	761
10	Seattle, Washington	4C	Mixed – Marine	104
11	Chicago, Illinois	5A	Cool – Humid	480
12	Boulder, Colorado	5B	Cool – Dry	383
13	Minneapolis, Minnesota	6A	Cold – Humid	425
14	Helena, Montana	6B	Cold – Dry	219
15	Duluth, Minnesota	7	Very Cold	117
16	Fairbanks, Alaska	8	Subarctic	40

CDD = cooling degree days

Table 7.2 Properties of the Simulated Buildings

	Large Office	**Medium Office**	**Midrise Apartment**	**Small Office**
Elevation				
Layout				
Building #	1	2	3	4
Total floor area (m^2)	46,320	4,982	3,135	511
Number of floors	12 plus basement	3	3	1
WWR	38%	33%	15%	21.20%
HVAC system type	MZ-VAV	MZ-VAV	N/A	PSZ-AC
Cooling	2 water cooled chillers	PACU	Split system DX	Unitary DX

WWR = Window-to-wall ratio; MZ-VAV = Multi Zone Variable Air Volume; PSZ-AC = Packaged Single Zone Air Conditioner; PACU = Packaged Air Conditioning Unit; DX = Direct Expansion.

7.1.1.2 Time-Series Clustering

Time-series clustering included four main steps: data organization, data normalization, time-series clustering, and cluster validation. First, for each of the 64 simulation cases, the 3-month period that has the highest CDD (and therefore the maximum cooling demand) was identified, and its corresponding weather condition and energy consumption data were extracted. Second, a magnitude-normalized energy consumption was calculated to eliminate the impact of the magnitude of the energy consumption. Z-normalization was used to calculate the magnitude-normalized energy consumption because it enables the evaluation of the shape of the energy consumption profile rather than its magnitude (Park et al. 2019). The magnitude-normalized energy consumption was used as the weather factor because it solely represents the impact of outdoor weather conditions on energy consumption. Third, the 3-month factor profiles were clustered – based on their patterns – using partitional time-series clustering. In clustering, DTW was used as the distance measure. Fourth, the clustering results were validated using the Silhouette index (Rousseeuw 1987). Although there is no universal procedure for this, in this study, the elbow method was used to determine the optimal number of clusters. Then, the prototype of each cluster was extracted accordingly.

7.1.1.3 Factor Prediction Model Development

For each cluster, a machine learning model that predicts the hourly weather factor based on outdoor weather conditions was developed. The following weather condition features were used: temperature, dewpoint temperature, relative humidity, wind speed, and solar radiation. The features were extracted from the TMY3 weather data of the corresponding cities. In developing the models, the following five machine-learning algorithms were tested: CART, Gaussian process regression (GPR), SVR, and ANN. The parameters of these algorithms were tuned through parameter grid search using cross validation to maximize the prediction performance.

7.1.2 Occupant-Behavior Factor Prediction Model Development

The development of the occupant-behavior factor prediction model included three primary steps: data preprocessing, weather normalization, and factor prediction model development.

7.1.2.1 Data Preprocessing

The real data collected from the PBTC building were preprocessed prior to developing machine learning model. This included five main steps: data aggregation, data integration, data cleaning and outlier filtering, data normalization, and data splitting. First, 15-min intervals of cooling energy consumption were aggregated into hourly consumption values. Occupant-behavior data from multiple occupants were aggregated using a majority voting strategy. The behavior of an occupant was assumed unchanged until another behavior is reported or until the end of the day.

Second, cooling energy consumption data, outdoor weather conditions data, and occupant behavior data were integrated using their date and time. Third, the missing values of the integrated data were replaced by the mean of the non-missing values. The outlier values, identified by Cook's distance, were removed. Non-summer months, in which cooling demand is very low, were also removed from the dataset. Fourth, each variable in the dataset was normalized between 0 and 1 to avoid overflowing of an individual variable. Finally, the resulting data were randomly split into two datasets (dataset #1 and #2), at a ratio of 8:2, respectively.

7.1.2.2 Weather Normalization

For the energy consumption, weather normalization was performed to remove the effect of weather conditions and better reveal the impact of occupant behavior. The daily weather normalization method proposed by Hydro One (Hydro One 2006) was adapted so that it can be used for hourly normalization. Accordingly, the weather normalization included three steps. First, a regression model, which calculates energy consumption using temperature, dewpoint temperature, relative humidity, wind speed, and solar radiation features was developed. Second, using this model, the expected energy consumption values for the actual outdoor weather conditions were calculated. Third, the R^2 of the regression model, which indicates the percentage of variance in the cooling energy consumption that can be explained by the aforementioned features, was calculated. The percentage of variance that cannot be explained by the aforementioned features was attributed to the other factors such as occupant behavior. Fourth, the actual energy consumption values were compared to the expected values, and the variance between the two consumptions was defined as the occupant-behavior factor. In this study, it was assumed that the variance that cannot be explained by the five weather features was attributed to only/mostly occupant behavior.

7.1.2.3 Factor Prediction Model Development

A machine learning model that predicts the hourly occupant-behavior factor based on occupant behavior was developed. The following occupant-behavior features were used: thermostat setpoint, portable heater status, door status, and window shade status. In developing the model, the aforementioned five machine-learning algorithms (see Section 7.1.1.3) were tested, and their parameters were tuned through parameter grid search.

7.1.3 Ensembler Model Development

Finally, an ensembler model was developed to predict hourly cooling energy consumption. The ensembler model takes the weather and occupant-behavior factors as features. The weather factor is predicted by the corresponding cluster's weather factor prediction model. The occupant-behavior factor is predicted by the occupant-behavior factor prediction model. Prior to the machine learning process, both features were normalized between 0 and 1 to avoid overflowing of a factor. In developing the ensembler model, the same algorithms (see Section 7.1.1.3) were tested using 10-fold cross validation, and their parameters were tuned through parameter grid search.

7.1.4 Performance Evaluation

The performance of the three constituent models, and hence the whole hybrid model, was evaluated as follows. The performance of the weather-factor prediction model was evaluated on the simulation-generated data using 10-fold cross validation. The performance of the occupant-behavior-factor prediction model was evaluated on the real dataset #1 using 10-fold cross validation. The performance of the ensembler model, and therefore the whole hybrid model, was evaluated using the real dataset #2.

The following metrics were utilized for performance evaluation, which were calculated using Eq. (1.1) to (1.3): CV, RMSE, and R^2. According to the ASHRAE Guideline 14, an hourly prediction model is considered as calibrated if hourly CV values fall below 30%.

7.2 Results and Discussion

7.2.1 Simulation, Magnitude Normalization, and Time-Series Clustering Results

The EnergyPlus simulations generated a cooling energy consumption dataset. For example, Figure 7.1 shows the energy consumption values for the first 100 hours of the 64 cases. The large office building in Houston consumed 608,058.7 kWh energy while the small office building in Fairbanks consumed 601.8 kWh only – a ratio of 1010.4 between the most consumer and the least consumer. This variation in energy consumption is caused by, both, the building types and weather conditions. When only the same building models in different locations were compared, the ratio between the most consumer and the least consumer case ranged from 7.6 (medium office building) to 17.4 (small office building), showing a much smaller variance when the difference is caused by weather conditions only. For example, Figure 7.1 illustrates the variance in consumption patterns due to both building types and weather conditions, while Figure 7.2, which shows the magnitude normalized values, reveals the variance due to weather conditions only. In Figure 7.2, the higher values indicate the weather conditions which have higher energy demand, and vice versa.

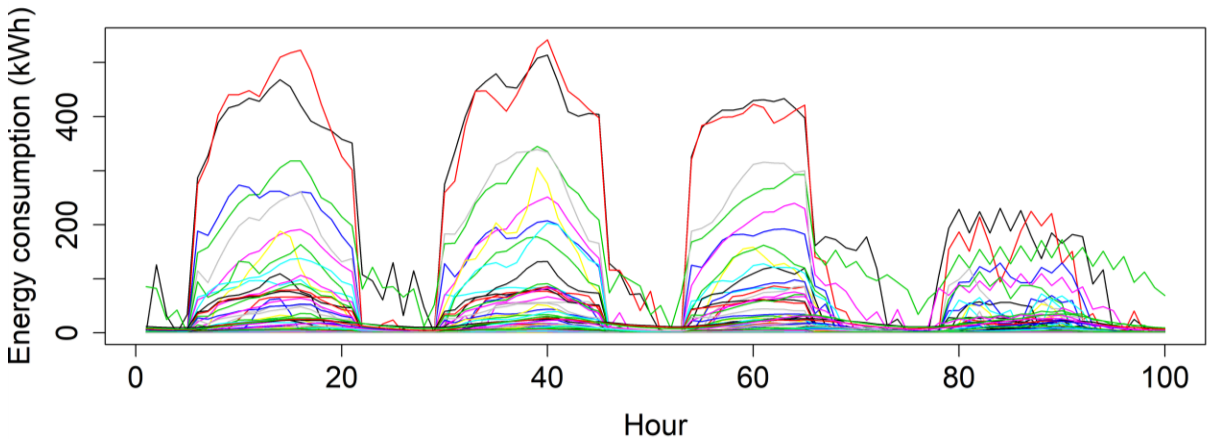


Figure 7.1 – Energy Consumption Values for the first 100 Hours

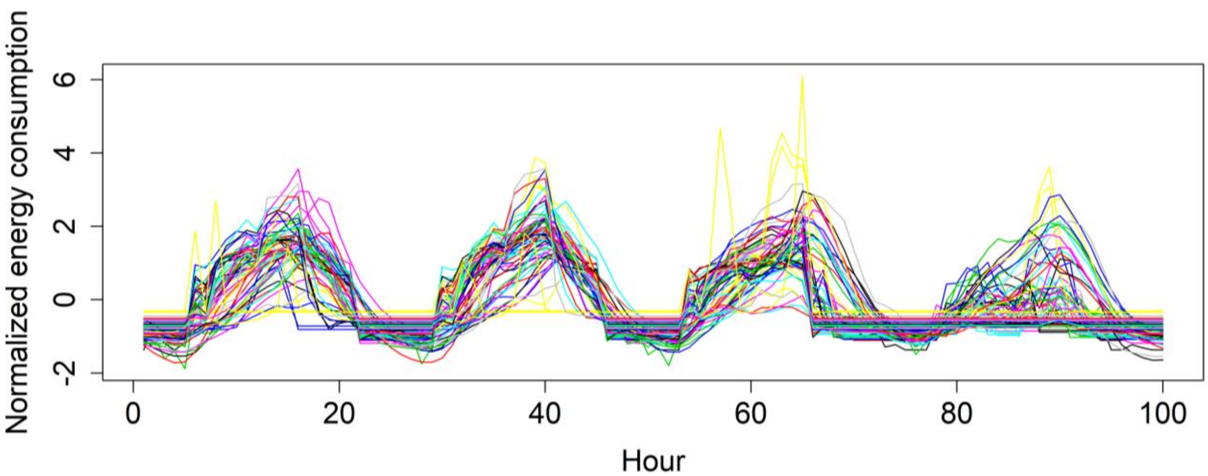


Figure 7.2 – Normalized Energy Consumption Values for the first 100 Hours

Figure 7.3 shows the Silhouette plot. The curve shows an "elbow" break around 11 clusters and therefore the 11 clusters extracted by the partitional algorithm were determined as the optimal clustering. Table 7.3 shows the resulting cluster sizes, i.e., which cases belong to which cluster. There is no cluster with a single member, which shows that the outliers were not clustered (Tureczek et al. 2019). In this study, clustering the outliers is undesirable, because the aim of the clustering was to group similar consumption patterns, rather than the outliers. Figure 7.4 shows the centroids (i.e., the cluster prototypes) of the resulting 11 clusters. For example, all the members of Cluster #5 are midrise apartments. A significant portion (90%) of Cluster #11 are from cold climates. All the members of Cluster #10 are from San Francisco, California. Cluster #9 is consisted of office buildings from Miami, Florida. Overall these results show that all locations and building types have some distinct consumption patterns and the resulting clusters and their centroids represents these patterns.

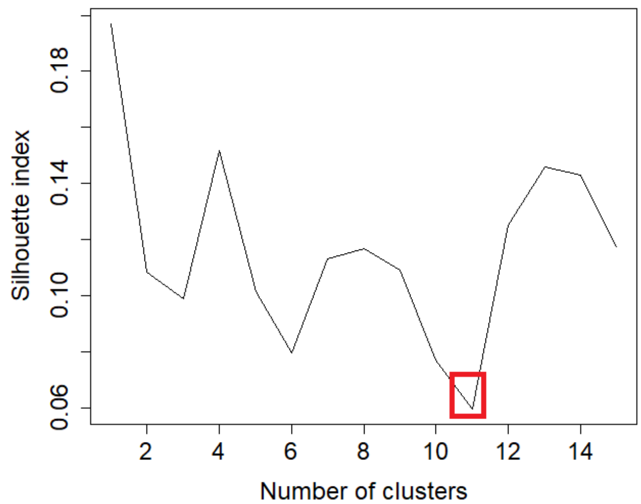


Figure 7.3 – Silhouette Plot

Table 7.3 Resulting Cluster Sizes

Cluster no	Size of cluster	Members of the clusters[1]
1	10	1.2, 1.4, 1.6, 2.2, 2.3, 2.4, 2.6, 4.2, 4.3, 4.6
2	5	3.10, 4.7, 4.10, 4.15, 4.16
3	6	2.12, 4.4, 4.5, 4.8, 4.9, 4.12
4	3	1.13, 2.13, 4.13
5	12	3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.8, 3.9, 3.11, 3.12, 3.13, 3.16
6	2	1.10, 2.10,
7	4	1.5, 1.11, 2.11, 4.11
8	6	1.8, 1.9, 1.12, 2.5, 2.8, 2.9
9	3	1.1, 2.1, 4.1
10	3	1.7, 2.7, 3.7
11	10	1.3, 1.14, 1.15, 1.16, 2.14, 2.15, 2.16, 3.14, 3.15, 4.14

[1]x.y: where x indicates the building #, y indicates the location #

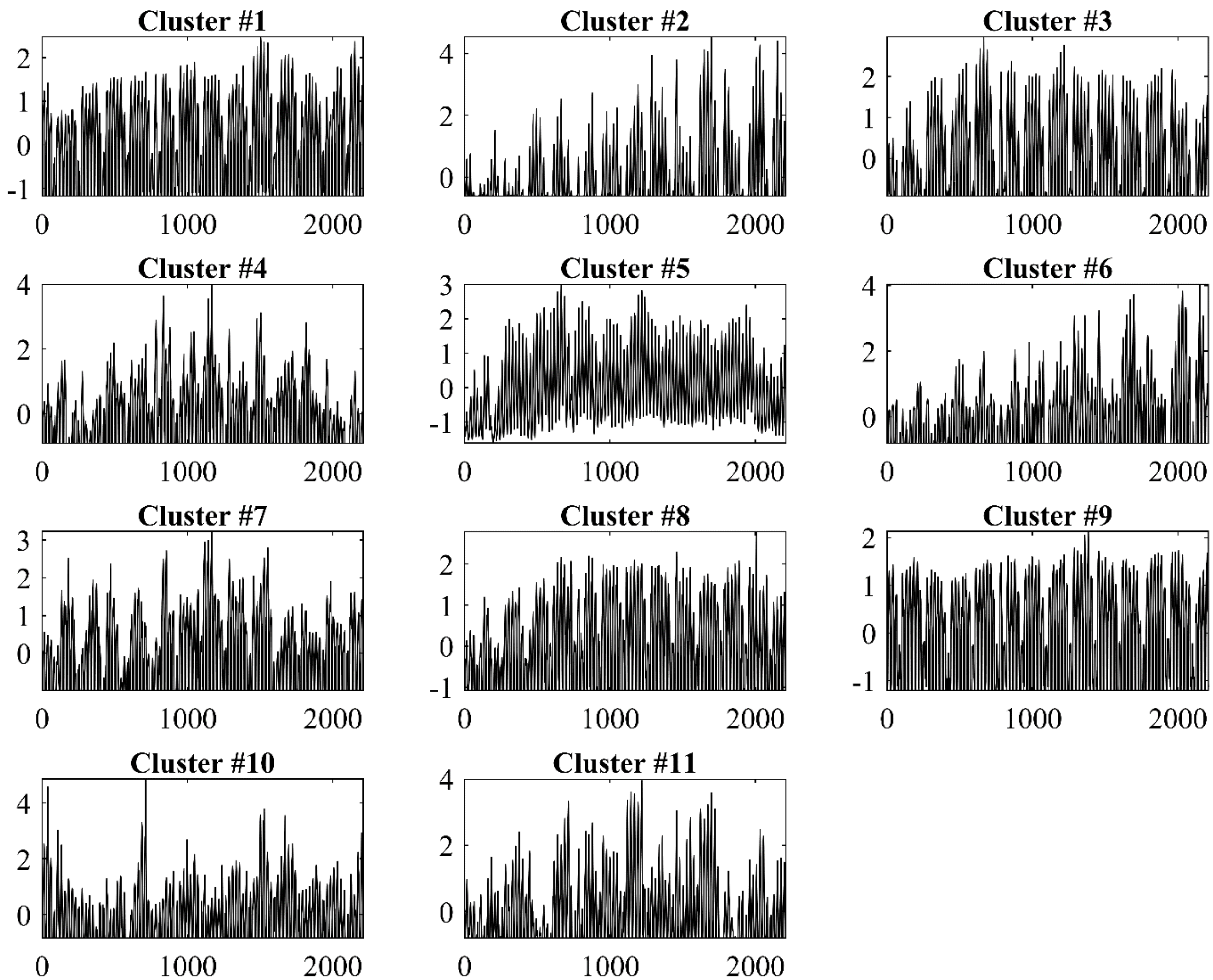


Figure 7.4 – Centroids of the Extracted Clusters

7.2.2 Weather-Factor Prediction Performance

Table 7.4 summarizes the performance results of the four machine-learning algorithms in predicting the weather-factor values for each cluster. The weather-factor prediction models achieved RMSEs and CVs in the ranges of 0.17 - 0.39 and 9.48% - 28.99%, respectively. Although there was no clear outperformer across the algorithms, the GPR models achieved the lowest CV values except for Cluster #1 and Cluster #5. On average the GPR models achieved 18.77% CV, which is lower than the other algorithms' average. The predicted weather-factor values also show a good agreement with the actual values. For example,

Figure 7.5 shows the regression between the predicted and the actual values for Cluster #6, using the GPR model. The GPR-based models were, therefore, chosen for predicting the weather-factor values for all clusters.

The GPR model with temperature, dewpoint temperature, relative humidity, wind speed, and solar radiation features achieved 99.3% R^2, which indicates that the model with the aforementioned features explains 99.3% of the variance in weather-factor values. The remaining variance is very small and can be caused by several factors such as other weather condition features (e.g., atmospheric pressure). In general, such high values of R^2 show that the developed weather-factor prediction models are able to explain the largest proportion of the variance in weather-factor values and therefore the models can be used for understanding the impact of outdoor weather conditions on energy consumption.

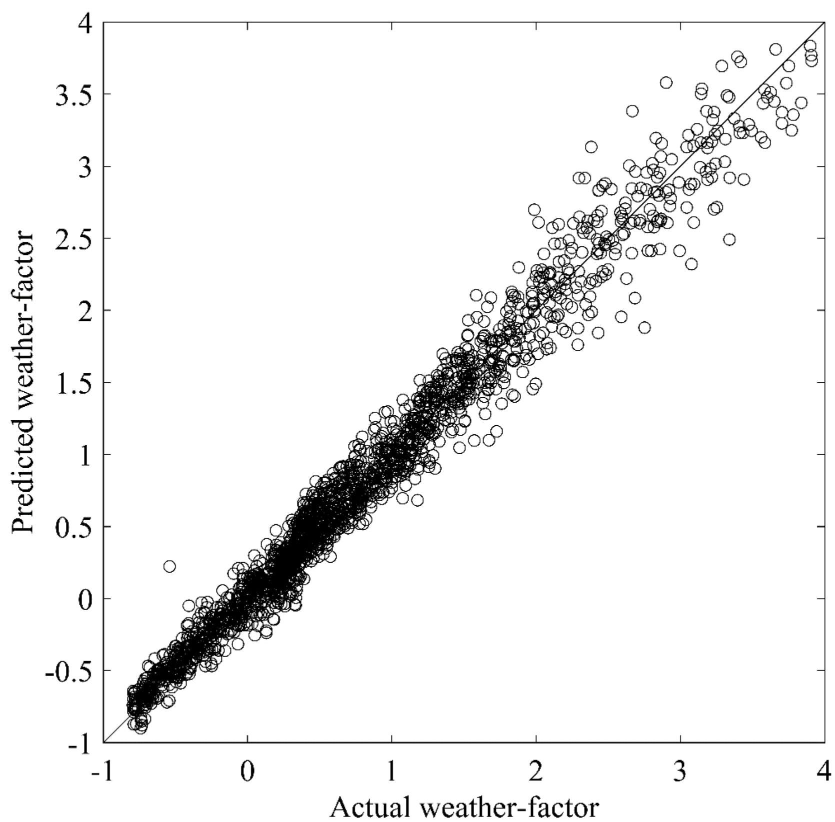


Figure 7.5 – Regression between Actual and Predicted Weather-Factor Values

Table 7.4 Performance of Weather-Factor Prediction

Cluster no	Machine learning algorithm							
	CART		GPR		SVR		ANN	
	RMSE	CV	RMSE	CV	RMSE	CV	RMSE	CV
1	0.26	17.13%	0.28	18.12%	0.30	19.91%	0.30	19.89%
2	0.35	26.85%	0.33	25.35%	0.34	25.77%	0.35	25.92%
3	0.26	17.42%	0.24	16.63%	0.27	18.51%	0.28	18.74%
4	0.28	18.88%	0.27	18.53%	0.30	19.99%	0.31	20.19%
5	0.37	27.11%	0.37	27.44%	0.38	28.85%	0.39	28.99%
6	0.19	10.37%	0.17	9.48%	0.17	9.62%	0.18	10.19%
7	0.27	17.75%	0.26	17.28%	0.28	18.83%	0.28	18.34%
8	0.27	18.02%	0.27	17.62%	0.29	19.12%	0.30	19.71%
9	0.25	16.56%	0.25	16.27%	0.28	19.04%	0.27	18.96%
10	0.32	20.30%	0.3	19.10%	0.31	19.76%	0.32	20.05%
11	0.34	21.37%	0.32	20.66%	0.35	22.61%	0.35	22.88%

CART = Classification and Regression Tree; GPR = Gaussian Process Regression; SVR = Support Vector Regression; ANN = Artificial Neural Networks; RMSE = Root Mean Square Error; CV = Coefficient of Variation.

7.2.3 Real Data Description and Weather Normalization

Figure 7.6 shows the raw and normalized energy consumption values of the PBTC building. The building consumed 11,978 kWh energy for cooling during the three months. There were significant differences in the hourly cooling energy consumption throughout the three months due to the changes in outdoor weather conditions and occupant behavior, and possibly other factors such as number and duration of occupancy. As discussed in Section 7.1.2.2, weather normalization was performed to remove the effect of weather conditions and better reveal the impact of occupant behavior. Similar to the weather-factor prediction, a GPR regression model was fitted and used to identify the variance between the actual and expected cooling energy consumption values, i.e., the occupant-behavior factor. Figure 7.7 shows the regression between the actual and expected cooling energy consumption. Figure 7.8 shows the expected energy consumption values. And, Figure 7.9 shows the hourly differences between the expected and actual normalized energy consumption (i.e., the occupant-behavior factors). The results also show that the GPR model with the same five outdoor weather features achieved 85.0% R^2, which indicates that 15% of the variance cannot be explained by this model. The unexplained portion of the variance could be attributed to non-weather parameters such as occupant behavior because the simulation-based model with outdoor weather features only (see Section 7.2.2), which was trained to predict consumption when there is no variation in occupant behavior, was able to explain 99.3% of the variance in cooling energy consumption.

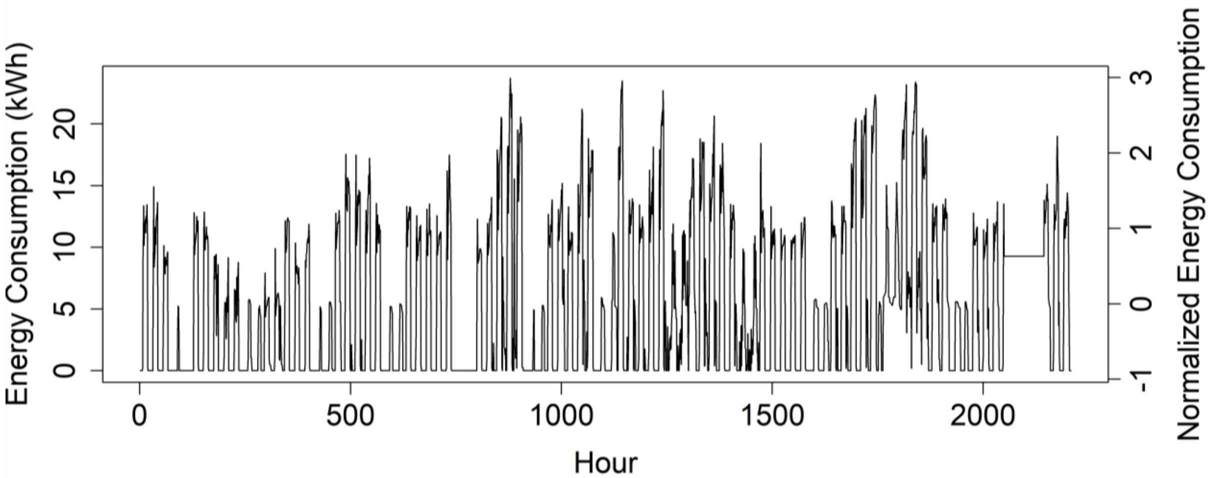


Figure 7.6 – Raw and Normalized Actual Energy Consumption of the Selected Building

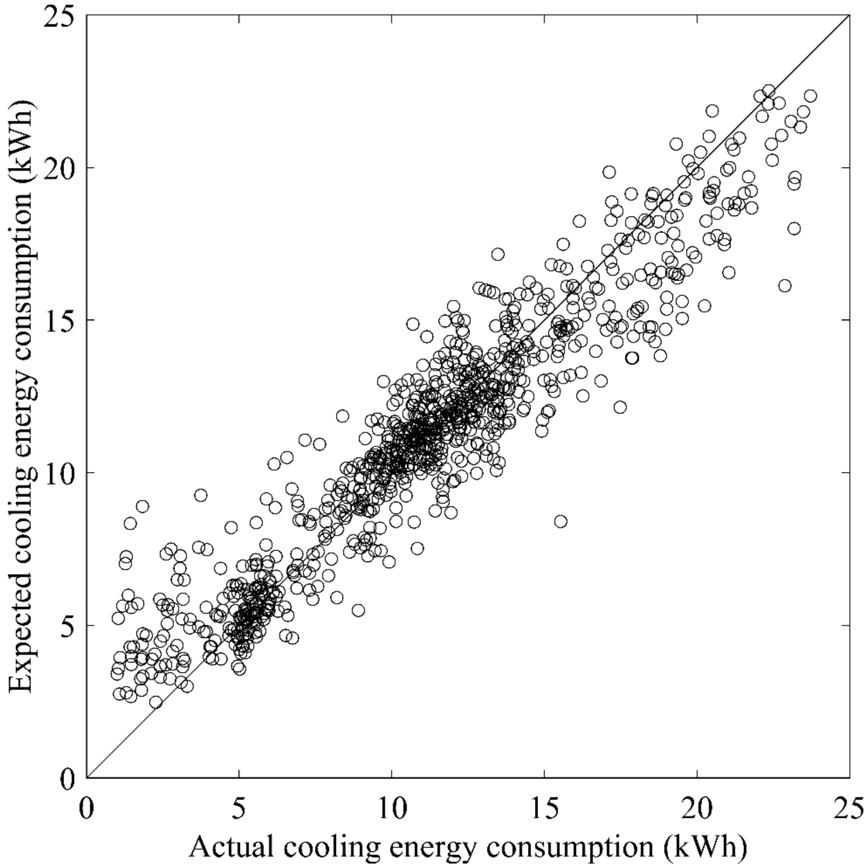


Figure 7.7 – Regression between Actual and Expected Cooling Energy Consumption

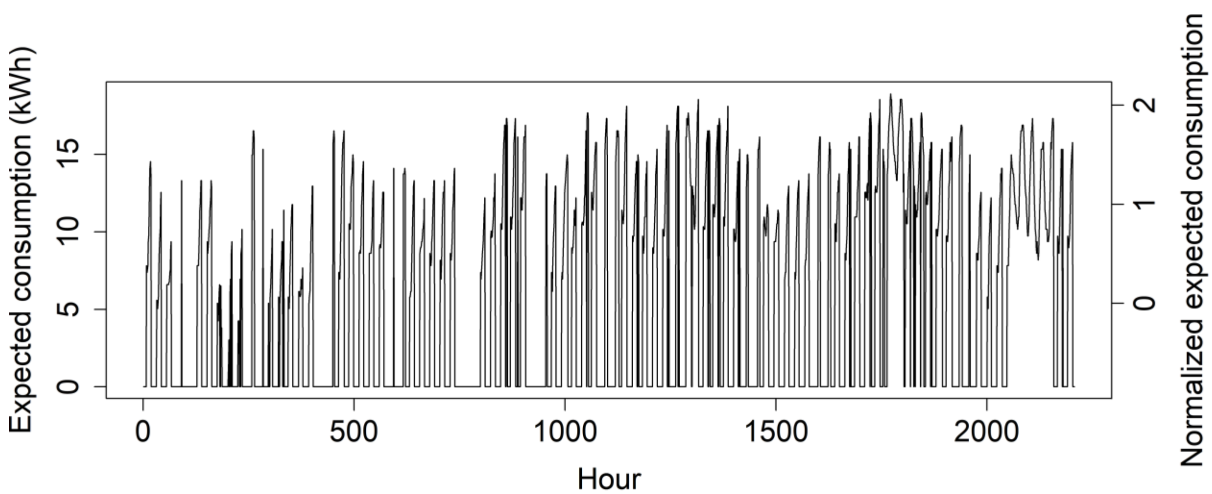


Figure 7.8 – Raw and Normalized Expected Energy Consumption for the Selected Real Building

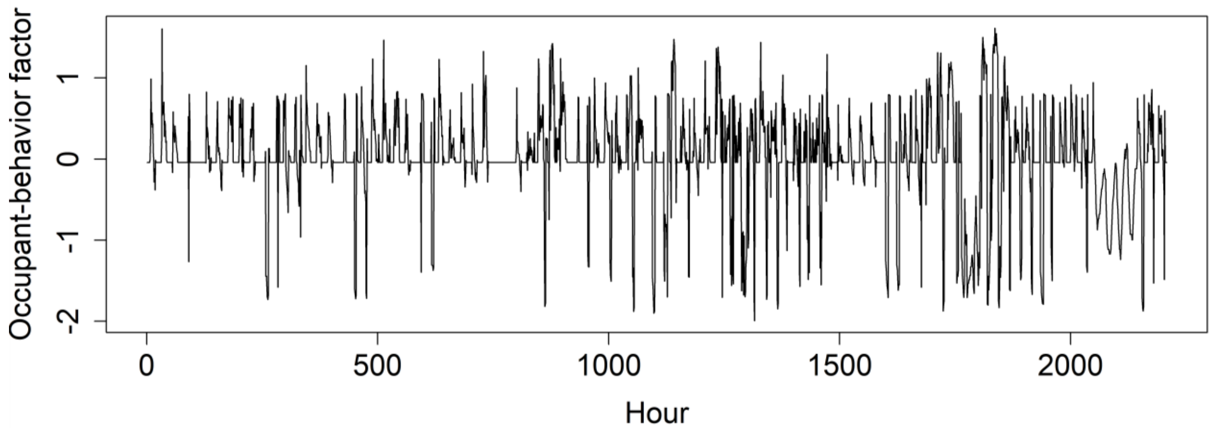


Figure 7.9 – Hourly Occupant-Behavior Factors

7.2.4 Occupant-Behavior-Factor Prediction Performance

Table 7.5 summarizes the prediction performance results of the four algorithms in predicting the values of the occupant-behavior factor. The occupant-behavior-factor models achieved RMSEs and CVs in the ranges of 0.75-0.87 and 9.19%-10.58%, respectively. Although there was no clear outperformer across the algorithms, the SVR model achieved the lowest CV and RMSE values, 0.75 RMSE and 9.19% CV. The predicted occupant-behavior-factor values show a good agreement with the actual values. Figure 7.10 shows the regression between the actual and predicted values of the occupant-behavior factor. The SVR-based model was, therefore, chosen for predicting the occupant-behavior-factor values.

Table 7.5 Performance of the Occupant-Behavior-Factor Prediction for all Algorithms

Algorithm							
CART		**GPR**		**SVR**		**ANN**	
RMSE	CV	RMSE	CV	RMSE	CV	RMSE	CV
0.87	10.58%	0.78	9.55%	0.75	9.19%	0.76	9.24%

CART = Classification and Regression Tree; GPR = Gaussian Process Regression; SVR = Support Vector Regression; ANN = Artificial Neural Networks; RMSE = Root Mean Square Error; CV = Coefficient of Variation.

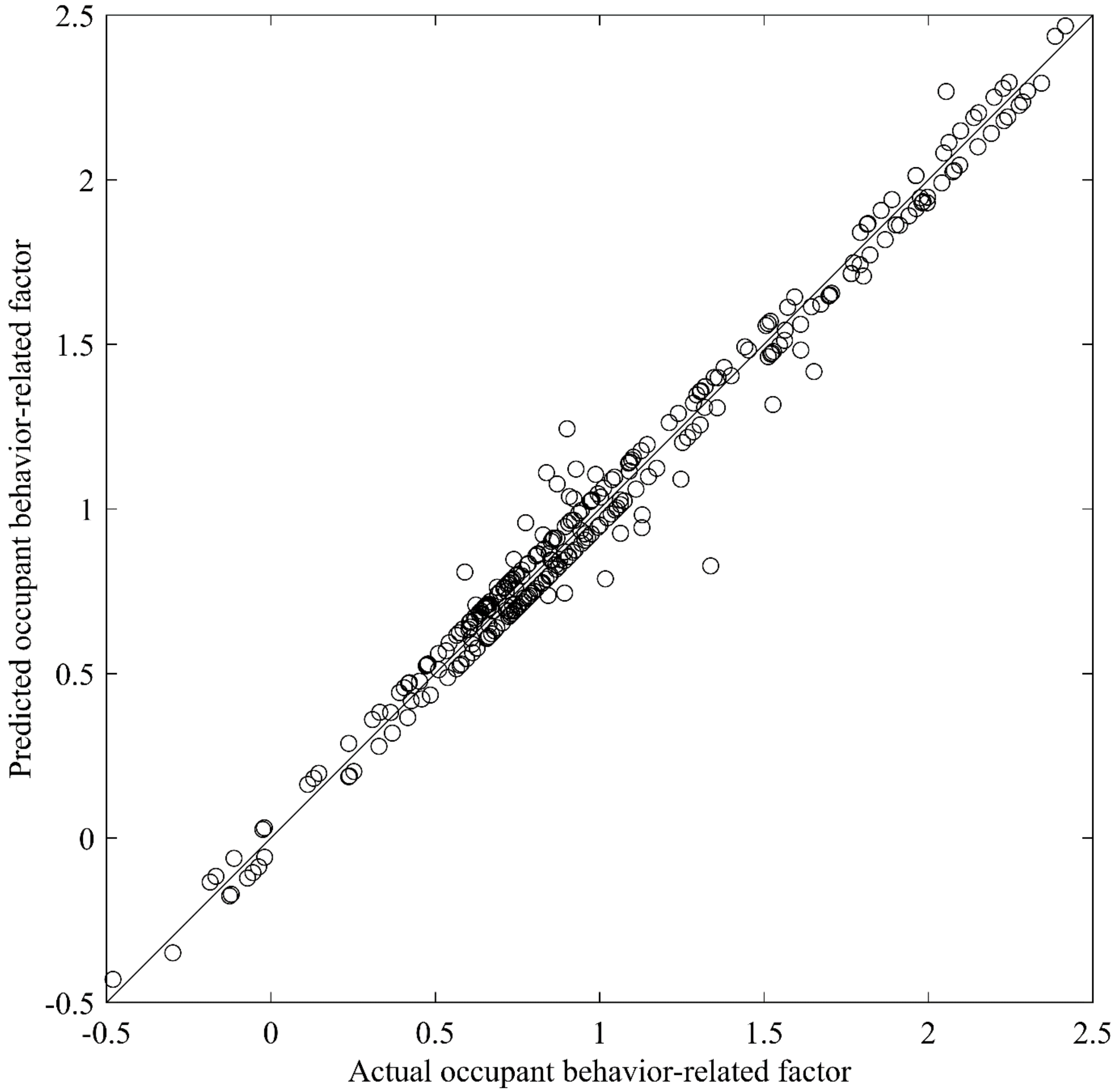


Figure 7.10 – Regression between Actual and Predicted Occupant-Behavior-Factor Values

7.2.5 Ensembler Model Performance

Table 7.6 summarizes the prediction performance results of the four algorithms in predicting the cooling energy consumption. The best prediction performance was achieved by the SVR model (0.73 kWh RMSE and 9.07% CV), with no clear outperformer across the GPR and ANN models, and with the CART models achieving the worst performance. Figure 7.11 shows the regression between the actual and predicted cooling energy consumption values of the SVR model. The predicted energy consumption values show a good agreement with the actual values.

The performance of the proposed hybrid model was compared to that of the real-data-driven model (Chapter 6). The real-data-driven prediction model achieved 0.98 RMSE and 16.11% CV, which is significantly lower than the performance of the proposed hybrid model. These results, thus, indicate that the proposed hybrid model is promising and can be utilized for better understanding and improvement of occupant behavior. The proposed model can be used to discover more efficient building operation and behavioral energy-saving strategies under a set of given weather conditions.

Table 7.6 Performance of the Ensemble Models for all Algorithms

Algorithm							
CART		**GPR**		**SVR**		**ANN**	
RMSE (kWh)	CV	RMSE (kWh)	CV	RMSE (kWh)	CV	RMSE (kWh)	CV
0.85	10.58%	0.76	9.55%	0.73	9.07%	0.74	9.24%

CART = Classification and Regression Tree; GPR = Gaussian Process Regression; SVR = Support Vector Regression; ANN = Artificial Neural Networks; RMSE = Root Mean Square Error; CV = Coefficient of Variation.

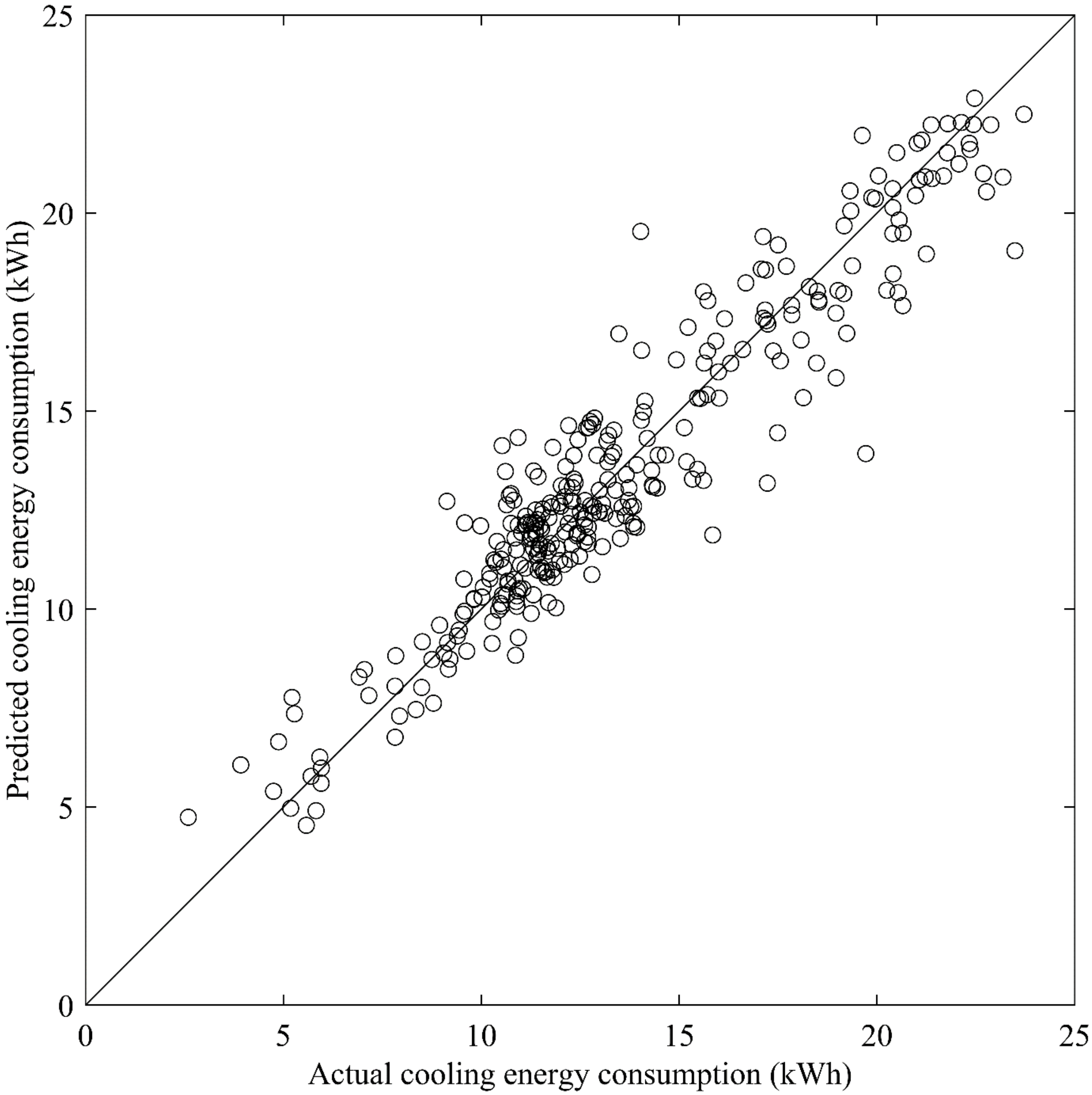


Figure 7.11 – Regression between Actual and Predicted Cooling Energy Consumption

CHAPTER 8 – CONCLUSIONS, CONTRIBUTIONS, LIMITATIONS, AND RECOMMENDATIONS FOR FUTURE RESEARCH

8.1 Conclusions

8.1.1 Conclusions for Energy-Related Values and Satisfaction Levels of Residential and Office Building Occupants

This thesis presented an empirical study to discover the human values that are related to energy-use behavior and energy consumption of building occupants. Seven energy-related values were identified and classified into three primary categories: thermal comfort, visual comfort, IAQ, health, personal productivity, environmental protection, and energy cost saving. The importance of these values to occupants and the satisfaction levels of occupants with these values were then investigated using two questionnaire surveys. The surveys focused on residential and office building occupants in AZ, IL, and PA. The results of the survey were statistically analyzed to identify similarities and differences across different types of occupants (residential and office), across different states (AZ, IL, PA), and across different potential energy-related factors (PEFs).

The results showed that all seven values were at least moderately important to the majority of residential and office building occupants. On average, health was ranked the highest in importance among the values – across both occupant types, and across the three states. Other than that, significant differences were shown in the importance rankings of residential and office building occupants, with energy cost saving being significantly more important for residential building occupants and visual comfort and personal productivity more important for office building occupants. No significant difference was shown in importance levels across the three states.

The results also showed that considerable percentages of occupants are unsatisfied with the fulfillment of one or more the values (11.2% to 21.4%) and that higher percentages of occupants

(24.8% to 38.6%) were thinking that their health and/or personal productivity are negatively affected by the current indoor environmental conditions. There were significant agreements on satisfaction rankings across both occupant groups and across the three states. But, significant differences were shown for a specific value, namely thermal comfort. Residential building occupants rated their satisfaction levels with thermal comfort (in both summer and winter) higher than office building occupants, and residential occupants in AZ showed higher satisfaction with thermal comfort in winter than those in PA.

For potential energy-related factors (PEFs), a number of differences in importance and satisfaction levels were shown across different gender groups, frequency of experiencing certain health symptoms, energy efficiency building features, energy-use behavior, and workspace and job characteristics. For example, (1) among residential building occupants, female occupants gave more importance to all seven values, and among office building occupants female occupants showed less satisfaction with all seven values; (2) residential building occupants with respiratory disease showed more importance to IAQ; (3) occupants who experienced higher frequencies of some health symptoms such as sore throat showed less satisfaction with several values such as IAQ; (4) occupants of Energy Star-certified buildings, for both residential and office buildings, expressed more satisfaction with environmental protection and energy cost saving than occupants of non-Energy Star buildings; (5) office building occupants who adjust thermostats, use/adjust room air conditioning units, and use/adjust ceiling fans were more satisfied with thermal comfort in summer; and (6) occupants of private workspaces attached higher importance to energy cost saving and showed higher satisfaction with thermal comfort (in winter and summer), visual comfort, and environmental protection.

The results also indicate some practical implications towards better understanding and improvement of energy-use behavior. For example, the difference in importance levels of energy cost saving across residential and office building occupants indicates the need to find more innovative ways to incentivize office building occupants to improve their energy-use behavior. Overall, the variability of importance and satisfaction levels across different PEFs indicates the need for a human-centered and value-sensitive approach for improving energy-use behavior and enhancing building energy efficiency.

8.1.2 Conclusions for Simulation-Data-Driven Occupant-Behavior-Sensitive Machine Learning-Based Energy Consumption Prediction

This thesis presented a study on taking a simulation-data-driven machine-learning approach for predicting building energy consumption in an occupant behavior-sensitive manner. In this approach, a model learns from a large set of energy-use cases that were modeled and simulated in EnergyPlus. In this study, the model was trained using 3-month hourly data for 5,760 energy-use cases. These cases represent different combinations of building characteristics, outdoor weather conditions, and occupant behaviors. In developing the model, the CART, EBT, ANN, and DNN algorithms were tested and compared in terms of their prediction accuracy, computational efficiency, and sensitivity to variations in sample sizes.

The results showed that all four algorithms were able to achieve accurate predictions when sufficient amounts of data were used. The ANN and DNN models were computationally expensive to train for some sample sizes, but were able to achieve very high prediction accuracies compared to the CART and EBT models. The neural network models with the optimal number of hidden layers outperformed both the CART and EBT models in terms of prediction accuracy, for all sample sizes. For example, the DNN model with four hidden layers, the most accurate model, achieved 2.97% CV. The CART and EBT algorithms required more training data than the ANN and DNN algorithms to be considered calibrated. For example, the ANN model required a minimum sample size of 1,000, compared to 5,000 for the EBT model. For all sample sizes, the training times of the CART and EBT algorithms were less than those for the ANN and DNN algorithms. For example, for a sample size of 100,000, the DNN model with four hidden layers required 4,122 seconds to train, while the CART model required only 1 second. The DNN models were able to achieve a CV of less than 5% for sample sizes larger than 50,000, but the other algorithms were never able to achieve such low CV

8.1.3 Conclusions for Real-Data-Driven Occupant-Behavior-Sensitive Machine Learning-Based Energy Consumption Prediction and Behavior Optimization

This thesis proposed a real-data-driven approach to determine the optimal occupant behavior that can simultaneously reduce energy consumption and improve occupant comfort. The proposed approach consists of two components: a set of machine learning-based occupant-behavior-sensitive models for predicting energy consumption and thermal and visual comfort, and a genetic algorithm-based optimization model for optimizing occupant behavior. To test and evaluate the proposed approach, an office building was instrumented and data about energy consumption, outdoor weather conditions, occupant behavior, and occupant comfort were collected for about three months. To verify that the behavior features are discriminating, the performance of the models were compared to others without occupant-behavior features. To consider the delayed effects of outdoor weather, prediction models with the past two-hour, one-hour, and no past-hour outdoor weather condition features were tested. A set of machine learning algorithms for classification and regression were also tested. For optimization, a set of occupant behavior variables were optimized for minimum energy consumption and maximum thermal and visual comfort for each hour. Three extreme solutions – energy-priority solution, thermal-comfort-priority solution, visual-comfort priority solution – and a balanced solution were also analyzed.

The results showed that the most accurate cooling and lighting energy consumption prediction models achieved 0.98 and 0.26 RMSE, respectively, and the thermal and visual comfort prediction models achieved 0.20 and 0.40 MAE, respectively. The prediction models with behavior features always outperformed those without behavior features, and the inclusion of past-hour outdoor weather condition features slightly improved the prediction performance of the models. For the machine learning algorithms, the best prediction performance was achieved by SVR models, with no clear outperformer across the ANN and CART models, and with the MLR models achieving the worst performance. For optimization, the energy-priority, thermal-comfort-priority, visual-comfort-priority, and balanced solutions achieved 21.8%, 10.8%, 14.2%, and 14.0% energy savings, respectively. After optimization, the thermal discomfort time (i.e., moderately unsatisfied or lower) decreased from 17.7% (actual level reported by occupants) to only 3.4%, 0.0%, 1.1%, and 0.1% of the time for the energy-priority, thermal-comfort-priority, visual-comfort-priority and balanced solutions, respectively. And the visual discomfort decreased from 60.7% to 17.5%, 18.9%, 11.3%, and 15.3% for the energy-priority, thermal-comfort-priority, visual-comfort-priority, and balanced solutions, respectively.

8.1.4 Conclusions for Hybrid Machine Learning-Based Energy Consumption Prediction: Coupling Data-Driven and Physical Approaches

This thesis proposed a hybrid machine-learning approach for occupant-behavior-sensitive energy consumption prediction, which is composed of three constituent models: (1) a machine learning model that learns from simulation-generated data and predicts the hourly values of a weather factor – a factor that represents the impact of outdoor weather conditions on cooling energy consumption, at a specific hour; (2) a machine learning model that learns from real data and predicts the hourly values of an occupant-behavior factor – a factor that represents the impact of occupant behavior on cooling energy consumption, at a specific hour; and (3) an ensembler model that predicts the hourly cooling energy consumption values based on the predicted values of the two factors.

The results showed that the proposed hybrid approach can be superior to the traditional approaches and that it has the potential to be successfully used for more accurate predictions. Both the weather-factor and the occupant-behavior-factor prediction models achieved accurate prediction results, which indicates that the constituent models were able to successfully predict the impact of outdoor weather conditions and occupant behavior on energy consumption. As a whole, the ensembler model and the proposed hybrid model achieved 0.73 kWh RMSE and 9.07% CV, which is considered calibrated according to ASHRAE Guideline 14.

8.2 Limitations

8.2.1 Limitations for Energy-Related Values and Satisfaction Levels of Residential and Office Building Occupants

Two main limitations are acknowledged. First, only those PEFs that can be solicited through questionnaire surveys were considered. For example, PEFs that affect thermal comfort such as activity level, age, gender, height, weight, and health conditions can be solicited through questionnaire surveys and were included; but metabolic rate, clothing insulation, air temperature, radiant temperature, air speed, and humidity are difficult to capture through questionnaire surveys and were thus excluded. Additional empirical studies will be conducted in future work to see if/how the survey and sensor data can be integrated to better understand the energy-related building occupant values. Second, the seven values considered in this study may not fully cover all energy-related values. Additional studies will be conducted in future work to further explore the energy-related occupant values.

8.2.2 Limitations for Simulation-Data-Driven Occupant-Behavior-Sensitive Machine Learning-Based Energy Consumption Prediction

Two main limitations are acknowledged. First, the proposed prediction model learned from energy simulations, which could be limited in representing the complexity and stochastic nature of occupant behavior. Only a limited number of occupant behaviors were considered. In addition, the proxy behavior variables used for the energy simulations are by nature simplified, lacking the real-world complexity that may be encountered with actual occupant behavior. Nevertheless, the results of this study are important and help demonstrate the profound impact of occupant behavior on building energy consumption, as well as the feasibility and potential success of an occupant behavior-sensitive energy consumption prediction approach. Additional validation will be conducted in future work to further validate the proposed prediction model, and the behavior modeling approach, using real data collected from real buildings and real occupants. Second, in line with its intended scope, this study did not consider the impact of different building characteristics and design decisions such as building shape and orientation, envelope properties and thermal insulation, and HVAC properties. Future research efforts could further study the combined impact of both building design and occupant behavior on building energy consumption.

8.2.3 Limitations for Real-Data-Driven Occupant-Behavior-Sensitive Machine Learning-Based Energy Consumption Prediction and Behavior Optimization

Three main limitations are acknowledged. First, the proposed approach was tested on real data collected from an office building. Although the use of real-life data helps better represent and understand the complex and stochastic nature of occupant behavior and its impact on energy consumption and comfort, the data used in this study are limited in terms of size and variability. Additional validation will be conducted in future work to see if/how the experimental results – in terms of prediction performance, feature analysis, and time series analysis – will change for different contexts, i.e., different buildings, different building operation plans, different occupant profiles and characteristics, different seasons, different locations and climates, different weather conditions, etc. Second, only a limited number of occupant-behavior features were considered in this study. And, like any other human feedback, the collected occupant feedback may involve some human error. Additional occupant-behavior features such as clothing type and activity level will be considered in future work, and improved technologies, procedures, and processes for collecting occupant feedback will be explored. Third, the potential correlations among the occupant-behavior features were not considered in this study. For example, a correlation between opening/closing shading devices and turning off/on lights, can logically be expected. In order to test such correlations, additional statistical analysis will be conducted in future work. The impact of such correlations, if they exist, on the feature analysis and machine learning will also be further studied in future work.

8.2.4 Limitations for Hybrid Machine Learning-Based Energy Consumption Prediction: Coupling Data-Driven and Physical Approaches

Three main limitations are acknowledged. First, the proposed approach was validated using a testing dataset that was collected from a single office building only. Although the proposed approach achieved accurate prediction results and showed that it has potential to be successfully used for coupling data-driven and physical approaches for more generalizable models and better representing the complex and stochastic nature of occupant behavior, the single office building used in this study to test the proposed approach may not be sufficient to fully validate this approach. Additional validation will be conducted in future work to see if/how the prediction results – in terms of prediction performance, clustering analysis, weather normalization – will change for different contexts, i.e., different buildings, different building operation plans, different occupant profiles and characteristics, different seasons, different locations and climates, different weather conditions, etc. Second, in this study, it was assumed that the variance in building energy consumption that cannot be explained by the outdoor weather conditions is mainly caused by occupant behavior. However, as mentioned above, in reality additional factors or errors could contribute to this variance. Additional experiments are, thus, required to better identify, understand, and consider the non-weather-related factors that cause variation in building energy consumption, and the extend of this variation, and to validate the weather normalization method that was used in this study to reveal the impact of occupant behavior. Third, some important variables such as building characteristics (e.g., materials, geometry, orientation) and building services (e.g., HVAC, lighting, and equipment) – which affect building energy consumption significantly – were not considered. This is because the real data collected for this study were from a single building and, thus, these building variables were constant. For consistency, these variables were also kept constant in the simulations. Thus, the scope of the resulting hybrid model was limited to the variables considered in the simulations and the experimental studies. To further improve the generalizability of the resulting model, additional simulations, experimental studies, and real-data collection from a large number of real buildings could be conducted in future work.

8.3 Contributions to the Body of Knowledge

8.3.1 Contributions of Energy-Related Values and Satisfaction Levels of Residential and Office Building Occupants

This research contributes to the body of knowledge on two main levels. First this research advances the theoretical and empirical knowledge in the area of energy-related human values by identifying importance levels of occupant values, satisfaction levels with the values, factors that are associated with higher/lower importance and/or satisfaction levels, and potential factors that could help predict occupant satisfaction levels. Second, the research indicates some practical implications towards better understanding and improvement of energy-use behavior for behavioral energy savings. For example, the difference in importance levels of energy cost saving across residential and office building occupants indicates the need to find more innovative ways to incentivize office building occupants to improve their energy-use behavior.

8.3.2 Contributions of Simulation-Data-Driven Occupant-Behavior-Sensitive Machine Learning-Based Energy Consumption Prediction

This research contributes to the body of knowledge on two main levels. First, this research offers a simulation-data-driven machine-learning approach for predicting building energy consumption in an occupant behavior-sensitive manner. The proposed approach uses a set of proxy variables to represent and account for the behavior, in a simplified manner, in the energy simulations. The proposed approach could help better understand the impact of occupant behavior on building energy consumption, as well as identify opportunities for behavioral energy-saving measures and efficient building-operation strategies. Second, this study provides important insights to the field of machine learning-based energy consumption prediction. The results showed that the neural network model with the optimal number of hidden layers always outperformed both the CART and EBT models in terms of prediction accuracy. Shallower neural network models outperformed deeper neural network models in smaller datasets; whereas deeper models outperformed shallower models in larger datasets. Thus, the increase in the number of hidden layers in neural networks does not always guarantee an increase in accuracy. The EBT models always outperformed the CART models in terms of prediction accuracy. The accuracies of the four algorithms always increased as the sample size increased. The training times of the CART and EBT algorithms were less than the ANN and DNN algorithms. Thus, bagging should be one of the ensembling methods to consider when developing a building energy consumption prediction model. Given that there is no one-size-fits-all machine learning algorithm, studying new algorithms in the context of building energy consumption prediction is critical.

8.3.3 Contributions of Real-Data-Driven Occupant-Behavior-Sensitive Machine Learning-Based Energy Consumption Prediction and Behavior Optimization

This research contributes to the body of knowledge in two primary ways. First, this research offers real-data-driven occupant-behavior-sensitive hourly energy consumption and comfort prediction models. The models learn from real building sensor data and real occupant feedback to help capture the real-life complexity of occupant behavior and their impact on occupant comfort and energy usage. Compared to the simulation-data-driven approach (Section 8.3.2), this real-data-driven approach aims to better capture and model the real-life behavior and comfort of occupants and the real-life energy-consumption patterns of buildings. Second, this research offers a data-driven approach to incorporate occupant behavior into building energy optimization. Combining the prediction models with optimization offers a powerful tool for finding the right energy-use behavioral changes that can achieve, both, energy saving and comfort improvement. The proposed approach paves the way for machine-learning-assisted behavioral energy efficiency and occupant engagement approaches that can incentivize occupants to save energy while improving their comfort and quality of life.

8.3.4 Contributions of Hybrid Machine Learning-Based Energy Consumption Prediction: Coupling Data-Driven and Physical Approaches

This research contributes to the body of knowledge in two primary ways. First, this research offers a novel hybrid modeling approach for energy consumption prediction. The proposed approach offers a direction towards leveraging the strengths and reducing the limitations of the traditional data-driven and physical modeling approaches by learning from both real data and simulation-generated data. Learning from both types of data aims to overcome two main limitations: the limited generalizability of data-driven approaches to different cases and the limited ability of physical modeling approaches to capture occupant behavior and real-life consumption patterns. Second, this research introduces two factors to better understand and quantify the impacts of outdoor weather conditions and occupant behavior on building energy consumption: weather factor and occupant-behavior factor. These factors can help represent the reasons for the changes in energy consumption, identify inefficiencies caused by occupant behavior, and hence support behavioral energy-saving decision making.

8.4 Recommendations for Future Research

8.4.1 Recommendations for Energy-Related Values and Satisfaction Levels of Residential and Office Building Occupants

Beyond occupant surveys, future research is recommended in two main directions. First, further empirical studies could be conducted in real buildings to better understand how occupant values impact occupant energy-use behavior, how indoor environmental conditions and energy-use behavior affect occupant satisfaction, and how to improve building energy efficiency while maintaining occupant satisfaction levels. Second, future studies could further explore how to use advanced technologies to automatically monitor and interpret occupant values and satisfaction levels, including the capturing and understanding of cues related to thermal comfort and discomfort.

8.4.2 Recommendations for Simulation-Data-Driven Occupant-Behavior-Sensitive Machine Learning-Based Energy Consumption Prediction

Future research is recommended in four main directions. First, future research could further study the combined impact of building characteristics, design decisions, and occupant behavior on building energy consumption. Second, the proposed approach could be tested in real-life contexts – using real data collected from real buildings and real occupants – to validate the model and study how the modeling approach could be improved to better capture the complexity and stochastic nature of occupant behavior. Understanding and modeling occupant behavior and its impact, in a realistic manner, is crucial to actually realizing the benefits of behavioral energy efficiency efforts. Third, the proposed behavior-sensitive energy prediction approach could be extended to a smart grid context – moving from the building level to the smart grid level, as well as incorporating additional key features in the modeling such as occupant needs, preferences, and satisfaction. Occupants are key to efficient energy utilization in buildings. A good understanding of the energy use, comfort needs and preferences, and behavior of occupants – at both the individual and aggregate level – is essential to identifying successful incentive schemes and designing effective demand-response programs. Four, the proposed approach could be implemented in an online fashion. Being able to update the prediction model using the real-time grid data would further improve the prediction accuracy and computational efficiency of the proposed approach.

8.4.3 Recommendations for Real-Data-Driven Occupant-Behavior-Sensitive Machine Learning-Based Energy Consumption Prediction and Behavior Optimization

Future research is recommended in six main directions. First, exploring the use of advanced sensing technologies and wearable devices to facilitate the collection of additional amounts and types of data – including additional data collection from a larger number of real buildings, as well as collection of additional types of data about the characteristics, behavior, and comfort of the occupants such as clothing, activities, and physiological parameters for improved monitoring, prediction, and understanding of occupant behavior and comfort. Second, identifying additional potential energy-saving behavioral changes, as well as validating the impact of these changes using real-life field studies. Third, understanding how group comfort can be optimized even in the presence of large individual variations. Fourth, understanding attitudes towards behavioral changes and how to incentivize, sustain, and reward behavior changes related to energy saving and sustainability. Fifth, better understanding how contextual factors – such as work settings and occupant characteristics – influence behavior, comfort expectations, and attitudes towards behavioral changes. Six, identifying the correlations among occupant-behavior features (e.g., identify if opening windows is correlated with switching off lights, and vice versa) and studying the impact of such correlations, if they exist, on the feature analysis and machine learning models.

8.4.4 Recommendations for Hybrid Machine Learning-Based Energy Consumption Prediction: Coupling Data-Driven and Physical Approaches

Future research is recommended in four main directions. First, a set of new buildings with different geometries and different building properties could be modeled and simulated to allow for representing the entire U.S. building stock. Accordingly, the clusters will be updated to represent the new buildings; and the weather-factor prediction models for each cluster will be retrained to accommodate the refinements in the clusters. Second, the occupant-behavior-factor prediction model could be further improved through including additional types of occupant behavior (e.g., opening/closing windows, occupant mobility). In this regard, an advanced data collection system could be developed via incorporating smart sensing technologies. As a result, the prediction model will be able to learn and predict the impact of new behaviors on energy consumption. Third, the proposed approach could be tested and adapted to different contexts (e.g., different buildings, different occupant profiles) by conducting additional experiments. Fourth, additional factors, such as building characteristics and other sources of energy consumption variance, could be studied and modeled to distinguish and capture their impacts separately.

REFERENCES

Abraham, Y., Zhao, Z., Anumba, C., & Asadi, S. (2017). Implementation of a Preference Monitoring Application for office Building Occupants. In *Lean and Computing in Construction Congress - Volume 1: Proceedings of the Joint Conference on Computing in Construction* (pp. 793–801). Heraklion: Heriot-Watt University. https://doi.org/10.24928/JC3-2017/0201

Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T. (2015). Time-series clustering – A decade review. *Information Systems*, *53*, 16–38. https://doi.org/10.1016/j.is.2015.04.007

Ahmad, M. W., Hippolyte, J.-L., Reynolds, J., Mourshed, M., & Rezgui, Y. (2016). Optimal scheduling strategy for enhancing IAQ , visual and thermal comfort using a genetic algorithm. Retrieved from http://orca.cf.ac.uk/92212/1/Ahmad_IAQ_Conference_Paper.V2.0.pdf

Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H., & Saidur, R. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, *33*, 102–109. https://doi.org/10.1016/j.rser.2014.01.069

Alahmad, M. A., Wheeler, P. G., Schwer, A., Eiden, J., & Brumbaugh, A. (2012). A comparative study of three feedback devices for residential real-time energy monitoring. *Industrial Electronics, IEEE Transactions on, 59*(4), 2002-2013.

Amasyali, K., & El-Gohary, N. (2015). Discovering the values of residential building occupants for value-sensitive improvement of building energy efficiency. *Proceedings of the*

Canadian Society for Civil Engineering's 5th International/11th Construction Specialty Conference (ICSC15), Vancouver, BC, Canada, June 7-10, 2015.

Amasyali, K., & El-Gohary, N. M. (2016). Energy-related values and satisfaction levels of residential and office building occupants. *Building and Environment*, *95*, 251-263.

Amasyali, K., & El-Gohary, N. (2017). Deep Learning for Building Energy Consumption Prediction. In *6th CSCE/CRC International Construction Specialty Conference*. Vancouver, Canada.

Amasyali, K., & El-Gohary, N. (2018). Machine learning-based occupant energy use behavior optimization. *Proceedings of the 2018 Construction Research Congress (CRC 2018),* New Orleans, LA, April 2- 5, 2018.

Amasyali, K., & El-Gohary, N. (2019). Predicting Energy Consumption of Office Buildings: A Hybrid Machine Learning-Based Approach. In *Advances in Informatics and Computing in Civil and Construction Engineering* (pp. 695–700). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-00220-6_83

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, *81*, 1192–1205. https://doi.org/10.1016/j.rser.2017.04.095

American Lung Association. (2015). Indoor air quality: Why does indoor air quality matter?. Retrieved from http://www.lung.org/associations/charters/mid-atlantic/air-quality/indoor-air-quality.html

An, N., Zhao, W., Wang, J., Shang, D., & Zhao, E. (2013). Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. *Energy*, *49*(1), 279–288. https://doi.org/10.1016/j.energy.2012.10.035

ANSI/ASHRAE/ASHE standard 55-2010: Thermal environmental conditions for human occupancy. Atlanta: ASHRAE,.

Aries, M. B. C., Veitch, J. A., & Newsham, G. R. (2010). Windows, view, and office characteristics predict physical and psychological discomfort. *Journal of Environmental Psychology*, *30*(4), 533-541.

Asadi, E., Silva, M. G. da, Antunes, C. H., Dias, L., & Glicksman, L. (2014). Multi-objective optimization for building retrofit: A model using genetic algorithm and artificial neural network and an application. *Energy and Buildings*, *81*, 444–456. https://doi.org/10.1016/j.enbuild.2014.06.009

Ascione, F., Bianco, N., De Masi, R. F., Mauro, G. M., & Vanoli, G. P. (2017a). Resilience of robust cost-optimal energy retrofit of buildings to global warming: A multi-stage, multi-objective approach. *Energy and Buildings*, *153*, 150–167. https://doi.org/10.1016/j.enbuild.2017.08.004

Ascione, F., Bianco, N., De Masi, R., Mauro, G., & Vanoli, G. (2015). Design of the Building Envelope: A Novel Multi-Objective Approach for the Optimization of Energy Performance and Thermal Comfort. *Sustainability*, *7*(8), 10809–10836. https://doi.org/10.3390/su70810809

Ascione, F., Bianco, N., De Stasio, C., Mauro, G. M., & Vanoli, G. P. (2017b). Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. *Energy*, *118*, 999–1017. https://doi.org/10.1016/j.energy.2016.10.126

ASHRAE. (2013). ASHRAE Standard 169-2013. *Climatic Data for Building Design Standards*. ASHRAE 2013.

ASHRAE. (2013). ASHRAE Standard 90.1-2013. *Energy Standard for Buildings Except Low-Rise Residential Buildings*.

Azadeh, A., Ghaderi, S. F., & Sohrabkhani, S. (2008). Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and Management, 49*(8), 2272-2278.

Azar, E., & Menassa, C. C. (2012a). A comprehensive analysis of the impact of occupancy parameters in energy simulation of office buildings. *Energy and Buildings, 55*(0), 841-853.

Azar, E., & Menassa, C. C. (2012b). Agent-based modeling of occupants and their impact on energy use in commercial buildings. *Journal of Computing in Civil Engineering, 26*(4), 506-518.

Azar, E., & Menassa, C. C. (2014). A comprehensive framework to quantify energy savings potential from improved operations of commercial building stocks. *Energy Policy*, *67*, 459–472. https://doi.org/10.1016/j.enpol.2013.12.031

Bagnasco, A., Fresi, F., Saviozzi, M., Silvestro, F., & Vinci, A. (2015). Electrical consumption forecasting in hospital facilities: An application case. *Energy and Buildings, 103*, 261-270.

Bakke, J. V., Moen, B. E., Wieslander, G., & Norback, D. (2007). Gender and the physical and psychosocial work environments are related to indoor air symptoms. *Journal of Occupational and Environmental Medicine / American College of Occupational and Environmental Medicine, 49*(6), 641-650.

Barlow, S., & Fiala, D. (2007). Occupant comfort in UK offices—How adaptive comfort theories might influence future low energy office refurbishment strategies. *Energy and Buildings*, *39*(7), 837-846.

Becerik-Gerber, B., Siddiqui, M. K., Brilakis, I., El-Anwar, O., El-Gohary, N., Mahfouz, T., et al. (2014). Civil engineering grand challenges: Opportunities for data sensing, information analysis, and knowledge discovery. *Journal of Computing in Civil Engineering, 28*(4), 1-13.

Beheshti, S., Sahebalam, A., & Nidoy, E. (2019). Structure dependent weather normalization. *Energy Science & Engineering*. https://doi.org/10.1002/ese3.272

Ben-Nakhi, A. E., & Mahmoud, M. A. (2004). Cooling load prediction for buildings using general regression neural networks. *Energy Conversion and Management*, *45*(13–14), 2127-2141.

Bluyssen, P. M., Aries, M., & van Dommelen, P. (2011). Comfort of workers in office buildings: The european HOPE project. *Building and Environment, 46*(1), 280-288.

Bonte, M., Thellier, F., & Lartigue, B. (2014). Impact of occupant's actions on energy building performance and thermal sensation. *Energy and Buildings*, *76*, 219–227. https://doi.org/10.1016/j.enbuild.2014.02.068

Borges, C. E., Penya, Y. K., & Fernandez, I. (2013b). Evaluating combined load forecasting in large power systems and smart grids. *Industrial Informatics, IEEE Transactions on*, *9*(3), 1570-1577.

Borges, C. E., Penya, Y. K., Fernández, I., Prieto, J., & Bretos, O. (2013a). Assessing tolerance-based robust short-term load forecasting in buildings. *Energies (19961073), 6*(4), 2110-2129.

Boston, G. (2013). Basal metabolic rate changes as you age. Retrieved from http://www.washingtonpost.com/lifestyle/wellness/basal-metabolic-rate-changes-as-you-age/2013/03/05/d26b1c18-80f1-11e2-a350-49866afab584_story.html

Boundless. (2015). How Values Influence Behavior. Retrieved from https://www.boundless.com/management/textbooks/boundless-management-textbook/organizational-behavior-5/drivers-of-behavior-44/how-values-influence-behavior-230-7046/

Cao, B., Ouyang, Q., Zhu, Y., Huang, L., Hu, H., & Deng, G. (2012). Development of a multivariate regression model for overall satisfaction in public buildings based on field studies in beijing and shanghai. *Building and Environment, 47*(0), 394-399.

Cárdenas, J., Romeral, J., García, A., & Giacometto, F. Short-term load forecasting based on dominant hidden patterns for an iEMS in the user-side.

Carrico, A. R., & Riemer, M. (2011). Motivating energy conservation in the workplace: An evaluation of the use of group-level feedback and peer education. *Journal of Environmental Psychology*, *31*(1), 1-13.

CBEI. (2017). Consortium for Building Energy Innovation. Retrieved from http://cbei.psu.edu/

Chae, Y. T., Horesh, R., Hwang, Y., & Lee, Y. M. (2016). Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy and Buildings*, *111*, 184–194. https://doi.org/10.1016/j.enbuild.2015.11.045

Chari, A., & Christodoulou, S. (2017). Building energy performance prediction using neural networks. *Energy Efficiency*, *10*(5), 1315–1327. https://doi.org/10.1007/s12053-017-9524-5

Chen, B. J., Chang, M. W., & Lin, C. J. (2004). Load forecasting using support vector machines: A study on EUNITE competition 2001. *Power Systems, IEEE Transactions on*, *19*(4), 1821-1830.

Cheng, S., Lei, L., & Zhe, K. (2015). Study on Lighting Energy Consumption Prediction Model of Office Buildings in Harbin,China using Diva. In *Proceedings of BS2015: 14th Conference of International Building Performance Simulation Association* (pp. 1221–1228). Hyderabad.

Cho, S. H., Kim, W. T., Tae, C. S., & Zaheeruddin, M. (2004). Effect of length of measurement period on accuracy of predicted annual heating energy consumption of buildings. *Energy Conversion and Management*, *45*(18), 2867-2878.

Chou, J.-S., & Bui, D.-K. (2014). Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy and Buildings*, *82*, 437–446. https://doi.org/10.1016/j.enbuild.2014.07.036

CIE. (2001). Lighting of indoor work places. Publication D-008/E-2001

Clevenger, C. M., Haymaker, J. R., & Jalili, M. (2014). Demonstrating the Impact of the Occupant on Building Performance. *Journal of Computing in Civil Engineering*, *28*(1), 99–102. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000323

Crawley, D. B., Hand, J. W., Kummert, M., & Griffith, B. T. (2008). Contrasting the capabilities of building energy performance simulation programs. *Building and Environment*, *43*(4), 661-673.

Dagnely, P., Reutte, T., Tourwe, T., Tsiporkova, E., & Verhelst, C. (2015). Predicting hourly energy consumption. Can you beat an autoregressive model?. *Belgian-Dutch Conference on Machine Learning (Benelearn)*.

Darby, S. (2006). The effectiveness of feedback on energy consumption. *A Review for DEFRA of the Literature on Metering, Billing and direct Displays*.

Daum, D., Haldi, F., & Morel, N. (2011). A personalized measure of thermal comfort for building controls. *Building and Environment*, *46*(1), 3-11.

Deb, C., Eang, L. S., Yang, J., & Santamouris, M. (2016). Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. *Energy and Buildings*, *121*, 284–297. https://doi.org/10.1016/j.enbuild.2015.12.050

Deb, K. (2001). *Multi-Objective Optimization Using Evolutionary Algorithms*. *John Wiley & sons, LTD*. https://doi.org/10.1109/TEVC.2002.804322

DeCicco, J., Yan, T., Keusch, F., Muñoz, D.H., Neidert, L. (2015). University of Michigan Energy Survey: Year One Report. Ann Arbor: University of Michigan; 2014, Retrieved from http://energy.umich.edu/file/esyearonereportlayoutversionpdf

Derrible, S., & Reeder, M. (2015). The cost of over-cooling commercial buildings in the United States. *Energy and Buildings*, *108*, 304–306. https://doi.org/10.1016/j.enbuild.2015.09.022

Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., … Crawley, D. (2011). *U.S. Department of Energy Commercial Reference Building Models of the National Building Stock*. Golden, CO (United States). https://doi.org/10.2172/1009264

DOE. (2012). Buildings Energy Data Book, Department of Energy, Retrieved from http://buildingsdatabook.eren.doe.gov/

Domingos, P. (2012). A few useful things to know about machine learning. *Commun.ACM*, *55*(10), 78-87.

Dong, B., Cao, C., & Lee, S. E. (2005). Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*, *37*(5), 545–553. https://doi.org/10.1016/j.enbuild.2004.09.009

Easterlin, Richard A. (1987) *Birth and fortune :the impact of numbers on personal welfare* Chicago : University of Chicago Press,

Edwards, L., Torcellini, P. A., & National Renewable Energy Laboratory (U.S.). (2002). *A literature review of the effects of natural light on building occupants*. Golden, CO : National Renewable Energy Laboratory,.

Edwards, R. E., New, J., & Parker, L. E. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. *Energy and Buildings*, *49*, 591-603.

EIA. (2013). Lower residential energy use reduces home energy expenditures as share of household income. Retrieved November 30, 2015 from https://www.eia.gov/todayinenergy/detail.cfm?id=10891

EIA. (2019). Energy Consumption by Sector. Retrieved April 15, 2019, from https://www.eia.gov/totalenergy/data/monthly/pdf/sec2_3.pdf

EIA. (2015). CBECS 2012: Building Stock Results. Retrieved from https://www.eia.gov/consumption/commercial/reports/2012/buildstock/

Ekici, B. B., & Aksoy, U. T. (2009). Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*, *40*(5), 356–362. https://doi.org/10.1016/j.advengsoft.2008.05.003

Ekonomou, L. (2010). Greek long-term energy consumption prediction using artificial neural networks. *Energy*, *35*(2), 512–517. https://doi.org/10.1016/j.energy.2009.10.018

EN. (2011). Light and lighting. Basic terms and criteria for specifying lighting requirements.

Energy Star (2015a). About ENERGY STAR. Retrieved from http://www.energystar.gov/about

Energy Star (2015b). Certified new homes. Retrieved from http://www.energystar.gov/index.cfm?c=new_homes.hm_index&s=mega

EPA. (1991). Indoor Air Facts No. 4 (revised) Sick Building Syndrome. Retrieved from http://www.epa.gov/iaq/pdfs/sick_building_factsheet.pdf

EPA. (2009). Buildings and their Impact on the Environment: A Statistical Summary. Retrieved from http://archive.epa.gov/greenbuilding/web/pdf/gbstats.pdf

EPA. (2015). An Introduction to Indoor Air Quality. Retrieved from http://www2.epa.gov/indoor-air-quality-iaq/introduction-indoor-air-quality

Escrivá-Escrivá, G., Álvarez-Bel, C., Roldán-Blay, C., & Alcázar-Ortega, M. (2011). New artificial neural network prediction method for electrical consumption forecasting based on building end-uses. *Energy and Buildings, 43*(11), 3112-3119.

European Commission. (2003). Rapid Press Release. *Indoor Air Pollution: New EU Research Reveals Higher Risks Than Previously Thought.* Reference IP/03/1278 of 22-09-2003

Fan, C., Xiao, F., & Wang, S. (2014). Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy, 127*, 1-10.

Fan, C., Xiao, F., & Zhao, Y. (2017). A short-term building cooling load prediction method using deep learning algorithms. *Applied Energy, 195*, 222–233. https://doi.org/10.1016/j.apenergy.2017.03.064

Faruqui, A., & Sergici, S. (2010). Household response to dynamic pricing of electricity-a survey of the empirical evidence. https://dx.doi.org/10.2139/ssrn.1134132

Faruqui, A., Sergici, S., & Sharif, A. (2010). The impact of informational feedback on energy consumption—A survey of the experimental evidence. *Energy, 35*(4), 1598-1608.

Farzana, S., Liu, M., Baldwin, A., & Hossain, M. U. (2014). Multi-model prediction and simulation of residential building energy in urban areas of chongqing, south west china. *Energy and Buildings, 81*, 161-169.

Ferlito, S., Atrigna, M., Graditi, G., De Vito, S., Salvato, M., Buonanno, A., et al. (2015). Predictive models for building's energy consumption: An artificial neural network (ANN) approach. *AISEM Annual Conference*.

Fernandez, I., Borges, C. E., & Penya, Y. K. (2011). Efficient building load forecasting. *Emerging Technologies & Factory Automation (ETFA), 2011 IEEE 16th Conference*.

Finkelstein, E. A., DiBonaventura, M. d., Burgess, S. M., & Hale, B. C. (2010). The costs of obesity in the workplace. *Journal of Occupational & Environmental Medicine, 52*(10), 971-976.

Fischer, C. (2008). Feedback on household electricity consumption: A tool for saving energy? *Energy Efficiency*, *1*(1), 79-104.

Fisk, W. J. (2000). Health and productivity gains from better indoor environments and their relationship with building energy efficiency. *Annual Review of Energy and the Environment*, *25*(1), 537-566.

Fisk, W. J., Black, D., & Brunner, G. (2011). Benefits and costs of improved IEQ in U.S. offices. *Indoor Air, 21*(5), 357-367.

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*, *23*, 272–288. https://doi.org/10.1016/j.rser.2013.03.004

Frey, P., Dunn, L., Cochran, R., Spataro, K., McLennan, J., DiNola, R., et al. (2012). The greenest building: Quantifying the environmental value of building reuse-a report by the US national trust for historic preservation. Retrieved from http://www.preservationnation.org/information-center/sustainable-communities/green-lab/lca/The_Greenest_Building_lowres.pdf

Frontczak, M., & Wargocki, P. (2011). Literature survey on how different factors influence human comfort in indoor environments. *Building and Environment*, *46*(4), 922-937.

Frontczak, M., Schiavon, S., Goins, J., Arens, E., Zhang, H., & Wargocki, P. (2012). Quantitative relationships between occupant satisfaction and satisfaction aspects of indoor environmental quality and building design. *Indoor Air*, *22*(2), 119-131.

Gaetani, I., Hoes, P.-J., & Hensen, J. L. M. (2016). Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy. *Energy and Buildings*, *121*, 188–204. https://doi.org/10.1016/j.enbuild.2016.03.038

Georgescu, M., Eccles, E., Manjunath, V., Swindle, E., & Mezic, I. Machine learning methods for site-level building energy forecasting and data rectification. Retrieved from http://www.bso14.org/BSO14_Papers/BSO14_Paper_034.pdf

González, P. A., & Zamarreño, J. M. (2005). Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings*, *37*(6), 595-601.

Gou, S., Nik, V. M., Scartezzini, J.-L., Zhao, Q., & Li, Z. (2017). Passive design optimization of newly-built residential buildings in Shanghai for improving indoor thermal comfort while reducing building energy demand. *Energy and Buildings*. https://doi.org/10.1016/j.enbuild.2017.09.095

Grygierek, K., & Ferdyn-Grygierek, J. (2018). Multi-Objective Optimization of the Envelope of Building with Natural Ventilation. *Energies*, *11*(6), 1383. https://doi.org/10.3390/en11061383

Guerra-Santin, O., Bosch, H., Budde, P., Konstantinou, T., Boess, S., Klein, T., & Silvester, S. (2018). Considering user profiles and occupants' behaviour on a zero energy renovation strategy for multi-family housing in the Netherlands. *Energy Efficiency*, *11*(7), 1847–1870. https://doi.org/10.1007/s12053-018-9626-8

Haldi, F. (2010). Towards a Unified Model of Occupants' Behaviour and Comfort for Building Energy Simulation (Doctoral dissertation, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE).

Halpern, M. T., Shikiar, R., Rentz, A. M., & Khan, Z. M. (2001). Impact of smoking status on workplace absenteeism and productivity. *Tobacco Control, 10*(3), 233-238.

Hawkins, D., Hong, S. M., Raslan, R., Mumovic, D., & Hanna, S. (2012). Determinants of energy use in UK higher education buildings using statistical and artificial neural network methods. *International Journal of Sustainable Built Environment, 1*(1), 50-63.

Haynes, B. P. (2007). Office productivity: A shift from cost reduction to human contribution. *Facilities, 25*(11), 452-462.

Hellerstein, J. M. (2008). Quantitative data cleaning for large databases. *United Nations Economic Commission for Europe (UNECE)*. Retrieved November 30, 2015 from http://db.cs.berkeley.edu/jmh/papers/cleaning-unece.pdf

Hong, S.-M., Paterson, G., Mumovic, D., & Steadman, P. (2014). Improved benchmarking comparability for energy consumption in schools. *Building Research & Information*, *42*(1), 47–61. https://doi.org/10.1080/09613218.2013.814746

Hong, T. (2014). Occupant behavior: impact on energy use of private offices. *ASim 2012-1st Asia conference of International Building Performance Simulation Association.*

Hong, T., & Lin, H.-W. (2014). Occupant Behavior: Impact on Energy Use of Private Offices. In *ASim 2012 - 1st Asia conference of International Building Performance Simulation Association*. Shanghai, China: Lawrence Berkeley National Laboratory. Retrieved from http://www.escholarship.org/uc/item/6jp5w8kn

Hong, T., Chen, Y., Belafi, Z., & D'Oca, S. (2018). Occupant behavior models: A critical review of implementation and representation approaches in building performance simulation programs. *Building Simulation*, *11*(1), 1–14. https://doi.org/10.1007/s12273-017-0396-6

Hong, T., D'Oca, S., Turner, W. J. N., & Taylor-Lange, S. C. (2015). An ontology to represent energy-related occupant behavior in buildings. Part I: Introduction to the DNAs framework. *Building and Environment*, *92*, 764–777. https://doi.org/10.1016/j.buildenv.2015.02.019

Hong, T., Sun, H., Chen, Y., Taylor-Lange, S. C., & Yan, D. (2016a). An occupant behavior modeling tool for co-simulation. *Energy and Buildings*, *117*, 272–281. https://doi.org/10.1016/j.enbuild.2015.10.033

Hong, T., Taylor-Lange, S. C., D'Oca, S., Yan, D., & Corgnati, S. P. (2016b). Advances in research and applications of energy-related occupant behavior in buildings. *Energy and Buildings*, *116*, 694–702. https://doi.org/10.1016/j.enbuild.2015.11.052

Hou, Z., Lian, Z., Yao, Y., & Yuan, X. (2006). Cooling-load prediction by the combination of rough set theory and an artificial neural-network based on data-fusion technique. *Applied Energy, 83*(9), 1033-1046.

Huizenga, C., Abbaszadeh, S., Zagreus, L., & Arens, E. A. (2006). Air quality and thermal comfort in office buildings: results of a large indoor environmental quality survey. *Center for the Built Environment*. Retrieved November 30, 2015 from http://cbe.berkeley.edu/research/pdf_files/Huizenga_HB2006.pdf

Hydro One. (2006). Hydro One Weather Normalization Methodology. Retrieved from https://www.oeb.ca/documents/cases/EB-2005-0317/phase3/jun15/handout-weathernormalization-honi.pdf

IEA. (2013). *Transition to Sustainable Buildings*. OECD. https://doi.org/10.1787/9789264202955-en

IEA. (2014). *World Energy Outlook 2014*. OECD. https://doi.org/10.1787/weo-2014-en.

IEA. (2011). *Energy-efficient Buildings Heating and Cooling Equipment*. OECD. https://www.oecd-ilibrary.org/docserver/9789264118492-en.pdf?expires=1555288970&id=id&accname=ocid43013819&checksum=321865ECE9723BEF45E365E62E7847F5

IEA. (2015). About lighting. Retrieved November 28, 2015 from https://www.iea.org/topics/energyefficiency/subtopics/lighting/

IPMVP. (2003). Concepts and Options for Determining Energy Savings in New Construction, *3*. Retrieved from www.ipmvp.org

Iwafune, Y., Yagita, Y., Ikegami, T., & Ogimoto, K. (2014). Short-term forecasting of residential building load for distributed energy management. *Energy Conference (ENERGYCON)*.

Jain, A., & Satish, B. (2009). Clustering based Short Term Load Forecasting using Support Vector Machines. In *2009 IEEE Bucharest PowerTech* (pp. 1–8). https://doi.org/10.1109/PTC.2009.5282144

Jain, R. K., Smith, K. M., Culligan, P. J., & Taylor, J. E. (2014a). Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy*, *123*, 168–178. https://doi.org/10.1016/j.apenergy.2014.02.057

Jain, R., Damoulas, T., & Kontokosta, C. (2014b). Towards data-driven energy consumption forecasting of multi-family residential buildings: Feature selection via the lasso. (pp. 1675-1682) American Society of Civil Engineers.

Jazizadeh, F., Ghahramani, A., Becerik-Gerber, B., Kichkaylo, T., & Orosz, M. (2014). Human-Building Interaction Framework for Personalized Thermal Comfort-Driven Systems in Office Buildings. *Journal of Computing in Civil Engineering*.

Jian, Y., Li, Y., Wei, S., Zhang, Y., & Bai, Z. (2015). A Case Study on Household Electricity Uses and Their Variations Due to Occupant Behavior in Chinese Apartments in Beijing. *Journal of Asian Architecture and Building Engineering*, *14*(3), 679–686. https://doi.org/10.3130/jaabe.14.679

Jinhu, L., Xuemei, L., Lixing, D., & Liangzhong, J. (2010). Applying principal component analysis and weighted support vector machine in building cooling load forecasting. *Computer and Communication Technologies in Agriculture Engineering (CCTAE)*.

Jovanović, R. Ž., Sretenović, A. A., & Živković, B. D. (2015). Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings, 94*, 189–199. https://doi.org/10.1016/j.enbuild.2015.02.052

Junfang, W., & Dongxiao, N. (2009). Short-Term Power Load Forecasting Using Least Squares Support Vector Machines (LS-SVM). *Computer Science and Engineering, 2009. WCSE'09. Second International Workshop on Computer Science and Engineering*.

Kamaev, V., Shcherbakov, M., Panchenko, D., Shcherbakova, N., & Brebels, A. (2012). Using connectionist systems for electric energy consumption forecasting in shopping centers. *Automation and Remote Control, 73*(6), 1075-1084.

Karatasou, S., Santamouris, M., & Geros, V. (2006). Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings, 38*(8), 949-958.

Karjalainen, S. (2007). Gender differences in thermal comfort and use of thermostats in everyday thermal environments. *Building and Environment, 42*(4), 1594-1603.

Karjalainen, S. (2012). Thermal comfort and gender: A literature review. *Indoor Air, 22*(2), 96-109.

Karsoliya, S. (2012). Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture. *International Journal of Engineering Trends and Technology, 3*(6), 714–717.

Kendall, M., & Gibbons, J. (1990). *Rank correlation methods*. London: E. Arnold.

Kim, J., & de Dear, R. (2013). Workspace satisfaction: The privacy-communication trade-off in open-plan offices. *Journal of Environmental Psychology, 36*(0), 18-26.

Kim, J., de Dear, R., Cândido, C., Zhang, H., & Arens, E. (2013). Gender differences in office occupant perception of indoor environmental quality (IEQ). *Building and Environment, 70*(0), 245-256.

Klein, L., Kwak, J., Kavulya, G., Jazizadeh, F., Becerik-Gerber, B., Varakantham, P., & Tambe, M. (2012). Coordinating occupant behavior for building energy and comfort management using multi-agent systems. *Automation in Construction*, *22*, 525–536. https://doi.org/10.1016/j.autcon.2011.11.012

Kneifel, J. (2010). Life-cycle carbon and cost analysis of energy efficiency measures in new commercial buildings. *Energy and Buildings, 42*(3), 333-340.

Kwok, S. S. K., & Lee, E. W. M. (2011). A study of the importance of occupancy to building cooling load in prediction by intelligent approach. *Energy Conversion and Management, 52*(7), 2555-2564.

Kwok, S. S. K., Yuen, R. K. K., & Lee, E. W. M. (2011). An intelligent approach to assessing the effect of building occupancy on building cooling load prediction. *Building and Environment, 46*(8), 1681-1690.

Lahouar, A., & Ben Hadj Slama, J. (2015). Day-ahead load forecast using random forest and expert input selection. *Energy Conversion and Management*, *103*, 1040–1051. https://doi.org/10.1016/j.enconman.2015.07.041

Lai, A. C. K., Mui, K. W., Wong, L. T., & Law, L. Y. (2009). An evaluation model for indoor environmental quality (IEQ) acceptance in residential buildings. *Energy and Buildings, 41*(9), 930-936.

Lai, F., Magoulès, F., & Lherminier, F. (2008). Vapnik's learning theory applied to energy consumption forecasts in residential buildings. *International Journal of Computer Mathematics, 85*(10), 1563-1588.

Lai, J. H. K., & Yik, F. W. H. (2009). Perception of importance and performance of the indoor environmental quality of high-rise residential buildings. *Building and Environment, 44*(2), 352-360.

Lam, J. C., Wan, K. K. W., Wong, S. L., & Lam, T. N. T. (2010). Principal component analysis and long-term building energy simulation correlation. *Energy Conversion and Management, 51*(1), 135-139.

Lam, K., Zhao, J., B Ydstie, E., Wirick, J., Qi, M., & Park, J. (2014). An energyplus whole building energy model calibration method for office buildings using occupant behavior data mining and empirical data. *2014 ASHRAE/IBPSA-USA Building Simulation Conference*.

Lan, L., Wargocki, P., & Lian, Z. (2012). Optimal thermal environment improves performance of office work. *REHVA Journal (January 2012)*, 12-17.

Lehrer, D. (2006). LEED post-occupancy evaluation: Taking responsibility for the occupants. Retrieved from http://www.cbe.berkeley.edu/research/pdf_files/Lehrer2006_BetterBricksPOE.pdf

Lewis, J. O., Hogain, S. N., & Borghi, A. (2013). Building energy efficiency in European cities. Retrieved from http://urbact.eu/file/7005/download?token=HvW1MunV

Li, K., Hu, C., Liu, G., & Xue, W. (2015). Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. *Energy and Buildings*, *108*, 106–113. https://doi.org/10.1016/j.enbuild.2015.09.002

Li, Q., Meng, Q., Cai, J., Yoshino, H., & Mochida, A. (2009a). Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, *86*(10), 2249-2256.

Li, Q., Meng, Q., Cai, J., Yoshino, H., & Mochida, A. (2009b). Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. *Energy Conversion and Management*, *50*(1), 90-96.

Li, Q., Ren, P., & Meng, Q. (2010). Prediction model of annual energy consumption of residential buildings. *Advances in Energy Engineering (ICAEE), 2010 International Conference*.

Li, Z., & Huang, G. (2013). Re-evaluation of building cooling load prediction models for use in humid subtropical area. *Energy and Buildings*, *62*, 442–449. https://doi.org/10.1016/j.enbuild.2013.03.035

Li, N., Yang, Z., Becerik-Gerber, B., Tang, C., & Chen, N. (2015a). Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures? *Applied Energy*, *159*, 196–205. https://doi.org/10.1016/j.apenergy.2015.09.001

Li, X., Wen, J., & Bai, E.-W. (2015b). Building energy forecasting using system identification based on system characteristics test. In *2015 Workshop on Modeling and Simulation of*

Cyber-Physical Energy Systems (MSCPES) (pp. 1–6). IEEE. https://doi.org/10.1109/MSCPES.2015.7115401

Liu, D., & Chen, Q. (2013). Prediction of building lighting energy consumption based on support vector regression. *2013 9th Asian Control Conference (ASCC)*.

Magalhães, S. M. C., Leal, V. M. S., & Horta, I. M. (2017). Modelling the relationship between heating energy use and indoor temperatures in residential buildings through Artificial Neural Networks considering occupant behavior. *Energy and Buildings*, *151*, 332–343. https://doi.org/10.1016/j.enbuild.2017.06.076

Magnier, L., & Haghighat, F. (2010). Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network. *Building and Environment*, *45*(3), 739–746. https://doi.org/10.1016/j.buildenv.2009.08.016

Maiti, R. (2014). PMV model is insufficient to capture subjective thermal response from indians. *International Journal of Industrial Ergonomics*, *44*(3), 349-361.

Masoso, O. T., & Grobler, L. J. (2010). The dark side of occupants' behaviour on building energy use. *Energy and Buildings*, *42*(2), 173-177.

Massana, J., Pous, C., Burgas, L., Melendez, J., & Colomer, J. (2015). Short-term load forecasting in a non-residential building contrasting models and attributes. *Energy and Buildings*, *92*, 322-330.

Matsumoto, K., Doumpos, M., & Andriosopoulos, K. (2018). Historical energy security performance in EU countries. *Renewable and Sustainable Energy Reviews*, *82*, 1737–1748. https://doi.org/10.1016/j.rser.2017.06.058

Mena, R., Rodríguez, F., Castilla, M., & Arahal, M. R. (2014). A prediction model based on neural networks for the energy consumption of a bioclimatic building. *Energy and Buildings, 82*, 142-155.

Meyer, S., & Rakotonirainy, A. (2003). A survey of research on context-aware homes. *Proceedings of the Australasian Information Security Workshop Conference on ACSW Frontiers*.

Miller, C., Nagy, Z., & Schlueter, A. (2018). A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings. *Renewable and Sustainable Energy Reviews*, *81*, 1365–1377. https://doi.org/10.1016/j.rser.2017.05.124

Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, *32*(9), 1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009

Mohandes, M. (2002). Support vector machines for short-term electrical load forecasting. *International Journal of Energy Research, 26*(4), 335-345.

Naganathan, H., Chong, W. O., & Chen, X. (2016). Building energy modeling (BEM) using clustering algorithms and semi-supervised machine learning approaches. *Automation in Construction*, *72*, 187–194. https://doi.org/10.1016/j.autcon.2016.08.002

National Climatic Data Center. (2015). Climate at a glance: Time series. Retrieved April 16, 2015 from http://ncdc.noaa.gov/cag/time-series

Neslin, S. A., Novak, T. P., Baker, K. R., & Hoffman, D. L. (2009). An optimal contact model for maximizing online panel response rates. *Management Science, 55*(5), 727-737.

Neto, A. H., & Fiorelli, F. A. S. (2008). Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings*, *40*(12), 2169–2176. https://doi.org/10.1016/j.enbuild.2008.06.013

New, J. R., Sanyal, J., Bhandari, M., & Shrestha, S. (2012). AUTOTUNE E+ BUILDING ENERGY MODELS. *Proceedings of SimBuild; 2012: SimBuild 2012*. Retrieved from http://ibpsa-usa.org/index.php/ibpusa/article/view/440/426

Newsham, G., Brand, J., Donnelly, C., Veitch, J., Aries, M., & Charles, K. (2009). Linking indoor environment conditions to job satisfaction: A field study. *Building Research & Information*, *37*(2), 129-147.

Nicol, J. F., & Humphreys, M. A. (2002). Adaptive thermal comfort and sustainable thermal standards for buildings. *Energy and Buildings*, *34*(6), 563-572.

O'Brien, W., & Gunay, H. B. (2015, December). Mitigating office performance uncertainty of occupant use of window blinds and lighting using robust design. In *Building Simulation*, *8*, 621-636.

Panapakidis, I., Alexiadis, M., & Papagiannis, G. (2015). Evaluation of the performance of clustering algorithms for a high voltage industrial consumer. *Engineering Applications of Artificial Intelligence*, *38*, 1–13. https://doi.org/10.1016/j.engappai.2014.10.013

Pan, E., Li, H., Song, L., & Han, Z. (2015). Kernel-based non-parametric clustering for load profiling of big smart meter data. In *2015 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 2251–2255). IEEE. https://doi.org/10.1109/WCNC.2015.7127817

Park, J. Y., Yang, X., Miller, C., Arjunan, P., & Nagy, Z. (2019). Apples or oranges? Identification of fundamental load shape profiles for benchmarking buildings using a large and diverse dataset. *Applied Energy*, *236*, 1280–1295. https://doi.org/10.1016/j.apenergy.2018.12.025

Parker, D., Mills, E., Rainer, L., Bourassa, N., & Homan, G. (2012). Accuracy of the Home Energy Saver Energy Calculation Methodology. In *ACEEE Summer Study on Energy Efficiency in Buildings* (pp. 206–222). Retrieved from http://www.fsec.ucf.edu/en/publications/pdf/FSEC-CR-1930-12.pdf

Pattipati, K., Kodali, A., Luo, J., Choi, K., Singh, S., Sankavaram, C., et al. (2008). An integrated diagnostic process for automotive systems.*132*, 191-218.

Paudel, S., Elmtiri, M., Kling, W. L., Corre, O. Le, & Lacarrière, B. (2014). Pseudo dynamic transitional modeling of building heating energy demand using artificial neural network. *Energy and Buildings*, *70*, 81–93. https://doi.org/10.1016/j.enbuild.2013.11.051

Paudel, S., Nguyen, P. H., Kling, W. L., Elmitri, M., Lacarriere, B., & Corre, O. L. (2015). Support vector machine in prediction of building energy demand using pseudo dynamic approach. *The 28th International Conference On Efficiency, Cost, Optimization, Simulation And Environmental Impact Of Energy Systems.*

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences Discussions Discussions*, *4*(2), 439-473.

Pennsylvania State Climatologist. (2017). Pennsylvania State Climatologist. Retrieved from http://climate.psu.edu

Penya, Y. K., Borges, C. E., & Fernandez, I. (2011a). Short-term load forecasting in non-residential buildings. *Africon*.

Penya, Y. K., Borges, C. E., Agote, D., & Fernandez, I. (2011b). Short-term load forecasting in air-conditioned non-residential buildings. *Industrial Electronics (ISIE)*.

Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. *Energy and buildings*, *40*(3), 394-398.

Peschiera, G., Taylor, J. E., & Siegel, J. A. (2010). Response–relapse patterns of building occupant electricity consumption following exposure to personal, contextualized and occupant peer network utilization data. *Energy and Buildings*, *42*(8), 1329-1336.

Petersen, J. E., Shunturov, V., Janda, K., Platt, G., & Weinberger, K. (2007). Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives. *International Journal of Sustainability in Higher Education*, *8*(1), 16-33.

Pierce, J., Fan, C., Lomas, D., Marcu, G., & Paulos, E. (2010). Some consideration on the (in)effectiveness of residential energy feedback systems. *DIS 2010 - Proceedings of the 8th ACM Conference on Designing Interactive Systems*, pp. 244-247.

Platon, R., Dehkordi, V. R., & Martel, J. (2015). Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis. *Energy and Buildings*, *92*, 10–18. https://doi.org/10.1016/j.enbuild.2015.01.047

Popescu, D., & Ungureanu, F. (2013). Prediction of space heating consumption in district heated apartments. *ASME 2013 International Mechanical Engineering Congress and Exposition*.

Qualtrics. 2014. “ESOMAR 28 28 questions to help research buyers of online samples” *E-mail message to author,* September 30.

Ramos Ruiz, G., Fernández Bandera, C., Gómez-Acebo Temes, T., & Sánchez-Ostiz Gutierrez, A. (2016). Genetic algorithm for building envelope calibration. *Applied Energy*, *168*, 691–705. https://doi.org/10.1016/j.apenergy.2016.01.075

Rastogi, P., Khan, M. E., & Andersen, M. (2017). Gaussian-Process-Based Emulators for Building Performance Simulation. *Proceedings of BS 2017*. https://doi.org/10.26868/25222708.2017.448

Ratanamahatana, C. A., & Keogh, E. (2004). Everything you know about dynamic time warping is wrong. In *Third workshop on mining temporal and sequential data*

Reynolds, J., Rezgui, Y., Kwan, A., & Piriou, S. (2018). A zone-level, building energy optimisation combining an artificial neural network, a genetic algorithm, and model predictive control. *Energy*, *151*, 729–739. https://doi.org/10.1016/j.energy.2018.03.113

Roelofsen, P. (2018). Time series clustering. Vrije Universiteit Amsterdam. Retrieved from https://beta.vu.nl/nl/Images/stageverslag-roelofsen_tcm235-882304.pdf

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Rupp, R. F., & Ghisi, E. (2014). What is the most adequate method to assess thermal comfort in hybrid commercial buildings located in hot-humid summer climate? *Renewable and Sustainable Energy Reviews*, *29*, 449–462. https://doi.org/10.1016/j.rser.2013.08.102

Santin, O. G. (2013). Occupant behaviour in energy efficient dwellings: evidence of a rebound effect. *Journal of Housing and the Built Environment*, *28*(2), 311-327.

Schiavon, S., & Altomonte, S. (2014). Influence of factors unrelated to environmental quality on occupant satisfaction in LEED and non-LEED certified buildings. *Building and Environment, 77*(0), 148-159.

Schmidt, M., & Åhlund, C. (2018). Smart buildings as Cyber-Physical Systems: Data-driven predictive control strategies for energy efficiency. *Renewable and Sustainable Energy Reviews*, *90*, 742–756. https://doi.org/10.1016/j.rser.2018.04.013

Sengupta, M., Xie, Y., Lopez, A., Habte, A., Maclaurin, G., & Shelby, J. (2018). The National Solar Radiation Data Base (NSRDB). *Renewable and Sustainable Energy Reviews*, *89*, 51–60. https://doi.org/10.1016/j.rser.2018.03.003

Shahzadeh, A., Khosravi, A., & Nahavandi, S. (2015). Improving load forecast accuracy by clustering consumers using smart meter data. In *2015 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–7). IEEE. https://doi.org/10.1109/IJCNN.2015.7280393

Sharifzadeh, M. (2013). Does fitness and exercises increase productivity? assessing health, fitness and productivity relationship. *American Journal of Management Vol, 13*(1), 33.

Skirbekk, V. (2004). Age and individual productivity: A literature survey. *Vienna Yearbook of Population Research, 2*, 133-153.

Solomon, D. M., Winter, R. L., Boulanger, A. G., Anderson, R. N., & Wu, L. L. (2011). Forecasting energy demand in large commercial buildings using support vector machine regression. Retrieved from http://academiccommons.columbia.edu/catalog/ac%3A143154

Sun, K., & Hong, T. (2017). A simulation approach to estimate energy savings potential of occupant behavior measures. *Energy and Buildings*, *136*, 43–62. https://doi.org/10.1016/j.enbuild.2016.12.010

Tablada, A., De Troyer, F., Blocken, B., Carmeliet, J., & Verschure, H. (2009). On natural ventilation and thermal comfort in compact urban environments – the old havana case. *Building and Environment, 44*(9), 1943-1958.

Tavakol, M., & Dennick, R. (2011). Making sense of cronbach's alpha. *International Journal of Medical Education, 2,* 53.

Tsanas, A., & Xifara, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, *49*, 560–567. https://doi.org/10.1016/j.enbuild.2012.03.003

Tureczek, A. M., Nielsen, P. S., Madsen, H., & Brun, A. (2019). Clustering district heat exchange stations using smart meter consumption data. *Energy and Buildings*, *182*, 144–158. https://doi.org/10.1016/j.enbuild.2018.10.009

Turhan, C., Kazanasmaz, T., Uygun, I. E., Ekmen, K. E., & Akkurt, G. G. (2014). Comparative study of a building energy performance software (KEP-IYTE-ESS) and ANN-based building heat load estimation. *Energy and Buildings, 85*, 115-125.

van Dam, S. S., Bakker, C. A., & van Hal, J. D. M. (2010). Home energy monitors: Impact over the medium-term. *Building Research & Information, 38*(5), 458-469.

Vapnik, V. N. (1995). The nature of statistical learning theory. New York, NY, USA: Springer-Verlag New York, Inc.

Wang, L., & Greenberg, S. (2015). Window operation and impacts on building energy consumption. *Energy and Buildings*, *92*, 313–321. https://doi.org/10.1016/j.enbuild.2015.01.060

Wang, N., Makhmalbaf, A., Srivastava, V., & Hathaway, J. E. (2017). Simulation-based coefficients for adjusting climate impact on energy consumption of commercial buildings. *Building Simulation*, *10*(3), 309–322. https://doi.org/10.1007/s12273-016-0332-1

Wang, Z., & Srinivasan, R. S. (2015). A review of artificial intelligence based building energy prediction with a focus on ensemble prediction models. In *2015 Winter Simulation Conference (WSC)* (pp. 3438–3448). IEEE. https://doi.org/10.1109/WSC.2015.7408504

Wang, Z., & Srinivasan, R. S. (2017). A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable and Sustainable Energy Reviews*, *75*, 796–808. https://doi.org/10.1016/j.rser.2016.10.079

Wang, Z., Srinivasan, R. S., & Shi, J. (2016). Artificial Intelligent Models for Improved Prediction of Residential Space Heating. *Journal of Energy Engineering*, *142*(4), 04016006. https://doi.org/10.1061/(ASCE)EY.1943-7897.0000342

Wang, Z., Wang, Y., & Srinivasan, R. S. (2018a). A novel ensemble learning approach to support building energy use prediction. *Energy and Buildings*, *159*, 109–122. https://doi.org/10.1016/j.enbuild.2017.10.085

Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018b). Random Forest based hourly building energy prediction. *Energy and Buildings*, *171*, 11–25. https://doi.org/10.1016/j.enbuild.2018.04.008

Warren Liao, T. (2005). Clustering of time series data—a survey. *Pattern Recognition*, *38*(11), 1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025

WBCSD. (2007). Energy efficiency in buildings: Business realities and opportunities. Retrieved from http://www.wbcsd.org/DocRoot/1QaHhV1bw56la9U0Bgrt/EEB -Facts-and-trends.pdf

Wong, S. L., Wan, K. K. W., & Lam, T. N. T. (2010). Artificial neural networks for energy analysis of office buildings with daylighting. *Applied Energy*, *87*(2), 551–557. https://doi.org/10.1016/j.apenergy.2009.06.028

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., et al. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems, 14*(1), 1-37.

Xing-Ping, Z., & Rui, G. (2007). Electrical energy consumption forecasting based on cointegration and a support vector machine in china. *WSEAS Transactions on Mathematics*, *6*(12), 878-883.

Xu, W., Gu, R., Liu, Y., & Dai, Y. (2015). Forecasting energy consumption using a new GM–ARMA model based on HP filter: The case of guangdong province of china. *Economic Modelling*, *45*, 127-135.

Xuemei, L., Jin-hu, L., Lixing, D., Gang, X., & Jibin, L. (2009). Building cooling load forecasting model based on LS-SVM. *Information Processing, 2009. APCIP 2009. Asia-Pacific Conference*.

Xuemei, L., Yuyan, D., Lixing, D., & Liangzhong, J. (2010a). Building cooling load forecasting using fuzzy support vector machine and fuzzy c-mean clustering. *Computer and Communication Technologies in Agriculture Engineering (CCTAE), 2010 International Conference*.

Xuemei, L., Lixing, D., Jinhu, L., Gang, X., & Jibin, L. (2010b). A novel hybrid approach of KPCA and SVM for building cooling load prediction. *Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference*.

Yan, D., O'Brien, W., Hong, T., Feng, X., Burak Gunay, H., Tahmasebi, F., & Mahdavi, A. (2015). Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and Buildings*, *107*, 264–278. https://doi.org/10.1016/j.enbuild.2015.08.032

Yang, J., Rivard, H., & Zmeureanu, R. (2005). On-line building energy prediction using adaptive artificial neural networks. *Energy and Buildings*, *37*(12), 1250-1259.

Yang, L., Yan, H., & Lam, J. C. (2014). Thermal comfort and building energy consumption implications – A review. *Applied Energy*, *115*, 164–173. https://doi.org/10.1016/j.apenergy.2013.10.062

Yang, W., Wang, K., & Zuo, W. (2012). Neighborhood component feature selection for high-dimensional data. *Journal of Computers*, *7*(1), 162–168. https://doi.org/10.4304/jcp.7.1.161-168

Yezioro, A., Dong, B., & Leite, F. (2008). An applied artificial intelligence approach towards assessing building performance simulation tools. *Energy and Buildings*, *40*(4), 612-620.

Yi, W., & Ying, L. (2010, November). Applying LS-SVM to predict primary energy consumption. *E-Product E-Service and E-Entertainment (ICEEE), 2010 International Conference*.

Yokoyama, R., Wakui, T., & Satake, R. (2009). Prediction of energy demands using neural network with model identification by global optimization. *Energy Conversion and Management*, *50*(2), 319-327.

Yousefi, F., Gholipour, Y., & Yan, W. (2017). A study of the impact of occupant behaviors on energy performance of building envelopes using occupants' data. *Energy and Buildings*, *148*, 182–198. https://doi.org/10.1016/j.enbuild.2017.04.085

Yu, J., Lee, B., & Park, D. (2014). Real-time cooling load forecasting using a hierarchical multi-class SVDD. *Multimedia Tools and Applications, 71*(1), 293-307.

Yu, Z., Fung, B. C. M., Haghighat, F., Yoshino, H., & Morofsky, E. (2011). A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy and Buildings*, *43*(6), 1409-1417.

Yu, Z., Haghighat, F., Fung, B. C. M., & Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings*, *42*(10), 1637–1646. https://doi.org/10.1016/j.enbuild.2010.04.006

Yun, G. Y., Kong, H. J., Kim, H., & Kim, J. T. (2012). A field survey of visual comfort and lighting energy consumption in open plan offices. *Energy and Buildings, 46*, 146-151.

Yusoff, Y., Ngadiman, M. S., & Zain, A. M. (2011). Overview of NSGA-II for optimizing machining process parameters. In *Procedia Engineering* (Vol. 15, pp. 3978–3983). https://doi.org/10.1016/j.proeng.2011.08.745

Zalejska-Jonsson, A., & Wilhelmsson, M. (2013). Impact of perceived indoor environment quality on overall satisfaction in swedish dwellings. *Building and Environment*, *63*(0), 134-144.

Zhang, H., Arens, E., Fard, S. A., Huizenga, C., Paliaga, G., Brager, G., & Zagreus, L. (2007). Air movement preferences observed in office buildings. *International Journal of Biometeorology*, *51*(5), 349-360.

Zhang, Y. M., & Qi, W. G. (2009). Interval forecasting for heating load using support vector regression and error correcting Markov chains. *Machine Learning and Cybernetics, 2009 International Conference*.

Zhang, Y., O'Neill, Z., Dong, B., & Augenbroe, G. (2015). Comparisons of inverse modeling approaches for predicting building energy performance. *Building and Environment, 86,* 177–190. https://doi.org/10.1016/j.buildenv.2014.12.023

Zhao, H. (2011). Artificial Intelligence Models for Large Scale Buildings Energy Consumption Analysis (Doctoral dissertation, Ecole Centrale Paris).

Zhao, H. X., & Magoules, F. (2009). A parallel statistical learning approach to the prediction of building energy consumption based on large datasets. *6th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES)*.

Zhao, H., & Magoulès, F. (2010). Parallel support vector machines applied to the prediction of multiple buildings energy consumption. *Journal of Algorithms & Computational Technology, 4*(2), 231-249.

Zhao, H., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, *16*(6), 3586-3592.

Zhou, X., Yan, D., Hong, T., & Ren, X. (2015). Data analysis and stochastic modeling of lighting energy use in large office buildings in China. *Energy and Buildings*, *86*, 275-287.