EXPLORATION OF INTELLIGENT HVAC OPERATION STRATEGIES FOR OFFICE BUILDINGS

by

Xiaoqi Liu

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Lyles School of Civil Engineering

West Lafayette, Indiana

December 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Panagiota Karava, Chair

Lyles School of Civil Engineering

Dr. Ilias Bilionis

School of Mechanical Engineering

Dr. Athanasios Tzempelikos

Lyles School of Civil Engineering

Dr. Jianghai Hu

School of Electrical and Computer Engineering

Approved by:

Dr. Dulcy M. Abraham

ACKNOWLEDGMENTS

I would like to express my gratitude to the people who have significantly helped me throughout my Ph.D. study.

My advisor Professor Panagiota Karava, with her insight and deep knowledge into the subject matter, steered me through this research and inspired me to advance this field. To address many of the challenges I encountered throughout this path, Professor Ilias Bilionis helped me build a strong background in the field of machine learning and data science, which made my research possible. Also, I want to thank Professors Jianghai Hu and Thanos Tzempelikos, who have provided invaluable inputs in the field of optimal control and building science, respectively.

My research would not have been possible without the significant contribution of the National Science Foundation and the Center for High Performance Buildings at Purdue University, which funded different aspects of my study.

During my years as a graduate student, my family, friends, and colleagues at Purdue and all over the world always supported me and encouraged me to keep going. A Ph.D. degree involves successes and failures, excitement, and frustration, and in that sense, their invaluable support helped me be resilient and determined.

Last but not least, I want to thank my husband Iason Konstantzos, who has always been by my side with his unparalleled support and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Commercial buildings not only have significant impacts on occupants' well-being, but also contribute to more than 19% of the total energy consumption in the United States. Along with improvements in building equipment efficiency and utilization of renewable energy, there has been significant focus on the development of advanced heating, ventilation, and air conditioning (HVAC) system controllers that incorporate predictions (e.g., occupancy patterns, weather forecasts) and current state information to execute optimization-based strategies. For example, model predictive control (MPC) provides a systematic implementation option using a system model and an optimization algorithm to adjust the control setpoints dynamically. This approach automatically satisfies component and operation constraints related to building dynamics, HVAC equipment, etc. However, the wide adaptation of advanced controls still faces several practical challenges: such approaches involve significant engineering effort and require site-specific solutions for complex problems that need to consider uncertain weather forecast and engaging the building occupants. This thesis explores smart building operation strategies to resolve such issues from the following three aspects.

First, the thesis explores a stochastic model predictive control (SMPC) method for the optimal utilization of solar energy in buildings with integrated solar systems. This approach considers the uncertainty in solar irradiance forecast over a prediction horizon, using a new probabilistic time series autoregressive model, calibrated on the sky-cover forecast from a weather service provider. In the optimal control formulation, we model the effect of solar irradiance as non-Gaussian stochastic disturbance affecting the cost and constraints, and the nonconvex cost function is an expectation over the stochastic process. To solve this optimization problem, we introduce a new approximate dynamic programming methodology that represents the optimal cost-to-go functions using Gaussian process, and achieves good solution quality. We use an emulator to evaluate the closed-loop operation of a building-integrated system with a solar-assisted heat pump coupled with radiant floor heating. For the system and climate considered, the SMPC saves up to 44% of the electricity consumption for heating in a winter month, compared to a well-tuned rule-based controller, and it is robust, imposing less uncertainty on thermal comfort violation.

Second, this thesis explores user-interactive thermal environment control systems that aim to increase energy efficiency and occupant satisfaction in office buildings. Towards this goal, we

present a new modeling approach of occupant interactions with a temperature control and energy use interface based on utility theory that reveals causal effects in the human decision-making process. The model is a utility function that quantifies occupants' preference over temperature setpoints incorporating their comfort and energy use considerations. We demonstrate our approach by implementing the user-interactive system in actual office spaces with an energy efficient model predictive HVAC controller. The results show that with the developed interactive system occupants achieved the same level of overall satisfaction with selected setpoints that are closer to temperatures determined by the MPC strategy to reduce energy use. Also, occupants often accept the default MPC setpoints when a significant improvement in the thermal environment conditions is not needed to satisfy their preference. Our results show that the occupants' overrides can contribute up to 55% of the HVAC energy consumption on average with MPC. The prototype user-interactive system recovered 36% of this additional energy consumption while achieving the same overall occupant satisfaction level. Based on these findings, we propose that the utility model can become a generalized approach to evaluate the design of similar user-interactive systems for different office layouts and building operation scenarios.

Finally, this thesis presents an approach based on meta-reinforcement learning (Meta-RL) that enables autonomous optimal building controls with minimum engineering effort. In reinforcement learning (RL), the controller acts as an agent that executes control actions in response to the real-time building system status and exogenous disturbances according to a policy. The agent has the ability to update the policy towards improving the energy efficiency and occupant satisfaction based on the previously achieved control performance. In order to ensure satisfactory performance upon deployment to a target building, the agent is trained using the Meta-RL algorithm beforehand with a model universe obtained from available building information, which is a probability measure over the possible building dynamical models. Starting from what is learned in the training process, the agent then fine-tunes the policy to adapt to the target building based on-site observations. The control performance and adaptability of the Meta-RL agent is evaluated using an emulator of a private office space over 3 summer months. For the system and climate under consideration, the Meta-RL agent can successfully maintain the indoor air temperature within the first week, and result in only 16% higher energy consumption in the 3rd month than MPC, which serves as the theoretical upper performance bound. It also significantly outperforms the agents trained with conventional RL approach.

1. INTRODUCTION

1.1 Background and motivation

Commercial buildings have significant impacts on humans and the environment. Not only do they affect occupants' comfort, health, and well-being, but they are also responsible for more than 19% of the total energy consumption in the US. Heating, Ventilation, and Air Conditioning (HVAC) systems account for 28% of energy consumption and 45% of peak electrical demand in commercial buildings and represent a substantial energy use reduction opportunity (EIA, 2019). Along with improvements in building equipment efficiency and utilization of renewable energy, deployment of sensors, actuators, and controllers, can achieve more than 30% aggregated annual energy savings (Fernandez *et al*., 2017), while 20% of commercial buildings peak load can be temporarily managed or curtailed to provide grid services (Kiliccote *et al*., 2016; Piette *et al*., 2007). Due to the promising results, there has been significant focus on the development of advanced HVAC controllers that incorporate predictions (e.g., occupancy patterns, weather forecasts) and current state information to execute optimization-based strategies. Such control methods are capable of planning building system operation over extended periods (e.g., hours and days rather than minutes) and multiple spatial scales (e.g., occupant, zone, whole-building, campus) (Braun, 1990; Bengea *et al*., 2012; Ma *et al*., 2012; Dong and Lam, 2014; Afram and Janabi-Sharifi, 2014; Tanner and Henze, 2014; Mirakhorli and Dong, 2016; Joe and Karava, 2019; Yang *et al*., 2020). Model Predictive Control (MPC) provides a systematic implementation option using a system model and an optimization algorithm to adjust the control setpoints dynamically. This control approach automatically satisfies component and operation constraints related to building dynamics, HVAC equipment, etc. (Garcia *et al*., 1989; Mayne *et al*., 2000; Oldewurtel *et al*., 2012). However, the wide adaptation of such control methods still faces several practical challenges.

The control objectives and constraints of MPC need to be customized for specific sites, considering the complex energy conversion schemes of the advanced building systems, which often leads to nonconvex optimal control problems that would impose challenges in finding high-quality control solutions (e.g., Kelman & Borrelli, 2011; Corbin *et al*., 2013; Candanedo & Athienitis, 2011; Li *et al*., 2015; Quintana & Kummert, 2015). In addition to that, for buildings with renewable energy systems whose performance also depend on stochastic environment

disturbances such as solar irradiance, the optimal utilization of renewable energy also requires control algorithms that make robust decisions under uncertain weather forecast (Petersen & Bundgaard, 2014; Garifi *et al*., 2018). Stochastic model predictive control (SMPC) has shown great potentials to address the latter (Oldewurtel *et al*., 2012; Ma *et al*., 2014). However, the challenge remains to extend the SMPC approach to efficiently solve nonconvex problems in order to be generalizable for optimizing the operation of building-integrated solar systems.

On the other hand, the success of many of these control strategies is heavily dependent on how occupants interact with the building (Schweiger *et al*., 2020). Occupant behaviors in high performance buildings may be affected by many factors including occupant comfort (or discomfort), social influences, or lack of knowledge surrounding building systems (Day *et al*., 2020). In that sense, it is essential to understand and possibly influence the way occupants interact with environment control systems when energy-efficient strategies such as MPC are implemented If users are neglected from building control systems, then energy use may increase if systems are overridden, or occupants may be less satisfied with their environment due to decreased thermal comfort. Alternatively, if occupants understand the building and feel that they are involved in environment control systems, then they may contribute to lower building energy use and they may increase their overall satisfaction with the work environment (Janda, 2011). This two-way communication between the occupants and thermal environment control systems can be enabled by user-interactive systems that transform building occupants into service users who participate, decide, provide, and receive feedback. With such systems in place, occupant satisfaction could be improved (Day and Heschong, 2016) and their behavior could be potentially influenced by implementing appropriate intervention techniques (Peschiera *et al*., 2010; Delmas and Kaiser, 2014; Xu *et al*., 2017; Li *et al*., 2019).

Another issue that prevents MPC from being widely adopted in building industry is the extensive engineering time and effort required to develop the control-oriented models that represent the building system dynamics (Henze, 2013; Cígler *et al*., 2013; Killian and Kozek, 2016; Li and Wen, 2014). For this reason, reinforcement learning (RL) has received attention due to its capacity to learn to improve control through interacting with the environment (by letting an agent execute control actions and receive feedback in terms of control performance and system states) without requiring a model (Vázquez-Canteli and Nagy, 2019; Wang and Hong, 2020). However, as conventional RL approach solely relies on learning from on-site data, and does not

takes advantage of physical knowledge of the building systems (e.g., construction and equipment specifications). As a result, the required time-consuming learning process can make the implementation of RL in buildings inefficient or even impractical (Liu and Henze, 2006a, Yang *et al*., 2015; Benedetti *et al*., 2016). Therefore, data-efficient RL algorithms that allow learning from existing building information need to be explored, while the recent theoretical advancements in the machine learning field towards this direction has made such options possible.

1.2 Objectives

The goal of this thesis is to explore intelligent operation schemes for smart buildings while addressing the following real-world adaptation challenges: 1) Uncertain weather forecast; 2) Engaging occupants to make informed decisions in their interactions with buildings; and 3) Achieving optimal controls without extensive engineering effort and cost. Towards this direction, the research is extended to the following specific objectives:

1. Develop a stochastic model predictive control framework that is robust to forecast uncertainty for optimal operation of buildings with integrated solar systems.
    i. Develop a computationally inexpensive solar irradiance forecast model that utilizes external weather forecast information and quantifies the prediction uncertainty.
    ii. Deploy the approximate dynamic programming (ADP) algorithm to effectively solve the optimal control problem at each prediction horizon, where the nonconvex cost function is an expectation over a stochastic process.
    iii. Examine the stochastic model predictive controller's performance in an emulator that represents the actual building-integrated solar system.
2. Develop a systematic approach to design interfaces of user-interactive systems that aim to increase energy efficiency and occupant satisfaction in office buildings.
    i. Understand and model the human decision-making process in their interactions with thermal environment control systems when energy efficiency strategies are implemented.
    ii. Conduct field experiments with human-subjects to reveal the causal effect of the factors (e.g., displayed energy use information, expected comfort level) involved in this process.

iii. Deploy a prototype user-interactive system with a novel web-interface in a building energy management system with a model-predictive controller and demonstrate its performance with regards to energy savings and occupant satisfaction.

3. Develop a meta-reinforcement learning approach to enable automated generation of optimal HVAC controls with minimum engineering effort.
    i. Identify a model universe, i.e. a probability measure over the possible building dynamical models, based on available building information to train the agent.
    ii. Evaluate the control performance and adaptability of the agent in a test-bed and compare the Meta-RL approach with conventional RL and MPC.

1.3 Document overview

Chapter 2 presents a state-of-the-art literature review on the building applications of model predictive control, reinforcement learning, and user-interactive systems.

Chapter 3 presents the stochastic model predictive control algorithm for the optimal operation of buildings with integrated solar systems under forecast uncertainty. The ADP algorithm that is used to solve the nonconvex stochastic optimal control problem, as well as the forecast model that quantifies solar irradiance uncertainty are discussed in detail.

Chapter 4 presents the prototype user-interactive system for private office thermal environment control. The field experiment to evaluate the impact of the energy use information on the occupants' thermostat setting behavior and energy saving potential is described. The occupants' decision-making model that reveals causal factors on the setpoint temperature selections considering their comfort and energy use is presented.

Chapter 5 presents the meta-reinforcement learning algorithm that allows the automated generation of optimal control policy based on available building information, minimizing the engineering effort and cost. The control performance and adaptability of such approach is presented and compared with conventional reinforcement learning and model predictive control in a test-bed office by emulation.

Chapter 6 includes potential extensions of this research and ideas for future work.

2. LITERATURE REVIEW

2.1 Model predictive control

Various studies from building research literature suggested that model predictive control (MPC) has shown great potential in energy saving and maintaining indoor thermal comfort, outperforming conventional control approaches such as rule-based control and night setback for building heating ventilation and air-conditioning (HVAC) systems (Oldewurtel *et al*., 2012; Široký *et al*., 2011; May-Ostendorp *et al*., 2011; Prívara *et al*., 2011; Hu and Karava, 2014). In MPC, an optimal control problem that minimizes an objective (e.g. energy consumption/cost, temperature bounds violation) over a prediction horizon is solved at the beginning of each control horizon given a building system dynamical model and updated future disturbance information. This results in a trajectory of optimal controls (e.g. heating/cooling rate) and states (e.g. temperatures) into the future satisfying the constraints on the equipment capacity, thermal comfort bounds, or any other given criteria (Oldewurtel *et al*., 2012).

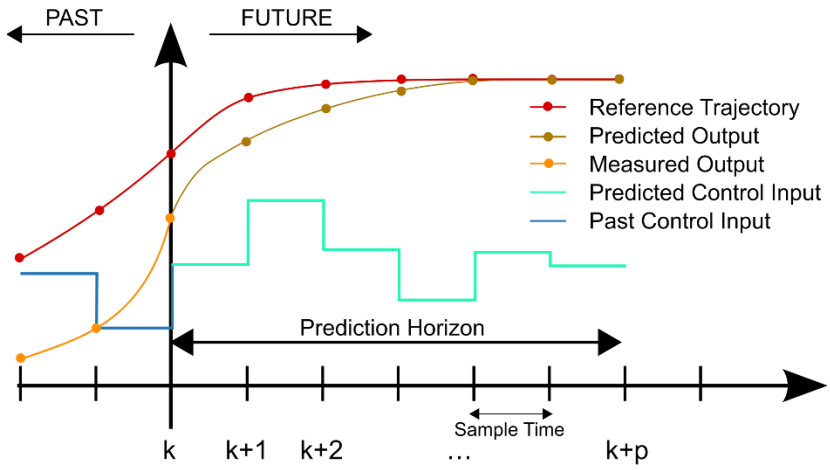


Figure 2.1. The 'moving horizon' approach of model predictive control (Wang, 2009).

The optimal control problems in MPC must include the key features of the building system, while being sufficiently simple to be computationally tractable. In some studies, the optimal control problems were formulated with linear (Oldewurtel *et al*., 2012; Zhang *et al*., 2013; Sturzenegger *et al*., 2014) or quadratic (Prívara *et al*., 2011, and Široky *et al*., 2011) cost functions, and linear models representing the building dynamics. In these cases, the problems could be solved

using linear or quadratic programming with guaranteed convergence to global minima. However, such simple problem formulations might not be applicable when the building energy systems include complex energy conversion schemes (Li *et al*. 2015; Kelman and Borrelli, 2011; Candanedo *et al*., 2013), or when other objectives or constraints are expressed by nonconvex functions (e.g. peak load, indoor comfort, uncertainty) (Corbin *et al*., 2013; Ma *et al*., 2012). For example, in the case of heat pumps, the coefficient of performance (COP) and capacity are multivariate polynomial functions of the source and load side temperatures (Verhelst *et al*., 2012; Gayeski *et al*., 2012), which can introduce nonconvexity in the cost and constraint functions.

Nonlinear programming solvers (Ma *et al*., 2012) and global search algorithms such as particle swarm (Corbin *et al*., 2013), pattern search (Li *et al*. 2015) and genetic algorithm (Wang and Jin, 2000) have been employed to solve the nonconvex optimal control problems, but only local optimality can be guaranteed. It is important that an MPC can provide high quality control solutions as it directly impacts energy savings, as well as improvements in thermal comfort. Attempts have been made to improve the solution quality of global search algorithms for nonconvex problems, such as tuning the hyperparameters by factorial experiment (Jaramillo *et al*., 2016), decision space discretization, seeding and taboo list (Corbin *et al*., 2013). However, these processes are time consuming and require system specific expertise that cannot be generalized for other applications.

Typically, the control-oriented models that represent the building system dynamics are physics-informed (i.e. grey-box, Braun, 1990; Oldewurtel *et al*., 2012; Cai, 2015; Cai and Braun, 2016; Joe and Karava, 2019; Andriamamonjy *et al*., 2019) or data-driven (i.e. black-box, Ferkl and Široký, 2010; Privara *et al*., 2013) linear models. Such models are developed through system identification experiments involving on-site collection of data including system temperatures, heat gains, and ambient environment conditions, etc. In recent studies, machine learning (ML)-based building modelling techniques are applied in MPC and can achieve adequate level of prediction accuracy without requiring domain-specific knowledge. However, a large amount of data from target buildings are still needed, and due to the complex model form such as deep neural networks, the optimal control problems solved at each prediction horizon are nonconvex, and only local optimality can be guaranteed (Ferreira *et al*., 2012; Aswani *et al*., 2012; Huang *et al*., 2014; Jain *et al*., 2017; Afram *et al*., 2017; Reynolds *et al*., 2018; Chen *et al*., 2018; Smarra *et al*., 2018; Bünning *et al*., 2020; Yang *et al*., 2020). ML-based MPC approaches also seek to improve the

controller adaptability by allowing continuous updates on the model with newly collected data (Yang *et al*., 2020; Rouchier *et al*., 2019). However, no study evaluated its adaptability yet.

Uncertainty can be introduced to MPC applications in buildings in different ways, such as uncontrolled disturbances, i.e. forecasted weather and random occupants' behavior, which can affect the cost and/or constraints; future states predicted by control-oriented models that are not perfectly accurate; and random sensing errors during real-time implementation. From the previous studies, we learnt that weather forecast (Henze *et al*., 2004) and occupants' behavior (Ma *et al*., 2012; Tanner, 2014; Oldewurtel *et al*., 2013) have significant impact on MPC performance. Therefore, uncertain disturbances forecast imposes challenges on MPC controller in predicting uncertain future states and selecting for the current time the optimal controls that minimize the cost and satisfy the constraints in the future.

Stochastic model predictive control (SMPC) allows disturbances in forms of probabilistic distribution and predicts the control strategies that minimize the expected cost and satisfy the constraints with a predefined probability (i.e. chance constraints, Charnes and Cooper, 1959; Oldewurtel *et al*., 2012; Zhang *et al*., 2013; Ma *et al*., 2014; Tanner, 2014). When the uncertain disturbances are assumed to follow uniform distribution or Gaussian distribution, chance constraints can be transformed to deterministic inequality constraints (Oldewurtel *et al*., 2012). However, in reality the probability distributions of the disturbances are often non-Gaussian. In those cases, sample-based approaches are employed to interpret chance constraints as deterministic for all the samples taken from the distributions of the disturbances, among which a selected number of constraint violations were allowed to happen (Zhang *et al*., 2013; Tanner and Henze, 2014; Ma *et al*., 2014).

2.1.1 Approximate dynamic programming

Optimal control problems at each prediction horizon of MPC can be solved using dynamic programming (DP) (Bellman, 1954), which is robust to the presence of non-Gaussian stochastic disturbances in the cost and constraints, and achieves good solution quality. DP is implemented in a receding horizon fashion and starts by estimating the optimal cost at the final time step and then moves backwards using the Jacobi-Bellman operator (Dobbs & Hencey, 2014; Dong & Lam, 2014; Lee *et al*., 2018b; Putta *et al*., 2015). In practice, DP requires solving recursively a series of optimization problems yielding the optimal cost at each time step (e.g., via the value iteration

algorithm). Also, instead of finding point estimates for optimal inputs, a major improvement by DP is finding the optimal inputs as a sequence of functions of system states (policy functions) over a prediction horizon. Dynamic programming can be solved in distributed way, where several processors participate simultaneously in the computation while maintaining coordination by information exchange via communication links (Bertsekas, 1995; Zhang *et al*., 2016). This approach requires less computation time for solving optimal control problems with high-dimensional state space, improving the efficiency, flexibility and scalability in the operation of large-scaled buildings or building clusters.

The technical difficulty of implementing DP arises from the need to approximate the optimal cost-to-go/policy functions based on a finite number of pairs of states and potentially noisy cost-to-go/optimal control observations. The optimal cost can be parameterized, e.g., using polynomials of a given degree, radial basis functions, neural networks (Bertsekas and Tsitkilis, 1995; Mnih *et al*., 2015). The choice of the approximating family is important as the limited expressivity of common function approximations may lead to suboptimal solutions. Previous research (Deisenroth *et al*., 2009; Scheidegger and Bilionis, 2019) employed approximation schemes for the optimal cost-to-go/control functions based on Gaussian process regression (GPR) (Rasmussen and Williams, 2006). GPR is a powerful Bayesian, non-parametric regression method robust to the presence of noise in the cost-to-go observations. In the proposed work, we employ approximate dynamic programming (ADP) to solve optimization problems, in which the nonconvex cost and constraint functions are subjected to stochastic disturbances, while using GPR to approximate the cost-to-go and policy functions.

2.1.2 Statistical weather forecast for predictive control

The operations of building energy systems are strongly impacted by the outdoor weather conditions. Therefore, especially for the case of predictive controls, a reliable weather forecast is essential for maintaining indoor thermal comfort and being energy efficient. Statistical forecast models have been widely used in building energy management applications. Typical weather forecast models predict the future weather based on simple historical patterns such as using the same data as the previous day, typical days of a month, etc. (Henze *et al*., 2004). However, such models do not capture nonlinear patterns such as the effect of cloud cover on solar irradiance (Lazos *et al*., 2014). On the contrary, machine learning models trained with historical weather data,

incorporate nonlinear patterns in weather variations and are better suited for predicting future weather (Dong and Lam, 2014; Lanza and Cosme, 2001).

Extracting information from past profiles works well for weather parameters with relatively small variation from one hour to the next, such as temperature and relative humidity. However, observable past patterns have limited influence on highly stochastic parameters such as solar irradiance (Mathiesen and Kleissl, 2011). In these cases, weather forecast services such as those from the National Oceanic and Atmospheric Administration (NOAA) that measure various meteorological parameters to generate predictions, can serve as baselines for predictive models (Pedersen and Petersen, 2017).

Previous studies developed weather forecast models that quantify the predictive uncertainty and take into account external weather forecast information or on-site measurements. Machine learning approaches such as Gaussian process regression (Zavala *et al*., 2009; Bilionis *et al*., 2014; Shann and Seuken, 2014), artificial neural networks (Chen *et al*., 2011; Yadav and Chandel, 2012) and support vector machines (Chakraborty *et al*., 2016) have shown promising results. However, implementation of such information in actual building controllers may require more straightforward approaches based on easily measurable and accessible data. Autoregressive models (Oldewurtel *et al*., 2012; Zhang *et al*., 2013) are computationally efficient and capture the physical nature of weather parameters (Lazos *et al*., 2015). In the proposed work, autoregressive process is utilized to model the cloud variability over time while the sky condition of clear, partly-cloudy, and overcast is classified using a probabilistic model based on the hourly-updated weather forecast.

2.2 User-interactive thermal environment control systems

In the literature, user-interactive systems refer to computer systems that support the interactions between humans and the computer, which occurs via the systems' user-interfaces (Preece *et al*., 1994). In recent years, such systems have been introduced to enable occupants' participation in building automation. For example, smart thermostats in residential buildings allow room setpoint scheduling based on occupancy, users' habit, or real-time utility rate, providing feedback on energy consumption/cost (Rau *et al*., 2016; Obinna *et al*., 2017). On the other hand, in commercial buildings, most of these systems are focused on receiving input from occupants (e.g., Daum *et al*., 2011; Erickson and Cerpa, 2012; Jazizadeh *et al*., 2014; Ghahramani *et al*.,

2014; West *et al*., 2014; Brager and Arens, 2016; Lee *et al*., 2019) rather than providing information. An exception is the study by Zeiler *et al*. (2009) which deployed a prototype user-interactive system with an indoor environment control interface that shares HVAC energy use information with occupants while collecting their feedback on thermal preferences. However, this study only focused on the development of hardware and software system architecture. Also, two pilot studies by Konstantakopoulos *et al*. (2015, 2019) implemented user-interactive systems with social games, in which occupants could vote for their desired lighting and HVAC setpoints, and get rewarded based on how energy efficient their voted strategies were. The real-time energy use data were accessible for the occupants from a web portal or mobile app. However, these studies focused on game formulation rather than the user-interface and energy feedback design.

2.2.1 User-interface design

Providing information (feedback) has long been regarded as a critical mechanism in motivating individual end-users to reduce their energy use voluntarily. While the vast majority of the studies focus on residential buildings (e.g., Siero *et al*., 1996; Emeakaroha *et al*., 2014; Vellei *et al*., 2016; Promann and Brunswicker, 2017), various disciplines, including human-computer interactions (HCI), ubiquitous computing (ubicomp), and environmental psychology, have contributed in-depth research on how feedback influences occupant interactions with commercial buildings regarding energy savings. From an environmental psychology perspective, a wide variety of behavioral intervention approaches have been used. These approaches range from education about energy use (Murtagh *et al*., 2013; Yun *et al*., 2013; Timm and Deal, 2016) to financial incentives (Konstantakopoulos *et al*., 2015), competitions (Ratliff *et al*., 2014; Gandhi and Brager, 2016), serious games (Orland *et al*., 2014), peer comparison (Peschiera *et al*., 2006; Zhang *et al*., 2013; Gulbinas *et al*., 2014; Gulbinas and Taylor, 2014), and engagement using social media (Lehrer *et al*., 2014). These studies primarily concentrated on the intervention's effect (e.g., the resulting energy savings) instead of the designed artifact. In contrast, the HCI/ubicomp studies focused on the design rather than conducting field studies of occupants' behavior (Froehlich, 2009; Froehlich *et al*., 2010; Karlin *et al*., 2017). Initial multidisciplinary work by Sanguinetti *et al*. (2018) proposed a design-behavior framework to guide the feedback design, highlighting the information provided to occupants, the timing of when this information is presented, and how the

information is displayed to the user, as critical components to consider during the design of building interfaces.

In this direction, drawing on insights from behavioral sciences and human-computer interface design, there has been an increasing emphasis on individual-level real-time (Gulbinas *et al*., 2014; Gulbinas and Taylor, 2014) and interactive (Zhuang and Wu, 2018) feedback tailored to specific behaviors. It is reported that occupants' daily interaction with office building thermostats is usually habit-driven rather than deliberate (Tetlow *et al*., 2015). Therefore, thermostat setting behavior can potentially be influenced through nudging, which is an intervention approach often adopted in the field of HCI (Caraban *et al*., 2019). Nudging means altering human behavior in a predictable way by subtly modifying the context of decision-making without forbidding any option or significantly changing their economic incentive (Thaler and Sunstein, 2009; Kasperbauer, 2017; Schweiger *et al*., 2020). One example of nudging is interactive feedback that provides the consequences of behavior (e.g., the potential increase of energy consumption) at the point of decision-making, that has been reported to encourage a more deliberate thermostat setting (Zhuang and Wu, 2018).

In summary, although the existing literature provides useful insights on the potential of energy use feedback, a systematic approach is needed to develop user-interactive systems that can be successfully deployed in smart building operation. Our goal in this paper is to fill this gap in knowledge by addressing the following objectives: i) identify the causal effect of the factors (e.g., displayed energy use information, expected comfort level) that describe the decision making process of occupant interactions with thermal environment control systems; ii) encode this knowledge in a human decision making model that can be used to design user-interactive systems that make energy-efficient behavior natural, easy, and intuitively understandable for the end-users resulting in HVAC energy savings and overall occupant satisfaction.

For modeling human decision-making, a classical decision theory that reveals the rationale behind human behavior is often adopted (Berger, 2013). This theory assumes that the criteria for choices among competing alternatives are based on user's (i.e., occupant's) preferences on the outcomes. The numerical representations of preferences are enabled by a utility function, which maps each choice to a scalar that quantifies the user's utility on the outcome of the choice, and decision-making can be realized as the maximization of the expected utility (Von Neumann and Morgenstern, 2007; Fishburn, 1970). Such approaches have been gaining attention in the HCI

community for user modeling (Payne and Howes, 2013; Jameson *et al*., 2014), and can be leveraged to design user-interactive systems for smart thermal environment control. For causal effects to be encoded in the model, control experiments, with and without implementing a treatment (energy use information), are needed (Rubin, 1974; Holland, 1986; Pearl, 2009).

2.3 Reinforcement learning

Reinforcement learning (RL) has been gaining attention as a promising approach in various applications (Silver *et al*., 2018; Levine *et al*., 2018), due to its capacity to learn through interacting with the environment without requiring an explicit mathematical model. Typically, in the formulation of an RL problem for building control, the building temperatures, energy system status, and exogenous variables (outdoor weather conditions, etc.) are treated as the states of the environment, and an RL agent learns by interacting with the environment. Such interaction includes the agent executing control actions (e.g. changing HVAC system heating/cooling rate or setpoint), which causes the transition of environmental states; then rewards are assigned to the agent based on the energy efficiency and/or occupants' satisfaction achieved by the control action. Based on the collected information on the past states, actions and rewards, the RL agent learns a policy that targets the maximization of the expected discounted sum of all future returns (Wang and Hong, 2020; Vázquez-Canteli and Nagy, 2019; Mason and Grijalva, 2019; Han *et al*., 2019). Recent research in RL indicated great potential of applicability in different levels of building systems, ranging from advanced energy systems (Yang *et al*., 2015; Lazic *et al*., 2018; Vázquez-Canteli *et al*., 2019a) or zone-level controls (Wang *et al*., 2017; Jia *et al*., 2019; Chen *et al*., 2019) in commercial buildings, to residential heat pumps (Ruelens *et al*., 2015; Peirelinck *et al*., 2018), holistic building systems control (adjusting HVAC, operable windows, ventilation, etc.) based on feedback of multiple indoor environment metrics (Chen *et al*., 2018; Ding *et al*., 2019; Park *et al*., 2019), and demand response in smart grid (Vázquez-Canteli *et al*., 2019b). Some important features in the mechanism of RL makes this approach appealing for use in building system controls: (i) Compared to MPC, it avoids the labor- and expertise-intensive process of developing (and customizing for each building) highly accurate control-oriented models, while achieving good control performance (Costanzo *et al*., 2016). (ii) The RL controller's adaptability to the environment could simplify the effort to maintain the controller once it is deployed in the buildings. Towards this direction, although more realistic scenarios need to be evaluated, the study by

Vázquez-Canteli *et al*. (2019a) initially demonstrated that RL controller is robust in terms of adapting to changes in electricity tariffs and building retrofits.

However, there are still significant barriers that prevent the wide adoption of RL controllers in real buildings. In the early stages of implementation, due to the lack of RL algorithms that can efficiently utilize historical data, the agent training process can be time consuming (multiple years), making it impractical to implement online (Liu and Henze, 2006a; Liu and Henze, 2006b; Dalamagkidis *et al*., 2007). Although the emergence of deep learning algorithms (Mnih *et al*., 2015; Mnih *et al*., 2016; Lillicrap *et al*., 2016; Schulman *et al*., 2017) helped reduce the required training time to multiple months, the stability of the control cannot be guaranteed when the training is still premature during these months(Zhang and Lam, 2018). For example, the agent's exploratory control actions might result in unnecessary energy waste or discomfort on occupants. To mitigate this issue, Liu and Henze (2006b) suggested to train the agent with an environment simulator before deploying it to actual buildings. Such approach has been adopted in majority of RL control studies in building applications, and different types of models such as grey-box (Lee *et al*., 2018a) and white-box (Wang *et al*., 2017; Jia *et al*., 2019) have been used as environment simulators. However, developing an environment simulator that can accurately predict the dynamical responses of the actual building would again require engineering effort and sufficient on-site measured operation data. On the other hand, although directly utilizing a standardized building model (e.g. generic grey-box model or reference buildings from EnergyPlus with empirically determined model parameters) for this purpose can eliminate such effort, the impact of the environment simulators' prediction quality on the agent's control performance in the actual building remains unstudied. Therefore, it is still unknown whether the adaptability of the agent can overcome the biasness induced by potentially inaccurate environment simulator within a reasonable amount of time, without causing occupant discomfort and excessive energy waste.

2.3.1 Meta-reinforcement learning

With the recent development of meta-reinforcement learning (Meta-RL, Finn, *et al*. 2017; Duan *et al*., 2017; Nichol *et al*., 2018; Sæmundsson *et al*., 2018; Xu *et al*., 2018; Rakelly *et al*., 2019; Kirsch *et al*., 2019), an RL agent is able to learn from a set of environments that share some common characteristics (sampled from the same prior probability distribution). Intuitively, if the agent can generalize well to such set of environments, it can be expected to perform well on another

environment that is sampled from the same prior probability distribution (Finn *et al*., 2017). So far, Meta-RL has been successfully demonstrated in the aforementioned studies to accelerate the learning and adaptation compared to conventional RL techniques by optimizing the initial parameters of the control policy executed by agent in gaming and robotics environments available at OpenAI gym (Brockman *et al*., 2016). In building control applications, although it is impractical to develop an environment simulator with high level of prediction quality, it would be feasible to identify the prior probability distribution of the environment (i.e., model universe) based on existing knowledge of the buildings available on-site (e.g., construction drawings, building information model, etc.) or from various public datasets (Deru *et al*., 2011; EIA, 2016; Miller and Meggers, 2017; Balaji *et al*., 2018; Miller *et al*., 2020). Therefore, with Meta-RL, learning over the model universe that describes similar building spaces can potentially improve the control quality, and achieve fast adaptation to a target building without developing an accurate simulator.

2.4 Research gaps

Based on the literature review from the previous sections, the following research gaps are identified and addressed by the work of this dissertation. First, in order to be generalizable for optimizing the building-integrated solar systems' operation, SMPC needs to (i) adopt ADP algorithm to efficiently solve nonconvex optimal control problems introduced by the complex energy conversion schemes of energy systems; and (ii) include a computationally inexpensive solar irradiance forecast model for predictive controls that quantifies the prediction uncertainty.

Second, user-interactive systems for commercial buildings need interface design that can (i) be intuitive for occupants to achieve energy saving; and (ii) incorporate behavior intervention approach that is easy to implement such as nudging. Also, a systematic approach supported by utility theory is needed to evaluate the interface design in terms of its predictable effect on occupant behavior and energy saving potential. To achieve that, the causal factors that affect occupants' decision-making process using the interface must be understood.

Third, Meta-RL algorithm needs to be evaluated due to its potential of enabling automated generation of optimal building control strategies with minimal engineering effort by learning from a model universe. To this end, identifying the model universe based on available building information such as building drawings also needs to be explored.

3. MODEL PREDICTIVE CONTROL UNDER FORECAST UNCERTAINTY FOR OPTIMAL OPERATION OF BUILDINGS WITH INTEGRATED SOLAR SYSTEMS

3.1 Overview

In this Chapter, we present a SMPC algorithm for buildings with solar systems coupled with HVAC and thermal energy storage. The algorithm is implemented in an emulator to predict the closed-loop response of the integrated system to control inputs, and to demonstrate optimal decisions under uncertainty. Our approach is unique in the following aspects (i) It quantifies the prediction uncertainty in solar irradiance using a new probabilistic time-series autoregressive model that takes sky-cover values from an external weather forecast service provider as input; (ii) It extends approximate dynamic programming (ADP) to solve optimization problems, in which the nonconvex cost function is an expectation over a stochastic process, and provides good solution quality using Gaussian process regression to approximate the cost-to-go functions.

We introduce the SMPC algorithm in Section 3.2, which includes the SMPC problem formulation, the solar irradiance forecast model and the ADP that solves the optimal control problem. In Section 3.3, we present the implementation of the SMPC for a building-integrated solar energy system. The performance evaluation of the SMPC and the uncertainty analysis are presented in Section 3.4.

3.2 Methodology

3.2.1 Stochastic model predictive control algorithm

Model predictive control for building energy systems aims to minimize the total heating or cooling energy consumption over a prediction horizon, while satisfying the constraints on equipment capacity and room conditions affecting occupant thermal comfort.

For each prediction horizon, the controller solves the following optimal control problem:

$$\min_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\ldots,\boldsymbol{\mu}_t} \mathbb{E}\left[\sum_{t=0}^{K-1} J_t(\mathbf{x}_t,\ \boldsymbol{\mu}_t(\mathbf{x}_t), \mathbf{w}_t)\right], \tag{3-1}$$

subject to the dynamics:

$$\mathbf{x}_{t+1} = \mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t), \tag{3-2}$$

and to the chance constraints:

$$\mathbb{E}\left[g_{i,t}(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t)\right] \geq 0 \text{ for } 1 \leq i \leq n_c, 0 \leq t \leq K-1. \tag{3-3}$$

The cost (equation (3-1)) is a function of the *state variables* ($\mathbf{x}_t$), which include the system temperatures at time t; the *control inputs* (given as $\mathbf{u}_t = \boldsymbol{\mu}_t(\mathbf{x}_t)$ at time t) such as heating or cooling power; and the *stochastic disturbances* (uncertain variables, $\mathbf{w}_t$), e.g. random factors that affect the solar irradiance that is an important energy source of building-integrated solar systems. The optimal input functions are known as *policy functions* ($\boldsymbol{\mu}_t$). The building and energy system dynamics are also functions of these three types of variables (equation (3-2)). The prediction horizon (K) over which optimal inputs are to be found via SMPC is selected based on the system properties.

In the presence of stochastic disturbances, chance constraints are deployed for handling constraints violations on the state variables (equation (3-3)) with low probability. In the cases of building-integrated solar systems, such as solar-assisted heat pumps, the feasible set of inputs is also expressed with chance constraints, as equipment capacities are subjected to stochastic disturbances. Therefore, n_c is the total number of chance constraints on the states or control inputs. Note that in equations (3-1) and (3-3), the expectations $\mathbb{E}[\cdot]$ are over $\mathbf{w}_t$. In our work, the stochastic disturbance, $\mathbf{w}_t$, to the integrated energy system includes the solar heat gain to the building and a 2-D Gaussian noise term from a solar irradiance forecast model, which is presented in the following section.

3.2.2 Solar irradiance forecast model

In this section, we present a model that predicts the global horizontal irradiance ($I_{g,t}$):

$$I_{g,t} = I_{\text{dir},t} + I_{\text{dif},t}, \tag{3-4}$$

where $I_{\text{dir},t}$ and $I_{\text{dif},t}$ are direct and diffuse components of the horizontal irradiance, respectively. From previous studies (e.g. Bilionis *et al.*, 2014), we know that the global horizontal irradiance is negatively correlated with the sky-cover, i.e., the fraction of the sky covered by clouds. Let the sky-cover at time t be defined as sc_t. Hourly forecasted values of sky-cover are obtained from NOAA. Our model predicts the future global horizontal irradiance given a sky-cover forecast. It is

stochastic in the sense that it quantifies the predictive uncertainty in solar irradiance. This uncertainty may be due to miss-classification of the sky condition (clear, partly-cloudy or overcast) given the sky-cover, and due to the variability of the cloud formations.

We assume that the sky-cover forecast values obtained from NOAA forecast are accurate for a 24-hour horizon. We model the cloud variability by a 2-D autoregressive process. Specifically, the first dimension of this autoregressive process models the atmospheric factors that block the direct part of the horizontal irradiance while the second-dimension models the atmospheric factors that block the diffuse part. In the autoregressive process, the future state is expressed as a linear function of the current state plus a Gaussian error term. Thus, the error can accumulate as the time step increases, and the autoregressive process captures the increase in solar irradiance uncertainty within a prediction horizon.

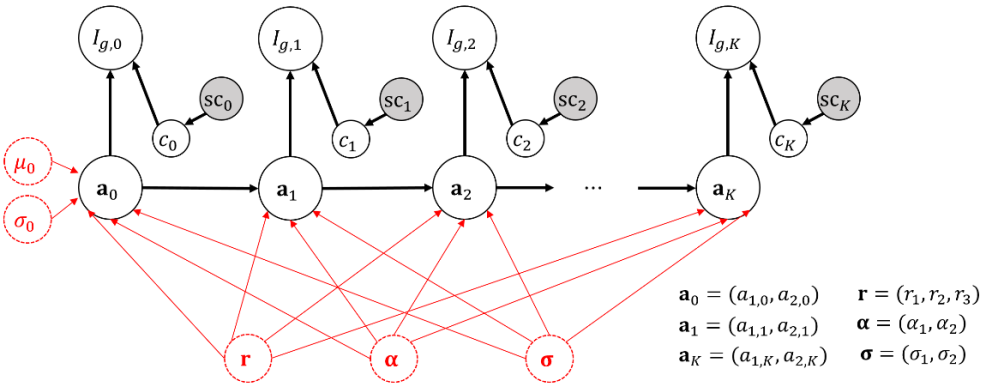


Figure 3.1. Solar irradiance forecast model.

The model we propose (Figure 3.1) has the following form:

$$I_{g,t}(\mathbf{a}_t, c_t) = I_{\text{clr,dir},t} \cdot \left(\text{sigm}\left(a_{1,t}\right)\right)^{1-l_1(c_t)} 1_{\{1,2\}}(c_t) + I_{\text{clr,dif},t} \cdot \left(\text{sigm}\left(a_{2,t}\right)\right)^{1-l_2(c_t)}, \qquad (3\text{-}5)$$

where $I_{\text{clr,dir},t}$ and $I_{\text{clr,dif},t}$ are the clear-sky direct and diffuse horizontal irradiance, which are analytically known and depend on the location, time of the day, day of the year, etc. (Bird and Hulstrom, 1981); $\mathbf{a}_t = \left(a_{1,t}, a_{2,t}\right)$ is a 2-D autoregressive process that modulates the clear-sky irradiances through the sigmoid function $\text{sigm}(\cdot)$; c_t is a discrete random process corresponding to the sky condition (clear (c_t =1), partly-cloudy (c_t =2), or overcast (c_t =3)), $1_A(\cdot)$ is the characteristic function of a set A ($1_A(x) = 1$ if $x \in A$ and 0 otherwise), and the indicator function

$\mathbf{l}(\cdot) = \big(l_1(\cdot), l_2(\cdot)\big)$ is defined by $\mathbf{l}(1) = (1,1)$, $\mathbf{l}(2) = (0,0)$, and $\mathbf{l}(3) = (1,0)$. More specifically, when the sky condition is clear ($c_t = 1$), the direct and diffuse horizontal solar irradiances are equal to clear-sky direct and diffuse horizontal solar irradiance, respectively. When the sky condition is partly-cloudy ($c_t = 2$), the direct and diffuse horizontal irradiances are only a fraction of the clear-sky direct and diffuse horizontal irradiance. When the sky condition is overcast ($c_t = 3$), the direct horizontal irradiance ($I_{\mathrm{dir},t}$) is 0. Thus, the global horizontal irradiance is only a fraction of the clear-sky diffuse horizontal irradiance.

For a certain value of the sky-cover sc_t, any of the three conditions, $c_t = 1,2,$ or 3, are possible as the sky-cover value is a spatial average of the fraction of the sky covered by clouds. To model this uncertainty, we assume that the probability of the sky condition c_t depends only on the sky-cover sc_t via a logistic regression expression:

$$p(c_t = i|\mathrm{sc}_t) = \frac{e^{r_i \mathrm{sc}_t}}{e^{r_1 \mathrm{sc}_t} + e^{r_2 \mathrm{sc}_t} + e^{r_3 \mathrm{sc}_t}}, \quad i = 1,2,3. \tag{3-6}$$

where the parameters r_1, r_2 and r_3 are to be inferred. The latent 2-D autoregressive process ($\mathbf{a}_t$) is given by

$$a_{1,t+1} = \alpha_1 a_{1,t} + \sigma_1 z_{1,t}, \tag{3-7}$$

$$a_{2,t+1} = \alpha_2 a_{2,t} + \sigma_2 z_{2,t}, \tag{3-8}$$

where $\mathbf{z}_t = (z_{1,t}, z_{2,t})$ is a 2-D Gaussian noise, and the parameters $\alpha_1, \alpha_2, \sigma_1$ and σ_2 are to be inferred. The initial probability distribution of the autoregressive process is:

$$p(a_{0,1}) = \mathcal{N}(a_{0,1}|\mu_0, \sigma_0^2), \tag{3-9}$$

$$p(a_{0,2}) = \mathcal{N}(a_{0,2}|\mu_0, \sigma_0^2), \tag{3-10}$$

where $\mathcal{N}(\cdot|\mu, \sigma^2)$ is the probability distribution function (PDF) of a univariate Gaussian distribution with mean μ and standard deviation σ; μ_0 and σ_0 are two additional parameters, which also need to be inferred from the data.

To train our model, we use sky-cover (sc_t, input) and global horizontal irradiance ($I_{g,t}$, output) measurements from the typical meteorological year (TMY3) dataset (Wilcox and Marion, 2008) at the location of interest (West Lafayette, IN). The unknown parameters to be estimated are $\boldsymbol{\theta} = (\mu_0, \sigma_0, \alpha_1, \alpha_2, \sigma_1, \sigma_2, r_1, r_2, r_3)$. Our model is a non-linear and non-Gaussian state space model (SSM). The hidden state is $\mathbf{s}_t = (c_t, a_{1,t}, a_{2,t})$ and the observed state is $I_{g,t}$. One of the key

challenges in estimation of parameters of a nonlinear and non-Gaussian SSM is the intractability of estimating the system state. Sequential Monte Carlo (SMC) methods (Gordon *et al.,* 1993; Kitagawa, 1996), known as particle filters, provide a robust solution to the nonlinear system identification problem (Schön *et al.*, 2015). The key challenge that drives the problem of parameter estimation is how to deal with the difficulty that the states are unknown (hidden). To handle this problem, we make use of the data augmentation strategy, which treats the states as auxiliary variables that are estimated together with the parameters. The expectation maximization (EM) algorithm solves the maximum likelihood formulation in this way (Dempster *et al.,* 1977). The maximum likelihood formulation amounts to finding a point estimate of the unknown parameters $\boldsymbol{\theta}$, for which the observed data is as likely as possible. This is done by maximizing the data likelihood function according to:

$$\widehat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}\,\epsilon\,\boldsymbol{\Theta}}\; p_{\theta}\left(I_{g,1:T}\right). \tag{3-11}$$

We wish to find the unknown parameters values $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ based on a batch of T measurements. For more details on how to use the EM algorithm for parameter estimation in SMC models, we refer readers to the paper by Schön *et al*. (2015). After maximizing the likelihood (Dahlin *et al*., 2015), we get point estimates of parameters $\boldsymbol{\theta}$ which best fits the observed data and are given as: = (0.2, 0.15, 0.1, 0.1, 0.6, 0.1, -7, 0, 7).

We present the evaluation of our model predictions on a validation dataset for three days (March 13th -15th, 2015). The dataset is considered representative as it contains a full range of values (0-100%) for sky-cover, which is the sole input to our model. Therefore, we can observe the predictive distribution of global horizontal irradiances given different sky-cover conditions from the validation dataset. As for factors such as seasonal variation, location, etc., we consider them to be encoded in the clear sky irradiance model by Bird and Hulstrom (1981). Figure 3.2 (left) shows the sky-cover forecast values for the aforementioned days obtained at 6 am of each day, respectively; the x-axis index points represent hours with high sky-cover forecast values (i.e., high probability of being overcast) on March 13th, moderate sky-cover forecast values (i.e., high probability of being partly-cloudy) on March 14th and low sky-cover forecast values (i.e., high probability of being clear) on March 15th. Figure 3.2 (right) compares model predictions with global solar irradiance values measured on the roof of the Herrick Laboratory building at Purdue University campus with a pyranometer (LI-COR LI-200). In this figure, the solid (blue) line represents a random solar irradiance time series sample from the predictive distribution of our

model. The model quantifies the uncertainty associated with predictions, represented in the graph by the shaded (blue) area (5-95th sample percentile from 1000 samples). The uncertainty of our model prediction shown in Figure 3.2 (right) is (i) low when the sky-cover forecast indicates clear (low sky-cover values, close to 0) or overcast (high sky-cover values, close to 100%) conditions; and (ii) high for partly-cloudy conditions (moderate sky-cover values, 30-70% in Figure 3.2, left).

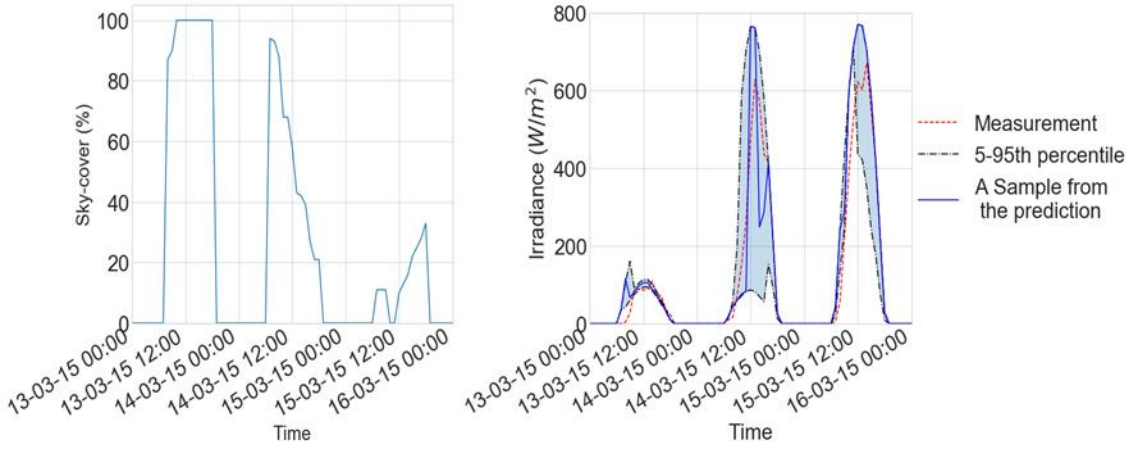


Figure 3.2. Sky-cover forecast obtained at 6 am of each day (left) and global horizontal irradiance prediction samples (right) (March 13-15th, 2015).

3.2.3 Approximate dynamic programming

In this section, we present the approximate dynamic programming methodology for the solution of the optimal control problem described with equations (3-1) to (3-3).

Consider a system with state space variables $\mathbf{x} \in \mathbf{X} \subset \mathbb{R}^{d_x}$ and dynamics given by equation (3-2), to which a set of feasible controls $\mathbf{u} \in \mathcal{U}(\mathbf{x})$ is applied. This set of admissible controls $\mathcal{U}(\mathbf{x})$ is described by equation (3-3). Let us define Π as the set of all admissible policies, i.e.,

$$\Pi = \{\pi = (\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_K) | \ \boldsymbol{\mu}_t : \mathbf{X} \to \mathbb{R}^{d_u} \text{ s.t. } \boldsymbol{\mu}_t(\mathbf{x}) \in \mathcal{U}(\mathbf{x}), \forall \mathbf{x} \in \mathbf{X}\}. \quad (3\text{-}12)$$

For a $\pi \in \Pi$, each element $\boldsymbol{\mu}_t$ of π defines the control function (or decision function) at time t. If the initial state is $\mathbf{x}_0$, then the expected additive cost over the time horizon corresponding to a policy $\pi \in \Pi$ is:

$$C_\pi(\mathbf{x}_0) = \mathbb{E}\left[\sum_{t=0}^{K-1} J_t(\mathbf{x}_t, \boldsymbol{\mu}_t(\mathbf{x}_t), \mathbf{w}_t)\right], \tag{3-13}$$

where the expectation is over the disturbance $\mathbf{w}_t$. A typical assumption for this disturbance, is that:

$$p_t(\mathbf{w}_t|\mathbf{w}_{0:t-1}, \mathbf{x}_{0:t}) = p_t(\mathbf{w}_t|\mathbf{x}_t). \tag{3-14}$$

This assumption is valid for the solar irradiance model presented in Section 3.2.1. The explicit dependence of this probability to time is due to the time-dependent NOAA sky-cover forecast, sc_t. The 2-D autoregressive process, $\mathbf{a}_t$, can be thought as part of the system state $\mathbf{x}_t$, and the stochastic disturbance $\mathbf{w}_t$ includes the solar heat gain to the building, and the independent Gaussian disturbances to $\mathbf{a}_t$.

We wish to minimize equation (3-13) over the policy set Π. The optimal cost function $C^*(\cdot)$ is defined by:

$$C^*(\mathbf{x}_0) = \min_{\pi\in\Pi} C_\pi(\mathbf{x}_0). \tag{3-15}$$

The optimal policy π^* may depend on the initial state $\mathbf{x}_0$, but under very general conditions, when an optimal policy exists, it is independent of the first state (Bertsekas, 1995). We define the optimal cost-to-go function at time t, $C_t^*(\mathbf{x}_t)$, as the cost incurred by the optimal policy from time t till the end of the prediction horizon, i.e.,

$$C_t^*(\mathbf{x}_t) = \min_{\pi\in\Pi} \sum_{s=t}^{K-1} \mathbb{E}[J_s(\mathbf{x}_s, \boldsymbol{\mu}_s(\mathbf{x}_s), \mathbf{w}_s)]. \tag{3-16}$$

It can be shown that $C_t^*(\mathbf{x}_t)$ must satisfies the Bellman equation:

$$C_t^*(\mathbf{x}_t) = \min_{\mathbf{u}_t\in\mathcal{U}(\mathbf{x}_t)} \mathbb{E}\big[J_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) + C_{t+1}^*\big(\mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t)\big)\big]. \tag{3-17}$$

This recursive relation suggests a powerful numerical scheme for solving dynamic programming problems. Specifically, one starts from the final cost-to-go, $C_K^*(\cdot) = 0$ in our case, and follows the recursion defined by equation (3-17) backwards for $t = K-1, \ldots, 0$, each time estimating the unknown function $C_t^*(\cdot)$ from the known $C_{t+1}^*(\cdot)$. This is known as the *value iteration* algorithm and it can only be implemented in an approximate way. First, we choose the function class within which the optimal cost-to-go functions are to be approximated. Second, we choose a finite, but well distributed, set of *M* collocation points using Latin hypercube sampling (LHS, Iman, 2008) in the state space, **X**, on which we evaluate the right-hand side of equation (3-

17). That is, we solve a separate non-linear constrained stochastic optimization problem on each one of these points. The expectation over $\mathbf{w}_t$ is approximated using a sampling average. Third, using the collected data, we approximate the next optimal cost-to-go function within the selected class and proceed with the recursion. In Algorithm 3.1, we present these three steps while in Appendix A, we discuss the mathematical details.

Algorithm 3.1: Value iteration algorithm

Inputs:	Plan horizon (*K*),
	Number of discrete points in state space (*M*),
	Number of irradiation samples (*N*)
	Time series samples of uncertain parameters $(\mathbf{W}_0, \mathbf{W}_1, \cdots, \mathbf{W}_{K-1})$
Outputs:	Optimal value/cost-to-go functions $(C^* = (C_0, C_1, \cdots, C_{K-1}))$
	Optimal policy functions $(\pi^* = (\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_{k-1}))$

Generate M discrete points in the state space $\mathbf{X}$ using LHS (given as $\mathbf{X}_M$).

$t = K - 1$

while $t \geq 0$ do:

for each $\mathbf{x}$ in $\mathbf{X}_M$*:*

Solve optimization problem over control variable $\mathbf{u}$ (Details in Appendix A)

end for

Learn cost-to-go function $\boldsymbol{C}_t$, and policy function $\boldsymbol{\mu}_t$ from the optimized control variables at points in $\mathbf{X}_M$ using Gaussian process regression (Details in Appendix A)

$t = t - 1$

3.3 Application to building-integrated solar system control

3.3.1 Building-integrated solar energy system

The building-integrated solar energy system is shown in Figure 3.3. It includes a building-integrated photovoltaic (BIPV/T) system with a corrugated unglazed transpired solar collector (UTC) that enables on-site generation of solar power and heat. The load side of an air-to-water heat pump (Swegon Maroon 2 MT29) is connected to a thermal energy storage (TES) tank,

providing hot water to a radiant floor heating (RHF) system that is used to condition an open-plan office space in Herrick Laboratories building at Purdue campus. The outlet air from the UTC serves as the source side for the heat pump. These three components (BIPV/T, TES tank, RFH) are the integrated solar system within the context of this paper. The thermal output of the BIPV/T system increases the COP of the heat pump and reduces the ventilation energy use. Assuming constant ventilation rate and supply air temperature, the benefits from the BIPV/T system are fixed, hence only the increase of the heat pump COP is considered in the optional control formulation. A model for the integrated energy system is developed in TRNSYS (Klein *et al*., 2011) and Table B.1 in Appendix B provides information for the basic settings and details are presented in Li *et al.* (2015).

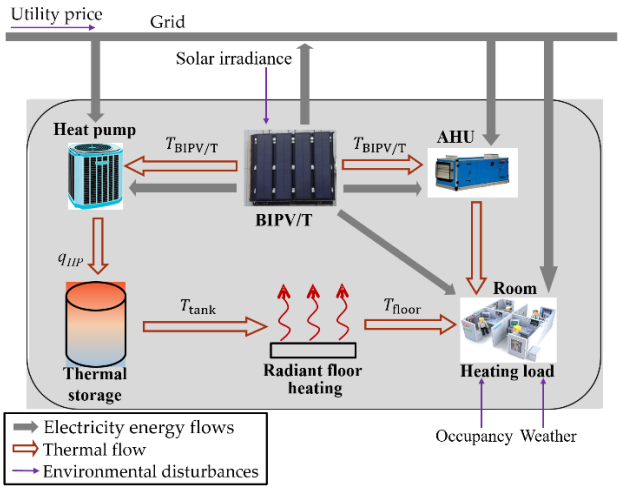


Figure 3.3. The building-integrated solar energy system.

The building is modeled with TRNSYS type 56. Settings for the ventilation, shading control, air and floor surface temperatures are provided in Table B.1 (Appendix B). The building envelope properties are extracted from drawings. The RFH system is modeled using component type 653 (mode 2) with a water flow rate of 400 kg/hr, coupled with a 10 m^3 TES tank (TRNSYS type 60) based on the recommendations provided by Li *et al*. (2015), which examined the interactions between design and control parameters. The BIPV/T system covers the top section of the south building façade (plenum area) to facilitate potential placement of the ducts, heat pump, and TES tank on the roof. The available area for the BIPV/T system is 65 m^2. The photovoltaic (PV) panels have a nominal power of 0.108 kW/m^2 (Day4 Energy Inc., model: DAY418MC). For the UTC configuration with PV panels, the PV panel coverage ratio is 90%, based on optimal design

recommendations by Li and Karava (2014), which provides 58.5 m^2 available PV area with 6.32 kWp (kilowatt-peak) capacity. The electricity generated by the BIPV/T system is used to cover the energy needs of the building or be sold back to the grid. The BIPV/T system is incorporated into TRNSYS as a user-defined component, using the energy models presented in Li *et al.* (2014).

3.3.2 Optimal control problem formulation

Based on the system details presented in the previous section, we formulate the system specific stochastic model predictive control problem. The control variable (u_t) is the total heating power provided by the air-to-water heat pump and the backup heater (when needed). The objective function is the expected value of the accumulated electric energy consumption over the prediction horizon, which is the sum of the electricity consumption from the air-to-water heat pump and the backup heater. The backup heater has a maximum capacity ($P_{\max}$) of 5000 watts and efficiency of 90% (η_{h}). It is installed in the TES tank in case of insufficient heating from the heat pump. Thus, the cost at a given time step in equation (3-1) is:

$$J_t(\mathbf{x}_t,\, u_t, I_{g,t}) =$$

$$\begin{cases} \frac{\mathrm{HC}_{\max,t}(\mathbf{x}_t,\mathbf{v}_t,I_{g,t})}{\mathrm{COP}_t(\mathbf{x}_t,\mathbf{v}_t,I_{g,t})} + \frac{u_t - \mathrm{HC}_{\max,t}(\mathbf{x}_t,\mathbf{v}_t,I_{g,t})}{\eta_{\mathrm{h}}}, \text{if } \mathrm{HC}_{\max,t}(\mathbf{x}_t,\mathbf{v}_t,I_{g,t}) < u_t \le u_{\max,t}(\mathbf{x}_t,\mathbf{v}_t,I_{g,t}) \\ \frac{u_t}{\mathrm{COP}_t(\mathbf{x}_t,\mathbf{v}_t,I_{g,t})}, \text{if } 0 \le u_t \le \mathrm{HC}_{\max,t}(\mathbf{x}_t,\mathbf{v}_t,I_{g,t}). \end{cases}$$

(3-18)

The COP and maximum heating capacity ($\mathrm{HC}_{\max}$) of the heat pump are functions of the solar irradiance ($I_{g,t}$) and outdoor dry bulb temperature (through the outlet air temperature of the BIPV/T collector, T_{bipvt}), and the tank temperature ($T_{\mathrm{tank},}$), which is one of the system states. A BIPV/T collector model (Li *et al.*, 2014) incorporated in the controller receives information on the predicted solar irradiance from the forecast model (Section 3.2.1), along with the outdoor dry bulb temperature forecast, and calculates T_{bipvt} ($T_{\mathrm{bipvt},t} = q(\mathbf{v}_t, I_{g,t})$) during the prediction horizon. Therefore, the COP and $\mathrm{HC}_{\max}$ are both functions of the system states, exogenous inputs ($\mathbf{v}_t$), and disturbances:

$$\mathrm{COP}_t(\mathbf{x}_t,\mathbf{v}_t,I_{g,t}) = 6.2504 + 0.1338T_{\mathrm{bipvt},t} - 0.0986T_{\mathrm{tank},t} + 0.005864T_{\mathrm{bipvt},t}^2 + 0.0004T_{\mathrm{tank},t}^2 - 0.0015T_{\mathrm{bipvt,t}}T_{\mathrm{tank,t}},$$

(3-19)

$$\mathrm{HC}_{\max,t}(\mathbf{x}_t, \mathbf{v}_t, I_{g,t}) = 25.3537 + 0.7337T_{\mathrm{bipvt},t} - 0.0623T_{\mathrm{tank},t} + 0.0056T_{\mathrm{bipvt},t}^2 + 0.0001T_{\mathrm{tank},t}^2 - 0.0041T_{\mathrm{bipvt},t}T_{\mathrm{tank},t}\ , \tag{3-20}$$

$$u_{\max,t}(\mathbf{x}_t, \mathbf{v}_t, I_{g,t}) = \mathrm{HC}_{\max,t}(\mathbf{x}_t, \mathbf{v}_t, I_{g,t}) + P_{\max}. \tag{3-21}$$

Equations (3-19) and (3-20) show that the efficiency and capacity of the heat pump increase as T_{bipvt} increases. The low-order system model used in the controller is shown in Appendix B (Figure B.1) while additional details are provided in Li *et al.* (2015). The system dynamics is given by:

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, u_t, \mathbf{w}_t) = \boldsymbol{A}\mathbf{x}_t + \boldsymbol{B}_u u_t + \boldsymbol{B}_v \mathbf{v}_t + \boldsymbol{B}_w \mathbf{w}_t, \tag{3-22}$$

where $\mathbf{v}_t = \begin{bmatrix} T_{\mathrm{a},t} \\ T_{\mathrm{a2}} \\ T_{\mathrm{a3}} \\ \mathrm{IG}(t) \end{bmatrix}$. T_{a} is the outdoor dry bulb temperature from the NOAA weather forecast. We do not consider the forecast uncertainty on T_{a} as it is typically small and would have negligible impact on this heavy thermal mass system. $\mathrm{IG}(t)$ is the internal heat gain, which is considered known based on the building operation schedule (Appendix B, Table B.1). The variables T_{a2} and T_{a3} represent the ambient temperature of the TES tank and air temperature of the adjacent zone, respectively, and are assumed to be constant. $\boldsymbol{A} \in \mathbb{R}^{6\times6}, \boldsymbol{B}_u \in \mathbb{R}^{6\times1}, \boldsymbol{B}_v \in \mathbb{R}^{6\times4}$ and $\boldsymbol{B}_w \in \mathbb{R}^{6\times4}$ are time invariant matrices. The state vector of our system is,

$$\mathbf{x}_t = \begin{bmatrix} T_{\mathrm{room},t} \\ T_{\mathrm{floor},t} \\ T_{\mathrm{tank},\,t} \\ T_{\mathrm{enve},\,t} \\ a_{1,t} \\ a_{2,t} \end{bmatrix},$$

where $\mathbf{a}_t = (a_{1,t}, a_{2,t})$ is the state of the solar irradiance model (Section 3.2.1), T_{enve} is the average envelope temperature of the room, T_{room} is the room air temperature, T_{floor} is the average floor slab temperature, T_{tank} is the average tank temperature. As discussed in Section 3.2.1, the stochastic disturbance $\mathbf{w}_t$ corresponds to the 2-D Gaussian noise, say $\mathbf{z}_t$, perturbing $\mathbf{a}_t$ as well as to the random sky condition c_t. Therefore, we have

$$\mathbf{w}_t = \begin{bmatrix} h_1\left(I_{g,t}(\mathbf{a}_t, c_t)\right) \\ h_2\left(I_{g,t}(\mathbf{a}_t, c_t)\right) \\ z_{1,t} \\ z_{2,t} \end{bmatrix},$$

where the function $I_{g,t}(\mathbf{a}_t, c_\mathrm{t})$ is the global horizontal irradiance (see Section 3.2.1), while $\mathbf{h}(I_{g,t}) = \begin{bmatrix} h_1(I_{g,t}) \\ h_2(I_{g,t}) \end{bmatrix} = \begin{bmatrix} q_{\mathrm{SG1},t} \\ q_{\mathrm{SG2},t} \end{bmatrix}$ gives the solar heat gain on the floor $\left(q_{\mathrm{SG2},t}\right)$ as well as other building interior surfaces $\left(q_{\mathrm{SG1},t}\right)$ (Klein *et al*., 2011).

We use seven chance constraints ($n_c = 7$) on the temperature states and feasible sets of control inputs. To determine their proper form, we examine the distributions of these variables under 1000 uncertain solar irradiance samples. For the temperature states, even in the most uncertain case when $\mathrm{sc} = 0.5$, most of the samples are concentrated around the expected value (Figure 3.4). Therefore, the following constraints impose minimum bounds on the expected room, floor and tank temperatures,

$$\mathbb{E}\left[x_{i,t+1}\right] - T_{\min,i,t+1} \geq 0, \tag{3-23}$$

yielding, with the notation of equation (3-3),

$$g_{i,t}(\mathbf{x}_t, u_t, \mathbf{w}_t) = f_{i,t}(\mathbf{x}_t, u_t, \mathbf{w}_t) - T_{\min,i,t+1}, \tag{3-24}$$

for $i = 1, \ldots, 3$. Similarly, the constraints below impose maximum bounds on the expected building temperatures,

$$T_{\max,i,t+1} - \mathbb{E}\left[x_{i,t+1}\right] \geq 0, \tag{3-25}$$

yielding

$$g_{4+i,t}(\mathbf{x}_t, u_t, \mathbf{w}_t) = T_{\max,i,t+1} - f_{i,t}(\mathbf{x}_t, u_t, \mathbf{w}_t), \tag{3-26}$$

where $\boldsymbol{T}_{t,\,\max}$ and $\boldsymbol{T}_{t,\,\min}$ are known based on the values and schedules given in Appendix B Table B.1. Finally, we enforce with high probability the control bounds with the following constraint:

$$\mathbb{P}\left[0 \leq u_t \leq u_{\max,t}\left(\mathbf{x}_t, \mathbf{v}_t, I_{g,t}(\mathbf{a}_t, c_t)\right)\right] \geq 1 - \alpha, \tag{3-27}$$

where α is a small number corresponding to our tolerance for violating this constraint. As an expectation, this probability can be expressed by:

$$\mathbb{E}\left[1_{\left[0, u_{\max,t}\left(\mathbf{x}_t, \mathbf{v}_t, I_{g,t}(\boldsymbol{a}_t, c_t)\right)\right]}(u_t)\right] - 1 + \alpha \geq 0, \tag{3-28}$$

where $1_A(\cdot)$ is the characteristic function of a set A. In the notation of equation (3-3), this constraint can be expressed as:

$$g_{7,t}(\mathbf{x}_t, u_t, \mathbf{w}_t) = 1_{\left[0, u_{\max,t}\left(\mathbf{x}_t, \mathbf{v}_t, I_{g,t}(\boldsymbol{a}_t, c_t)\right)\right]}(u_t) - 1 + \alpha. \quad (3\text{-}29)$$

In the most uncertain scenario when $\mathrm{sc} = 0.5$, the distribution of the heating capacity value range is from 26 kW to 31 kW with most of the samples on the two ends (Figure 3.4). Therefore, in equation (3-29), a small value of α=1% is used to ensure that, with 99% of the probability, the control input u_t does not exceed the equipment capacity.

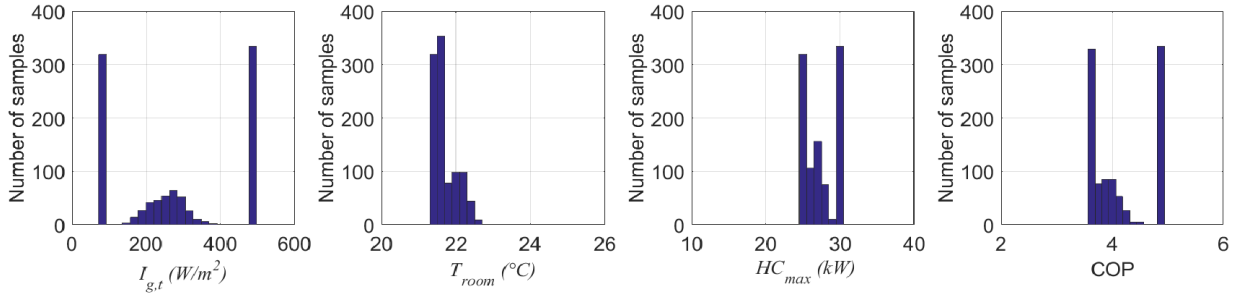


Figure 3.4. Histograms of global horizontal irradiance, room temperature, heat pump capacity and COP under 1000 irradiance samples (sc = 0.5, u = 15 kW, T_{a} = 15℃, $T_{\mathrm{room},0}$=21℃).

Figure 3.5 shows the flow chart of the control algorithm. At the beginning of each simulation time step of 1 hour, the algorithm reads the initial temperature states from TRNSYS and it also receives weather forecast information (sky-cover, outdoor dry bulb temperature, etc.) for the future K =24 hours of the prediction horizon. Optimal control decisions are made every 1 hour (control horizon) between 6:00 a.m. to 20:00 p.m.

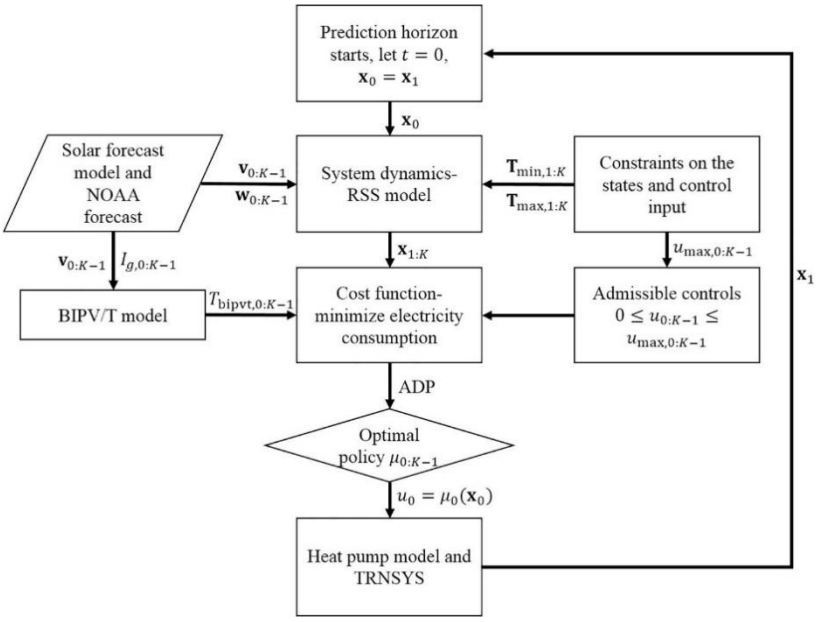


Figure 3.5. Optimal control algorithm.

3.3.3 Optimal control problem solution

This section presents the implementation of the value iteration algorithm for solving the optimal control problem detailed in the previous section. Each run yields policy functions that predict the optimal control for each time step in the prediction horizon.

Following the procedure outlined in Algorithm 3.1 (Section 3.2.2), the state-space has been discretized for computing point estimates of optimal 'cost-to-go' and policy functions. The first four variables in the state space (tank, room, floor and envelope temperatures) are used for calculating the cost at each time step, while the dependence of the cost-to-go functions on other variables is negligible, and thereby not considered.

To obtain the collocation points in 4-dimensional state space, T_{room} and T_{tank} are sampled using Latin hypercube sampling (LHS) as a two-dimensional vector varying between their temperature bounds (Appendix B Table B.1). As shown in Figure 3.6, samples are uniform for T_{room} and T_{tank}. T_{floor} is generally higher than T_{room} while T_{enve} is usually lower. Therefore, the state variables representing T_{floor} and T_{enve} are sampled from exponential distributions with location parameters (1 for the floor temperature; -1 for the envelope temperature) and scale parameters (1 for both) chosen to keep deviations around 1-3ºC from T_{room}. A sample size of 500

was deemed sufficient to represent the state space for these simulations. T_{room} and T_{tank} have the most significant impact on the energy consumption, and are thus used for visualizing the value iteration algorithm.

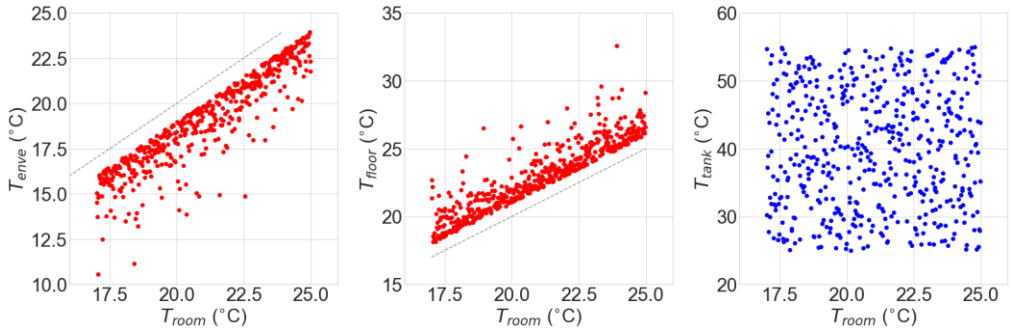


Figure 3.6. Distribution of 500 collocation points in 4-dimensional temperature space for implementation of value iteration algorithm at a time step.

The solar irradiance model presented in Section 3.2.1 generates N =100 irradiation samples at each time step of 1 hour for a prediction horizon of K = 24 hours. At each time step t in a prediction horizon, we evaluate the right-hand side of the Bellman equation (3-17) at each collocation point using all N irradiance samples at the time step (see Appendix A equation (A-3)). Each evaluation requires solving a stochastic optimization problem (details in Section 3.3.2) with respect to the optimal control at that time step. We parallelize the M=500 optimization problems to solve at time step using MPI4Py (Dalcín *et al.*, 2008); while employ a gradient based 'pyOpt' solver (Perez *et al*, 2012) to achieve further efficiency by providing the analytical derivatives of objective function and constraints. After evaluating all the points at a time step, we collect those collocation points as inputs and the corresponding optimal cost-to-go values and optimal controls as outputs to approximate the next optimal cost-to-go and policy functions, respectively. This is carried out via GPR in GPy module (Hensman *et al.*, 2012). We use squared exponential covariance functions in the GPR and we maximize the marginal likelihood to find the optimal hyperparameters following the method described in Chapter 5 of (Rasmussen and Williams, 2006). More details regarding GPR and the evaluation of right-hand side of the Bellman equation are presented in Appendix A. After completing the approximation of cost-to-go and policy functions at time t, we move to time step $t-1$ and repeat the evaluation and approximation procedures with updated disturbances (see Algorithm 3.1). In this way, the ADP algorithm is implemented in

receding horizon fashion to obtain sets of policy functions, while the effect of future predictions is captured in the approximated value and policy functions. The policy function at the first time step of the prediction horizon is used for generating the optimal control for the building system.

The algorithm deals with non-controllable scenarios via min-max controls, which means if no feasible control input can keep the temperature states within constraints for the next time step, we apply the minimum (for over-heating) or maximum (for under-heating) feasible input. To visualize the cost-to-go functions, we only show results for T_{room} and T_{tank} as the function values are found to vary primarily in these dimensions.

Figure 3.7 details the evolution of the cost-to-go functions across the prediction horizon. We observe that the use of min-max type control inputs reduces as the simulation moves to lower time steps. This elicits the effect of longer time horizon in reducing the energy consumption in the system. The cost increases as the number of time step decreases because it includes energy costs incurred by the system at future points of time when it receives optimal control inputs. At lower time steps, the estimated cost increases as room and tank temperatures decrease.

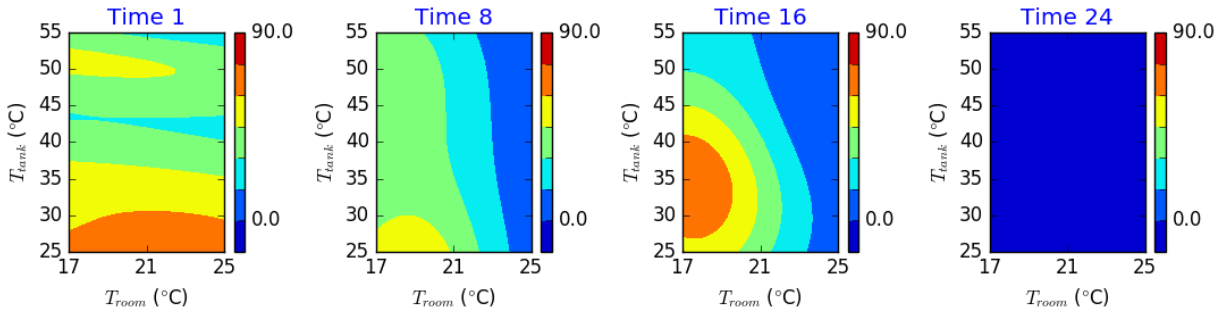


Figure 3.7. Contour plots demonstrating evolution of cost-to-go function at time step 1, 8, 16, 24 as computed using value iteration algorithm (irradiation samples at 0:00 of 18 January 2017, T_{floor}=20℃, T_{enve}=20℃).

The expensive part of policy function computation at each step is the optimization at each collocation point, which requires several evaluations of the right-hand side of Bellman's function. The optimization problems at the 500 collocation points are parallelized to reduce the computation time. A single evaluation of the right-hand side of the Bellman equation for one collocation point can take about 0.4 minutes with our current Python implementation. After parallelizing the 500 collocation points to 100 nodes of the Rice supercomputing cluster at Purdue University, the

computation time of ‘cost-to-go’ evaluation at a time step takes about 2 to 3 minutes. Therefore, a complete ADP solution for a 24-hour prediction horizon takes about 30-40 minutes in average considering the system operation schedule from 6 a.m. to 20 p.m. All the processes can be sped up by implementing in a lower level language and augmenting the parallelization ability. For this control approach to be implementable in real-time building operation, access to cloud computation services is required. The source code for the ADP we implemented in this study can be found in Paritosh *et al*. (2017).

3.4 Performance analysis

In this section, we present the emulation process that we use to evaluate the SMPC based on two aspects: (i) Comparing its performance, in terms of energy savings and comfort maintenance; and (ii) Analyzing the uncertainty on the energy consumption and thermal comfort violation associated with the stochastic disturbance.

3.4.1 Emulator

We deploy the emulation framework shown in Figure 3.8 to evaluate the performance of the SMPC for the integrated solar system. Physical models for the building, BIPV/T system, RFH, and TES tank are built using TRNSYS. The data-driven heat pump model is developed in MATLAB. The predictive controller is developed in Python and it is coupled with TRNSYS Type 155 using MATLAB as the mid-ware. Real time actual weather data are used as inputs to the physical models in TRNSYS. At every control horizon between 6:00 a.m. to 20:00 p.m., the controller predicts the optimal heating system operation by running a 24-hour-horizon ADP solution and sends the control signal to the heat pump and the backup heater. Every 1-hour emulation time-step in TRNSYS, it takes about 31 to 42 minutes to complete. This includes about 1 to 2 minutes for the communication between MATLAB and Python, and 30 to 40 minutes required for an ADP solution. Therefore, considering 1 hour of control horizon, our solution can be implemented to an actual controller for the integrated solar system.

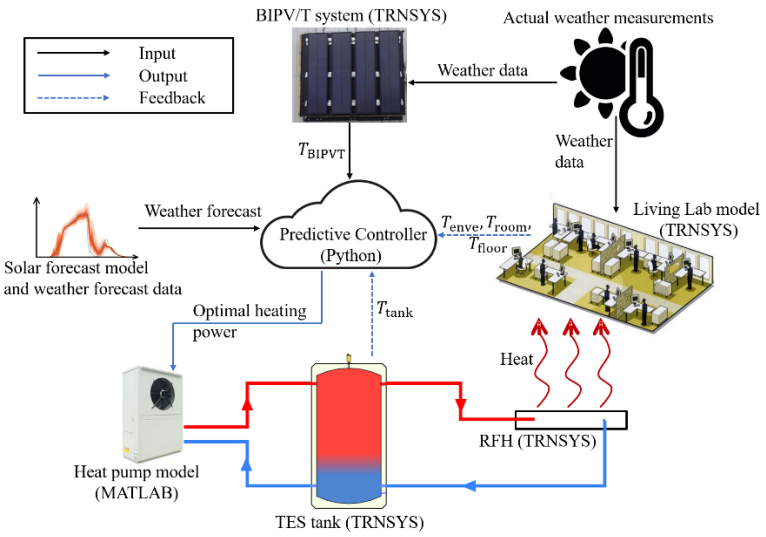


Figure 3.8. System emulation diagram.

3.4.2 Control performance

In this section, we present the performance evaluation results. The SMPC uses the solar forecast model to quantify the uncertainty in solar irradiance while the physical models in TRNSYS receive measured weather data. A benchmark control strategy is the theoretical performance bound (PB), in which we assume that the actual weather condition is perfectly known in advance. Therefore, both the controller and TRNSYS receive measured weather data. The ADP algorithm is implemented to obtain optimal control solutions. However, PB is a theoretical concept rather than an actual controller. A well-tuned rule-based control (RBC) with control decisions based on the outdoor dry bulb temperature and sky-cover forecast values is also used as baseline. It follows the solar energy availability so that the energy system achieves high efficiency (Candanedo, 2011). The details of the RBC are presented in Appendix C. A 24-hour prediction horizon is implemented for the SMPC and PB. The same initial temperature states are used for all cases. To eliminate the effect of initial states, we use a pre-simulation period of five days.

The temperature exceedance (in ℃-hr) according to ASHRAE Standard 55 (ASHRAE & ANSI, 2017) and electricity consumption (in kWh), are used as performance metrics. In this study, both the total temperature exceedance (including occupied and unoccupied hours) and the temperature exceedance at occupied hours are considered:

$$\Delta T_{\mathrm{op}} = \sum\left(\left|T_{\mathrm{op}}^{t} - T_{\mathrm{set}}^{t}\right|\Delta t\right), \tag{3-30}$$

where T_{op} is the operative temperature in ℃; T_{set} is the setpoint temperature in ℃; Δt is the time step in hour. The occupied hours we considered in this study are from 8:00 a.m. to 18:30 p.m.

We consider a three-day emulation, based on the weather data shown in Figure 3.9. During this period (Feb. 1st – Feb. 3rd, 2017), the outdoor dry bulb temperature varies from -10℃ to 4.5℃; the first day is relatively warm and partly-cloudy with high uncertainty on the solar irradiance predictions (for details see Section 3.2.1) and the following two days are relatively cold and sunny (high probability of sky condition being clear).

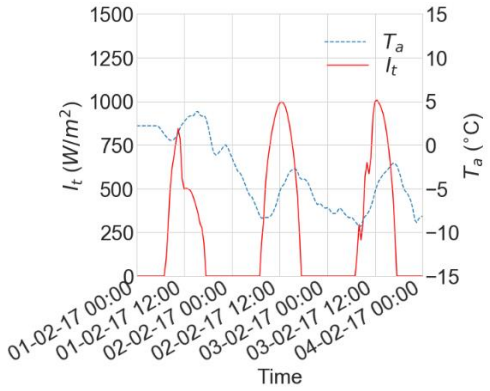


Figure 3.9. Outdoor dry bulb air temperature and incident solar irradiance on the south façade during the three-day emulation (Feb. 1st – Feb. 3rd, 2017).

The emulation results for the SMPC (Figure 3.10, top) show that the heat pump operation starts at 6 a.m. on Feb. 1st with nearly maximum system capacity due to the anticipated increase in the set-point temperature during the occupied hours and, therefore, the tank is charged in advance. Along with a slight tank charge in the afternoon, the stored energy is sufficient to maintain the temperature for the rest of the day. Another reason for the intense charge at 6 a.m. of Feb. 1st is that, the uncertainty on the solar irradiance forecast is high on the upcoming hours based on the sky-cover forecast (range from 40% to 80%) received at 6 a.m. In order to meet the lower setpoint bound on the temperature states under uncertain disturbances, the SMPC controller follows a more conservative operation schedule. In contrast, for PB (Figure 3.10, middle), the heat pump operates with less power at 6 a.m. on Feb. 1st. Based on the perfectly accurate weather information, the cost is reduced when the heat pump operation is postponed till the afternoon when sufficient energy can be stored even for the following day.

Due to the different pump operation on the previous day, the starting tank temperature on Feb. 2nd at 6 a.m. in SMPC is lower than that in PB. Therefore, for SMPC, the 6 a.m. charge is repeated on Feb 2nd with less intensity, and the heat pump is also ON during the sunny hours in anticipation of the outdoor temperature decrease in the evening. While in PB, the heat pump operation on Feb. 2nd is not required as the TES tank has been charged already during the previous day.

On Feb. 3rd, the starting tank temperatures at 6 a.m. are similar in both SMPC and PB. The sky-cover forecast indicates partly-cloudy condition (high uncertainty in irradiance) in the morning and high probability of being sunny in the afternoon (low uncertainty in irradiance). Therefore, the heat pump operates mostly in the afternoon sunny hours for both cases to store energy at increased efficiency for discharge at night. The higher heat pump power in SMPC in the morning can be explained by the high solar irradiance uncertainty at the time.

In RBC (Figure 3.10, bottom), the operations are designed to follow the solar availability to take advantage of the increased system COP, while also considering the outdoor dry bulb temperature (details in Appendix C). Therefore, the heat pump is continuously ON from 6 a.m. to 20 p.m. every day at the power rate ranging from 3 to 6 kW.

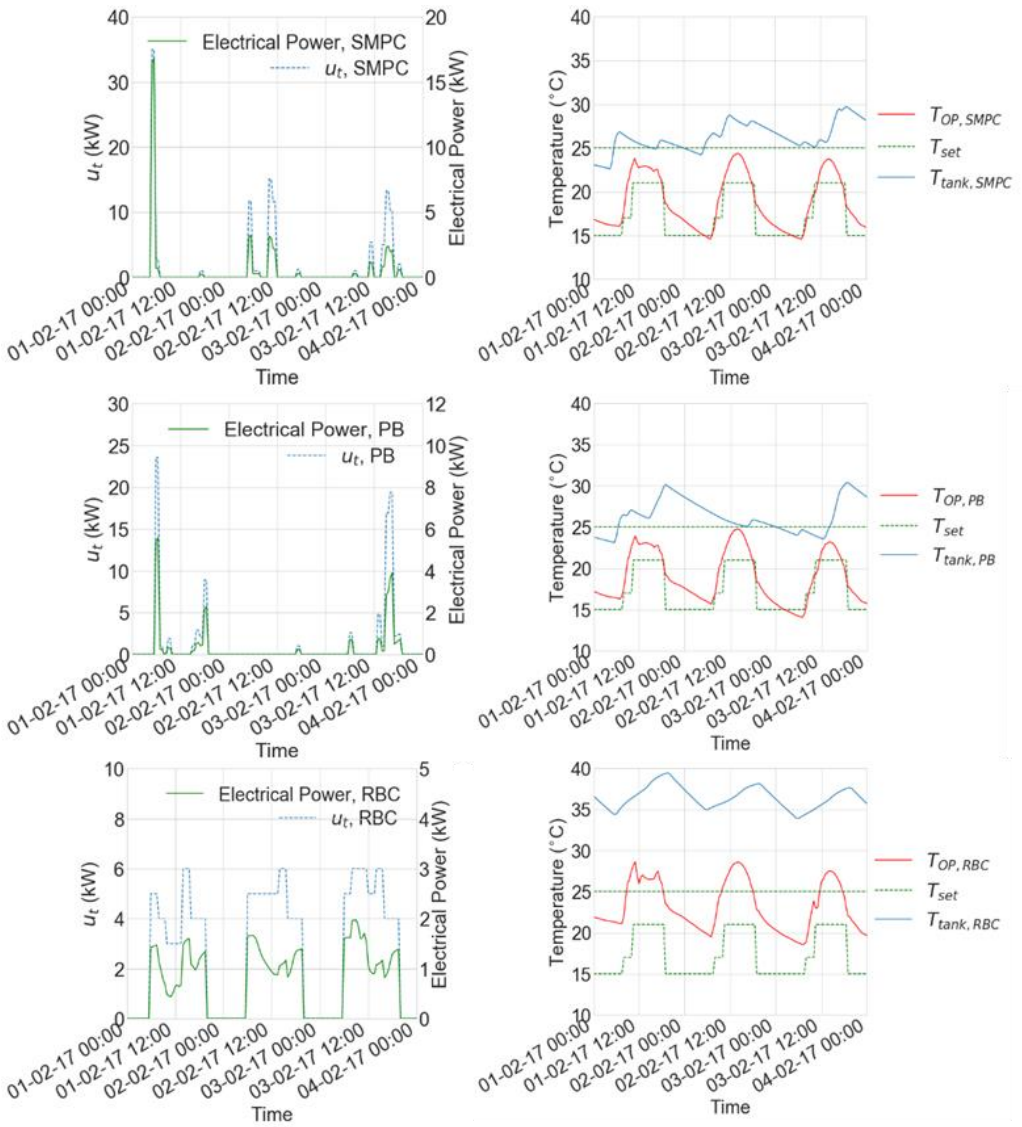


Figure 3.10. Temperatures, heating and electrical power for the three-day emulation of SMPC, PB and RBC (Feb. 1st – Feb. 3rd, 2017).

Emulations were also performed for a winter month (Jan. 16th to Feb. 16th, 2017) for the three cases discussed above and the results are shown in Table 3.1. During this period, the outdoor dry bulb temperature varies from -15℃ to 18℃. Overall, SMPC results in slightly less temperature exceedance (3.22℃-hr in occupied hours) but higher electricity consumption (57.28 kWh, 34.7%) over a month compared to PB.

Table 3.1. Performance metrics comparison for the winter month emulation (Jan. 16th – Feb. 16th, 2017).

Metrics	SMPC	PB	RBC
Temperature exceedance (all hours) (ºC-hr)	Lower-setpoint: 145.13 Upper-setpoint: 104.50 Total: 249.63	Lower-setpoint: 154.01 Upper-setpoint: 100.73 Total: 254.74	Lower-setpoint: 0.05 Upper-setpoint: 727.54 Total: 727.58
Temperature exceedance (occupied hours) (ºC-hr)	Lower-setpoint: 77.86 Upper-setpoint: 104.50 Total: 182.36	Lower-setpoint: 84.85 Upper-setpoint: 100.73 Total: 185.58	Lower-setpoint: 0 Upper-setpoint: 727.54 Total: 727.54
Total heating energy (kWh)	664.80	566.80	1574.00
Total electricity (kWh)	222.41	165.13	399.50

Compared to RBC, SMPC saves around 44.0% (177.09 kWh) electricity consumption. Also, SMPC improves room thermal comfort, reducing the temperature exceedance during occupied hours to 25.1% of RBC (primarily improvements occur during the occupied hours). Figure 3.11 presents the histogram of the operative temperature exceedance during the occupied hours. To count for the difference between over-heating and under-heating, at each time step, the exceedance metric is calculated as follows:

$$\text{Exceedance}_{\text{occ}} = \left(T_{\text{op}}^{t} - T_{\text{set}}^{t}\right)\Delta t. \tag{3-31}$$

It is seen that SMPC and PB rarely result in temperature exceedance to be greater than 1ºC-hr or lower than -1ºC-hr; while RBC shows clearly more over-heating compared to the other two cases.

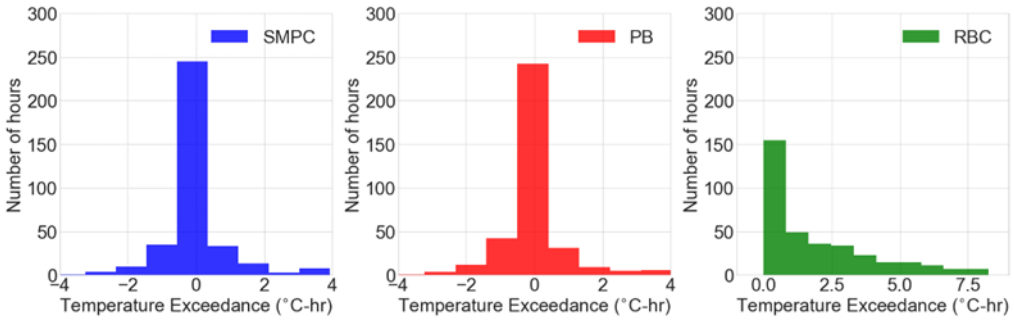


Figure 3.11. Operative temperatures exceedance histogram for the long-term simulation (Jan. 16th – Feb. 16th, 2017).

3.4.3 Uncertainty analysis on energy efficiency and thermal comfort maintenance

In SMPC, the ADP algorithm returns a policy function μ_t at each time step, which determines the optimal control for the upcoming control horizon based on the current states. At a time step, when system disturbances vary, the optimal control that the policy predicts also varies (as disturbances affect the states). In this way, the uncertainty in the system disturbances can introduce uncertainty on the cost as it is a function of the optimal control. To evaluate the impact of the uncertainty, we draw 100 samples of the time-series solar irradiances data (based on the solar irradiance model in Section 3.2.1) during a period from January 16th, 2017 to February 5th, 2017 under the measured sky-cover values. A set of hourly-updated policy functions is obtained based on the hourly-updated weather forecast information. To examine the performance, we implement this set of policy functions in the controller to predict the optimal controls for the system over 100 solar irradiance samples. Each sample is an input to TRNSYS assuming it represents a possible scenario of 'actual' irradiance. Therefore, 100 emulations are performed with this certain set of policy functions to predict 100 sets of optimal controls. In comparison, we also predict a series of control inputs with RBC based on the same weather forecast information. This series of control inputs are also implemented in 100 emulations under the 100 solar irradiance samples. The initial states for both cases are kept the same for fair comparison.

The operative temperature profiles from SMPC and RBC for the first 10 days (Jan. 16th, 2017 -Jan. 26th, 2017) are presented in Figure 3.12 and 3.13, respectively. Figure 3.12 shows that the policy functions can well control the operative temperature (T_{op}) during the occupied hours over different solar irradiance samples. The upper and lower bounds of the operative temperature are mostly maintained within the setpoint bounds, with a reasonably small amount of violations. In comparison, RBC is less capable of controlling the operative temperature within the setpoint bounds (Figure 3.13).

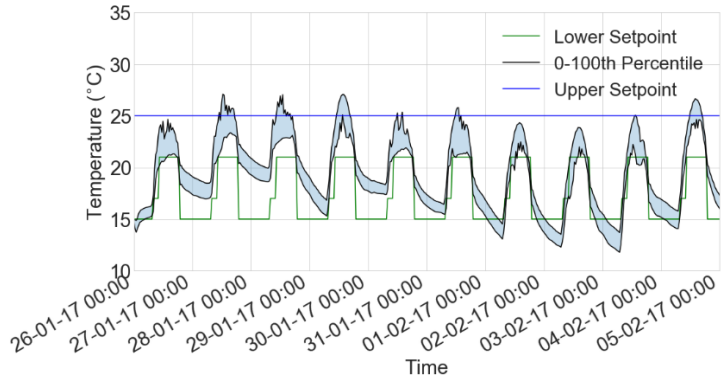


Figure 3.12. Operative temperature profile with uncertainty for SMPC (Jan. 16th, 2017 - Jan. 26th, 2017, 100 solar irradiance samples).

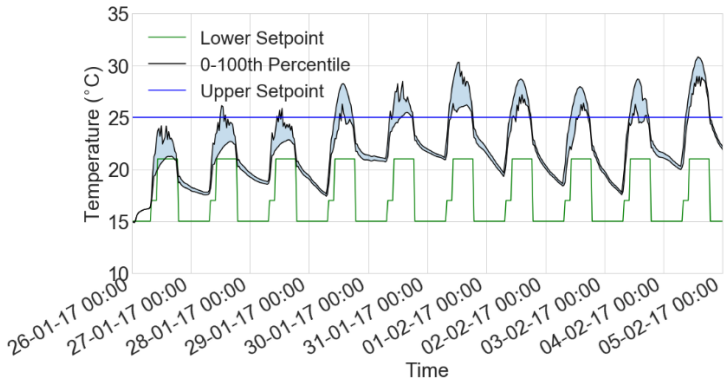


Figure 3.13. Operative temperature profile with uncertainty for RBC (Jan. 16th, 2017 - Jan. 26th, 2017, 100 solar irradiance samples).

The cumulative cost at a time is the total cost from the starting time. For example, the cumulative cost at the end of day 3 is the total cost from the starting time to the end of day 3. It measures control performance in reducing cost over time. According to Figure 3.14, the mean cumulative cost of the 20th day for RBC is 241.44 kWh with an uncertain range of [239.09 kWh, 244.56 kWh] considering the 0th- 100th sample percentile; while it is 141.48 kWh for SMPC with an uncertain range of [114.99 kWh, 179.32 kWh] considering the 0th- 100th sample percentile.

Although SMPC introduces higher uncertainty on the cost compared to RBC, the optimal control at every prediction horizon is robust in terms of cost saving under various solar irradiance scenarios, because even the 100th percentile cumulative cost of SMPC is lower than the 0th percentile cumulative cost of RBC at the 20th day. The reason behind the higher cost uncertainty in SMPC is that, the policy functions predict different optimal controls based on the states under

different irradiance samples to satisfy the constraints. Therefore, there are 100 different optimal control trajectories from the 100 different irradiance samples; while in RBC, there is only one control input trajectory, which is obtained from the hourly weather forecast, for all the samples. This reduces the major uncertainty source on the cost in RBC to the variation of system efficiency (COP), while the uncertainty sources in SMPC are the different optimal control trajectories as well as the variations of COP.

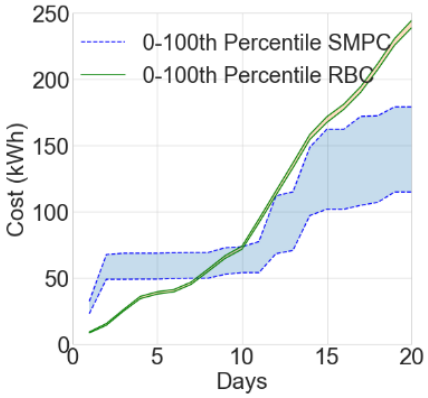


Figure 3.14. 0-100th Sample percentile of cumulative cost (Jan. 16th, 2017 - Feb. 5th, 2017).

Similarly, the cumulative temperature exceedance during the occupied hours in both cases over the emulation period are presented in Figure 3.15. SMPC clearly outperforms RBC in terms of thermal comfort maintenance, and at the same time, it imposes less risk on comfort violation. The mean cumulative temperature exceedance of the 20th day for RBC is 136.21 °C-h with an uncertain range of [121.85 °C-h, 148.72 °C-h] considering the 0th- 100th sample percentile; while it is 56.56 °C-h for SMPC with an uncertain range of [48.42 °C-h, 67.06 °C-h] considering the 0th-100th sample percentile. The reason behind the low comfort violation uncertainty in SMPC is that the policy function predicts different optimal controls over different irradiance samples aiming at maintaining the system states from all the samples within the constraints. Therefore, the temperature exceedance in SMPC is significantly lower than that in RBC, in which the only control trajectory may fail to maintain thermal comfort under different irradiance scenarios.

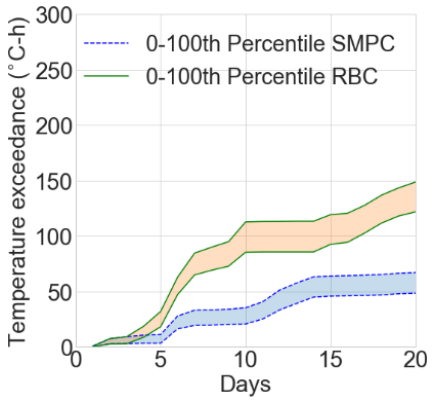


Figure 3.15. 0-100th sample percentile of cumulative temperature exceedance during the occupied hours (Jan. 16th, 2017 -Feb. 5th, 2017).

3.5 Summary

In this Chapter, we presented a SMPC algorithm augmented with a new probabilistic autoregressive model that is used to quantify solar irradiance uncertainty using sky-cover forecasts. Based on that, we introduced a new ADP methodology to solve the resulting nonconvex stochastic optimization problem. The SMPC controller was implemented in an emulator to demonstrate optimal control decisions under uncertainty for a building with integrated solar system.

The results show that SMPC outperforms RBC in terms of both energy savings and temperature control. For the integrated system and climate considered in this work, it reduces the electricity consumption by 44% in a winter month and reduces thermal comfort violations by 75%. Also, SMPC predicts optimal policies that satisfy the constraints, considering a wide range of possible outcomes of the uncertain solar irradiance forecast. More specifically, SMPC: (i) Achieves similar performance on comfort control and results in more realistic energy savings compared to the PB, which assumes perfect knowledge of the future disturbances; (ii) It is robust in comfort maintenance, imposing less uncertainty on comfort violation compared to RBC. Under uncertain solar forecast, the highest cumulative cost resulting from SMPC is less than the lowest cumulative cost of RBC after 20 days of emulation. Therefore, in summary, the developed SMPC approach has shown promising results for the operation of building-integrated solar systems. It should be noted that its performance in actual implementation also depends on the accuracy of the process model and input data.

4. A USER-INTERACTIVE SYSTEM FOR SMART THERMAL ENVIRONMENT CONTROL IN OFFICE BUILDINGS

4.1 Overview

In this Chapter, the objective is to develop a systematic approach to design interfaces of user-interactive systems that aim to increase energy efficiency and occupant satisfaction in office buildings. Towards this end, we present a prototype user-interactive system with a novel web-interface. Using this interface, occupants can adjust their temperature setpoints while providing feedback on their comfort preferences while receiving energy use information that is real-time, personalized, and directly related to specific actions. Furthermore, we implement the interface in a building management system with a model predictive HVAC controller. Finally, we design a set of experiments to reveal the causal factors of occupant's thermostat adjustment behavior on HVAC energy use. Our field study aims to test the hypothesis that user-interactive systems integrated into smart building operation make energy-efficient behavior natural, easy, and intuitively understandable for the end-users resulting in HVAC energy savings and overall occupant satisfaction. Using the data collected, we formulate a human decision-making model based on utility theory and infer the model parameters with a Bayesian approach that quantifies the uncertainty induced by the limited amount of data that can be observed in realistic settings. We argue that the proposed utility model can become a systematic approach to evaluate the design of similar user-interactive systems for different office layouts and building operation scenarios. This will allow to quantify the increase of smart office buildings' efficiency and flexibility by employing occupant-engaged controls.

In Section 4.2, we describe the experimental study with human test subjects and the test-bed building with our user-interactive system. The experiment's key observations are summarized in Section 4.3, providing the basis for the decision-making model presented in Section 4.4.

4.2 Experimental study

This section presents a prototype user-interactive system that was implemented in an office building (Section 4.2.1). The system includes a model predictive HVAC controller (Section 4.2.2) to determine the default setpoints that minimize energy consumption and a web-based thermostat

interface with personalized and real-time energy use information for the occupants (Section 4.2.3). The experimental study was conducted in the summer of 2018 to evaluate the effect of the energy use information on occupants' thermostat adjustment behavior, and HVAC energy use is presented in Section 4.2.4.

4.2.1 Building and HVAC system

Three identical south-facing private offices (3.3 × 3.7 × 3.2m high) in a high-performance building located in West Lafayette, Indiana, were used as test-beds for this study (Figure 4.1). The offices have one exterior curtain wall façade with a 54% window-to-wall ratio. The windows are equipped with high-performance glazing units with selective low-emissivity coating (visible transmittance: 70%, solar transmittance: 33%). Dark-colored motorized interior roller shades are installed in the offices with a total visible transmittance of 2.53% and an openness factor of 2.18%. Each office has two electric lighting fixtures with two 32-W T5 fluorescent lamps (total of 128 W).

Heating and cooling are delivered to the spaces through a variable air volume (VAV) system with a central air handling unit (AHU), which supplies cooled air to the offices at a constant temperature of 16°C, but with a variable flow rate (between 140 and 550cfm). Each office has a VAV box with a zone damper that can modulate the supply airflow rate in the cooling mode and a reheat coil (capacity 762W) to increase the supply air temperature as needed. The cooling and heating source (chilled water and steam) in the actual test-bed are provided from the campus plant. For this work, we assume that an air-cooled chiller is the cooling source that provides chilled water to the cooling coil in the AHU in compliance with typical office building settings. The chiller's performance data were adopted from the catalog of an actual product (Trane CGAM20), and the nominal capacity and coefficient of performance (COP) are 68.9 kW and 2.67, respectively. In this study, the capacity was scaled down to 12% (8.27 kW) based on the cooling load of the offices. The energy input ratio (EIR) method was used to determine the real-time efficiency and power (DoE, 2010). The hot water in the reheat coils was assumed to come from a gas boiler with 90% efficiency. A building management system (BMS) is available through the installed Tridium JACE controllers and Niagara/AX software framework.

Figure 4.1. The three private offices in West Lafayette, IN.

4.2.2 MPC algorithm for HVAC control

MPC is based on the premise that data-driven building models can be created using monitored data. These models can be used to determine the most energy-efficient or cost-effective control strategies. At each control time-step, using weather and internal load measurements and predictions, an open-loop optimal control problem is solved over a finite horizon, and the values within the control horizon in the optimal input trajectory are implemented to the system. In the next control time-step, a new optimal control problem is formulated and solved with updated information on weather/internal gains forecast. In this section, we present the MPC controller developed for the specific test-bed and HVAC system. The objective function minimizes the total HVAC energy consumption over a prediction horizon of 12 hours:

$$\min_{\mathbf{u}_0,\mathbf{u}_1,\ldots,\mathbf{u}_{K-1}} \sum_{t=0}^{K-1} J_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t). \tag{4-1}$$

The time step is 0.5 hours, so $K = 24$ in equation (4-1) and the control horizon is also 0.5 hours. J_t is the sum of HVAC energy consumption at time step t. It is a function of $\mathbf{x}_t$, the vector of states; $\mathbf{u}_t$, the vector of control inputs; and $\mathbf{w}_t$, the vector of disturbances.

4.2.2.1 Building model

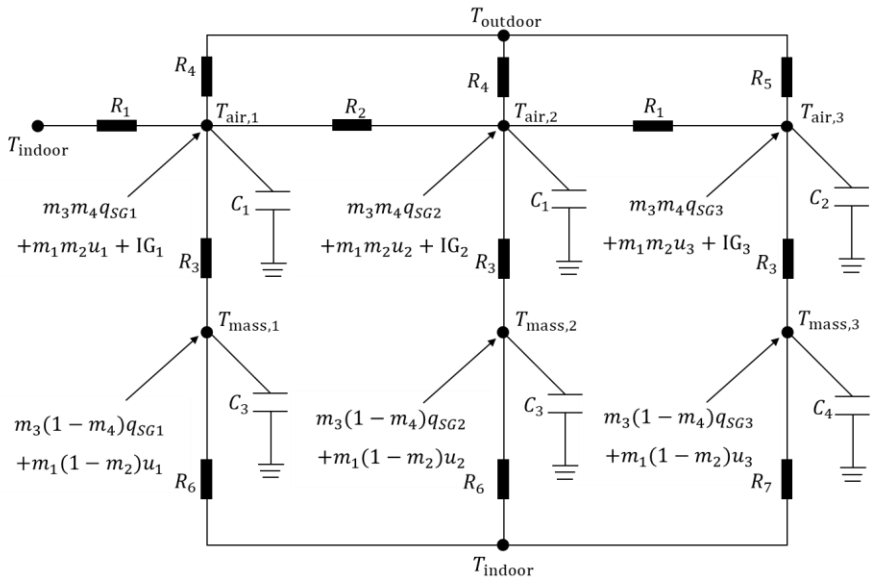


Figure 4.2. The RSS model for the building.

At every prediction horizon, the evolution of the building temperature states is predicted by a reduced-order state space (RSS) model, the graphical representation of which is shown in Figure 4.2.The building dynamics are given by the following linear equations,

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}_u\mathbf{u}_t + \mathbf{B}_w\mathbf{w}_t, \tag{4-2}$$

where $\mathbf{A} \in \mathbb{R}^{6\times6}$, $\mathbf{B}_u \in \mathbb{R}^{6\times3}$, and $\mathbf{B}_w \in \mathbb{R}^{6\times8}$ are time-invariant matrices. Solving equation (4-2) recursively over the prediction horizon gives,

$$\underbrace{\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\\ \mathbf{x}_3\\ \vdots\\ \mathbf{x}_K\end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix}\mathbf{B}_u & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{A}\mathbf{B}_u & \mathbf{B}_u & \cdots & \mathbf{0}\\ \mathbf{A}^2\mathbf{B}_u & \mathbf{A}\mathbf{B}_u & \cdots & \mathbf{0}\\ \vdots & \vdots & \cdots & \mathbf{0}\\ \mathbf{A}^{K-1}\mathbf{B}_u & \mathbf{A}^{K-2}\mathbf{B}_u & \cdots & \mathbf{B}_u\end{bmatrix}}_{\mathbf{G}_u}\underbrace{\begin{bmatrix}\mathbf{u}_0\\ \mathbf{u}_1\\ \mathbf{u}_2\\ \vdots\\ \mathbf{u}_{K-1}\end{bmatrix}}_{\mathbf{U}} + \underbrace{\begin{bmatrix}\mathbf{B}_w & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{A}\mathbf{B}_w & \mathbf{B}_w & \cdots & \mathbf{0}\\ \mathbf{A}^2\mathbf{B}_w & \mathbf{A}\mathbf{B}_w & \cdots & \mathbf{0}\\ \vdots & \vdots & \cdots & \mathbf{0}\\ \mathbf{A}^{K-1}\mathbf{B}_w & \mathbf{A}^{K-2}\mathbf{B}_w & \cdots & \mathbf{B}_w\end{bmatrix}}_{\mathbf{G}_w}\underbrace{\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_1\\ \mathbf{w}_2\\ \vdots\\ \mathbf{w}_{K-1}\end{bmatrix}}_{\mathbf{W}}$$

$$+ \underbrace{\begin{bmatrix}\mathbf{A}\\ \mathbf{A}^2\\ \mathbf{A}^3\\ \vdots\\ \mathbf{A}^K\end{bmatrix}}_{\mathbf{H}}\mathbf{x}_0. \tag{4-3}$$

The state variables $\mathbf{x}_t = \begin{bmatrix} \mathbf{T}_{\text{air},t} \\ \mathbf{T}_{\text{mass},t} \end{bmatrix}$ include the room air temperatures $\mathbf{T}_{\text{air},t}$ and mass temperatures $\mathbf{T}_{\text{mass},t}$ at time t,

$$\mathbf{T}_{\text{air},t} = \begin{bmatrix} T_{\text{air},1,t} \\ T_{\text{air},2,t} \\ \vdots \\ T_{\text{air},N,t} \end{bmatrix}, \mathbf{T}_{\text{mass},t} = \begin{bmatrix} T_{\text{mass},1,t} \\ T_{\text{mass},2,t} \\ \vdots \\ T_{\text{mass},N,t} \end{bmatrix},$$

where $i = 1, 2, \cdots N$, is the room index, with $N = 3$ for the present study. The control variables $\mathbf{u}_t$ are the heating/cooling rates to the offices,

$$\mathbf{u}_t = \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{N,t} \end{bmatrix}.$$

The elements in the disturbance vector $\mathbf{w}_t$ are: (i) The solar gain to the offices ($\mathbf{q}_{SG}$); (ii) The internal heat gains ($\mathbf{IG}$); (iii) The outdoor air temperature (T_{out}); and (iv) The corridor temperature (T_{indoor}) constantly set to 21°C.

$$\mathbf{w}_t = \begin{bmatrix} \mathbf{q}_{SG,t} \\ \mathbf{IG}_t \\ T_{\text{out},t} \\ T_{\text{indoor}} \end{bmatrix}, \mathbf{q}_{SG,t} = \begin{bmatrix} q_{SG1,t} \\ q_{SG2,t} \\ \vdots \\ q_{SGN,t} \end{bmatrix}, \mathbf{IG}_t = \begin{bmatrix} \text{IG}_{1,t} \\ \text{IG}_{2,t} \\ \vdots \\ \text{IG}_{N,t} \end{bmatrix}.$$

The parameters identified in the RSS model are the elements of state matrix **A,** input matrix $\mathbf{B}_u$, and disturbance matrix $\mathbf{B}_w$. Each element in these matrices has a form of multiplication of the thermal capacities ($C_{1:4}$), resistances ($R_{1:7}$), as well as the coefficients multiplied to the heat flux input for the transmitted solar radiation (m_3, m_4) and heating/cooling rate (m_1, m_2). To estimate those parameters, we collected the room air temperature, heating/cooling rate and disturbance measurements from the offices every 5 minutes for 15 days (10 days for training, 5 days for validation). To ensure sufficient excitation, the setpoint temperatures were ranging from 19 ℃ to 27 ℃. Using this dataset, we solved an optimization problem minimizing the mean absolute error (MAE) between the air temperature measurements and the air temperatures predicted by Eq. (2) for all the rooms. The validation root mean square error (RMSE) for $\mathbf{T}_{\text{air}}$ is 0.65°C, while $\mathbf{T}_{\text{mass}}$ are treated as unobserved states because they are impractical to measure in actual office building settings.

4.2.2.2 Cost function and constraints

At each time step, the cost is the sum of the energy consumption from the fan, chiller and the boiler that provides hot water for the reheat coil:

$$J_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) = \left(P_{\text{fan},t} + P_{\text{chiller},t} + P_{\text{reheat},t}\right)\Delta t, \tag{4-4}$$

where Δt is the length of a time step (0.5 hours). The details of the fan, chiller and reheat power models are presented in Appendix D.

Constraints are imposed on the control inputs based on the HVAC heating and cooling capacity:

$$u_{i,t,\min} \le u_{i,t} \le u_{i,t,\max}, \tag{4-5}$$

$$\underbrace{\begin{bmatrix} \mathbf{u}_{0,\min} \\ \mathbf{u}_{1,\min} \\ \mathbf{u}_{2,\min} \\ \vdots \\ \mathbf{u}_{K-1,\min} \end{bmatrix}}_{\mathbf{U}_{\min}} \le \underbrace{\begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_{K-1} \end{bmatrix}}_{\mathbf{U}} \le \underbrace{\begin{bmatrix} \mathbf{u}_{0,\max} \\ \mathbf{u}_{1,\max} \\ \mathbf{u}_{2,\max} \\ \vdots \\ \mathbf{u}_{K-1,\max} \end{bmatrix}}_{\mathbf{U}_{\max}}, \tag{4-6}$$

where $u_{i,t,\max}$ is the upper control bound for office i at time t, when the reheat coil runs at its maximum capacity while the damper is at the minimum position. $u_{i,t,\min}$ is the lower control bound at time t when the damper is at the maximum open position while the reheat coil is off.

Also, the room air temperatures must be kept within certain bounds to maintain thermal comfort:

$$T_{i,t,\min} \le T_{\text{air},i,t} \le T_{i,t,\max}, \tag{4-7}$$

where $T_{i,t,\max} = 25℃$ and $T_{i,t,\min} = 20℃$ are upper and lower bounds of the office setpoint temperatures, respectively based on ASHRAE Standard 55 (ASHRAE & ANSI, 2017). Given equation (4-3) we have:

$$\underbrace{\begin{bmatrix} \mathbf{T}_{1,\min} \\ \mathbf{T}_{2,\min} \\ \vdots \\ \mathbf{T}_{K,\min} \end{bmatrix}}_{\mathbf{T}_{\min}} \le \mathbf{M}(\mathbf{G}_u\mathbf{U} + \mathbf{G}_w\mathbf{W} + \mathbf{H}\mathbf{x}_0) \le \underbrace{\begin{bmatrix} \mathbf{T}_{1,\max} \\ \mathbf{T}_{2,\max} \\ \vdots \\ \mathbf{T}_{K,\max} \end{bmatrix}}_{\mathbf{T}_{\max}}, \tag{4-8}$$

where $\mathbf{T}_{t,\min} = \begin{bmatrix} T_{1,t,\min} \\ T_{2,t,\min} \\ \vdots \\ T_{N,t,\min} \end{bmatrix}$; $\mathbf{T}_{t,\max} = \begin{bmatrix} T_{1,t,\max} \\ T_{2,t,\max} \\ \vdots \\ T_{N,t,\max} \end{bmatrix}$; $\mathbf{M} = \begin{bmatrix} \mathbf{I}_N & \mathbf{0}_{N,N} \end{bmatrix}_{1\times K}$ consists of identity matrices $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ with ones on the main diagonal and zeros elsewhere, and zero matrices $\mathbf{0}_{N,N} \in \mathbb{R}^{N\times N}$ with all the entries being zeros.

Therefore, based on equations (4-6) and (4-8), the constraints on the control input and room air temperature can be written as:

$$\begin{bmatrix} \mathbf{M}\mathbf{G}_u \\ -\mathbf{M}\mathbf{G}_u \\ \mathbf{I}_{N\cdot K} \\ -\mathbf{I}_{N\cdot K} \end{bmatrix} \mathbf{U} \leq \begin{bmatrix} \mathbf{T}_{\max} - \mathbf{M}\mathbf{G}_w\mathbf{W} - \mathbf{M}\mathbf{H}\mathbf{x}_0 \\ \mathbf{M}\mathbf{G}_w\mathbf{W} + \mathbf{M}\mathbf{H}\mathbf{x}_0 - \mathbf{T}_{\min} \\ \mathbf{U}_{\max} \\ -\mathbf{U}_{\min} \end{bmatrix}. \tag{4-9}$$

At each prediction horizon, a constrained nonlinear optimization problem described with the cost function (4-1) and constraint (4-9) is solved to obtain the sequence of optimal control inputs **U**. We use the nonlinear programming solver 'fmincon' with sequential quadratic programming algorithm in MATLAB environment.

4.2.2.3 MPC implementation

Figure 4.3 presents the schematics of data communication for the MPC implementation. The calculations were performed using a server computer with MATLAB. For the disturbance prediction at every prediction horizon, we downloaded the NOAA weather forecast data including the outdoor air temperature, sky-cover and relative humidity. The predicted solar gains to the offices were calculated using these data based on the model by Seo (2010). The occupancy schedule was from 10 a.m. to 5 p.m. The internal heat gain from the occupants, lighting, and equipment were set to 75, 128, and 105 W/office, respectively, based on the data from the test-bed. The outputs from the MPC algorithm are the optimal heating/cooling rates for the offices. Subsequently, we compute the setpoint temperatures resulting from the optimal heating/cooling rates using equation (4-2), and send these setpoints to BMS through Modbus protocol for implementation. After implementing the new setpoints at each control horizon, we send to the server computer real-time sensor measurements for the air temperature, supply air temperature and flow rate, and exogenous inputs (transmitted solar irradiance and outdoor air temperature) to estimate the unobserved initial state ($\mathbf{T}_{\text{mass}}$) using Kalman filter (Stengel, 1994).

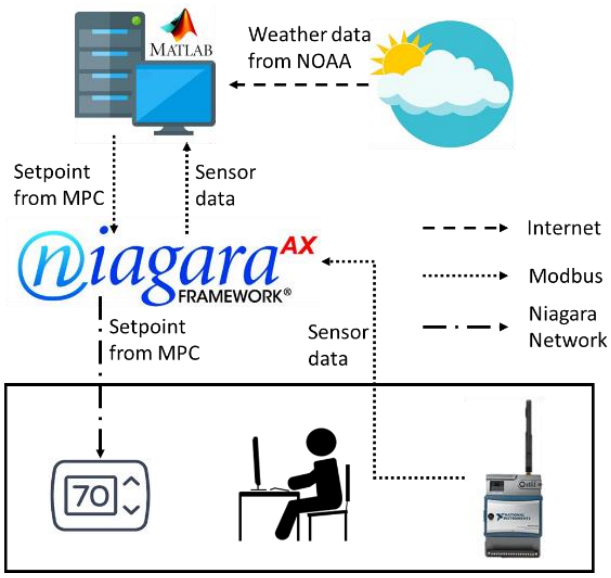
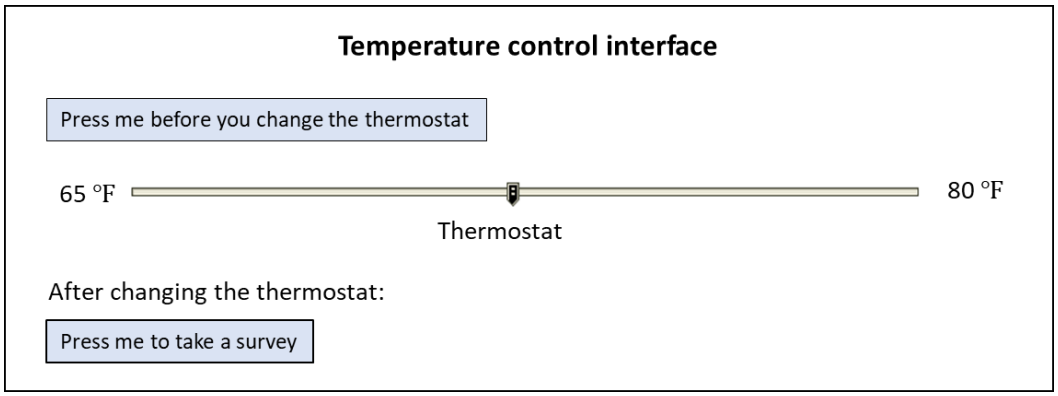


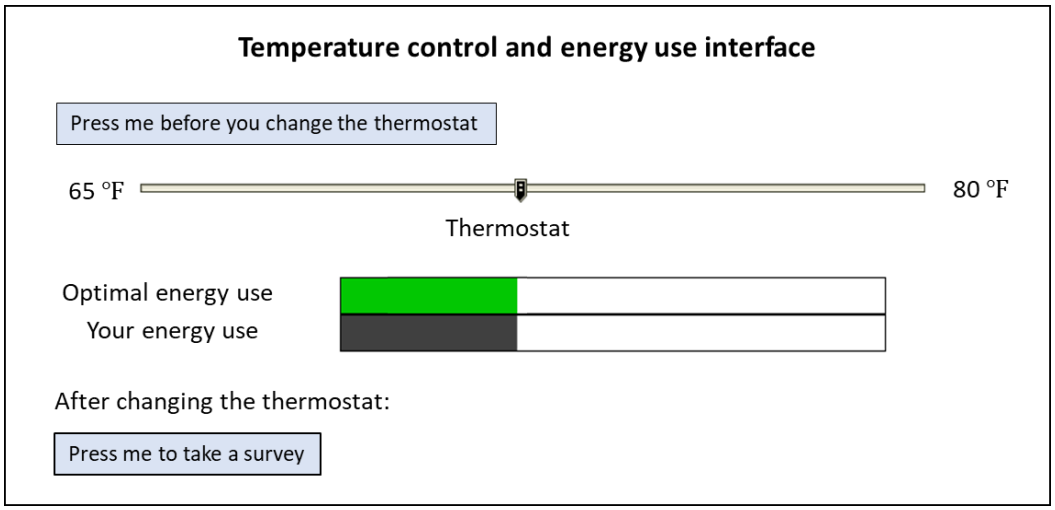
Figure 4.3. Data communication for MPC implementation.

4.2.3 User-interface

As discussed in the previous section, the setpoint temperatures determined by the MPC algorithm were implemented in the building management system. Two user-interfaces were developed using Niagara GUI Workbench. The temperature control interface in Setup 1 (Figure 4.4 (a)) serves as the interface for the control group in the field experiment. The interface for Setup 2 (Figure 4.4 (b)) displays the energy use information in addition to providing temperature control. In both interfaces, occupants could temporarily override the default setpoints with each override session lasting for up to 1 hour. In this way, minimal effort was required from the occupants to maximize energy savings (by accepting the default setpoint) in Setup 2. After changing the setpoint, occupants can follow a link at the bottom of the interfaces to provide feedback regarding their thermal preference (prefer cooler, slightly cooler, slightly warmer, warmer, or satisfied with the current condition).



(a) Temperature control interface (Setup 1)



(b)Temperature control and energy use interface (Setup 2)

Figure 4.4. Web-interfaces implemented in the field study.

The MPC algorithm calculates the real-time hypothetical, optimal energy use for the following hour with equation (4-4) using the current temperatures, environmental disturbances, and the optimal controls as inputs. This information is provided to the occupants as a motivational 'goal' (green bar in Figure 4.4), allowing occupants to directly compare with their energy use (dark grey bar). The length of the bar visualizes the current accumulative energy usage (up to 10kWh) of a day. It is displayed right below the thermostat and updated in real-time so occupants can see this information when they wish to adjust the thermostat. Details of the functionality of the energy use interface are explained in Figure 4.5, which presents a user-interaction scenario.

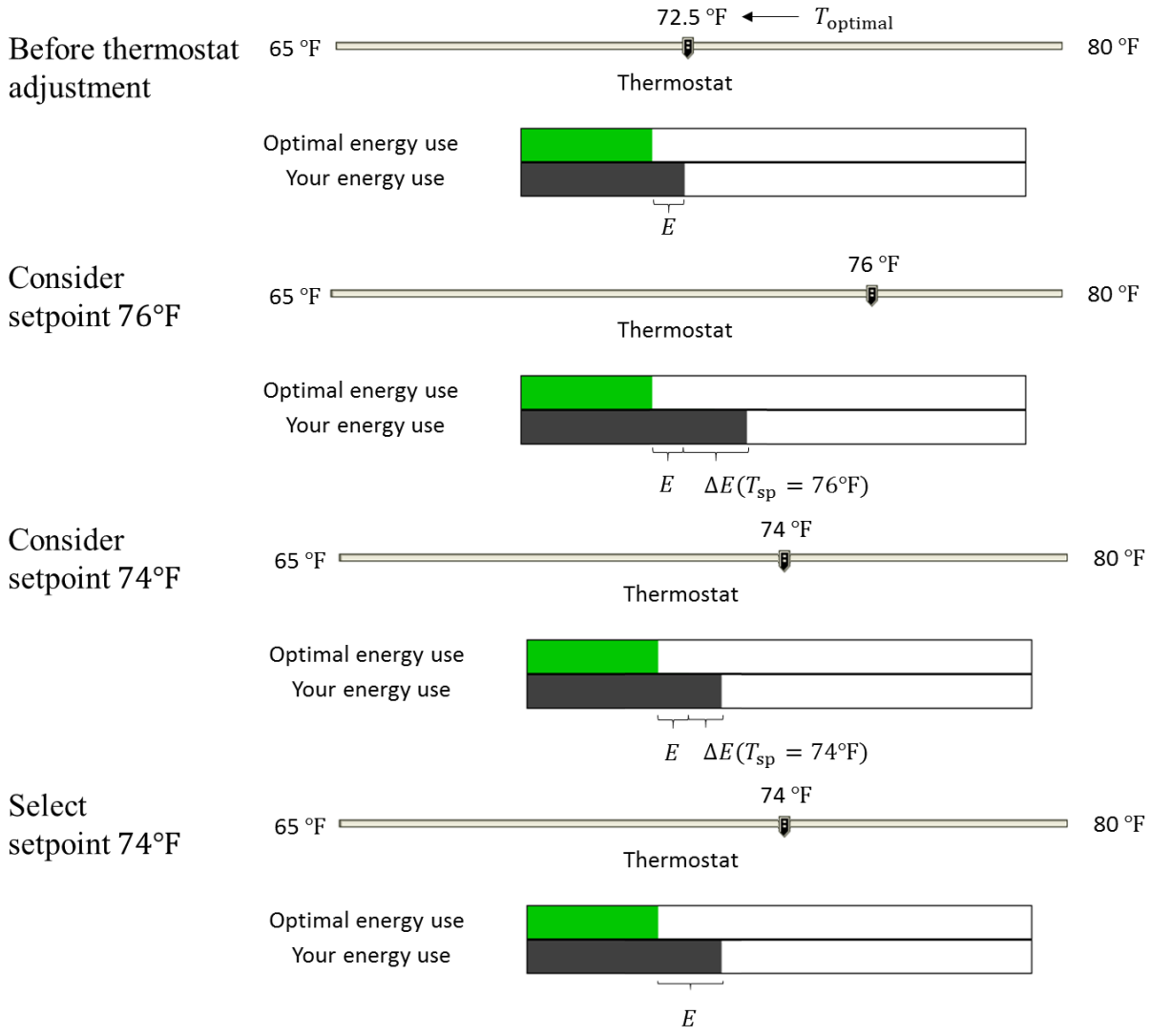


Figure 4.5. Functionality of the energy use interface.

If an occupant always accepts the default setpoint temperature T_{optimal}, the length of the dark grey bar would remain the same with the green bar. However, when an occupant overrides the default T_{optimal} and tries to select from the potential setpoints on the thermostat (T_{sp}), we instantly display to the dark grey bar the expected increase of energy use of the following hour (ΔE, which varies as T_{sp} varies) while the occupant moves the indicator on the thermostat slider. Through this design, occupants can visualize the energy impact of their temperature selection, and consider it along with their thermal preference to select a new setpoint. ΔE is approximated as follows:

$$\Delta E_t = \xi_t \cdot \left| T_{\text{sp}} - T_{\text{optimal},t} \right|. \quad (4\text{-}10)$$

As the building dynamics are linear, by plugging T_{sp} and T_{optimal} to the left-hand side of the linear equation (4-2) respectively, subtracting each other, we can establish that the difference between HVAC thermal inputs to achieve T_{sp} and $T_{\text{optimal},t}$ is proportional to $\left| T_{\text{sp}} - T_{\text{optimal},t} \right|$. Therefore, ΔE, which is approximately the difference between the HVAC thermal inputs divided by the overall HVAC system efficiency, is also proportional to $\left| T_{\text{sp}} - T_{\text{optimal},t} \right|$. The coefficient ξ_t depends on the real-time overall energy efficiency of the HVAC system and, ultimately, on the weather condition and system operating state. It is the inverse of the product of the HVAC system efficiency and the corresponding element of heat transfer coefficient in the input matrix $\mathbf{B}_u$.

4.2.4 Experimental procedure

The field experiment was conducted in the summer of 2018 (Jul. 3rd, 2018–Sept. 13th, 2018). 23 office occupants (12 males and 11 females) were recruited for the study. The participants were university students or staff (between 22 and 36 years old) who were not familiar with this research. Each private office was occupied by one occupant every day between 10 a.m. to 5 p.m. All occupants were asked to perform their usual office work (computer-related work, reading, writing, etc.) during the day, and they were free to take short breaks (within 10 minutes) and a lunch break for about 1 hour. They were asked to wear the clothes they would normally wear in the office in the summer, and not to change the clothes between 10 a.m. to 5 p.m. Details regarding the experimental setups were explained to the occupants when they arrived in the morning to help them become quickly familiar on their first days of each setup.

To identify the casual effect of the energy use information, each occupant used the interface shown in Figure 4.4 (a)- Setup 1 during the first for two days, and then another two days, the interface in Figure 4.4 (b)-Setup 2. In other words, Setup 1 provides the control group data. In each office, the web-interface was displayed on a monitor placed on the desks to be easily accessible (Figure 4.6). The occupants were given access to remote controls for the roller shades and electric lights. They were advised to interact with remote controls and the web-interfaces as they would typically do in their offices. The instrumentation was installed so that there was no interference with the occupants' regular position and task, and they were instructed to avoid any direct contact with the monitoring instrumentation. The occupants were asked to complete the override surveys

whenever they adjusted the thermostat (link available on the web-interface) and an exit survey at the end of each test-day to obtain information on their overall thermal satisfaction. The detailed questions for the surveys are provided in Appendix E. The study was approved by the Institutional Review Board (IRB Protocol #: 1503015873).

4.2.5 Data acquisition and instrumentation

The office layout and locations of the sensors are presented in Figure 4.6. All data were measured and recorded every minute. The following physical variables were monitored during the experiment.

- Air flow rate, supply air temperature, reheat coil mode, occupancy, thermostat override status, setpoint temperature, shading position, and electric light level. The sensors/actuators for these variables were connected with the existing BMS using BACnet protocol.
- Room air temperature (J-type thermocouples, resolution: 0.01 °C, accuracy: 0.4%), transmitted solar irradiance (LI-COR 200-SL pyranometer, resolution: 0.1 W/m^2, accuracy: 3%) on the façade. The sensors for these variables were connected with the National Instrument wireless data acquisition system, which communicated with the BMS using Modbus protocol.

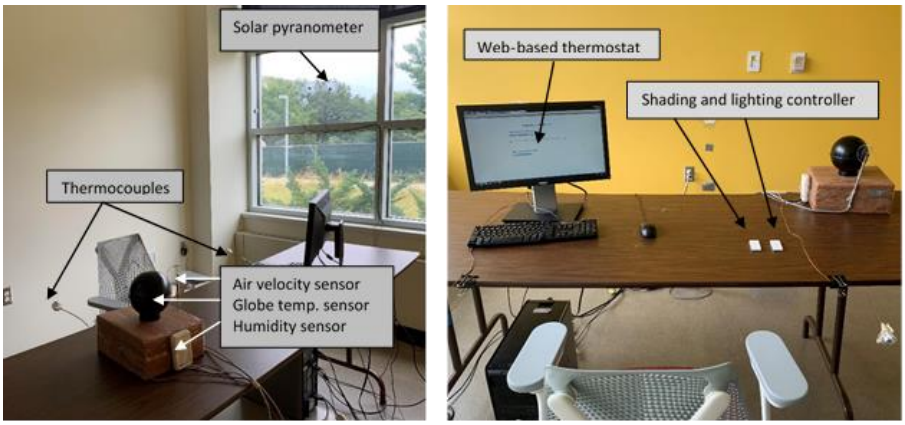


Figure 4.6. Sensor locations and office layout.

4.3 Experimental results

This section presents the field experiment results, analyzing the occupants' thermostat adjustment behavior, survey responses for Setup 1 and Setup 2, and the corresponding HVAC energy consumption.

The thermal preference votes for Setup 1 and 2 are presented in Figure 4.7. Overall, the mean setpoint temperature determined by MPC was 21.4℃ for Setup 1 and 21.5 for Setup 2; the median was 21.4℃, and the standard deviation was 0.8℃ for either setup. Under these conditions, most occupants reported preferring slightly warmer or warmer, while less preferred marginally cooler in the thermal preference votes. In Setup 1, we observed 108 thermostat adjustments (or overrides), and this value was 62 in Setup 2. Noticeably, around 71% less 'prefer slightly warmer' votes in Setup 2, which means fewer thermostat adjustments from the occupants due to their preference for slightly warmer conditions (51 overrides in Setup 1 compared to 15 in Setup 2). Since the occupants were exposed to similar thermal conditions in the two setups, they chose not to adjust the thermostats when deemed unnecessary considering the energy consumption and potential improvement in their comfort. Also, the survey responses in Setup 2 (Question 2 in Table E.1 Override survey, Appendix E) show that all the occupants considered the energy use information in the thermostat adjustments.

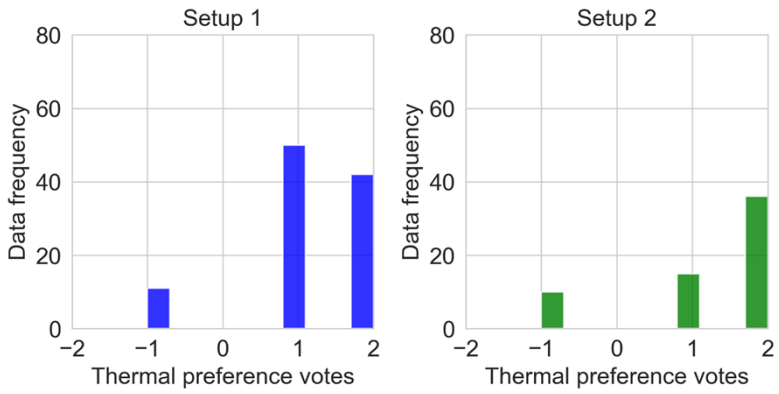


Figure 4.7. Occupants' thermal preference votes for Setup 1 and Setup 2 (-2: prefer cooler, -1: prefer slightly cooler, 0: satisfied with current condition, 1: prefer slightly warmer, 2: prefer warmer).

Figure 4.8 presents the setpoint temperatures selected by the occupants and the setpoint temperatures determined by the MPC controller before the thermostat adjustments per thermal

preference vote for both setups. In both setups, when the occupants preferred slightly cooler, they reduced the setpoint temperatures from 21.9℃ to 20.8℃ on average, and the number of overrides was similar (11 overrides in Setup 1 compared to the 10 in Setup 2, see Figure 4.7). The expected increase of energy use of such an adjustment was often as low as approximately 0.2 kWh (i.e., $\xi_t \approx 0.2$ kWh/℃) due to the relatively higher cooling energy efficiency. Therefore, the difference in energy use resulting from thermostat adjustment might be inconspicuous to the occupants, which could explain the similar behavior in Setup 1 and Setup 2.

When the occupants preferred slightly warmer, they increased the setpoint temperatures from 21.3℃ to 22.9℃ in Setup 1, and 22.3℃ in Setup 2 on average. Usually, higher setpoint temperatures require reheat operations, which would result in higher energy use because of the relatively lower efficiency. On average, the energy use would increase 1kWh for 1℃ increment in the setpoint (i.e. $\xi_t \approx 1$ kWh/℃). As this information was presented to the occupants, the difference in the setpoint selections and the reduced frequency of overrides (Figure 4.7) in the two setups indicates that they attempted to reduce their energy use in Setup 2. It also means that the occupants considered both their comfort and energy use in the decision-making.

Similarly, when the occupants preferred warmer conditions, the occupants increased the setpoints from 21.1℃ to 23.6℃ in Setup 1, and 23℃ in Setup 2 on average. However, the number of overrides only reduced from the 42 in Setup 1 to the 36 in Setup 2 (Figure 4.7). The reason could be that the occupants considered it necessary to increase the setpoint to improve their comfort, while attempting to mitigate their energy impact. It also needs to be noted that the standard deviation of the occupants' selected setpoint temperatures decreased from the 1.1℃ in Setup 1 to the 0.7℃ in Setup 2. The reason could be that the occupants' selections were more consistent in the deliberate decision-making, rather than randomly setting the thermostats to a higher setpoint. The setpoint temperature profiles from some representative occupants are presented in Appendix F.

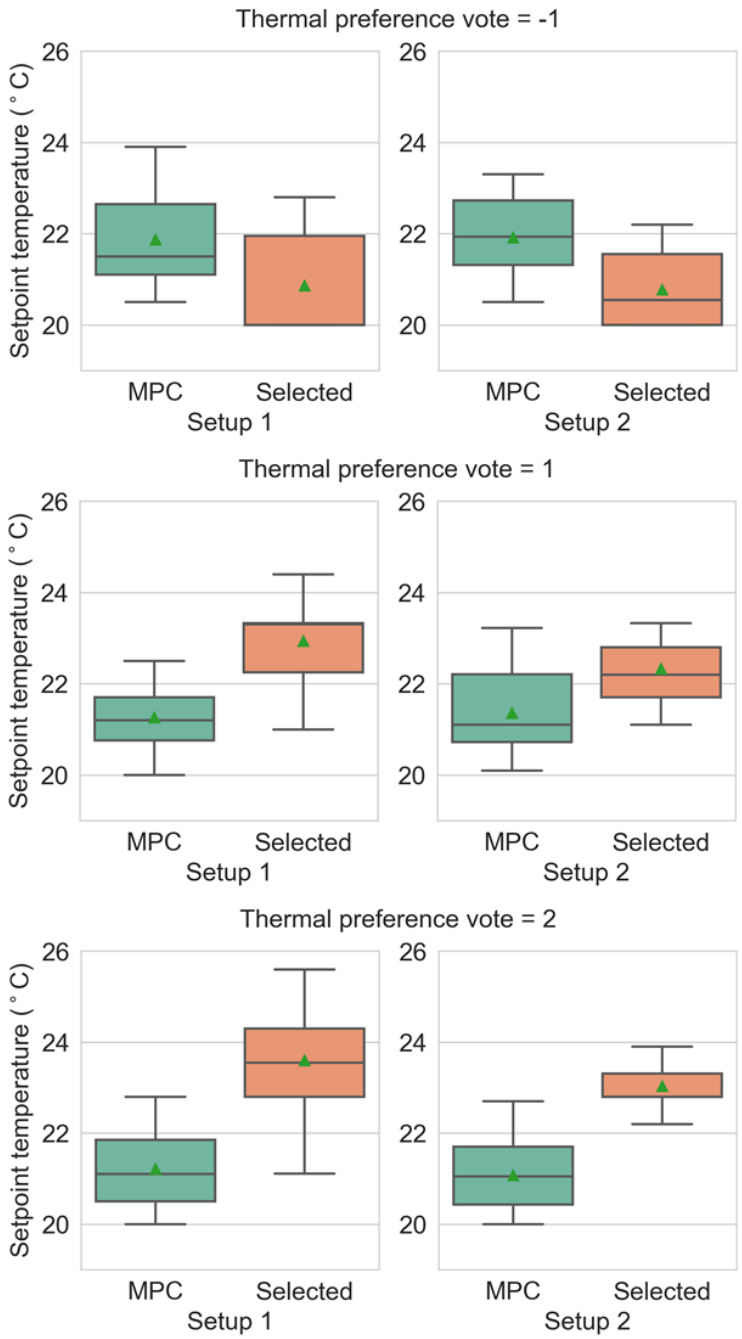


Figure 4.8. The default (MPC) and occupants' selected setpoint temperatures per thermal preference vote for Setup 1 and 2.

Figure 4.9 compares the daily HVAC energy use until 5 p.m. for Setup 1 and 2. The mean optimal daily HVAC energy consumption for both setups was similar (2.1 kWh for Setup 1 and 2.2 kWh for Setup 2). Therefore, the differences in the two setups' actual consumption were mainly

attributed to the occupant overrides. The mean actual daily HVAC energy consumption per office in Setup 1 was 4.9 kWh, while in Setup 2 it was reduced to 3.9 kWh. By comparing the mean optimal and actual consumption for Setup 1, it can be seen that the occupants' overrides resulted in significant addition to the energy use of the smart HVAC control (i.e., 55% of the energy use in Setup 1). This confirms that occupant interactions with thermostats should be taken into account in the realistic energy saving estimations of advanced control strategies, as it significantly affects energy use. However, 36% of the additional energy consumption associated with the occupants' overrides can be recovered with the web-interface implemented in Setup 2. The recovered energy results from occupants' acceptance of the default setpoints, as well as the deliberate decision making and consideration of energy use when selecting the new setpoints. In addition, the standard deviation of the daily actual energy consumption was reduced from the 1.8 kWh in Setup 1 to 0.7 kWh in Setup 2, suggesting that the energy saving in Setup 2 could be consistent.

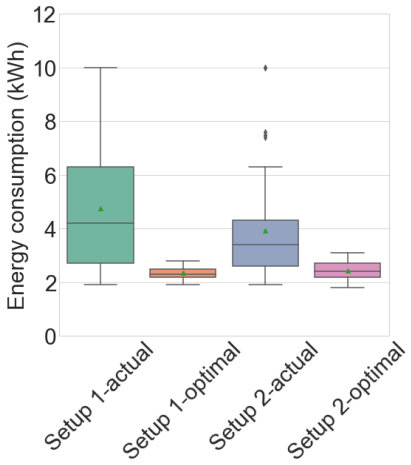


Figure 4.9. Daily HVAC energy use (until 5 p.m.) comparison in Setup 1 and 2.

Occupants' survey responses on their overall discomfort experience during a day (see the exit survey on Table E.1, Appendix E) are presented in Figure 4.10. It can be seen that there is no significant difference between Setup 1 and Setup 2, and in both setups, occupants only 'rarely' experienced discomfort conditions for more than 30 minutes on average. Therefore, the thermal conditions in Setup 2 achieved lower energy consumption with the same level of overall occupant satisfaction compared to Setup 1.

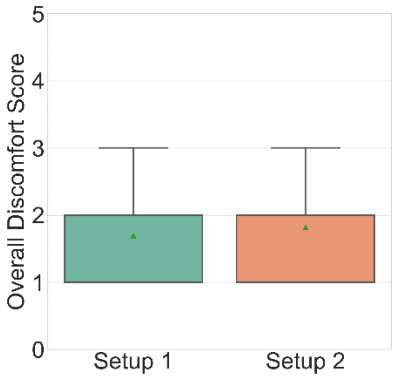


Figure 4.10. Occupants' survey responses about overall discomfort (1: never, 2: rarely, 3: a few times, 4: most of the time, 5: always).

4.4 Human decision-making model for thermal environment control

4.4.1 Modeling the decision-making process

Our experimental data show that occupants select a setpoint temperature by evaluating (i) the expected comfort level (in Setup 1 and Setup 2) and (ii) the resulting energy use (in Setup 2 only). We assume that there are occupant-specific utility functions U_1 and U_2 that quantify the subjects' preference over the potential choices in Setup 1 and 2, respectively. Occupants prefer choices with higher utility values. Therefore, the occupants' decision-making process for setpoint selections can be seen as maximizing their utilities. Note that these two utility functions are not completely independent. Specifically, U_1 is a special case of U_2 and thus the two share some parameters. However, we use two separate functions in order to simplify the mathematical notation.

4.4.1.1 Decision-making model for Setup 1

Let us derive a plausible mathematical form for the utility of an occupant under Setup 1. In this setup, the occupant can choose the setpoint T_{sp} from the set of setpoints $\mathcal{T} = [18.3℃, 26.7℃]$ without any consideration of the energy consumption. So, the utility U_1 is a function of T_{sp}, i.e., $U_1 = U_1(T_{\text{sp}})$. Assuming all other ambient conditions are relatively constant, e.g., humidity, air velocity, there must exist a setpoint T_{sp}^* which the occupant prefers because it delivers the maximum comfort. The simplest mathematical form with this property is:

$$U_1(T_{\mathrm{sp}}) = U_1(T_{\mathrm{sp}}; a, T_{\mathrm{sp}}^*) = -a(T_{\mathrm{sp}} - T_{\mathrm{sp}}^*)^2, \tag{4-11}$$

where a is a positive parameter that describes how sensitive the occupants' preference is to the square distance between the choice T_{sp} and the ideal T_{sp}^*. In this way, the utility U_1 exhibits a maximum at:

$$\underset{T_{\mathrm{sp}}}{\operatorname{argmax}}\, U_1(T_{\mathrm{sp}}; a, T_{\mathrm{sp}}^*) = T_{\mathrm{sp}}^*. \tag{4-12}$$

4.4.1.2 Decision-making model for Setup 2

We now derive the mathematical form for the utility function in Setup 2. The difference between the two setups is that the occupant also receives information from the energy use portal. Denote this energy use information by F for feedback. We assume that the feedback can be summarized by the expected increase of energy use (ΔE, see equation (4-10)) and the difference between the optimal energy use and the occupant's energy use before adjusting the thermostat (E).

$$F(T_{\mathrm{sp}}, E, T_{\mathrm{optimal}}, \xi) = E + \Delta E = E + \xi |T_{\mathrm{sp}} - T_{\mathrm{optimal}}|. \tag{4-13}$$

So, $U_2 = U_2(T_{\mathrm{sp}}, F)$. We expect that for any fixed F the occupant would prefer the setpoint that makes them most comfortable. That is, the function $U_2(T_{\mathrm{sp}}, F)$ exhibits a maximum at $T_{\mathrm{sp}} = T_{\mathrm{sp}}^*$ when F is held constant. Furthermore, we expect that the occupant will always prefer to consume less energy for any T_{sp}. That is, we expect that $U_2(T_{\mathrm{sp}}, F)$ is a decreasing function of F when T_{sp} is kept constant. In other words, $U_2(T_{\mathrm{sp}}, F)$ exhibits a maximum at $F = 0$. However, even though the occupant would always prefer $F = 0$ and $T_{\mathrm{sp}} = T_{\mathrm{sp}}^*$, this choice is not possible as F is also affected by E, T_{optimal} and ξ, which are not controlled by the occupant. Instead, during each thermostat adjustment, the occupant can only choose a T_{sp} that is between T_{optimal} ($\Delta E = 0$ when $T_{\mathrm{sp}} = T_{\mathrm{optimal}}$) and T_{sp}^*. Given that F is correlated to the distance between T_{sp} and T_{optimal} via equation (4-13), the simplest mathematical form of U_2 is:

$$U_2(T_{\mathrm{sp}}, E, T_{\mathrm{optimal}}, \xi; a, c, T_{\mathrm{sp}}^*) = -a(T_{\mathrm{sp}} - T_{\mathrm{sp}}^*)^2 - cF(T_{\mathrm{sp}}, E, T_{\mathrm{optimal}}, \xi)^2 \tag{4-14}$$

where c is a positive parameter that captures the importance that the occupant attributes to the feedback. In this way, the maximum of U_2 at given $E, T_{\mathrm{optimal}}, \xi$ is exhibited at:

$$\left.\frac{\partial U_2}{\partial T_{\mathrm{sp}}}\right|_{T_{\mathrm{sp}} = \underset{T_{\mathrm{sp}}}{\operatorname{argmax}}\, U_2} = 0. \tag{4-15}$$

Writing down the details of the left-hand side of equation (4-15), which is the partial derivative of U_2 regarding T_{sp}, we have:

$$\frac{\partial U_2}{\partial T_{\mathrm{sp}}} = -2aT_{\mathrm{sp}} + 2aT_{\mathrm{sp}}^* - 2c\xi^2 T_{\mathrm{sp}} + c\xi^2 T_{\mathrm{optimal}} - 2cE\xi = 0 \tag{4-16}$$

Therefore, given equations (4-16), re-arranging to a function of T_{sp}, and (4-15), the occupants' setpoint selections in Setup 2 can be predicted by,

$$\underset{T_{\mathrm{sp}}}{\operatorname{argmax}}\, U_2\left(T_{\mathrm{sp}}, E, T_{\mathrm{optimal}}, \xi; a, c, T_{\mathrm{sp}}^*\right) = \frac{-aT_{\mathrm{sp}}^* - c\xi^2 T_{\mathrm{optimal}} + cE\xi}{-a - c\xi^2}. \tag{4-17}$$

4.4.2 Calibrating the model using the experimental data

In the field experiment, we collected datasets $\boldsymbol{\mathcal{D}} = (\mathcal{D}_1, \mathcal{D}_2)$, in which there are M and N observations from Setup 1 and Setup 2, respectively. The observed data are $\mathcal{D}_1^{(i)} = \left(T_{\mathrm{sp,obs},1}^{(i)}\right)$, and $\mathcal{D}_2^{(j)} = \left(T_{\mathrm{sp,obs},2}^{(j)}, E^{(j)}, T_{\mathrm{optimal}}^{(j)}, \xi^{(j)}\right)$, where $i = 1,2,\cdots M$; $j = 1,2,\cdots N$ are the data indices for the two setups. $\mathbf{T}_{\mathrm{sp,obs}} = (T_{\mathrm{sp,obs},1}, T_{\mathrm{sp,obs},2})$ are the observed setpoint selections from the occupants in Setup 1 and Setup 2, respectively. The likelihood of observing $T_{\mathrm{sp,obs},1}^{(i)}$ and $T_{\mathrm{sp,obs},2}^{(j)}$ can be expressed as:

$$p\left(T_{\mathrm{sp,obs},1}^{(i)} \middle| a, T_{\mathrm{sp}}^*\right) = \mathcal{N}\left(T_{\mathrm{sp,obs},1}^{(i)} \middle| \mu_1^{(i)}, \sigma_1^2\right), \tag{4-18}$$

$$p\left(T_{\mathrm{sp,obs},2}^{(j)} \middle| E^{(j)}, T_{\mathrm{optimal}}^{(j)}, \xi^{(j)}, a, c, T_{\mathrm{sp}}^*\right) = \mathcal{N}\left(T_{\mathrm{sp,obs},2}^{(j)} \middle| \mu_2^{(j)}, \sigma_2^2\right), \tag{4-19}$$

where $\mathcal{N}(\cdot\,|\mu, \sigma^2)$ is the probability density function (PDF) of a univariate Gaussian distribution with mean μ and standard deviation σ. Given equations (4-12) and (4-17) that predict the occupants' setpoint selections, the means on the right-hand side of equations (4-18) and (4-19) are:

$$\mu_1^{(i)} = \underset{T_{\mathrm{sp}}}{\operatorname{argmax}}\, U_1(T_{\mathrm{sp}}; a, T_{\mathrm{sp}}^*), \tag{4-20}$$

$$\mu_2^{(j)} = \underset{T_{\mathrm{sp}}}{\operatorname{argmax}}\, U_2\left(T_{\mathrm{sp}}, E^{(j)}, T_{\mathrm{optimal}}^{(j)}, \xi^{(j)}; a, c, T_{\mathrm{sp}}^*\right). \tag{4-21}$$

In reality, due to the occupants' limited cognitive capacity (i.e. the occupants do not have the exact utility function in their mind but rather some internal criteria for the energy use and

comfort consideration), and other constraints (e.g. the limited amount of time that occupants can spend on the selection, etc.), the occupants would not accurately evaluate their preferences on the setpoints options (Simon, 1990). Therefore, we assume that $\mathbf{T}_{\text{sp,obs}}$ are 'near' rather than exactly equal to the setpoint at the maximum of the utility. The standard deviation $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$ on the right-hand side of equations (4-18) and (4-19) accounts for such differences, as well as the personal differences among a group of occupants. σ_1 and σ_2 are setup specific, as in Setup 2 this difference could be attributed to hidden factors related to the occupants' understanding of energy use, which do not exist in Setup 1.

Figure 5.11 is the graphical representation of the causal factors affecting the occupants' setpoint selection. The variables in green represent the factors related to energy use consideration. In green solid circles is the information from the energy use portal, in which T_{optimal} and ξ reflect the smart HVAC operation based on the building and energy system dynamics, as well as the environmental disturbances. The variables in orange circles are the factors related to comfort consideration; while the utility function parameters are in dashed circles. Along with $\boldsymbol{\sigma}$, all these variables together affected the occupants' setpoint selections. Based on these data $\boldsymbol{\mathcal{D}}$ collected from the experiment (shaded circles), we can infer the unobserved variables $\boldsymbol{\theta} = (a, c, T^*_{\text{sp}}, \boldsymbol{\sigma})$ in the blank circles.

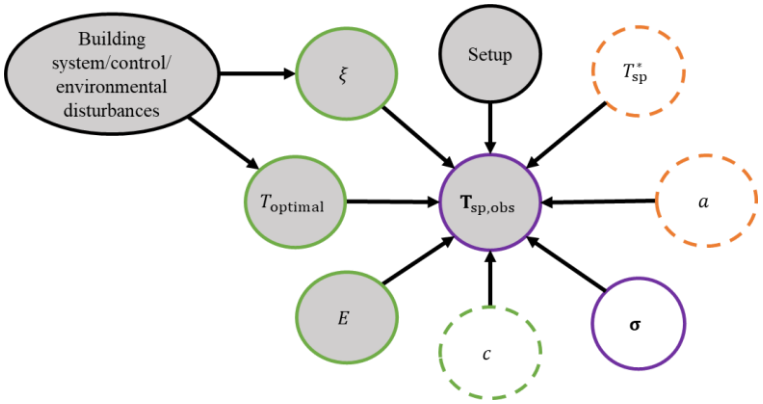


Figure 4.11. Causal factors affecting the occupants' setpoint selection.

We estimated the unobserved variables using a Bayesian modeling approach. This allows the quantification of uncertainty considering the proposed model form and limited amount of observed data. According to Bayes rule, we have:

$$p(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}}) \propto p(\boldsymbol{\mathcal{D}}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}), \tag{4-22}$$

where $p(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}})$ is the posterior distribution of the unobserved variables; $p(\boldsymbol{\mathcal{D}}|\boldsymbol{\theta})$ is the data likelihood described in equations (4-18) and (4-19); $p(\boldsymbol{\theta})$ is the prior distribution of the unobserved variables, and we assigned the following uninformative priors:

$$p(\vartheta) = \exp(\vartheta|0.1), \tag{4-23}$$

$$p\left(T_{\mathrm{sp}}^{*}\right) = \mathcal{N}\left(T_{\mathrm{sp}}^{*}\middle|23, 3^{2}\right), \tag{4-24}$$

where $\vartheta = a, c, \boldsymbol{\sigma}$; and $\exp(\cdot\,|\lambda)$ is the PDF of an exponential distribution with rate parameter λ. Python PyMC 2.3.7 package (Patil *et al*., 2010) was employed to code the model and to sample from the posterior of the unobserved variables with Markov chain Monte Carlo (MCMC) method. Adaptive Metropolis-Hastings sampler was used. After analyzing the traces and autocorrelations of the chain, we use in total 500000 samples and discarded the first 100000. 2000 samples are gathered by keeping one MCMC sample out of every 200. It should be noted that we excluded the data from one occupant who is highly likely to belong to a different thermal preference cluster (Lee *et al*., 2017) than others (see Appendix F). However, we believe that the forms of the decision-making model and utility function considered in this study are still applicable to occupants from other clusters as well.

4.4.3 Modeling results

Table 4.1. Descriptive statistics of the inferred unobserved variables.

Variable	Mean	Median	Standard deviation	95% credible interval
T_{sp}^{*}	23.095	23.095	0.106	[22.884, 23.297]
a	0.582	0.511	0.245	[0.162, 0.965]
c	0.221	0.195	0.122	[0.056, 0.433]
σ_1	1.045	1.039	0.082	[0.898, 1.211]
σ_2	1.035	1.028	0.079	[0.912, 1.123]

Based on the inferred parameters in Table 4.1, we present the posterior median contour of the utility function in Figure 4.12. The utility value becomes higher as the setpoint temperature (x-axis) approaches around 23.1℃ (equals to the median and mean of inferred T_{sp}^{*}), and as the difference with the optimal energy use (y-axis) gets close to 0. Also, for any potential setpoint the

utility value becomes higher as the additional energy use approaches 0, and for the same amount of additional energy use the utility value is higher as the setpoint gets closer to T_{sp}^{*}. Therefore, it represents the fact that, ideally occupants would prefer to select T_{sp}^{*} with as less additional energy use as possible.

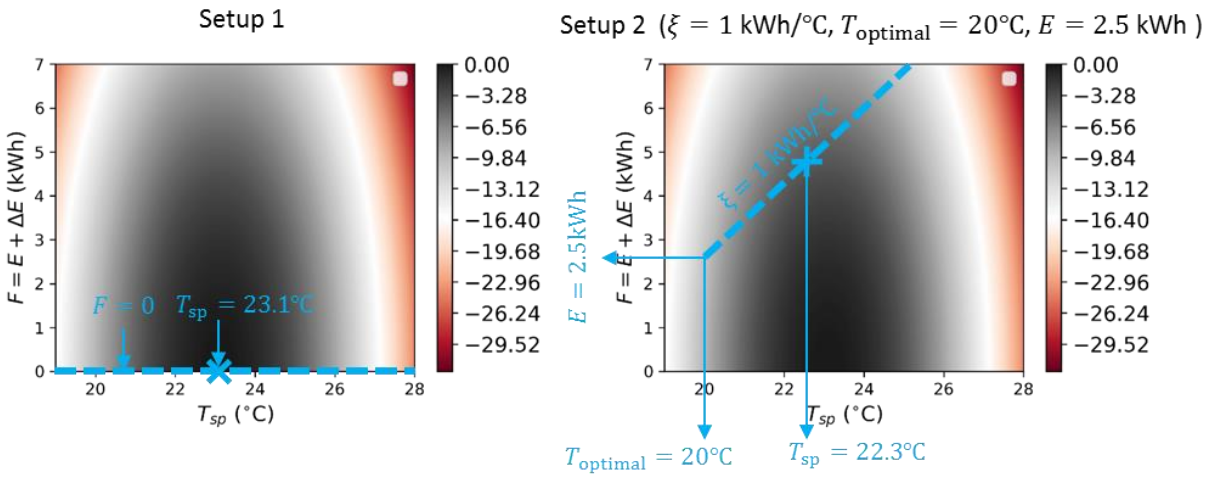


Figure 4.12. Posterior median contour of the utility with intersection lines representing Setup 1 (left) and a scenario in Setup 2 (right).

The dash line in Figure 4.12 (left) represents the scenario for Setup 1 in which energy use information is not displayed to the occupants ($F = 0$ from the utility contour). Figure 4.13 (top left) details the utility value for all the potential setpoint temperatures for Setup 1 considering 95% credible interval for the utility function parameters. Using the setpoint (median) with the maximum utility value for Setup 1 ($T_{\text{sp}} = 23.1$°C) as a reference, we compute the probability of any potential setpoint being preferred by the occupants than the reference condition in Setup 1. This result is presented in Figure 4.13 (top right) and it shows that, when the occupants do not consider energy use, their preference on the setpoints are sensitive to the distance from T_{sp}^{*}.

A typical scenario is represented with the dash line from the utility contour in Figure 4.12 (right). In this scenario the current temperature determined by MPC is 20°C (T_{optimal}) and occupants want to increase the setpoint using the web-interface in Setup 2. In the meantime, before changing the setpoint the occupants observe a visual presentation of the difference with the optimal energy use (E) at 2.5kWh, while the expected energy use increases at the rate of $\xi = 1$ kWh/°C

based on the system and environmental condition. Figure 4.13 (bottom left) details the utility value for the potential setpoint temperatures for the scenario considered, taking into account 95% credible interval of the utility model parameters. The median of setpoint temperatures with the maximum utility value is 22.3℃, which is increased from the default temperature of 20℃, resulting in 2.3kWh of additional energy use. Compared to selecting 23.1℃ (T_{sp}^{*}) with 3.1kWh of additional energy use, the occupants are more likely to save 0.8 kWh.

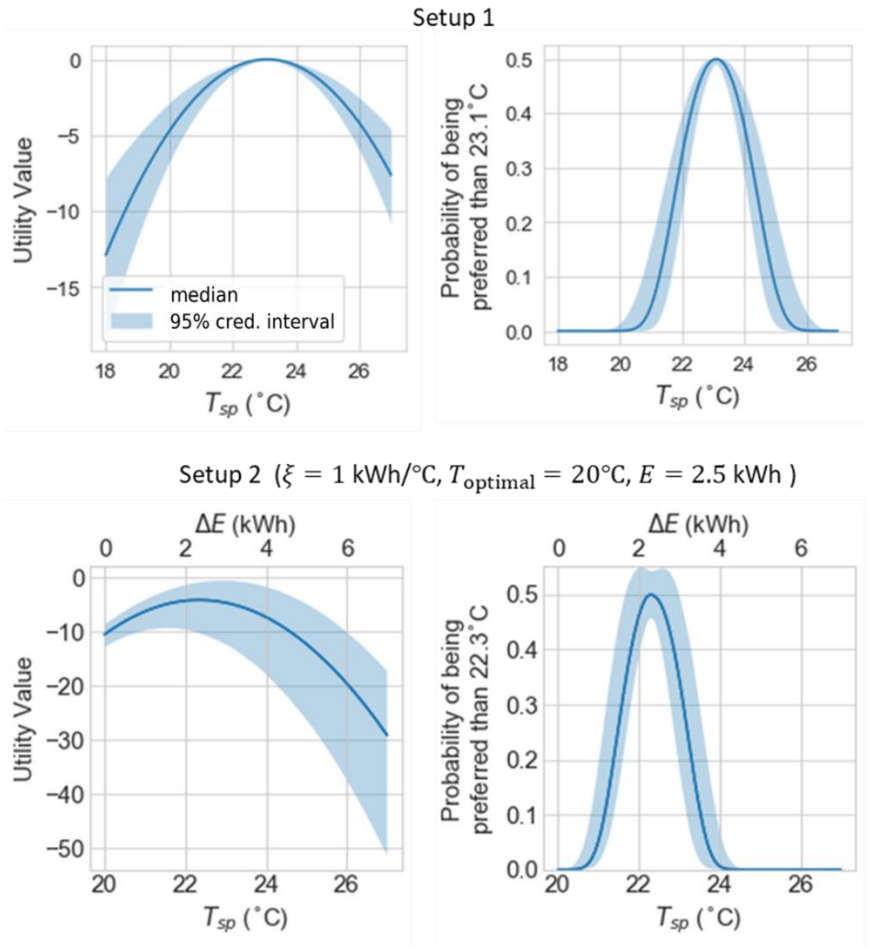


Figure 4.13. Occupants' preference on the setpoint temperatures in Setup 1 (top) and Setup 2 (bottom).

4.4.4 Model validation

In order to demonstrate that the decision-making model can effectively predict the occupants' setpoint selections given the data we observed in the field experiment, we perform 10-folds cross validation by splitting the observed data into training set (90% of the data) and validation set (10% of the data). Given E, ξ, and T_{optimal} we predict the probability distribution of the setpoints at the maximum of the utility. Then we compare the observed selected setpoints in the validation data set with the mean of the predicted setpoints. In Figure 4.14, the x-axis represents the prediction error, which is the difference between the observed selected setpoints and the mean of the predicted setpoints temperatures. The y-axis is the percentile of the observed setpoints in the predicted distributions. As all the points fall around the 45° line, it means that the predictive distributions can well capture the observed data points.

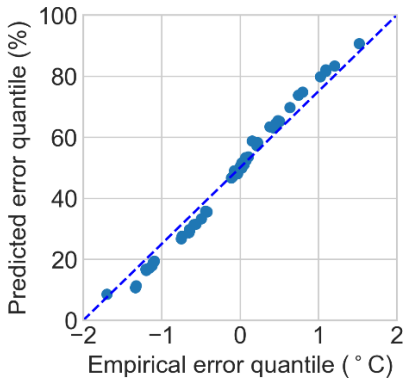


Figure 4.14. Quantile plot of the prediction error for model validation.

4.5 Summary

In this Chapter, we introduced a prototype user-interactive system integrated with a model predictive HVAC controller. Its novel web-interface enables private office occupants to consider real-time and personalized energy use information in their setpoint temperature selections. We implemented the system in an actual building and conducted a field experiment with human subjects to test our hypothesis that energy use information leads to more energy-efficient thermostat adjustment behavior by office building occupants. The experimental results showed that for the specific system and climate, occupants' overrides can contribute up to 55% of the

energy consumption on average with the implementation of smart HVAC control based on MPC. However, by providing real-time energy use information, 36% of the additional energy consumption (on average) can be recovered while achieving the same level of overall occupant satisfaction.

Subsequently, we presented a new modeling approach that reveals the causal effects in human decision-making process on occupant interactions with thermal environment control systems when energy efficient strategies are implemented. We demonstrated our approach using an actual building as test-bed and developed a utility model that quantifies occupants' preference on temperature setpoint incorporating their comfort and energy use considerations. The results showed that with the designed interface, the same level of overall occupant satisfaction was achieved and occupants (i) selected setpoints that were closer to temperatures determined by MPC to reduce energy use and (ii) often accepted the default setpoints when they did not need a significant improvement in the thermal environment condition. Based on these findings, we propose that the utility model can become a systematic approach to evaluate the design of similar user-interactive systems and web interfaces for different office layouts and building operation scenarios.

In order to achieve this, further investigation is needed to overcome the limitations of this initial study. More specifically, the energy use information presented to occupants could vary if the user-interactive system was implemented in a different building and climate. Since previous studies (Wilhite and Ling, 1995; Vellei et al., 2016) suggest that the effect of energy feedback may vanish with time, the proposed user-interactive system should be evaluated in a long-term field experiment with a large number of participants. Similar field studies are needed in different offices types and locations around the world for a larger database and diverse population to infer a more generalized utility. The prototype web-interface and human decision-making model presented in this paper is a first step towards developing standard user-interactive systems in a consistent and reliable way. In addition, other development tools such as Python GUI could be employed to provide more options in terms of data visualization and collection.

5. A META-REINFORCEMENT LEARNING APPROACH FOR OPTIMAL HVAC CONTROL

5.1 Overview

In this Chapter, we present a new Meta-RL approach to automate the discovery of building HVAC control policy. Using a model universe, identified with available building information, the agent is trained to learn policy that makes control decisions to improve energy efficiency and provide occupant comfort that can quickly adapt to the target building from. The agent is deployed to an emulator of a private office space as test-bed to evaluate the control performance and adaptability. In order to substantiate the necessity of our approach, we also demonstrate the impact of the environment simulators' prediction quality on the control agent's performance in the actual building with conventional RL approach.

We introduce the RL and Meta-RL algorithms in Section 5.2. The case study is presented in Section 5.3, and the performance evaluation of the proposed Meta-RL is presented in Section 5.4.

5.2 Methodology

Reinforcement learning problem for building control can be formulated as a Markov decision process (MDP). The states from the state space $s_t \in \mathcal{S}$ contains variables that describe the environment status at time t including building system temperatures, and exogenous variables such as outdoor weather conditions, etc. The action from the action space $a_t \in \mathcal{A}$ taken by the agent at time t is the control variable (e.g. the thermal input from HVAC system). After an action taken, the environment transits into a new state following the state transition probability function $p(s_{t+1}|s_t, a_t)$. It can also be written as $p(s'|s, a)$.

Then, the agent receives a reward as feedback from the environment (E) based on the energy efficiency and occupants' comfort achieved by the action. The reward $R(s_t, a_t)$ or R_t is dependent on the current states and actions. The accumulated future rewards (return) is:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k},$$

(5-1)

where the rewards are multiplied to discount factor γ by how far off in the future they're obtained. As some of the state variables may not be measured in real buildings (e.g. building envelope temperatures), we introduce the term o_t to represent the part of the states that are observable, $o_t \in \mathcal{O}$ which is the observation space. The action of the agent is determined by a policy function $\pi(o_t)$ based on the current observations.

5.2.1 Reinforcement learning algorithm

Many approaches in reinforcement learning utilize the action-value function (Q-function) to describe the expected return after taking an action a_t given the current observations o_t, and thereafter following policy π. The learning objective for the agent is to maximize the expected future return represented by the optimal Q-function:

$$Q^*(o_t, a_t) = \max_{\pi} \mathbb{E}_{\pi}[G_t | o_t, a_t]. \tag{5-2}$$

The Q-function is approximated by a deep neural network, and it can be written in a recursive form, Bellman equation, which suggests iterative updates based on the latest observations, rewards and actions:

$$Q^*(o_t, a_t) = \mathbb{E}[R_t + \gamma Q^*_{\pi}(o_{t+1}, \pi(o_{t+1}))]. \tag{5-3}$$

We employed deep deterministic policy gradient algorithm (DDPG, Lillicrap *et al*., 2016) to train the RL agent to learn by interacting with the environment. DDPG is (i) a model-free algorithm so it does not require the state transition function to be known; (ii) suitable for control problems where the control variable is continuous; and (iii) an actor-critic algorithm that not only learns the policy function (actor) that can be directly implemented as controller, but also learns the Q-function (critic) to update the policy in a direction of performance improvement. In this way, the algorithm can be substantially more data efficient and stable by take advantage of the past experience.

In DDPG, the actions taken by the agent are chosen based on the policy:

$$\pi_\theta(o_t) = \text{clip}\big(\mathcal{N}(\cdot\,|\mu_\theta(o_t), \sigma^2), a_{t,\min}, a_{t,\max}\big), \tag{5-4}$$

where $\mathcal{N}(\cdot\,|\mu_\theta(o_t), \sigma^2)$ is the probability distribution function (PDF) of a univariate Gaussian distribution. Its mean function $\mu_\theta(o_t)$ is represented by a deep neural network, the input of which are the observations o_t, and the weights (θ) of the deep neural network are updated during the learning process. Agent's explorations are enabled by the standard deviation σ, which is selected

according to the scale of the action space. If an action value sampled from the Gaussian distribution exceeds the lower limit $a_{t,\min}$ or the higher limit $a_{t,\max}$, The function clip$(\cdot)$ overrides the action value to $a_{t,\min}$ or $a_{t,\max}$, respectively.

All the transitions (o_t, a_t, r_t, o_{t+1}) are saved in a replay buffer $\mathcal{D}$. At every time step, a mini-batch of B transitions representing the past experience are sampled to perform updates on the policy and the Q-function. DDPG approximates the right-hand side of equation (5-3) represented with y by minimizing the mean-squared Bellman error (MSBE) with gradient descent:

$$\phi_{t+1} = \phi_t - \beta_1 \nabla_{\phi_t} \frac{1}{B} \sum_i (y_i - Q(o_i, a_i|\phi_t))^2, \tag{5-5}$$

where ϕ represents the weighting parameters of Q-function network,

$$y_i = r_i + \gamma Q'\left(o_{i+1}, \pi'\left(o_{i+1}\middle|\theta_{\text{targ},t}\right)\middle|\phi_{\text{targ},t}\right). \tag{5-6}$$

$i = 1, \cdots, B$ is the index of the sampled transitions from a mini-batch. β_1 is the step size of the gradient descent for the Q-function. After the Q-function is updated, the weighting parameters θ of the policy are optimized towards achieving the maximum Q-function value by gradient ascent:

$$\theta_{t+1} = \theta_t + \beta_2 \nabla_{\theta_t} \frac{1}{B} \sum_i Q(o_i, \pi(o_i|\theta_t)|\phi_{t+1}), \tag{5-7}$$

where β_2 is the step size of the gradient ascent for the policy function. Q' and μ' are the target networks with weighting parameters ϕ_{targ} and θ_{targ}, respectively. They are the temporal difference backups of the Q-function and policy function networks to direct the learning towards performance improvement. The target networks are updated by Polyak (ρ) averaging to improve the stability of learning. The learning can be repeated on the environment over the entire training data (with length T) for multiple epochs until convergence. The pseudo code of the algorithm is described in Algorithm 5.1.

Algorithm 5.1: DDPG algorithm

Inputs: initial θ, ϕ

Outputs: policy function π_θ

Initialize an empty replay buffer $\mathcal{D}$

Initialize $\theta_{\text{targ}} = \theta$, $\phi_{\text{targ}} = \phi$

for epoch $j = 1, \cdots, J$

  for $t = 1, \cdots, T$

    Observe o_t and select action a_t using equation (5-4)

    Execute a_t, get reward r_t and o_{t+1}

    Store (o_t, a_t, r_t, o_{t+1}) in $\mathcal{D}$

    Sample a random mini-batch of B transitions (o_i, a_i, r_i, o_i') from $\mathcal{D}$, $i = 1, \cdots, B$

    Update the weighting parameters of the Q-function with equation (5-5)

    Update the weighting parameters of the policy function with equation (5-7)

    Update the weighting parameters of the target networks:

$$\theta_{\text{targ},t+1} = \rho\theta_{\text{targ},t} + (1-\rho)\theta_{t+1}$$

$$\phi_{\text{targ},t+1} = \rho\phi_{\text{targ},t} + (1-\rho)\phi_{t+1}$$

    $t = t + 1$

  end for

  $j = j + 1$

end for

5.2.2 Meta-RL algorithm

We utilized Meta-RL in order to train the agent with the model universe $\mathcal{E}$ obtained based on existing knowledge of a building space. Our implementation is based on the study by Finn *et al*. (2017). First, a sufficient set of N environments that can well represent the possible environment models given the knowledge on a building is randomly sampled from $\mathcal{E}$. A Meta-RL policy function (π_{θ_e}) with weighting parameters θ_e is randomly initialized. The Meta-RL policy is adopted in each sampled environment initially, but updated with the DDPG algorithm described in Section 5.2.1. The policy in a sampled environment E_n is noted as π_{θ_n} with weighting parameters θ_n, and $n = 1, \cdots, N$. To ensure compatibility, the deep neural networks representing

π_{θ_n} and π_{θ_e} share the same architecture. After updating the policy in each sampled environment, a new trajectory of transitions based on the updated policy is collected. With such trajectories from all the sampled environments, the Meta-RL policy is updated with gradient ascent maximizing the expected accumulated rewards from all sampled environments:

$$\theta_{e,m+1} = \theta_{e,m} + \beta_3 \nabla_{\theta_{e,m}} \sum\nolimits_n \mathbb{E}\left[\sum_{t=1}^{T} r(s_{n,t}, \pi(o_{n,t}|\theta_n))\right], \tag{5-8}$$

where β_3 is the step size of the gradient ascent. In this study, the implementation of the meta policy gradient descent is based on a recently developed proximal policy optimization (PPO) by Schulman *et al*. (2017). The pseudo code for Meta-RL is described in Algorithm 5.2.

Algorithm 5.2: Meta-RL

Input: $E_1, E_n \cdots, E_N \in \mathcal{E}$
randomly initialized θ_e and $\phi_{1:N}$

Outputs: Meta-RL policy function π_{θ_e}

for episode $m = 1, \cdots, M$

  for $n = 1, \cdots, N$

    Initialize the weighting parameters of π_{θ_n}: $\theta_n = \theta_{e,m}$

    Update θ_n and ϕ_n with DDPG (Algorithm 5.1) for $t = 1, \cdots, T$

    Sample a trajectory of transitions $\mathcal{O}_n = \{(s_t, a_t, r_t, s_{t+1})\}$ for $t = 1, \cdots, T$ using π_{θ_n}

    $n = n + 1$

  end for

  Update θ_e with equation (5-8) using all trajectories $\mathcal{O}_{1:N}$

  $m = m + 1$

end for

5.3 Case study

One of the south-facing private offices in a high-performance building located in West Lafayette, Indiana was used as test-beds for this study (see Figure 4.1 in Chapter 4). The detailed descriptions of the test-bed and its HVAC system can be found in Section 4.2.1. The only

difference is that the capacity of the assumed chiller was scaled down to 3% (2.07 kW) based on the cooling load of one private office.

5.3.1 Environment and agent

For RL agent training, we use the dynamical model that predicts building thermal response to describe part of the state transition function that simulates the behavior of the environment. Please note that this model is unknown to the agent. We adopted a generic second-order thermal network model structure (3R2C) as it is recognized as a simple yet accurate representation of a building zone similar to our test-bed. A graphical representation of the thermal network is shown in Figure 5.1.

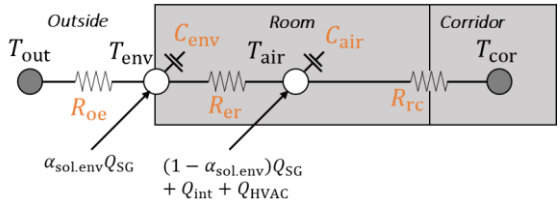


Figure 5.1. The thermal network for the 3R2C model.

The model universe specific to the test-bed can be described with uncertain parameters of the model $\xi = (C_{\text{env}}, C_{\text{air}}, R_{\text{oe}}, R_{\text{er}}, R_{\text{rc}}, \alpha_{\text{sol.env}})$, as based on the typically available building information such as construction drawings, the values of the parameters cannot be accurately determined without the process of system identification. In this study, we obtain the probability distributions of these parameters presented in Figure 5.2 using the process described in Appendix G. The parameter $\alpha_{\text{sol.env}}$ has the value between 0 and 1, and thereby a non-informative prior of uniform distribution $\mathcal{U}(0,1)$ is assumed.

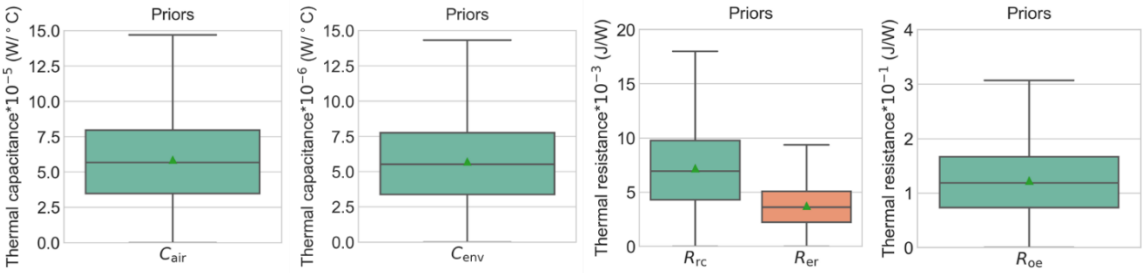


Figure 5.2. Probability distribution of the uncertain parameters in the subset of model universe.

The states of the environment include the envelop temperature $T_{\mathrm{env},t}$, indoor air temperature T_{air}, the corridor temperature T_{cor}, the outdoor air temperature T_{out}, solar heat gain Q_{SG}, and internal heat gain Q_{int}:

$$s_t = \left[T_{\mathrm{env},t}, T_{\mathrm{air},t}, T_{\mathrm{out},t}, T_{\mathrm{cor},t}, Q_{\mathrm{SG},t}, Q_{\mathrm{int},t}\right]. \tag{5-9}$$

The time step is 30-minutes, and the control horizon is 1 time step for the system considered. During the training, T_{out} and the global horizontal solar irradiance are represented by historical data that are available at the national solar radiation data base (Sengupta *et al*., 2018), while the solar heat gain Q_{SG} can be calculated based on the global horizontal irradiance, window and shading properties, time of the year and building orientation using the model by Klein *et al*. (2011). For implementation in the target building space, T_{out} and global horizontal irradiance can be either measured at a local weather station, or accessed from the real-time weather information at the National Oceanic and Atmospheric Administration (NOAA). Q_{int} is known based on the existing equipment and building operation schedule described in Section 4.2.1. T_{air} and T_{cor} are typically measured from the thermostats placed in the office space and adjacent zone. As T_{env} is not a typically available measurement, it is not considered as an observable state by the RL agent in this study. Therefore,

$$o_t = \left[T_{\mathrm{air},t}, T_{\mathrm{out},t}, T_{\mathrm{cor},t}, Q_{\mathrm{SG},t}, Q_{\mathrm{int},t}\right]. \tag{5-10}$$

The action taken by the agent $a_t \in [0,1]$ determines the thermal input to the office space from the HVAC system (Q_{HVAC} in Figure 5.1). When $a_t > 0.5$, the HVAC system operates on reheat mode, otherwise the cooling mode is on. The rewards are assigned based on the HVAC electricity consumption and occupants' comfort:

$$r_t = -\left(\frac{P_{\text{cool},t}}{\text{COP}\left(T_{\text{out},t}\right)} + \frac{P_{\text{heat},t}}{\eta} + \text{penalty}_t\right), \tag{5-11}$$

where,

$$P_{\text{cool},t} = \begin{cases} -\left[\dot{m}_{\min} + \dfrac{0.5 - a_t}{0.5}\left(\dot{m}_{\max} - \dot{m}_{\min}\right)\right] \cdot c_p\left(T_{\text{sup}} - T_{\text{air},t}\right), \text{if } a_t \le 0.5 \\ -\dot{m}_{\min} c_p\left(T_{\text{sup}} - T_{\text{air},t}\right), \text{otherwise.} \end{cases} \tag{5-12}$$

$$P_{\text{heat},t} = \begin{cases} \dfrac{a_t - 0.5}{0.5} \cdot \text{Cap}_{\text{heat}}, \text{if } a_t > 0.5 \\ 0, \text{otherwise.} \end{cases} \tag{5-13}$$

$$\text{COP}\left(T_{\text{out},t}\right) = c_0 + c_1 T_{\text{out},t} + c_2 T_{\text{lv}} + c_3 T_{\text{out},t}^2 + c_4 T_{\text{lv}}^2 + c_5 . T_{\text{lv}} T_{\text{out},t}, \tag{5-14}$$

$\dot{m}_{\min}$ is the supply air flow rate at the minimum damper position (0.081 kg/s, equivalent to 140 cfm), while $\dot{m}_{\max}$ is the supply air flow rate at the maximum damper position (0.318 kg/s, equivalent to 550 cfm). $T_{\text{sup}} = 16℃$ is the supply air temperature from the AHU. c_p is the specific heat of air at 1004 J/kg · K. Cap_{heat} is the maximum reheat coil capacity at 1462W. η is the efficiency of the boiler at 90%. The coefficients $c_{0:5}$ in equation (5-14) are obtained based on the data available in the chiller manufacturer's catalog (DoE, 2010). $T_{\text{lv}} = 7℃$ is the leaving water temperature from the chiller. The value of the penalty is selected to be on the same scale as the other component in the reward, but also large enough for the agent to treat it as a perceivable loss in the learning process:

$$\text{penalty}_t = \begin{cases} 5, \text{ if } T_{\text{air},t+1} \le 22℃ \text{ or } T_{\text{air},t+1} \ge 24℃ \text{ at occupied hours,} \\ 0, \text{ otherwise.} \end{cases} \tag{5-15}$$

5.3.2 RL agent training

The RL agent is trained with environments simulated with the aforementioned models and historical weather data from June 1st –August 31st in the year of 2015-2017 for West Lafayette, IN. The simulated environments are made compatible with OpenAI gym (Brockman *et al*., 2016) format, so that they are standardized for algorithm implementation. The DDPG and Meta-RL algorithm is coded in Python based on Tensorflow v1.0 (Abadi *et al*., 2016), with which the key

variables, functions such as the deep neural networks, and the computational procedure from the algorithm are defined. The hyperparameters in the algorithms are fine-tuned to find the appropriate settings for the specific application. The key settings applied in the algorithms for agent training are presented in Table 5.1 for DDPG and Table 5.2 for Meta-RL.

Table 5.1. Agent training settings for DDPG.

Neural network structure (for both Q-function network and policy function network)	# of hidden layers	4
	# of hidden nodes in each layer	64
	Activation function	Rectified linear unit (ReLU)
Number of training epochs J		10
Number of time steps in an epoch T		13248 time steps (9 months)
Mini-batch size I		200
Discount factor γ		0.99
Standard deviation σ of the noise term in action selection		0.05
Step sizes $\beta_{1:2}$		0.001
Polyak ρ		0.95

Table 5.2. Agent training settings for Meta-RL.

Neural network structure for and policy function network	# of hidden layers	4
	# of hidden nodes in each layer	64
	Activation function	Rectified linear unit (ReLU)
Number of training episodes K		5
Number of epochs in DDPG		1
Other settings in DDPG		See Table 5.1
Number of time steps in an epoch T		13248 time steps (9 months)
Number of sampled environments N		100
Step size β_3		0.001

5.4 Performance analysis

In order to evaluate the control performance of the trained agent, we deploy it to the test-bed office represented by an 3R2C model, the parameters of which are estimated based on sufficient on-site measurements (see Section 4.2.2.1). Therefore, the model is assumed to accurately represent the dynamical thermal response of the test-bed. During the testing deployment, the

weather data from June 1st-August 31st 2018 at West Lafayette, IN are used, and we allow the policy to be updated online with DDPG algorithm (settings complied with Table 5.1) as the testing progresses, so that the agent can adapt to the test-bed environment. Three scenarios are compared:

(i) Agent learned from the environment represented by the model with parameters obtained from system identification experiment. The model can predict the responses of the test-bed with high level of accuracy;

(ii) Agent learned from the environment represented by the models with parameters that are empirically selected without calibration based on on-site measurements. The models do not guarantee the prediction accuracy on the responses of the test-bed;

(iii) Agent learned from the subset of model universe identified with existing knowledge of the test-bed. Samples of models that represent the environments were drawn from the probability distribution described in Figure 5.2.

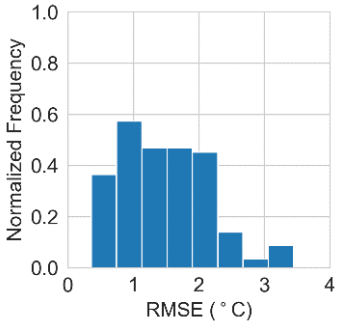


Figure 5.3. Inaccurate models' prediction RMSE for indoor air temperature.

In Scenario (i) and (ii) the agent was trained with DDPG algorithm using the settings presented in Table 5.1, while in Scenario (iii) the algorithm was Meta-RL with the settings presented in Table 5.2. For Scenario (ii), we tested on 100 models with different levels of accuracy, and their prediction root mean square errors (RMSE) on the air temperature (T_{air}) are shown in the Figure 5.3. The performance metrics considered are the energy consumption from the HVAC system (in kWh), and the temperature exceedance in °C-hr according to ASHRAE Standard 55 (ASHRAE & ANSI, 2017) during the occupied hours. We also compared the control performance with MPC with equivalent control objective and constraints, serving as the theoretical performance bound (see Appendix G).

Figure 5.4 details the control performances in two days with typical summer weather at West Lafayette, IN (July 16th, 2018-July 17th, 2018). The yellow dashed lines, red solid lines, and green dash-dotted lines represent the performance of the agents (controllers) in Scenario (i), (iii), and MPC, respectively. The shaded blue area shows the performance of the agents in Scenario (ii) considering the 0-100th percentile of the samples. The outdoor air temperature varies from 17°C to 31°C with sunny sky conditions (see Figure 5.5).

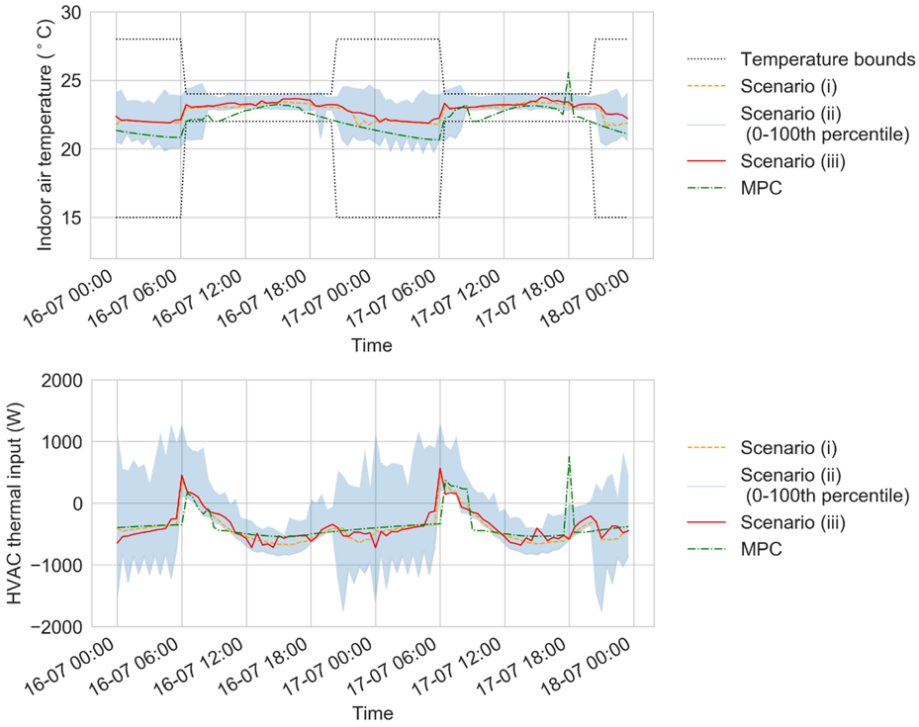


Figure 5.4. Indoor air temperatures, HVAC thermal input for Scenario (i-iii) and MPC (July 16th –July 17th, 2018).

Overall, in Scenario (i), (iii) and MPC, the HVAC thermal input determined by the controllers share similar trends (Figure 5.4 bottom). In the morning, the HVAC system operates at reheat mode in order to maintain the indoor air temperature within the setpoint bounds, due to the relatively lower outdoor air temperature and solar heat gain. The heating rates are higher at the beginning of the reheat operation to quickly increase the indoor air temperature to meet the lower setpoint bound. Then less heating power are applied gradually as the outdoor air temperature and solar heat gain increase. In MPC, the reheat starts at 6 a.m. and ends at 9 a.m., while the indoor air temperatures (Figure 5.4, top) are kept close to the lower setpoint bounds. This is because the MPC

controller is able to plan for the operation based on the dynamical model, achieving the minimum energy consumption required to maintains the indoor air temperatures. In comparison, the agents in Scenario (i) and (iii) do not know the model, the reheat is scheduled more conservatively (from around 5:30 a.m. to 10:30 a.m.). The resulting indoor air temperatures are around 23℃ (in the middle of the setpoint bounds). For the same reason, MPC controller proactively switched to reheat at 18:00 p.m. on July 17th to avoid violations on the lower setpoint bound, while the agents in Scenario (i) and (iii) started to reheat at around 20:00 p.m. when the outdoor air temperature declined. The agents in Scenario (ii) also learn to use reheat to maintain the indoor air temperature in the morning, although the control policies trained from un-calibrated environment models cannot guarantee the optimal starting/ending time and heating rate for energy saving.

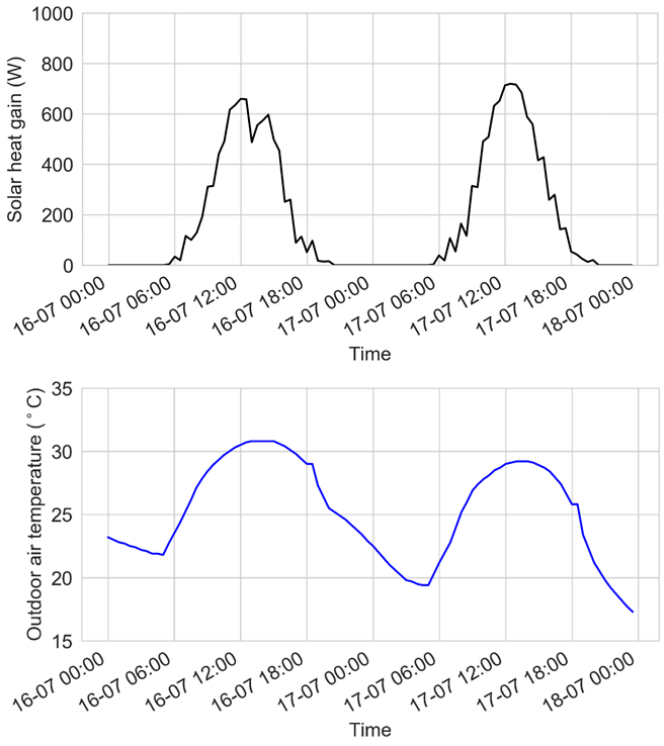


Figure 5.5. Outdoor air temperature and solar heat gain (July 16th –July 17th, 2018).

After the reheat operations in the morning, when the outdoor air temperatures and solar heat gain reach the high levels of a day, the MPC controllers typically keeps the zone damper at the minimum position (i.e. minimum cooling rate, usually around -500W to -400W given the AHU supply air temperature and flow rate) until the end of occupation time. Similar strategy is followed

by the agents in Scenario (i), (ii) and (iii) with only slight fluctuations due to the agents' exploratory actions and changes in the observed environment states. From 19:30 p.m. (the beginning of the unoccupied hour) until the following morning, the controllers in Scenario (i), (iii) and MPC still keep the minimum cooling rate in order to minimize the energy consumption (as the setpoint temperature bounds are relaxed during these hours). However, the policies learned from un-calibrated environment models are likely to misguide the agents in Scenario (ii) to unnecessarily preheat or precool the office space during this time.

Table 5.3. Performance metrics comparison for all scenarios and MPC (June 1st –August 31st, 2018).

	Scenario (i)	Scenario (ii) average	Scenario (iii)	MPC
Total energy consumption (kWh)	436.92	664.68	489.03	404.30
Total temperature exceedance (°C-hr)	0.1	21.44	4.81	13.96

The comparison on the performance metrics for all scenarios and MPC are shown in Table 5.3. Overall, despite slightly better performance in terms of maintaining the indoor air temperature with negligibly higher electricity consumption, the agent learned from the environment represented by the model with accurate parameters is able to achieve similar performance compared to MPC. The agent learned from the subset of model universe can also successfully maintain the indoor air temperature, with 14% higher energy consumption than Scenario (i), and 21% higher energy consumption than MPC. While in average, over the 3 months of adaptation to the test-bed, the agent in Scenario (ii) can result in 52% higher energy consumption than Scenario (i), and 64% higher energy consumption than MPC. The average total temperature exceedance is also significantly higher in Scenario (ii).

Figure 5.6 presents the performance of the agent in Scenario (ii) (see the blue dots) with different levels of model prediction RMSE. The x-axis is the total energy consumption, and the y-axis shows the temperature exceedance. With RMSE < 0.5°C, the agent in Scenario (ii) can achieve the performance that is closer to the agent learned from the subset of model universe. However, without spending the time and engineering effort to calibrate with on-site measurements, it is hardly likely to obtain models with low prediction RMSE. On the other hand, it is highly possible to result in models with high prediction RMSE by selecting the parameters empirically, but the

agent trained from such model can less likely maintain the indoor air temperature and reduce the energy consumption.

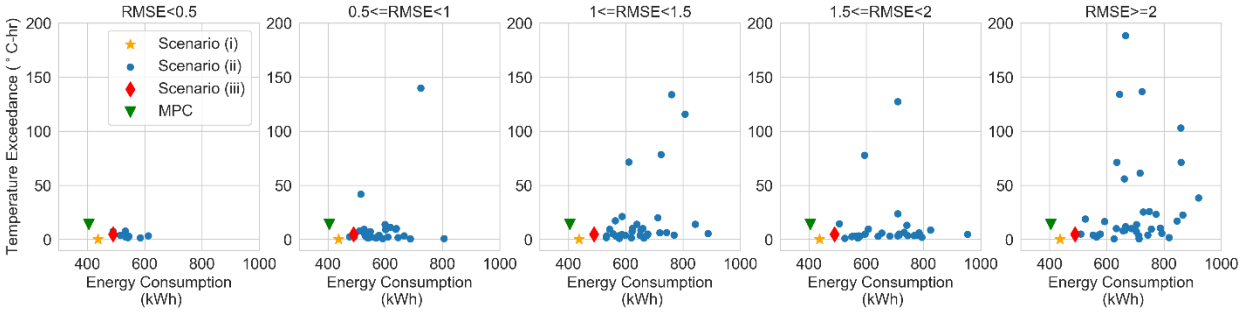


Figure 5.6. Performance metrics comparison for different levels of model RMSE in Scenario (ii) (June 1st –August 31st, 2018).

Figure 5.7 shows the performance comparisons for the 1st week (June 1st-7th, 2018), the 1st month (June 1st-30th, 2018), the 2nd month (July 1st-31st, 2018), and the 3rd month (August 1st -31st, 2018) during the testing deployment. As time passes, the agent in all scenarios showed adaptation to the test-bed, reducing the energy consumption and temperature exceedance in the direction of approaching the results from MPC. However, it takes 3 months for the agent in Scenario (ii) to be able to manage the indoor temperature within the bounds, while more time will be needed to learn to minimize the energy consumption. The agent learned from the subset of model universe in Scenario (iii) is able to consistently maintain the indoor temperature even in the first week of deployment, and thereby eliminate the potential occupant discomfort. In the 3rd month, the agent in Scenario (iii) can achieve the energy consumption only 16% higher than MPC, reducing significantly the potential energy waste that could result in Scenario (ii).

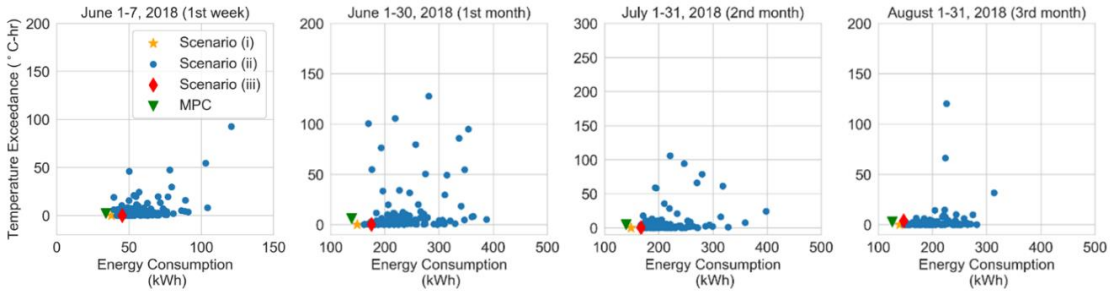


Figure 5.7. Performance metrics comparison for June 1st-7th, June 1st-30th, July 1st-31st, August 1st -31st, 2018.

5.5 Summary

In this study, we developed an autonomous HVAC control approach based on Meta-RL. The Meta-RL agent was trained with a model universe represented by uncertain parameters obtained from available information of the target building to facilitate the agent's adaptation. The control performance of the agent was evaluated by deploying the agent to an emulator represented by a low-order model of the test-bed office space.

For the test-bed considered, the results show that using the Meta-RL approach the control agent successfully maintains the indoor air temperature within the first week of deployment while consumes only 16% higher energy consumption compared to MPC by the 3rd month without requiring extensive engineering effort for model tuning. In contrast, with conventional RL approach, the agents learned from environments represented by the models with empirically determined parameters without on-site calibration can result on average in 61% higher HVAC energy consumption compared to MPC in the 3rd month of deployment, while imposing considerable risk in the temperature maintenance in the first 2 months. Also, the agent's control performance in terms of both indoor air temperature maintenance and energy consumption deteriorated as the prediction accuracy of the models decrease. In the near future, the developed Meta-RL algorithm will be deployed in the actual test-offices at Herrick building.

6. FUTURE WORK

The studies presented in this thesis explore intelligent building operation strategies that are tested with specific building systems. However, the components of the developed methodology are generalizable: the uncertain solar irradiance forecast model can be used for predictive control of solar energy systems; the SMPC algorithm with ADP can be applicable for operating other energy systems with stochastic environmental disturbances and complex energy conversion schemes; the user-interactive system design approach is transferrable to other private offices with integrated smart controls; the Meta-RL control approach can be generalized for other types of building zones. Nevertheless, future research can be extended into the following directions towards the realization of smart and user-interactive building system operations that can be widely adopted in building industry.

The model universe can be extended to include comprehensive options of building zone types. To this end, assisted by open-source datasets (Deru *et al*., 2011; EIA, 2016; Miller and Meggers, 2017; Balaji *et al*., 2018; Miller *et al*., 2020), a platform needs to be established that contains building metadata with standardized format including construction and energy system specifications as well as time-series building operation data. Based on these, a statistically-informed taxonomy of building system dynamical models can be created to facilitate the automated generation of Meta-RL agents or MPC controllers for any type of building zones, under different scenarios of information availability.

In order to expand the applicability to occupant-centric buildings and renewable energy systems, occupants' preference and uncertain weather forecast need to be considered in the optimal control decision making with RL approach. Occupants' preference on the indoor environment conditions and associated impact on energy use can be inferred from their feedback, as well as their interactions with building systems. Such preferences can be encoded as rewards to the RL agent, while the uncertain weather forecast could be treated as parts of environment states. Recently, new RL algorithms that allow agents to learn from environments described by stochastic states and rewards are being developed in the machine learning research community (Christiano *et al*., 2017; Ibarz *et al*., 2018; Wang *et al*., 2020). The potential applications of these methods need to be investigated for smart building controls.

Although further research is still required, the easiness to learn from human feedback has been regarded as one of the most desirable characteristics of RL algorithms. RL agent can receive human feedback from user-interactive systems. In order to establish efficient communication between occupants and the agent, the effect of other forms of energy feedback (e.g., monetary incentive, peer comparison and instructive feedback) presented in the user-interface need to be systematically evaluated under different building operation scenarios, such as demand response and social game implementation. Although occupants' preference on comfort and energy use in such context can be quantified with similar methods with those presented in Chapter 4, their preference needs to be translated into the rewards to the RL agent in a way that facilitates the learning.

APPENDIX A. FUNCTION APPROXIMATION IN DYNAMIC PROGRAMMING

A.1 SPECIFYING THE APPROXIMATING FUNCTION CLASS FOR OPTIMAL COST-TO-GO FUNCTIONS

In this work, we use Gaussian process regression (GPR) to approximate the optimal cost-to-go and policy functions (Rasmussen and Williams, 2006). GPR offers various benefits: (i) It can quantify the approximation error (Bayesian epistemic uncertainty); (ii) It can deal with noisy observations (robustness to averaging errors when approximating the expectation on the right-hand side of Bellman equation); and (iii) It is nonparametric and, thus, it does not impose a restrictive functional form to the optimal cost-to-go function.

GPR assigns a Gaussian process probability measure on the space of optimal cost-to-go functions and then it conditions this probability measure on pairs of input-output observations. This probability measure corresponds to our prior beliefs about the regularity, length scale, and, in general, the variability of each C_t^*, before seeing any actual function evaluations. Without loss of generality, we may select this probability measure to have a zero mean. Mathematically, we write:

$$C_t^* \sim \mathrm{GP}(0, k_t), \tag{A-1}$$

where $k_t: \mathbf{X} \times \mathbf{X} \to \mathbb{R}_+$ is a covariance function encoding our prior beliefs, see Chapter 4 of Rasmussen and Williams (2006). In this work, we use the squared exponential (SE) covariance function:

$$k_t(\mathbf{x}, \mathbf{x}') = s_t^2 \exp\left\{-\frac{1}{2}\sum_{i=1}^{d_x} \frac{\left(x_i - x_i'\right)^2}{\ell_{t,i}^2}\right\}, \tag{A-2}$$

where s_t^2 and $\ell_{t,i}$ correspond to the variance and to the length scale of the i-th input of C_t^*, respectively. These parameters are not predetermined. They are estimated as part of the regression process. Note that the SE covariance corresponds to the belief that C_t^* is infinitely differentiable. For problems with kinks or discontinuities in the optimal cost-to-go function, it is preferable to use a Matérn or an exponential covariance function.

A.2 SELECTING THE NUMBER AND CHOICE OF THE COLLOCATION POINTS

Let $\mathbf{x}_{t,1:M} = \left(\mathbf{x}_{t,1}, \dots, \mathbf{x}_{t,M}\right)$ be a set of *M collocation* points in the state space **X**. It is vitally important that the collocation points $\mathbf{x}_{t,1:M}$ cover the state space **X** as well as possible. We ensure this by using Latin hypercube sampling (LHS), which has exhibited better properties than fixed or random sampling strategies (Iman, 2008). The number of collocation points, M, does not have to be predetermined. One can add points sequentially, train the model, estimate the residual epistemic uncertainty, and add more collocation points if required. That latter can be done, for example, by selecting the points about which one is maximally uncertain (MacKay, 1992). To keep the implementation of our scheme simple, in our application we select an adequately large and fixed M, see Section 3.3.3.

A.3 EVALUATING THE RIGHT-HAND SIDE OF THE BELLMAN EQUATION AT AN ARBITRARY POINT

Assume that we have already an estimate of the $(t+1)$-th optimal cost-to-go function, say $\hat{C}^*_{t+1}$. We show how the right-hand side of equation (3-17) can be evaluated at an arbitrary point $\mathbf{x}_t$. This evaluation requires solving a nonconvex and nonlinear constrained stochastic optimization problem with respect to the optimal control $\mathbf{u}_t$. To remove the stochasticity of the problem, we estimate all expectations with sampling averages. To this end, let $\mathbf{w}_{t,1:N} = \left(\mathbf{w}_{t,1}, \dots, \mathbf{w}_{t,N}\right)$ be N samples from $p(\mathbf{w}_t|\mathbf{x}_t)$. We replace the stochastic optimization problem with the following deterministic one (modified right-hand side of equation (3-17)):

$$\min_{\mathbf{u}_t} \frac{1}{N} \sum_{n=1}^{N} \left[J_t\left(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_{t,n}\right) + \hat{C}^*_{t+1}\left(\mathbf{f}_t\left(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_{t,n}\right)\right)\right] \tag{A-3}$$

subject to the constraints (modified equation (3-3)):

$$\frac{1}{N} \sum_{n=1}^{N} g_{i,t}\left(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_{t,n}\right) \geq 0 \text{ for } 1 \leq i \leq n_c. \tag{A-4}$$

In our application, we solve this problem with the 'pyOpt' solver (Perez *et al.*, 2012) that wraps around Python's SciPy optimize (Jones *et al*., 2001) which utilizes the subroutine sequential least square programming (SLSQP) originally developed by (Kraft, 1988). The optimization over $\mathbf{u}$ is solved with 'pyOpt' solver allows easier integration with MPI (Dongarra, 1994) for parallel execution over points in state space.

A.4 ITERATING BACKWARDS USING GAUSSIAN PROCESS REGRESSION

The collocation points $\mathbf{x}_{t,1:M}$ are the observed *function inputs*. For each $\mathbf{x}_{t,j}$, $j = 1, \dots, M$, we follow the methodology of the previous section to compute an approximation $y_{t,j}$ of the optimal cost-to-go function at $\mathbf{x}_{t,j}$, i.e., of $C_t(\mathbf{x}_{t,j})$. Collectively, we write $\mathbf{y}_{t,1:M} = (y_{t,1}, \dots, y_{t,M})$. We refer to $\mathbf{y}_{t,1:M}$ as the function outputs. Gaussian process regression, Chapter 2 of Rasmussen and Williams, (2006), learns an approximation $\hat{C}_t^*$ from the observed inputs and outputs $(\mathbf{x}_{t,1:M}, \mathbf{y}_{t,1:M})$. This approximation corresponds to the posterior mean of the Gaussian process obtained after conditioning our prior beliefs to the observed data $(\mathbf{x}_{t,1:M}, \mathbf{y}_{t,1:M})$.

The mathematical form of the approximation is:

$$\hat{C}_t^*(\mathbf{x}_t) = \mathbf{k}_t(\mathbf{x}_t, \mathbf{x}_{t,1:M})(\mathbf{K}_t + \sigma_t^2 \mathbf{I})^{-1} \mathbf{y}_{t,1:M}, \tag{A-5}$$

where $\mathbf{K}_t = \left(k_t(\mathbf{x}_{t,i}, \mathbf{x}_{t,j})\right)_{i,j=1}^{M}$ is the covariance matrix of the observed inputs, $\mathbf{k}_t(\mathbf{x}_t, \mathbf{x}_{t,1:M}) = \left(k_t(\mathbf{x}_t, \mathbf{x}_{t,1}), \dots, k_t(\mathbf{x}_t, \mathbf{x}_{t,M})\right)$ is the cross-covariance vector between a test input and the observed ones, and σ_t^2 is variance modeling the uncertainty introduced by sampling average of equation (A-3). The parameters of this approximation are $\boldsymbol{\phi} = (s_t^2, \ell_{1,t}, \dots, \ell_{d_x,t}, \sigma_t^2)$ and their optimal values are identified by maximizing the data likelihood, Chapter 5 of Rasmussen and Williams, (2006). Our implementation is based on the Python module GPy (Hensman *et al*., 2012). The same approach can be used for learning approximate policy functions.

APPENDIX B. BUILDING AND SYSTEM SPECIFICATIONS

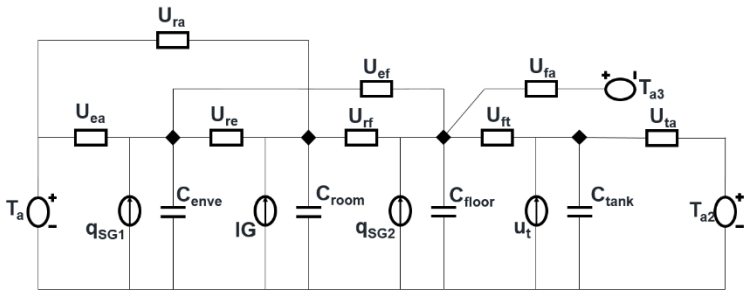


Figure B.1. The thermal network for the state-space model (Li *et al.*, 2015).

Table B.1. Building properties and operation settings.

ROOM TEMP.	• 17 – 25 °C, 8:00 a.m. – 10:30 a.m. • 21 – 25 °C, 10:30 a.m. – 18:30 p.m. • >15 °C, 18:30 p.m. – 8:00 a.m.
FLOOR TEMP.	• 19 – 29 °C based on ASHRAE Standard 55 (ASHRAE & ANSI, 2017)
TANK TEMP.	• 20 – 55 °C
BIPV/T DESIGN	• Total area: 65 m^2 • PV area: 58.5 m^2, 6.32 kWp capacity • Air flow rate: 3500 m^3/h corresponding suction velocity: 0.015 m/s
SHADING CONTROL	• ON, when the incident solar radiation on the window exceeds 180 W/m^2 • OFF, when the incident solar radiation on the window below 160 W/m^2
ENVELOPE PROPERTIES	• Exterior wall: U = 0.122 W/m^2·K, Gypsum board 0.016 m • Interior wall: U = 0.207 W/m^2·K, Gypsum board 0.032 m • Window: U = 0.63 W/m·K, solar transmittance = 0.228, absorbance = 0.487, SHGC = 0.327 • Roof: U = 0.368 W/m^2·K • Floor: U = 0.689 W/m^2·K, Concrete 0.165 m
INTERNAL HEAT GAIN	• Lighting: 10.76 W/m^2, for 8:00 a.m. – 18:00 p.m.; 0 otherwise • Equipment: 21.52 W/m^2, for 8:00 a.m. – 18:00 p.m.; 0 otherwise • Occupants: 75 W/person, for 8:00 a.m. – 18:00 p.m.; 0 otherwise
INFILTRATION	• Air change rate = 0.737/hr
VENTILATION	• 0.06 cfm/ft^2 or 5 cfm/person, whichever is greater, supply air temperature: 22 °C according to ASHRAE Standard 62.1 (ASHRAE & ANSI, 2013)

APPENDIX C. RULE-BASED CONTROL

The rule-based control strategy considers weather forecast information including outdoor dry bulb air temperature (T_a) and sky-cover (sc). It predicts heat pump power (Q_{hp}) aiming at utilizing solar energy to improve the system efficiency. Therefore, the schedule strictly follows solar availability considering some thresholds of outdoor dry bulb temperature.

Table C.1. Rule-based control schedule.

Time of the day	Heat pump operations
0-6 a.m.	• Q_{hp}= 0 kW.
6-8 a.m.	• Q_{hp}= 5 kW, if $T_a \leq$ 8°C for the following period (8-10 a.m.) in average. • Q_{hp}= 0 kW, otherwise.
8-10 a.m.	• Find Q_{hp} values from Table C.2, if $T_a \leq$ -8°C for the following period (10 a.m.-12 p.m.) in average. • Find Q_{hp} values from Table C.3, if -8°C $< T_a \leq$ 0°C for the following period (10 a.m.-12 p.m.) in average. • Find Q_{hp} values from Table C.4, if 0°C $< T_a \leq$ 8°°C for the following period (10 a.m.-12 p.m.) in average. • Q_{hp}= 0 kW, otherwise.
10 a.m. -12 p.m.	• Find Q_{hp} values from Table C.2, if $T_a \leq$ -8°C for the following period (12-14 p.m.) in average. • Find Q_{hp} values from Table C.3, if -8°C $< T_a \leq$ 0°C for the following period (12-14 p.m.) in average. • Find Q_{hp} values from Table C.4, if 0°C $< T_a \leq$ 8°C for the following period (12-14 p.m.) in average. • Q_{hp}= 0 kW, otherwise.
12-14 p.m.	• Find Q_{hp} values from Table C.2, if $T_a \leq$ -8°C for the following period (14-16 p.m.) in average. • Find Q_{hp} values from Table C.3, if -8°C $< T_a \leq$ 0°C for the following period (14-16 p.m.) in average. • Find Q_{hp} values from Table C.4, if 0°C $< T_a \leq$ 8°C for the following period (14-16 p.m.) in average. • Q_{hp}= 0 kW, otherwise.
14-16 p.m. and 16-20 p.m.	• Find Q_{hp} values from Table C.5.
20 p.m. -12 a.m.	• Q_{hp}= 0 kW.

Table C.2. Heat pump power input look-up table (8 a.m., 10 a.m. and 12 p.m., T_a ≤-8°C).

Q_{hp} **in kW**					
		Average sc **at the following period**			
		>0.8	**0.5-0.8**	**0.2-0.5**	**<0.2**
Average sc **at the current period**	**>0.8**	10	10	12	15
	0.5-0.8	8	8	10	12
	0.2-0.5	6	8	10	12
	<0.2	6	6	8	10

Table C.3. Heat pump power input look-up table (8 a.m., 10 a.m. and 12 p.m., -8°C $< T_a \leq$ 0°C).

Q_{hp} **in kW**					
		Average sc **at the following period**			
		>0.8	**0.5-0.8**	**0.2-0.5**	**<0.2**
Average sc **at the current period**	**>0.8**	5	6	8	10
	0.5-0.8	5	5	6	8
	0.2-0.5	4	4	5	6
	<0.2	4	4	4	5

Table C.4. Heat pump power input look-up table (8 a.m., 10 a.m. and 12 p.m., 0°C $< T_a \leq$ 8°C).

Q_{hp} **in kW**					
		Average sc **at the following period**			
		>0.8	**0.5-0.8**	**0.2-0.5**	**<0.2**
Average sc **at the current period**	**>0.8**	2	3	3	5
	0.5-0.8	2	2	3	3
	0.2-0.5	2	2	2	3
	<0.2	2	2	2	2

Table C.5. Heat pump power input look-up table for 14-16 p.m. and 16-20 p.m.

Q_{hp} **in kW**			
		14-16 p.m.	**16-20 p.m.**
Average T_a **for the next 28 hours**	T_a**> 8°C**	0	0
	0°C $< T_a \leq$ **8°C**	4	2
	-8°C $< T_a \leq$ **0°C**	6	4
	$T_a \leq$**-8°C**	8	6

APPENDIX D. ENERGY CONSUMPTION AND EFFICIENCY MODELS FOR HVAC SYSTEM COMPONENTS

This appendix describes how the efficiency and power of the fan, chiller and reheat are determined in the cost function equation (4-4) in Section 4.2.2.

At time t, the fan power is computed by the following equation:

$$P_{\text{fan},t} = f_{\text{fan}}\left(\sum_{i=1}^{N} \dot{m}_{i,t}\right), \tag{D-1}$$

where,

$$f_{\text{fan}}(\dot{m}) = \frac{\dot{m}_{\text{design}} \cdot \Delta P}{\eta_{\text{fan}} \rho_{\text{air}}} \text{Curve}_{\text{cub}}\left(\frac{\dot{m}}{\dot{m}_{\text{design}}}\right). \tag{D-2}$$

$\dot{m}_{\text{design}}$ is the design flow rate of the fan (3000 m^3/h); ΔP is the design pressure difference between the inlet and outlet of the fan (500Pa) ; η_{fan} is the nominal fan efficiency (0.61), and ρ_{air} is the air density (1.225 kg/m^3). $i = 1,2,3$ is the room index.

The average air mass flow rate (in kg/s) for each office at a time step t is,

$$\dot{m}_{i,t} = \begin{cases} \text{if } u_{i,t} > \dot{m}_{\min} c_p\left(T_{\text{sup}} - T_{\text{air},i,t}\right), \dot{m}_{\min}, \\ \text{if } u_{i,t} \leq \dot{m}_{\min} c_p\left(T_{\text{sup}} - T_{\text{air},i,t}\right), \frac{u_{i,t}}{c_p\left(T_{\text{sup}} - T_{\text{air},i,t}\right)}. \end{cases} \tag{D-3}$$

$\dot{m}_{\min}$ is the minimum air flow rate of 0.081 kg/s when the room damper is at the minimum open position, assuming the density of the air is 1.225 kg/m^3; c_p is the specific heat of air, which is 1.004 kJ/kg · K; T_{sup} is the AHU supply air temperature of 16°C.

The chiller power at time t is computed with chiller electric input ratio model (DoE, 2010):

$$P_{\text{chiller},t} = f_{\text{PLR}}\left[\frac{-P_{\text{coil},t}}{f_{\text{CAP}}(T_{\text{out},t})}\right] \cdot f_{\text{CAP}}\left(T_{\text{out},t}\right) \cdot f_{\text{COP}}\left(T_{\text{out},t}\right), \tag{D-4}$$

where,

$$\begin{cases} f_{\text{CAP}} = Q_{\text{ref,Cap}} \cdot \text{Curve}_{\text{biquad},1}\left(T_{\text{leaving}}, T_{\text{out}}\right) \\ f_{\text{COP}} = \dfrac{1}{C_{\text{ref,COP}}} \cdot \text{Curve}_{\text{biquad},2}\left(T_{\text{leaving}}, T_{\text{out}}\right) \\ f_{\text{PLR}} = \text{Curve}_{\text{quad}}\left(\dfrac{-P_{\text{cool}}}{f_{\text{CAP}}}\right) = \text{Curve}_{\text{quad}}(\text{PLR}) \end{cases} \tag{D-5}$$

$Q_{\text{ref,Cap}}$ is the reference capacity (68.9kW) of the chiller downscaled to 12%; $C_{\text{ref,COP}}$ is the reference chiller COP of 2.67; T_{leaving} is the leaving water temperature of the chiller, which is 7℃. The cooling rate at the cooling coil at time t is the sum of the cooling rate at each office divided by a constant $\gamma = 0.8$, assuming a fixed outdoor air ratio of 10% according ASHRAE Standard 62.1 (ASHRAE & ANSI, 2013) and sensible load ration of 85% based on the climate condition in West Lafayette, IN.

$$P_{\text{coil},t} = \frac{1}{\gamma}\sum_{i=1}^{N} P_{\text{cool},i,t}, \tag{D-6}$$

The cooling rate at each office at time t is calculated by:

$$P_{\text{cool},i,t} = \begin{cases} \text{if } u_{i,t} > \dot{m}_{\min} c_p \left(T_{\text{sup}} - T_{\text{air},i,t}\right), \dot{m}_{\min} c_p \left(T_{\text{sup}} - T_{\text{air},i,t}\right), \\ \text{if } u_{i,t} \le \dot{m}_{\min} c_p \left(T_{\text{sup}} - T_{\text{air},i,t}\right), u_{i,t}. \end{cases} \tag{D-7}$$

The reheat power at time t is the sum of heating rate at each office divided by the boiler efficiency η, which is assumed to be 0.9 in this study:

$$P_{\text{reheat},t} = \sum_{i=1}^{N} \frac{P_{\text{heat},i,t}}{\eta}, \tag{D-8}$$

where,

$$P_{\text{heat},i,t} = \begin{cases} \text{if } u_{i,t} > \dot{m}_{\min} c_p \left(T_{\text{sup}} - T_{\text{air},i,t}\right), u_{i,t} - \dot{m}_{\min} c_p \left(T_{\text{sup}} - T_{\text{air},i,t}\right), \\ \text{if } u_{i,t} \le \dot{m}_{\min} c_p \left(T_{\text{sup}} - T_{\text{air},i,t}\right), 0. \end{cases} \tag{D-9}$$

APPENDIX E. SURVEY QUESTIONS

Table E.1. Survey questions.

OVERRIDE SURVEY	
Questions	**Options**
1. Please enter your subject number	
2. Please select the answer that best describes the energy performance information and its impact on your latest thermostat adjustment (displayed in Setup 2 only).	• I considered the energy performance information • I did not consider the energy performance information
3. How satisfied are you with current thermal conditions?	• I prefer warmer • I prefer slightly warmer • I'm satisfied with current condition • I prefer slightly cooler • I prefer cooler
EXIT SURVEY	
Questions	**Options**
1. Please enter your subject number	
2. During your stay in the office today, were there any occasions where the thermal condition was continuously unpleasant and/or interfering with your ability to focus on your work for more than 30 minutes?	• Never • Rarely • A few times • Most of the time • Always

APPENDIX F. SETPOINT TEMPERATURE PROFILES FROM 4 OCCUPANTS

Figure F.1 shows 4 examples of the setpoint temperature profiles from the occupants. The orange dash lines represent the MPC setpoints, which varied during a day based on the outdoor air temperature and solar irradiance forecast. The MPC setpoints were usually low in the morning (around 20 to 21℃), and increased in the afternoon. The blue line represents the actual setpoint temperatures in the offices between 10 a.m. to 5 p.m. When the actual setpoints and MPC setpoints are not equal, it means that the occupants made thermostat adjustments. Occupants 9, 20 and 11 reported to prefer slightly warmer or warmer conditions in the thermal preference votes. Most of the other occupants shared similar patterns in thermostat interaction with Occupant 9 and 20, although the choice of setpoint temperatures varied. Both of them adjusted the thermostats less frequently in Setup 2 compared to Setup 1. Also, the setpoints selected by Occupant 9 were often over 24℃ and reached 25.5℃ in Setup 1, while in Setup 2 the setpoint temperatures were less than 23℃. Occupant 20 selected setpoint temperatures around 24℃ most of the time both in Setup 1 and 2. Occupant 11 were mostly satisfied with the setpoint temperatures determined by MPC and rarely adjusted the thermostat in both setups, which is similar to the behavior of 5 other occupants participated in the experiment. However, Occupant 3 demonstrated different behavior, which is consistently reducing the setpoint to 20℃ in every thermostat adjustment, while all others mostly increased the setpoints. Therefore, it is highly likely that Occupant 3's thermal preference is distinctively different from others and belongs to a different preference cluster.

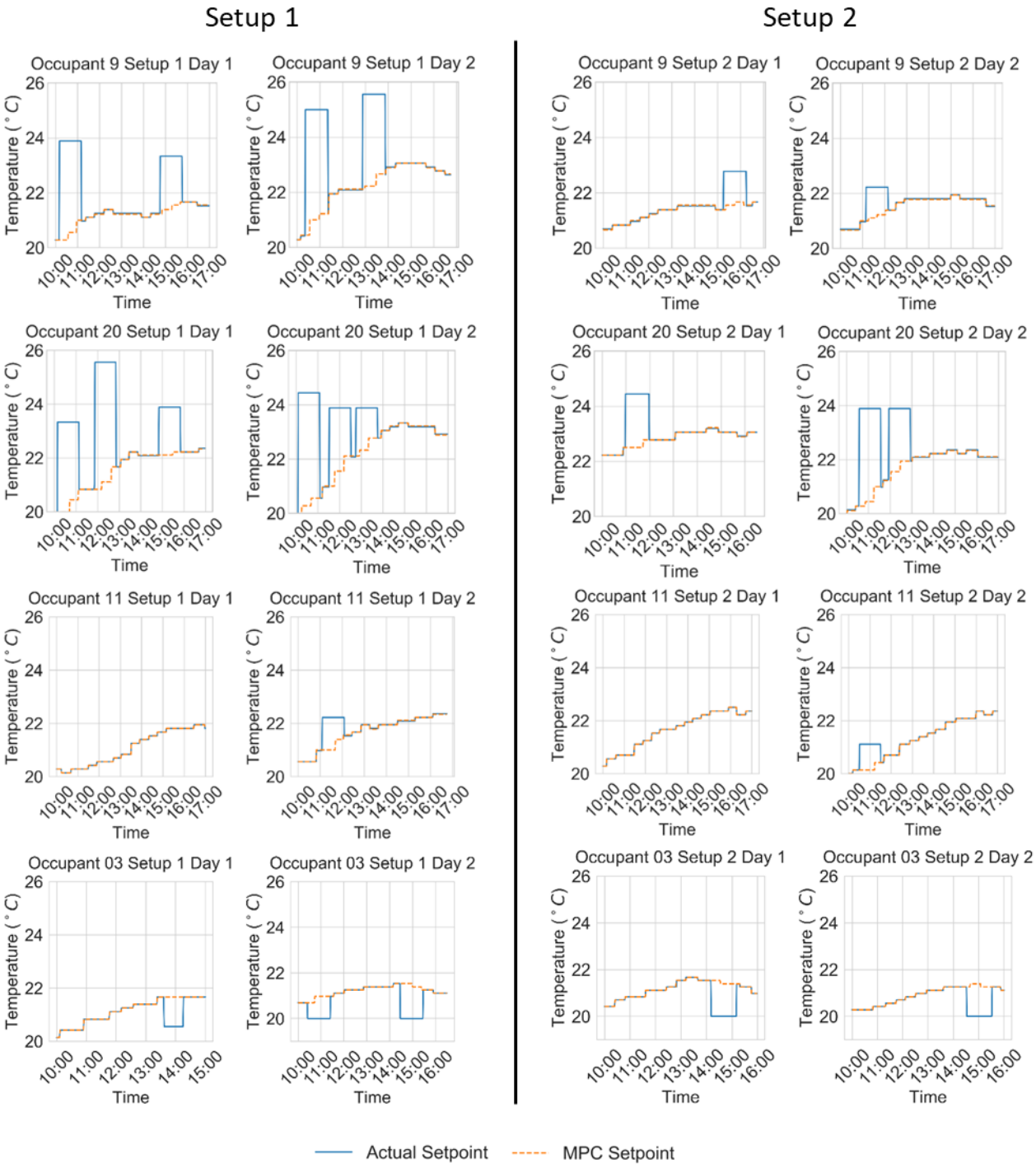


Figure F.1. Setpoint temperature profiles from 4 typical occupants.

APPENDIX G. DETERMINE THE UNCERTAIN PARAMETERS IN THE MODEL UNIVERSE

We assume that the thermal capacitances and resistances follow truncated Gaussian distributions $\mathcal{N}(\cdot\,|\mu_{\mathrm{C}}, \sigma_c^2, 0, \infty)$ and $\mathcal{N}(\cdot\,|\mu_{\mathrm{R}}, \sigma_R^2, 0, \infty)$, respectively. The mean values for the parameters are determined by:

$$\mu_{\mathrm{C}} = A\left(\sum_{i=1}^{L} x_i \rho_i \mathrm{cp}_i\right), \tag{G-1}$$

$$\mu_{\mathrm{R}} = \frac{\left(r_{\mathrm{si}} + r_{\mathrm{so}} + \sum_{i=1}^{L} \frac{x_i}{k_i}\right)}{A}. \tag{G-2}$$

x_i and k_i are the thickness and the thermal conductivity of wall layers respectively. A is the overall area of the construction element. r_{si} and r_{so} are the inside and outside surface convection resistance. ρ_i and cp_i designate the density and the specific heat capacity of the layer. L is the total number of the construction layers. The construction elements considered for each capacitor and resistor are listed in Table G.1. The material, area and thickness of construction elements are available from the construction drawing of the test-bed building, while the thermal properties of the materials are available at ASHRAE Handbook (ASHRAE, 2017). As we assume uninformative priors, the standard deviation considered is 0.5 multiplied by the mean values ($\sigma_C = 0.5\mu_C$, $\sigma_R = 0.5\mu_R$), and the all the parameters have lower bound of 0.

Table G.1. Construction elements considered for each resistance and capacitance parameter.

R_{oe}	• Exterior wall (aluminum board, concrete)
R_{er}	• Air (room volume) • Gypsum board surface
R_{rc}	• Interior partition (gypsum board, air gap) • Ceiling (ceiling board, air gap)
C_{env}	• Exterior wall • Floor (concrete)
C_{air}	• Air (room volume) • Interior partition (gypsum board, air gap) • Ceiling (ceiling board, air gap)

APPENDIX H. EQUIVALENT MPC PROBLEM FOR REINFORCEMENT LEARNING CONTROL CASE STUDY

In the equivalent MPC problem, the control variable is still the action a_t, and the objective is to minimize the negative of the reward over a prediction horizon of 24 time-steps (12 hours). The control horizon remains 1 time-step, and the building dynamics follow equation (H-3), which is equivalent to what Figure 5.1 described.

$$\min_{a_0, a_1, \cdots, a_{23}} \sum_{t=0}^{24} -r_t \,, \tag{H-1}$$

$s.t.$

$$0 \leq a_t \leq 1, \tag{H-2}$$

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}_a a_t + \mathbf{B}_w \mathbf{w}_t, \tag{H-3}$$

Where $\mathbf{x}_t = \begin{bmatrix} T_{\text{env},t} \\ T_{\text{air},t} \end{bmatrix}$, $\mathbf{w}_t = \begin{bmatrix} T_{\text{out},t} \\ T_{\text{cor},t} \\ Q_{\text{SG},t} \\ Q_{\text{int},t} \end{bmatrix}$, and $\mathbf{s}_t = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{w}_t \end{bmatrix}$. $\mathbf{A} \in \mathbb{R}^{2\times 2}$, $\mathbf{B}_a \in \mathbb{R}^{2\times 1}$, $\mathbf{B}_w \in \mathbb{R}^{2\times 4}$ are time invariant matrices. The parameters of the dynamical model (i.e. elements in matrices $\mathbf{A}$, $\mathbf{B}_a$, and $\mathbf{B}_w$) are obtained from a system identification experiment described in Section 4.2.2.1.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Ghemawat, S. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Afram, A., & Janabi-Sharifi, F. (2014). Theory and applications of HVAC control systems–A review of model predictive control (MPC). *Building and Environment*, *72*, 343-355.

Afram, A., Janabi-Sharifi, F., Fung, A. S., & Raahemifar, K. (2017). Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system. *Energy and Buildings*, *141*, 96-113.

American Society of Heating, Refrigerating, Air-Conditioning Engineers, & American National Standards Institute (ASHRAE & ANSI). (2017). Standard 55-2017. *Thermal environmental conditions for human occupancy*.

American Society of Heating, Refrigerating, Air-Conditioning Engineers, & American National Standards Institute (ASHRAE & ANSI). (2013). Standard 62.1-2013. *Ventilation for Acceptable Indoor Air Quality*.

American Society of Heating, Refrigerating, Air-Conditioning Engineers, & American National Standards Institute (ASHRAE & ANSI). (2019). Standard 90.1-2013. *Energy Standard for Buildings Except Low-Rise Residential Buildings.*

American Society of Heating, Refrigerating, Air-Conditioning Engineers (ASHRAE). (2017). ASHRAE Handbook. *Fundamentals.*

Andriamamonjy, A., Klein, R., & Saelens, D. (2019). Automated grey box model implementation using BIM and Modelica. *Energy and Buildings*, *188*, 209-225.

Aswani, A., Master, N., Taneja, J., Culler, D., & Tomlin, C. (2011). Reducing transient and steady state electricity consumption in HVAC using learning-based model-predictive control. *Proceedings of the IEEE*, *100*(1), 240-253.

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., ... & Berges, M. (2018). Brick: Metadata schema for portable smart building applications. *Applied energy*, *226*, 1273-1292.

Bellman, R. (1954). *The theory of dynamic programming* (No. RAND-P-550). Rand Corp., Santa Monica, CA.

Benedetti, M., Cesarotti, V., Introna, V., & Serranti, J. (2016). Energy consumption control automation using Artificial Neural Networks and adaptive algorithms: Proposal of a new methodology and case study. *Applied Energy*, *165*, 60-71.

Bengea, S., Kelman, A., Borrelli, F., Taylor, R., & Narayanan, S. (2012, August). Model predictive control for mid-size commercial building hvac: Implementation, results and energy savings. In *Second international conference on building energy and environment* (pp. 979-986).

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Bertsekas, D. P., & Tsitsiklis, J. N. (1995, December). Neuro-dynamic programming: an overview. In *Proceedings of 1995 34th IEEE Conference on Decision and Control* (Vol. 1, pp. 560-564). IEEE.

Bertsekas, D. P., (1995). *Dynamic programming and optimal control* (Vol. 1, No. 2, p. 4). Belmont, MA: Athena scientific.

Bilionis, I., Constantinescu, E. M., & Anitescu, M. (2014). Data-driven model for solar irradiation based on satellite observations. *Solar energy*, *110*, 22-38.

Bird, R. E., & Hulstrom, R. L. (1981). *Simplified clear sky model for direct and diffuse insolation on horizontal surfaces* (No. SERI/TR-642-761). Solar Energy Research Inst., Golden, CO (USA).

Brager, G., & Arens, E. (2016). Center for the Built Environment: tools & technologies for performance. *Room One Thousand*, *4*(4).

Braun, J. E. (1990). Reducing energy costs and peak electrical demand through optimal control of building thermal storage. *ASHRAE transactions*, *96*(2), 876-888.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). Openai gym. *arXiv preprint arXiv:1606.01540*.

Bünning, F., Huber, B., Heer, P., Aboudonia, A., & Lygeros, J. (2020). Experimental demonstration of data predictive control for energy optimization and thermal comfort in buildings. *Energy and Buildings*, *211*, 109792.

Cai, J. (2015). Gray-box modeling of multistage direct-expansion units to enable control system optimization. *ASHRAE Transactions*, *121*, 203.

Cai, J., & Braun, J. (2016). An inverse hygrothermal model for multi-zone buildings. *Journal of Building Performance Simulation*, *9*(5), 510-528.

Candanedo Ibarra, J. (2011). *A study of predictive control strategies for optimally designed solar homes* (Doctoral dissertation, Concordia University).

Candanedo, J. A., & Athienitis, A. K. (2010). Investigation of Anticipatory Control Strategies in a Net-Zero Energy Solar House. *ASHRAE Transactions*, *116*(1).

Candanedo, J. A., & Athienitis, A. K. (2011). Predictive control of radiant floor heating and solar-source heat pump operation in a solar house. *HVAC&R Research*, *17*(3), 235-256.

Candanedo, J. A., Dehkordi, V. R., & Stylianou, M. (2013). Model-based predictive control of an ice storage device in a building cooling system. *Applied Energy*, *111*, 1032-1045.

Caraban, A., Karapanos, E., Gonçalves, D., & Campos, P. (2019, May). 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).

Chakraborty, D., Elzarka, H., & Bhatnagar, R. (2016). Generation of accurate weather files using a hybrid machine learning methodology for design and analysis of sustainable and resilient buildings. *Sustainable Cities and Society*, *24*, 33-41.

Charnes, A., & Cooper, W. W. (1959). Chance-constrained programming. *Management science*, *6*(1), 73-79.

Chen, B., Cai, Z., & Bergés, M. (2019, November). Gnu-rl: A precocial reinforcement learning solution for building hvac control using a differentiable mpc policy. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 316-325).

Chen, C., Duan, S., Cai, T., & Liu, B. (2011). Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Solar energy*, *85*(11), 2856-2870.

Chen, J., Augenbroe, G., & Song, X. (2018). Lighted-weighted model predictive control for hybrid ventilation operation based on clusters of neural network models. *Automation in Construction*, *89*, 250-265.

Chen, Y., Norford, L. K., Samuelson, H. W., & Malkawi, A. (2018). Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. *Energy and Buildings*, *169*, 195-205.

Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems* (pp. 4299-4307).

Cígler, J., Gyalistras, D., Široky, J., Tiet, V., & Ferkl, L. (2013, June). Beyond theory: the challenge of implementing model predictive control in buildings. In *Proceedings of 11th Rehva world congress, Clima* (Vol. 250).

Corbin, C. D., Henze, G. P., & May-Ostendorp, P. (2013). A model predictive control optimization environment for real-time commercial building application. *Journal of Building Performance Simulation*, *6*(3), 159-174.

Costanzo, G. T., Iacovella, S., Ruelens, F., Leurs, T., & Claessens, B. J. (2016). Experimental analysis of data-driven control for a building heating system. *Sustainable Energy, Grids and Networks*, *6*, 81-90.

Dahlin, J., Lindsten, F., & Schön, T. B. (2015). Particle Metropolis–Hastings using gradient and Hessian information. *Statistics and computing*, *25*(1), 81-92.

Dalamagkidis, K., Kolokotsa, D., Kalaitzakis, K., & Stavrakakis, G. S. (2007). Reinforcement learning for energy conservation and comfort in buildings. *Building and environment*, *42*(7), 2686-2698.

Dalcín, L., Paz, R., Storti, M., & D'Elía, J. (2008). MPI for Python: Performance improvements and MPI-2 extensions. *Journal of Parallel and Distributed Computing*, *68*(5), 655-662.

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

Day, J. K., McIlvennie, C., Brackley, C., Tarantini, M., Piselli, C., Hahn, J., ... & Pritoni, M. (2020). A review of select human-building interfaces and their relationship to human behavior, energy use and occupant comfort. *Building and Environment*, 106920.

Day, J., & Heschong, L. (2016). Understanding behavior potential: the role of building interfaces. In *ACEEE Summer Study on Energy Efficiency in Buildings* (Vol. 8, pp. 1-10).

Deisenroth, M. P., Rasmussen, C. E., & Peters, J. (2009). Gaussian process dynamic programming. *Neurocomputing*, *72*(7-9), 1508-1524.

Delmas, M. A., & Kaiser, W. (2014). *Behavioral responses to real-time individual energy usage information: A large scale experiment*. California Environmental Protection Agency, Air Resources Board, Research Division.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1-22.

Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., ... & Yazdanian, M. (2011). US Department of Energy commercial reference building models of the national building stock.

Ding, X., Du, W., & Cerpa, A. (2019, November). OCTOPUS: Deep reinforcement learning for holistic smart building control. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 326-335).

Dobbs, J. R., & Hencey, B. M. (2014). Model predictive HVAC control with online occupancy model. *Energy and Buildings*, *82*, 675-684.

DoE, U. S. (2010). Energyplus engineering reference. *The reference to energyplus calculations*.

Dong, B., & Lam, K. P. (2014, February). A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting. In *Building Simulation* (Vol. 7, No. 1, pp. 89-106). Springer Berlin Heidelberg.

Dongarra, J., Walker, D., Lusk, E., Knighten, B., Snir, M., Geist, A., ... & Cownie, J., (1994). Special issue-mpi-a message-passing interface standard. *International Journal of Supercomputer Applications and High Performance Computing*, *8*(3-4), 165.

Duan, Y., Schulman, J., Chen, X., Bartlett, P. L., Sutskever, I., & Abbeel, P. (2016). Rl $^ 2$: Fast reinforcement learning via slow reinforcement learning. *arXiv preprint arXiv:1611.02779*.

Emeakaroha, A., Ang, C. S., Yan, Y., & Hopthrow, T. (2014). A persuasive feedback support system for energy conservation and carbon emission reduction in campus residential buildings. *Energy and buildings*, *82*, 719-732.

Erickson, V. L., & Cerpa, A. E. (2012, November). Thermovote: participatory sensing for efficient building hvac conditioning. In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings* (pp. 9-16).

Ferkl, L., & Široký, J. (2010). Ceiling radiant cooling: Comparison of ARMAX and subspace identification modelling methods. *Building and Environment*, *45*(1), 205-212.

Fernandez, N. E., Katipamula, S., Wang, W., Xie, Y., Zhao, M., & Corbin, C. D. (2017). *Impacts of commercial building controls on energy savings and peak load reduction* (No. PNNL-25985). Pacific Northwest National Lab.(PNNL), Richland, WA (United States).

Ferreira, P. M., Ruano, A. E., Silva, S., & Conceicao, E. Z. E. (2012). Neural networks based predictive control for thermal comfort and energy savings in public buildings. *Energy and buildings*, *55*, 238-251.

Finn, C., Abbeel, P., & Levine, S. (2017, July). Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *International Conference on Machine Learning* (pp. 1126-1135).

Fishburn, P. C. (1970). *Utility theory for decision making* (No. RAC-R-105). Research analysis corp McLean VA.

Froehlich, J. (2009, February). Promoting energy efficient behaviors in the home through feedback: The role of human-computer interaction. In *Proc. HCIC Workshop* (Vol. 9, pp. 1-11).

Froehlich, J., Findlater, L., & Landay, J. (2010, April). The design of eco-feedback technology. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1999-2008).

Gandhi, P., & Brager, G. S. (2016). Commercial office plug load energy consumption trends and the role of occupant behavior. *Energy and Buildings*, *125*, 1-8.

Garcia, C. E., Prett, D. M., & Morari, M. (1989). Model predictive control: theory and practice—a survey. *Automatica*, *25*(3), 335-348.

Garifi, K., Baker, K., Touri, B., & Christensen, D. (2018, August). Stochastic model predictive control for demand response in a home energy management system. In *2018 IEEE Power & Energy Society General Meeting (PESGM)* (pp. 1-5). IEEE.

Gayeski, N. T., Armstrong, P. R., & Norford, L. K. (2012). Predictive pre-cooling of thermo-active building systems with low-lift chillers. *HVAC&R Research*, *18*(5), 858-873.

Ghahramani, A., Jazizadeh, F., & Becerik-Gerber, B. (2014). A knowledge based approach for selecting energy-aware and comfort-driven HVAC temperature set points. *Energy and Buildings*, *85*, 536-548.

Gordon, N. J., Salmond, D. J., & Smith, A. F. (1993, April). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE proceedings F (radar and signal processing)* (Vol. 140, No. 2, pp. 107-113). IET Digital Library.

Gulbinas, R., & Taylor, J. E. (2014). Effects of real-time eco-feedback and organizational network dynamics on energy efficient behavior in commercial buildings. *Energy and buildings*, *84*, 493-500.

Gulbinas, R., Jain, R. K., & Taylor, J. E. (2014). BizWatts: A modular socio-technical energy management system for empowering commercial building occupants to conserve energy. *Applied Energy*, *136*, 1076-1084.

Han, M., May, R., Zhang, X., Wang, X., Pan, S., Yan, D., ... & Xu, L. (2019). A review of reinforcement learning methodologies for controlling occupant comfort in buildings. *Sustainable Cities and Society*, *51*, 101748.

Hensman, J., Fusi, N., Andrade, R., Durrande, N., Saul, A., Zwiessele, M., & Lawrence, N. D. (2012). GPy: A gaussian process framework in python.

Henze, G. P., Kalz, D. E., Felsmann, C., & Knabe, G., (2004). Impact of forecasting accuracy on predictive optimal control of active and passive building thermal storage inventory. *HVAC&R Research*, *10*(2), 153-178.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, *81*(396), 945-960.

Hu, J., & Karava, P. (2014). Model predictive control strategies for buildings with mixed-mode cooling. *Building and Environment*, *71*, 233-244.

Huang, H., Chen, L., & Hu, E. (2014, June). Model predictive control for energy-efficient buildings: An airport terminal building study. In *11th IEEE International Conference on Control & Automation (ICCA)* (pp. 1025-1030). IEEE.

Ibarz, B., Leike, J., Pohlen, T., Irving, G., Legg, S., & Amodei, D. (2018). Reward learning from human preferences and demonstrations in Atari. In *Advances in neural information processing systems* (pp. 8011-8023).

Iman, R. L. (2008). *Latin hypercube sampling*. John Wiley & Sons, Ltd.

Jain, A., Behl, M., & Mangharam, R. (2017, May). Data Predictive Control for building energy management. In *2017 American Control Conference (ACC)* (pp. 44-49). IEEE.

Jameson, A., Berendt, B., Gabrielli, S., Cena, F., Gena, C., Vernero, F., & Reinecke, K. (2014). *Choice architecture for human-computer interaction*.

Jaramillo, R. C., Braun, J. E., & Horton, W. T. (2016). A multi-agent control approach for optimization of central cooling plants.

Jazizadeh, F., Ghahramani, A., Becerik-Gerber, B., Kichkaylo, T., & Orosz, M. (2014). User-led decentralized thermal comfort driven HVAC operations for improved efficiency in office buildings. *Energy and Buildings*, *70*, 398-410.

Jia, R., Jin, M., Sun, K., Hong, T., & Spanos, C. (2019). Advanced building control via deep reinforcement learning. *Energy Procedia*, *158*, 6158-6163.

Joe, J., & Karava, P. (2019). A model predictive control strategy to optimize the performance of radiant floor heating and cooling systems in office buildings. *Applied Energy, 245*, 65-77.

Jones, E., Oliphant, T., & Peterson, P., (2016). SciPy: Open source scientific tools for Python. 2001. http://www. scipy. org.

Karlin, B., Koleva, S., Kaufman, J., Sanguinetti, A., Ford, R., & Chan, C. (2017, July). Energy UX: Leveraging Multiple Methods to See the Big Picture. In *International Conference of Design, User Experience, and Usability* (pp. 462-472). Springer, Cham.

Kasperbauer, T. J. (2017). The permissibility of nudging for sustainable energy consumption. *Energy Policy, 111*, 52-57.

Kelman, A., & Borrelli, F. (2011). Bilinear model predictive control of a HVAC system using sequential quadratic programming. *IFAC Proceedings Volumes*, *44*(1), 9869-9874.

Kiliccote, S., Olsen, D., Sohn, M. D., & Piette, M. A. (2016). Characterization of demand response in the commercial, industrial, and residential sectors in the United States. *Wiley Interdisciplinary Reviews: Energy and Environment*, *5*(3), 288-304.

Killian, M., & Kozek, M. (2016). Ten questions concerning model predictive control for energy efficient buildings. *Building and Environment, 105*, 403-412.

Kirsch, L., van Steenkiste, S., & Schmidhuber, J. (2019, September). Improving Generalization in Meta Reinforcement Learning using Learned Objectives. In *International Conference on Learning Representations*.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of computational and graphical statistics*, *5*(1), 1-25.

Klein, S., Beckman, A., Mitchell, W., & Duffie, A. (2011). TRNSYS 17-A TRansient SYstems Simulation program. *Solar Energy Laboratory, University of Wisconsin, Madison*.

Konstantakopoulos, I. C., Barkan, A. R., He, S., Veeravalli, T., Liu, H., & Spanos, C. (2019). A deep learning and gamification approach to improving human-building interaction and energy efficiency in smart infrastructure. *Applied energy, 237*, 810-821.

Konstantakopoulos, I., Spanos, C. J., & Sastry, S. S. (2015). *Social game for building energy efficiency: Utility learning, simulation, analysis and incentive design*. Tech. rep. Technical Report UCB/EECS-2015-3, EECS Department, University of California, Berkeley.

Kraft, D. (1988). A software package for sequential quadratic programming.

Lanza, P. A. G., & Cosme, J. M. Z., (2001). A short-term temperature forecaster based on a novel radial basis functions neural network. *International journal of neural systems*, *11*(01), 71-77.

Lazic, N., Boutilier, C., Lu, T., Wong, E., Roy, B., Ryu, M. K., & Imwalle, G. (2018). Data center cooling using model-predictive control. In *Advances in Neural Information Processing Systems* (pp. 3814-3823).

Lazos, D., Sproul, A. B., & Kay, M. (2014). Optimisation of energy management in commercial buildings with weather forecasting inputs: A review. *Renewable and Sustainable Energy Reviews*, *39*, 587-603.

Lazos, D., Sproul, A. B., & Kay, M. (2015). Development of hybrid numerical and statistical short term horizon weather prediction models for building energy management optimisation. *Building and Environment*, *90*, 82-95.

Lee, D., Lee, S., Karava, P., & Hu, J. (2018a, June). Simulation-based policy gradient and its building control application. In *2018 Annual American Control Conference (ACC)* (pp. 5424-5429). IEEE.

Lee, D., Lee, S., Karava, P., & Hu, J. (2018b, June). Approximate dynamic programming for building control problems with occupant interactions. In *2018 Annual American Control Conference (ACC)* (pp. 3945-3950). IEEE.

Lee, S., Bilionis, I., Karava, P., & Tzempelikos, A. (2017). A Bayesian approach for probabilistic classification and inference of occupant thermal preferences in office buildings. *Building and Environment*, *118*, 323-343.

Lee, S., Joe, J., Karava, P., Bilionis, I., & Tzempelikos, A. (2019). Implementation of a self-tuned HVAC controller to satisfy occupant thermal preferences and optimize energy use. *Energy and Buildings, 194*, 301-316.

Lehrer, D. R., Vasudev, J., & Kaam, S. (2014). A usability study of a social media prototype for building energy feedback and operations.

Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., & Quillen, D. (2018). Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *The International Journal of Robotics Research*, *37*(4-5), 421-436.

Li, D., Xu, X., Chen, C. F., & Menassa, C. (2019). Understanding energy-saving behaviors in the American workplace: A unified theory of motivation, opportunity, and ability. *Energy Research & Social Science, 51*, 198-209.

Li, S., & Karava, P., (2014). Energy modeling of photovoltaic thermal systems with corrugated unglazed transpired solar collectors–Part 2: Performance analysis. *Solar Energy*, *102*, 297-307.

Li, S., Joe, J., Hu, J., & Karava, P., (2015). System identification and model-predictive control of office buildings with integrated photovoltaic-thermal collectors, radiant floor heating and active thermal storage. *Solar Energy*, *113*, 139-157.

Li, S., Karava, P., Currie, S., Lin, W. E., & Savory, E., (2014). Energy modeling of photovoltaic thermal systems with corrugated unglazed transpired solar collectors–Part 1: Model development and validation. *Solar Energy, 102*, 282-296.

Li, X., & Wen, J. (2014). Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews*, *37*, 517-537.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., ... & Wierstra, D. (2016, January). Continuous control with deep reinforcement learning. In *ICLR (Poster)*.

Liu, S., & Henze, G. P. (2006a). Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory: Part 1: Theoretical foundation. Energy and buildings, 38(2), 142-147.

Liu, S., & Henze, G. P. (2006b). Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory: Part 2: Results and analysis. *Energy and buildings*, *38*(2), 148-161.

Liu, X., Paritosh, P., Awalgaonkar, N. M., Bilionis, I., & Karava, P. (2018). Model predictive control under forecast uncertainty for optimal operation of buildings with integrated solar systems. *Solar energy*, *171*, 953-970.

Ma, Y., Kelman, A., Daly, A., & Borrelli, F. (2012). Predictive control for energy efficient buildings with thermal storage: Modeling, stimulation, and experiments. *IEEE control systems magazine*, *32*(1), 44-64.

Ma, Y., Matuško, J., & Borrelli, F. (2014). Stochastic model predictive control for building HVAC systems: Complexity and conservatism. *IEEE Transactions on Control Systems Technology*, *23*(1), 101-116.

MacKay, D. J. (1992). Information-based objective functions for active data selection. *Neural computation*, *4*(4), 590-604.

Mason, K., & Grijalva, S. (2019). A review of reinforcement learning for autonomous building energy management. *Computers & Electrical Engineering*, *78*, 300-312.

Mathiesen, P., & Kleissl, J., (2011). Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. *Solar Energy, 85*(5), 967-977.

Mayne, D. Q., Rawlings, J. B., Rao, C. V., & Scokaert, P. O. (2000). Constrained model predictive control: Stability and optimality. *Automatica*, *36*(6), 789-814.

May-Ostendorp, P., Henze, G. P., Corbin, C. D., Rajagopalan, B., & Felsmann, C. (2011). Model-predictive control of mixed-mode buildings with rule extraction. *Building and Environment, 46*(2), 428-437.

Miller, C., & Meggers, F. (2017). The Building Data Genome Project: An open, public data set from non-residential building electrical meters. *Energy Procedia*, *122*, 439-444.

Miller, C., Kathirgamanathan, A., Picchetti, B., Arjunan, P., Park, J. Y., Nagy, Z., ... & Meggers, F. (2020). The Building Data Genome Project 2, energy meter data from the ASHRAE Great Energy Predictor III competition. *Scientific Data*, *7*(1), 1-13.

Mirakhorli, A., & Dong, B. (2016). Occupancy behavior based model predictive control for building indoor climate—A critical review. *Energy and Buildings*, *129*, 499-513.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., ... & Kavukcuoglu, K. (2016, June). Asynchronous methods for deep reinforcement learning. In *International conference on machine learning* (pp. 1928-1937).

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. *nature, 518*(7540), 529-533.

Murtagh, N., Nati, M., Headley, W. R., Gatersleben, B., Gluhak, A., Imran, M. A., & Uzzell, D. (2013). Individual energy use and feedback in an office setting: A field trial. *Energy Policy, 62*, 717-728.

National Oceanic and Atmospheric Administration (NOAA). (2020). Climate Data Online Data Tools (https://www.ncdc.noaa.gov/cdo-web/datatools). National Oceanic and Atmospheric Administration.

Nichol, A., Achiam, J., & Schulman, J. (2018). On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*.

Obinna, U., Joore, P., Wauben, L., & Reinders, A. (2017). Comparison of two residential Smart Grid pilots in the Netherlands and in the USA, focusing on energy performance and user experiences. *Applied energy*, *191*, 264-275.

Oldewurtel, F., Parisio, A., Jones, C. N., Gyalistras, D., Gwerder, M., Stauch, V., ... & Morari, M., (2012). Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, *45*, 15-27.

Oldewurtel, F., Sturzenegger, D., & Morari, M. (2013). Importance of occupancy information for building climate control. *Applied energy*, *101*, 521-532.

Orland, B., Ram, N., Lang, D., Houser, K., Kling, N., & Coccia, M. (2014). Saving energy in an office environment: A serious game intervention. *Energy and Buildings*, *74*, 43-52.

Paritosh, P., Bilionis, I., Liu, X., (2017). Dynamic programming review. Retrieved from https://bitbucket.org/parth_paritosh/dynamic_programming_review.

Park, J. Y., Dougherty, T., Fritz, H., & Nagy, Z. (2019). LightLearn: An adaptive and occupant centered controller for lighting based on reinforcement learning. *Building and Environment*, *147*, 397-414.

Patil, A., Huard, D., & Fonnesbeck, C. J. (2010). PyMC: Bayesian stochastic modelling in Python. *Journal of statistical software*, *35*(4), 1.

Payne, S. J., & Howes, A. (2013). Adaptive interaction: A utility maximization approach to understanding human interaction with technology. *Synthesis Lectures on Human-Centered Informatics*, *6*(1), 1-111.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics surveys*, *3*, 96-146.

Pedersen, T. H., & Petersen, S. (2018). Investigating the performance of scenario-based model predictive control of space heating in residential buildings. *Journal of Building Performance Simulation*, *11*(4), 485-498.

Peirelinck, T., Ruelens, F., & Decnoninck, G. (2018, June). Using reinforcement learning for optimizing heat pump control in a building model in Modelica. In *2018 IEEE International Energy Conference (ENERGYCON)* (pp. 1-6). IEEE.

Perez, R. E., Jansen, P. W., & Martins, J. R. (2012). pyOpt: a Python-based object-oriented framework for nonlinear constrained optimization. *Structural and Multidisciplinary Optimization*, *45*(1), 101-118.

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, S., & Bundgaard, K. W., 2014. The effect of weather forecast uncertainty on a predictive control concept for building systems operation. *Applied Energy, 116,* 311-321.

Piette, M. A., Watson, D., Motegi, N., & Kiliccote, S. (2007). *Automated Critical PeakPricing Field Tests: 2006 Pilot ProgramDescription and Results* (No. LBNL-62218). Ernest Orlando Lawrence Berkeley NationalLaboratory, Berkeley, CA (US).

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-computer interaction*. Addison-Wesley Longman Ltd..

Privara, S., Cigler, J., Váňa, Z., Oldewurtel, F., Sagerschnig, C., & Žáčeková, E. (2013). Building modeling as a crucial part for building predictive control. *Energy and Buildings, 56,* 8-22.

Privara, S., Široký, J., Ferkl, L., & Cigler, J. (2011). Model predictive control of a building heating system: The first experience. *Energy and Buildings, 43*(2-3), 564-572.

Promann, M., & Brunswicker, S. (2017). Affordances of eco-feedback design in home energy context. In *Proceedings of Twenty-third Americas Conference on Information Systems*.

Putta, V. K., Kim, D., Cai, J., Hu, J., & Braun, J. E., (2015). Dynamic programming based approaches to optimal rooftop unit coordination. *Science and Technology for the Built Environment, 21*(6), 752-760.

Quintana, H. J., & Kummert, M., (2015). Optimized control strategies for solar district heating systems. *Journal of Building Performance Simulation, 8*(2), 79-96.

Rakelly, K., Zhou, A., Finn, C., Levine, S., & Quillen, D. (2019, May). Efficient off-policy meta-reinforcement learning via probabilistic context variables. In *International conference on machine learning* (pp. 5331-5340).

Rasmussen, C. E., & Williams, C. K., (2006). *Gaussian processes for machine learning (Vol. 1)*. Cambridge: MIT press.

Ratliff, L. J., Jin, M., Konstantakopoulos, I. C., Spanos, C., & Sastry, S. S. (2014, September). Social game for building energy efficiency: Incentive design. In *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (pp. 1011-1018). IEEE.

Rau, P. L. P., Gong, Y., Huang, H. J., & Wen, J. (2016). A Systematic Study for Smart Residential Thermostats: User Needs for the Input, Output, and Intelligence Level. *Buildings*, *6*(2), 19.

Rayati, M., Sheikhi, A., & Ranjbar, A. M. (2015, February). Applying reinforcement learning method to optimize an Energy Hub operation in the smart grid. In *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)* (pp. 1-5). IEEE.

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.

Roth, A., & Reyna, J. (2019). *Grid-Interactive Efficient Buildings Technical Report Series: Whole-Building Controls, Sensors, Modeling, and Analytics* (No. NREL/TP-5500-75478; DOE/GO-102019-5230). National Renewable Energy Lab.(NREL), Golden, CO (United States).

Rouchier, S., Jiménez, M. J., & Castaño, S. (2019). Sequential Monte Carlo for on-line parameter estimation of a lumped building energy model. *Energy and Buildings*, *187*, 86-94.

Rubin, D. B. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, *81*(396), 961-962.

Ruelens, F., Claessens, B. J., Vandael, S., De Schutter, B., Babuška, R., & Belmans, R. (2016). Residential demand response of thermostatically controlled loads using batch reinforcement learning. *IEEE Transactions on Smart Grid*, *8*(5), 2149-2159.

Ruelens, F., Iacovella, S., Claessens, B. J., & Belmans, R. (2015). Learning agent for a heat-pump thermostat with a set-back strategy using model-free reinforcement learning. *Energies*, *8*(8), 8300-8318.

Sæmundsson, S., Hofmann, K., & Deisenroth, M. P. (2018, August). Meta reinforcement learning with latent variable Gaussian processes. In *34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018* (Vol. 34, pp. 642-652). Association for Uncertainty in Artificial Intelligence (AUAI).

Sanguinetti, A., Dombrovski, K., & Sikand, S. (2018). Information, timing, and display: A design-behavior framework for improving the effectiveness of eco-feedback. *Energy Research & Social Science*, *39*, 55-68.

Scheidegger, S., & Bilionis, I. (2019). Machine learning for high-dimensional dynamic stochastic economies. *Journal of Computational Science*, *33*, 68-82.

Schön, T. B., Lindsten, F., Dahlin, J., Wågberg, J., Naesseth, C. A., Svensson, A., & Dai, L. (2015). Sequential Monte Carlo methods for system identification. *IFAC-PapersOnLine*, *48*(28), 775-786.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Schweiger, G., Eckerstorfer, L., Hafner, I., Fleischhacker, A., Radl, J., Glock, B., ... & Corcoran, K. (2020). Active Consumer Participation in Smart Energy Systems. Energy and Buildings, 110359.

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.

Seo, D. (2010). Development of a universal model for predicting hourly solar radiation—Application: Evaluation of an optimal daylighting controller (Doctoral dissertation, University of Colorado at Boulder).

Shann, M., & Seuken, S. (2014, May). Adaptive home heating under weather and price uncertainty using GPs and MDPs. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems* (pp. 821-828).

Siero, F. W., Bakker, A. B., Dekker, G. B., & Van Den Burg, M. T. (1996). Changing organizational energy consumption behaviour through comparative feedback. *Journal of environmental psychology*, *16*(3), 235-246.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... & Lillicrap, T. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*(6419), 1140-1144.

Simon, H. A. (1990). Bounded rationality. In *Utility and probability* (pp. 15-18). Palgrave Macmillan, London.

Široký, J., Oldewurtel, F., Cigler, J., & Prívara, S. (2011). Experimental analysis of model predictive control for an energy efficient building heating system. *Applied energy*, *88*(9), 3079-3087.

Smarra, F., Jain, A., de Rubeis, T., Ambrosini, D., D'Innocenzo, A., & Mangharam, R. (2018). Data-driven model predictive control using random forests for building energy optimization and climate control. *Applied energy*, *226*, 1252-1272.

Stengel, R. F. (1994). Optimal control and estimation. Courier Corporation.

Sturzenegger, D., Gyalistras, D., Semeraro, V., Morari, M., & Smith, R. S. (2014, June). BRCM Matlab toolbox: Model generation for model predictive building control. In *2014 american control conference* (pp. 1063-1069). IEEE.

Tanner, R. A. (2014). Stochastic Optimization of Building Control Systems for Mixed-Mode Buildings. *University of Colorado, Boulder, CO*.

Tanner, R. A., & Henze, G. P. (2014). Stochastic control optimization for a mixed mode building considering occupant window opening behaviour. *Journal of Building Performance Simulation*, *7*(6), 427-444.

Tetlow, R. M., van Dronkelaar, C., Beaman, C. P., Elmualim, A. A., & Couling, K. (2015). Identifying behavioural predictors of small power electricity consumption in office buildings. *Building and Environment*, *92*, 75-85.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Timm, S. N., & Deal, B. M. (2016). Effective or ephemeral? The role of energy information dashboards in changing occupant energy behaviors. *Energy Research & Social Science*, *19*, 11-20.

U.S. Energy Information Administration (EIA). (2016). Commercial buildings energy consumption survey (CBECS), U.S. Department of Energy.

US Energy Information Administration (EIA). (2019). *Annual Energy Outlook 2019: With Projections to 2050*. Government Printing Office.

Vázquez-Canteli, J. R., & Nagy, Z. (2019). Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Applied energy*, *235*, 1072-1089.

Vázquez-Canteli, J. R., Kämpf, J., Henze, G., & Nagy, Z. (2019b, November). Citylearn v1. 0: An openai gym environment for demand response with deep reinforcement learning. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 356-357).

Vázquez-Canteli, J. R., Ulyanin, S., Kämpf, J., & Nagy, Z. (2019a). Fusing TensorFlow with building energy simulation for intelligent energy management in smart cities. *Sustainable cities and society*, *45*, 243-257.

Vellei, M., Natarajan, S., Biri, B., Padget, J., & Walker, I. (2016). The effect of real-time context-aware feedback on occupants' heating behaviour and thermal adaptation. *Energy and Buildings*, *123*, 179-191.

Verhelst, C., Logist, F., Van Impe, J., & Helsen, L. (2012). Study of the optimal control problem formulation for modulating air-to-water heat pumps connected to a residential floor heating system. *Energy and buildings*, *45*, 43-53.

Von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior (commemorative edition)*. Princeton university press.

Wang, J., Liu, Y., & Li, B. (2020). Reinforcement Learning with Perturbed Rewards. In *AAAI* (pp. 6202-6209).

Wang, L. (2009). *Model predictive control system design and implementation using MATLAB®*. Springer Science & Business Media.

Wang, S., & Jin, X. (2000). Model-based optimal control of VAV air-conditioning system using genetic algorithm. *Building and Environment*, *35*(6), 471-487.

Wang, Y., Velswamy, K., & Huang, B. (2017). A long-short term memory recurrent neural network based reinforcement learning controller for office heating ventilation and air conditioning systems. *Processes*, *5*(3), 46.

Wang, Z., & Hong, T. (2020). Reinforcement learning for building controls: The opportunities and challenges. *Applied Energy*, *269*, 115036.

West, S. R., Ward, J. K., & Wall, J. (2014). Trial results from a model predictive control and optimisation system for commercial building HVAC. *Energy and Buildings*, *72*, 271-279.

Wilcox, S., & Marion, W. (2008). *Users manual for tmy3 data sets (revised)* (No. NREL/TP-581-43156). National Renewable Energy Lab.(NREL), Golden, CO (United States).

Wilhite, H., & Ling, R. (1995). Measured energy savings from a more informative energy bill. *Energy and buildings*, *22*(2), 145-155.

Xu, T., Liu, Q., Zhao, L., & Peng, J. (2018). Learning to explore with meta-policy gradient. *arXiv preprint arXiv:1803.05044*.

Xu, X., Maki, A., Chen, C. F., Dong, B., & Day, J. K. (2017). Investigating willingness to save energy and communication about energy use in the American workplace with the attitude-behavior-context model. *Energy research & social science*, *32*, 13-22.

Yadav, A. K., & Chandel, S. S. (2012). Artificial neural network based prediction of solar radiation for Indian stations. *International Journal of Computer Applications*, *50*(9).

Yang, L., Nagy, Z., Goffin, P., & Schlueter, A. (2015). Reinforcement learning for optimal control of low exergy buildings. *Applied Energy*, *156*, 577-586.

Yang, S., Wan, M. P., Chen, W., Ng, B. F., & Dubey, S. (2020). Model predictive control with adaptive machine-learning-based model for building energy efficiency and comfort optimization. *Applied Energy*, *271*, 115147.

Yun, R., Lasternas, B., Aziz, A., Loftness, V., Scupelli, P., Rowe, A., ... & Zhao, J. (2013, April). Toward the design of a dashboard to promote environmentally sustainable behavior among office workers. In *International Conference on Persuasive Technology* (pp. 246-252). Springer, Berlin, Heidelberg.

Zavala, V. M., Constantinescu, E. M., Krause, T., & Anitescu, M. (2009). *Weather forecast-based optimization of integrated energy systems* (No. ANL/MCS-TM-305). Argonne National Lab.(ANL), Argonne, IL (United States).

Zeiler, W., van Houten, R., & Boxem, G. (2009). SMART buildings: Intelligent software agents. In *Sustainability in Energy and Buildings* (pp. 9-17). Springer, Berlin, Heidelberg.

Zhang, W., Xu, Y., Li, S., Zhou, M., Liu, W., & Xu, Y. (2016). A distributed dynamic programming-based solution for load management in smart grids. *IEEE Systems Journal*, *12*(1), 402-413.

Zhang, X., Schildbach, G., Sturzenegger, D., & Morari, M. (2013, July). Scenario-based MPC for energy-efficient building climate control under weather and occupancy uncertainty. In *2013 European Control Conference (ECC)* (pp. 1029-1034). IEEE.

Zhang, Y., Wang, Z., & Zhou, G. (2013). Antecedents of employee electricity saving behavior in organizations: An empirical study based on norm activation model. *Energy Policy*, *62*, 1120-1127.

Zhang, Z., & Lam, K. P. (2018, November). Practical implementation and evaluation of deep reinforcement learning control for a radiant heating system. In *Proceedings of the 5th Conference on Systems for Built Environments* (pp. 148-157).

Zhuang, X., & Wu, C. (2019). The effect of interactive feedback on attitude and behavior change in setting air conditioners in the workplace. *Energy and Buildings*, *183*, 739-748.

VITA

XIAOQI (CLARE) LIU

Ph.D., Lyles School of Civil Engineering, Purdue University (2015 – 2020)

M.S., Lyles School of Civil Engineering, Purdue University (2013 – 2015)

B.S., Department of Built Environment and Building Service Engineering, Chongqing University (2009 – 2013)

PUBLICATIONS

In Journals

Liu, X., Paritosh, P., Awalgaonkar, N. M., Bilionis, I., & Karava, P. (2018). Model predictive control under forecast uncertainty for optimal operation of buildings with integrated solar systems. *Solar Energy*, *171*, 953-970.

Liu, X., Lee, S., Bilionis, I., Karava, P., Joe, J., & Sadeghi, A., A user-interactive system for smart thermal environment control in office buildings. *Applied Energy*. Submitted.

Liu, X., Karava, P., Bilionis, I., A meta-reinforcement learning approach for optimal HVAC control. *Journal of Building Performance Simulation*. Under final preparation.

Awalgaonkar, N., Bilionis, I, **Liu, X.**, Karava, P., Tzempelikos, T., Learning personalized thermal preferences via Bayesian active learning with unimodality constraints. *Engineering Applications of Artificial Intelligence*. Submitted.

In Conference Proceedings

Liu, X., Paritosh, P., Awalgaonkar, N. M., Bilionis, I., & Karava, P. (2018, July). Optimal solar energy utilization in building operation under weather uncertainty. In *Proceedings of 5th High Performance Buildings Conference.*

Zhang, H., Kim, M., **Liu, X.**, Tzempelikos, T. (2021, January). A comparison of sensing type and control complexity techniques for personalized thermal comfort. In *Proceedings of 2021 ASHRAE Winter Conference*. Accepted.