Model Predictive Control for Energy Efficient Buildings

by

Yudong Ma

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

in

Control and Design

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Francesco Borrelli, Chair
Professor J. Karl Hedrick
Professor Andrew Packard
Professor Alper Atamtürk

Fall 2012

The dissertation of Yudong Ma, titled Model Predictive Control for Energy Efficient Buildings, is approved:

Chair _____ Date _____

_____ Date _____

_____ Date _____

_____ Date _____

University of California, Berkeley

Model Predictive Control for Energy Efficient Buildings

Abstract

Model Predictive Control for Energy Efficient Buildings

by

Yudong Ma

Doctor of Philosophy in Control and Design

University of California, Berkeley

Professor Francesco Borrelli, Chair

The building sector consumes about 40% of energy used in the United States and is responsible for nearly 40% of greenhouse gas emissions. Energy reduction in this sector by means of cost-effective and scalable approaches will have an enormous economic, social, and environmental impact. Achieving substantial energy reduction in buildings may require to rethink the entire processes of design, construction, and operation of buildings. This thesis focuses on advanced control system design for energy efficient commercial buildings.

Commercial buildings are plants that process air in order to provide comfort for their occupants. The components used are similar to those employed in the process industry: chillers, boilers, heat exchangers, pumps, and fans. The control design complexity resides in adapting to time-varying user loads as well as occupant requirements, and quickly responding to weather changes. Today this is easily achievable by over sizing the building components and using simple control strategies.

Building controls design becomes challenging when predictions of weather, occupancy, renewable energy availability, and energy price are used for feedback control. Green buildings are expected to maintain occupants comfort while minimizing energy consumption, being robust to intermittency in the renewable energy generation and responsive to signals from the smart grid. Achieving all these features in a systematic and cost-effective way is challenging. The challenge is even greater when conventional systems are replaced by innovative heating and cooling systems that use active storage of thermal energy with critical operational constraints.

Model predictive control (MPC) is the only control methodology that can systematically take into account future predictions during the control design stage while satisfying the system operating constraints. This thesis focuses on the design and implementation of MPC for building cooling and heating systems. The objective is to develop a control methodology that can 1) reduce building energy consumption while maintaining indoor thermal comfort by using predictive knowledge of occupancy loads and weather information, (2) easily and systematically take into account the presence of storage devices, demand response signals from the grid, and occupants feedback, (3) be implemented on existing inexpensive and

distributed building control platform in real-time, and (4) handle model uncertainties and prediction errors both at the design and implementation stage.

The thesis is organized into six chapters. Chapter 1 motivates our research and reviews existing control approaches for building cooling and heating systems.

Chapter 2 presents our approach to developing low-complexity control oriented models learned from historical data. Details on models for building components and spaces thermal response are provided. The thesis focuses on the dynamics of both the energy conversion and storage as well as energy distribution by means of heating ventilation and air conditioning (HVAC) systems.

In Chapter 3, deterministic model predictive control problems are formulated for the energy conversion systems and energy distribution systems to minimize the energy consumption while maintaining comfort requirement and operational constraints. Experimental and simulative results demonstrate the effectiveness of the MPC scheme, and reveal significant energy reduction without compromising indoor comfort requirement.

As the size and complexity of buildings grow, the MPC problem quickly becomes computationally intractable to be solved in a centralized fashion. This limitation is addressed in Chapter 4. We propose a distributed algorithm to decompose the MPC problem into a set of small problems using dual decomposition and fast gradient projection. Simulation results show good performance and computational tractability of the resulting scheme.

The MPC formulation in Chapter 3 and 4 assumes prefect knowledge of system model, load disturbance, and weather. However, the predictions in practice are different from actual realizations. In order to take into account the prediction uncertainties at control design stage, stochastic MPC (SMPC) is introduced in Chapter 5 to minimize expected costs and satisfy constraints with a given probability. In particular, the proposed novel SMPC method applies feedback linearization to handle system nonlinearity, propagates the state statistics of linear systems subject to finite-support (non Gaussian) disturbances, and solves the resulting optimization problem by using large-scale nonlinear optimization solvers.

I dedicate this dissertation to my family and friends, especially

– to my parents and grandparents wholeheartedly for their endless love, support, and encouragement,

– to Yaoqiong Du for all the joy and laughter she brought into my life,

– to my best friends including Andi Setiawan and Jason Reid for their enormous help to my driving skill and spoken English,

– to all friends, "sisters", colleagues, "brothers", and researchers - who knowingly and unknowingly- that have been a great source of motivation and inspiration.

Contents

List of Figures

List of Tables

Acknowledgments

I express my deepest appreciation to Prof. Francesco Borrelli, chair of my PhD thesis committee. Prof. Borrelli not only guides me through rigorous research, but also leads me in a positive attitude in my daily life. His passion, patience, diligence, and wisdom always influence me from the past to the future. He is truly my professor, mentor, and friend.

I express my sincerely gratitude to Brandon Hencey, Scott Bortoff, Brian Coffey, Philip Haves, Sorin Bengea, and John D. Elliott for their tremendous help and contributions during the experimental tests at UC Merced. I also want to thanks Miroslav Baric and Jadranko Matuško who supported me definitely to go through the roughest time of my graduate study.

I would like to thank the Department of Mechanical Engineering at UC Berkeley, especially those members of my doctoral committee for their inputs, valuable discussions, and accessibility. In particular, I would like to thank Professor Andrew Packard and Professor J. Karl Hedrick who constantly encourage me over years.

I thank many friends that weave into my everyday life and offer my family enormous support. I would also like to thank the teachers who took part in this study for generously sharing their time and ideas. I have learned much through our conversations. Finally, my deepest love and appreciation goes to my parents.

Chapter 1

Introduction

1.1 Backgrounds and Motivations

According to the data book compiled by D&R international for the U.S. Department of Energy [49], commercial buildings represent almost one-fifth of the U.S. energy consumption. The top three end users in the commercial building sector are space heating, lighting, and space cooling, which represent close to half of commercial site energy consumption. Between 1980 and 2009, commercial floor space and primary energy consumption increased by 58% and 69%, respectively. In aggregate, commercial buildings consumed 17.9 quads of primary energy in 2009, representing 46.0% of building energy consumption and 18.9% of U.S. energy consumption. Energy reduction in this sector by means of cost-effective and scalable approaches will have an enormous economic, social, and environmental impact.

Commercial buildings are plants that process air and water in order to provide comfort for their occupants. The components used are similar to those employed in the process industry: chillers, boilers, heat exchangers, pumps, and fans. The set points for each individual component have to be controlled so that the system delivers demanded cooling and heating energy to building spaces. Over the last 25 years, the flexibility and sophistication of buildings control have greatly increased due to rapidly declining costs of embedded processors and deployment of wireless sensors. The study in [117] shows that more sophisticated controls approaches that consider a wider range of variables and automated control functions have significant national energy saving potentials. In particular, in [117] a selection of advanced control logic is analyzed, and their market-achievable energy saving potentials are reported. One of the conclusions in [117] is that the demand controlled ventilation could potentially achieve national energy saving of 0.3 quad. Note that a quad is a unit of energy equal to 1015 BTU or 1.055×10^{18} Joules. This potential energy saving motivates the development and implementation of advanced control logic for building energy systems.

Most modern buildings employ some level of automated control. In certain cases the control logic may be complex and optimized, but in the majority of cases building systems are controlled by basic control logic that errs on the side of simplicity over subtlety. Distinct

and interconnected proportional-integral-derivative (PID) loops and switching rules are the industry standard. This logic responds to setpoints and schedules for building components such as chillers and cooling towers.

Advanced decision systems are available on the market to optimize the system operation based on component modeling, feedback, and forecasts. A variety of proprietary control sequences for chillers, boilers, and cooling towers are available in the building industry. However, to the best of the authors' knowledge, their implementation is not widespread and often limited to specific configurations and components of the cooling and heating systems.

In addition, advanced controls design becomes challenging when predictions of weather, occupancy, renewable energy availability, and energy price are used for feedback control. High-performance green buildings are expected to maintain comfort and satisfaction of their occupants while optimizing energy, being robust to intermittencies in the renewable energy generation and responsive to signals from the smart grid. Achieving all these in often under-monitored environments is challenging. The challenge is even greater when conventional systems are replaced by innovative heating and cooling systems that use active storage of thermal energy with critical operational constraints [65, 88].

This thesis focuses on model predictive control of building cooling and heating systems. Model predictive control (MPC) is the only control methodology that can systematically take into account future predictions during the controller design stage while satisfying the system operating constraints. In buildings, model predictive control has two mechanisms to improve performance and reduce energy consumption of heating and cooling systems. The first mechanism is referred to as *load shifting* or *active storage*. Load shifting consists of shaping the energy profile delivered to a building, and exploiting the possibility of storing energy for later use. Thermal storage is inherent to a building structure and can be increased by including additional external energy storage devices. The optimal profile of the delivered energy depends on various factors such as time varying utility prices, availability of renewable energy, ambient temperature variation, and load shedding signals received from the utility grid. The second mechanism is component optimization. Buildings can be large systems with many control variables and degrees of freedom. Predictive models of building thermal dynamics and energy costs of control actuators allow computation of the optimal inputs to each actuator in order to deliver the desired energy profile.

This thesis focuses on the development and implementation of MPC for building heating and cooling systems. In particular, the main contributions of this thesis include,

1. The design of deterministic model predictive control for building heating and cooling energy systems with thermal storage. First, simple yet descriptive models are developed and validated for each building components. The MPC formulations are then presented for building heating and cooling systems, and extensive simulation and experiment results are presented to show the advantages of MPC over baseline controllers.

2. The design of a distributed MPC scheme. The centralized MPC problem is decomposed into small subproblems. The optimal control law is computed by iteratively

solving small problems in local units and communicating information between units. The proposed method makes use of sequential quadratic programming (SQP) [104, 62], proximal minimization [9], and dual decomposition [83] to handle the system nonlinearities and the decentralization, respectively. The key advance of this approach is the ability to solve MPC problem for large scale building heating and cooling systems in a distributed fashion. The resulting scheme is suitable to be implemented on a set of distributed low-cost processors.

3. The design of a stochastic MPC scheme for building energy systems so that the expected electricity energy is minimized while guaranteeing a desired (low) probability of comfort constraint violations. The proposed stochastic MPC takes into account the prediction uncertainties and model errors at design stage, and the uncertainties are modeled as random variables with finitely supported probability density distributions. The proposed approach handles system nonlinearities by feedback linearization, computes tightening offsets for chance constraints, and solves the resulting non-convex optimization problem by off-shelf optimization software.

In the following section, existing control logic for building heating and cooling systems in literature is reviewed.

1.2 Literature Review

Over the last decades, the costs of embedded processors and wireless sensors have been rapidly declining while their performance has been significantly improved. This trend enables the cost-effective introduction of more sophisticated building automation systems and building energy management control systems in order to facilitate the operations, maintenance, and monitoring of building systems. The flexibility of the building automation platforms and the complexity of building systems have made control a very active area of research and development. The ultimate goal is to reduce overall energy consumptions, ensure comfort of occupants, and satisfy indoor air quality requirement. A variety of control logic for building cooling and heating systems has been proposed and reported in literature. The authors in [133] and the ASHRAE handbook [1] offer a thorough review on existing control methodologies for building energy systems. Classical control has been widely adopted in building energy systems due to its simplicity to design and its low computational complexity to determine the control signals [1]. For example, ON/OFF or bang-bang controllers are ubiquitously used in old building systems without digital control, and proportional-integral-derivative (PID) control loops are extensively implemented in current building heating and cooling systems equipped with digital control panels and variable frequency drivers [1]. However, for large-scale coupled buildings, tuning nested PID loops controlling systems with multiple inputs and multiple outputs is very challenging and time consuming. Often the PID loops are coordinated by rule-based control. Rule-based control defines a sequence of if-then-else commands for actuators based on experience of building operators and managers [1].

Rule-based control design is widely used in building heating and cooling systems including chiller sequencing control, cooling tower sequencing control, pump sequencing control, and fan sequencing control.

This thesis focuses on model predictive control techniques. The main idea of model-based control is to use the model of the plant to predict the future evolution of the system [98]. At each sampling time, starting at the current state, an open-loop optimal control problem is solved over a finite horizon. The optimal command signal is applied to the process only during the following sampling interval. At the next time step a new optimal control problem based on new measurements of the state is solved over a shifted horizon.

For complex constrained multivariable control problems, MPC has become the accepted standard in the process industries [14]: its success is largely due to its unique ability to simply and effectively handle hard constraints on control and states.

MPC eliminates the drawbacks of traditional control techniques including: 1) excessive parameter tuning, 2) the lack of prediction capability, 3) the incapability of optimizing the energy consumption at a system level, and (4) the incapability to maintaining satisfactory performance at all operating conditions.

MPC has two challenges. 1) The development cycle for MPC is longer than traditional control methods as MPC requires analytical building models at the design stage, and 2) MPC requires more powerful computation units to solve an optimization problem in real-time. For these reasons, the acceptance of MPC in building community has been limited to research purpose. This thesis will show that there can be substantially benefits associated with MPC, and that both challenges can be addressed by automated modeling and proper problem formulation to enable real-time implementation on low cost computation platform.

Model Predictive Control

Model-based predictive control (MPC) is the only control methodology that can handle large-scale multi-input multi-output (MIMO) dynamically coupled systems subject to state and input constraints, with performance guarantees. J. Braun [19] firstly proposed an optimal control strategy to reduce energy costs and peak electrical demand by making use of building thermal energy storage, and this work inspired a number of model predictive control methodologies for building heating and cooling control.

The basic idea of MPC is to solve a finite time optimal control problem at each time step t,

$$\min_{u} \sum_{k=0}^{k=N-1} \text{Energy}(x_{t+k|t}, u_{t+k|t}, d_{t+k|t}), \tag{1.1a}$$

$$\text{subj. to} \quad x_{t+k+1|t} = f(x_{t+k|t}, u_{t+k|t}, d_{t+k|t}), \forall k = 0, 1, \ldots, N-1, \tag{1.1b}$$

$$y_{t+k|t} = g(x_{t+k|t}, u_{t+k|t}, d_{t+k|t}), \forall k = 0, 1, \ldots, N, \tag{1.1c}$$

$$u_{t+k|t} \in \mathcal{U}, \forall k = 0, 1, \ldots, N-1, \tag{1.1d}$$

$$y_{t+k|t} \in \mathcal{Y}, \forall k = 0, 1, \ldots, N, \tag{1.1e}$$

where the symbol $v_{t+k|t}$ is read as "the variable v at time $t+k$ predicted at time t", N is the prediction horizon, y is the output of the system, and x is the system states of building cooling and heating systems including zone temperatures, humidity level, CO_2 concentrations, and other factors affecting the occupants perceptions of comfort. The vector u collects the control inputs of building systems regulating the heating and cooling plants, and the vector d collects exogenous signals (called "disturbances" in this thesis) such as ambient temperatures and thermal load induced by occupancy, solar radiations, and electrical devices. The function $f(x, u, d)$ in (1.1b) allows to predict the future building states based on the current states, control inputs, and disturbances. The objective of the model-based optimal control problem (1.1) is to minimize the total energy consumptions (1.1a) while satisfying the operational constraints (1.1d) and comfort constraints (1.1e).

All MPC approaches appeared in the literature to building control solve a version of problem (1.1). Their difference resides in the type of

1) comfort constraints or state constraints (1.1e),

2) dynamic models to predict the building states (1.1b), energy consumptions (1.1a), and the disturbances $d_{t+k|t},\ \forall k = 0, 1, \ldots, N-1$,

3) the implementation of the optimal control inputs.

In the following sections, the MPC for building energy systems are reviewed and classified based on the above three ingredients in MPC design. Figure 1.1 reports a graphics of the classification scheme.

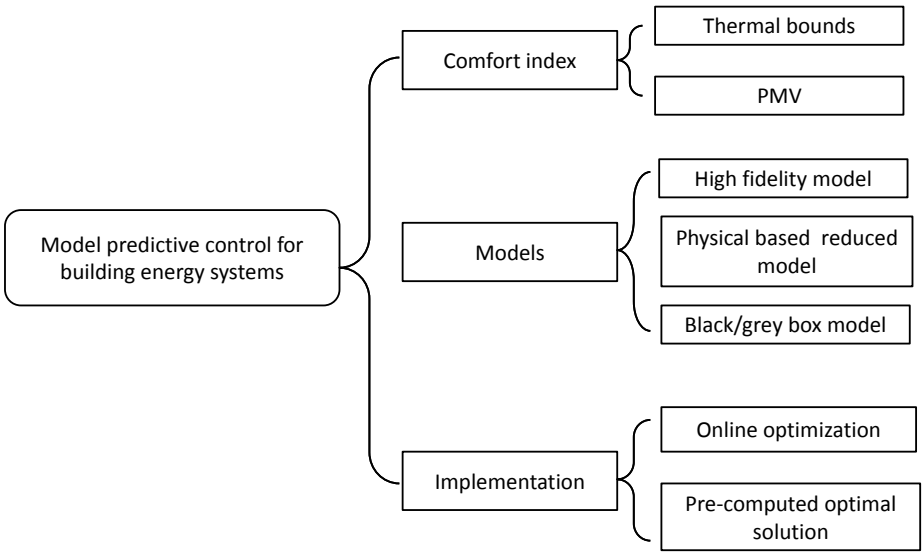


Figure 1.1: Classification schematic for building MPC logic.

Table 1.1: Relationship between PMV and thermal sensation.

PMV	+3	+2	+1	0	-1	-2	-3
Thermal sensation	Hot	Warm	Slightly warm	Neutral	Slightly cool	Cool	Cold

Comfort Constraints

Thermal comfort in buildings is generally difficult to define and measure since thermal satisfaction depends on many parameters related to sensible, latent, conductive, and radiative heat transfer processes governing the thermodynamic of human body. Extensive studies over the last decades have tried to establish a number of thermal comfort indices for indoor climate analysis and HVAC control system design [129]. Next we discuss two types of simplified comfort constraints: the predicted mean vote (PMV) and a combination of linear constraints on building measured variables.

The concept of PMV was introduced by Fanger [53] in 1970. PMV is not only based on temperature and relative humidity, but also depends on mean radiant temperature, air velocity, and individual factors such as metabolism rate and thermal resistance of clothing. PMV is a real scaler indicating the occupants' thermal comfort sensation.

PMV is computed as a function of four environmental variables [53] including temperature T, relative humidity h_r, mean radiant temperature T_{mr}, air velocity V and two individual parameters (metabolic rate M_r and clothing index I_{cl}),

$$\text{PMV} = G(T, h_r, T_{mr}, V, M_r, I_{cl}). \tag{1.2}$$

Equation (1.2) is based on the assumption that for long exposures to a constant thermal environment with a constant metabolic rate, a heat balance can be established for the human body and the bodily heat production is equal to its heat dissipation [53]. Table 1.1 shows the correspondence between PMV and thermal sensation. In 1994, this formula (1.2) was included in ISO Standard 7730 and a PMV-based criterion has been established between -0.5 and +0.5 as acceptable for thermal comfort in air conditioned environments.

The concept of PMV has been adopted in model-based optimal control design [55, 48, 35, 38] to guarantee occupants thermal sensation in buildings with minimum energy consumption. In particular, the work in [55, 48] focused on the study of indoor thermal comfort control problem in building HVAC systems. The HVAC heating and cooling systems are simplified as an auto-regressive with exogenous input model. A model predictive control problem of the form (1.1) is proposed to minimize energy consumption while maintaining the indoor thermal comfort criterion (PMV) at an adequate level. Simulation results are presented to validate the energy savings and thermal comfort. Also the model-based optimal control with PMV index is tested in a solar energy research center in [35]. Optimal control problem based on PVM is non-convex due to the PMV equation (1.2). The authors in [38] attempted to address this issue by introducing a convex approximation when computing the PMV (1.2), and simulation results are presented to demonstrate the effectiveness and computational efficiency of the proposed method.

In order to reduce the complexity of the control problem and enable real time implementation, simpler yet descriptive comfort indices can be formulated as a set of linear constraints on the zone temperatures, CO_2 concentrations, and relative humidity. One example is the thermal comfort index named "effective temperature", which is a function of indoor temperature and relative humidity [129]. Thermal comfort specified by ASHRAE can be identified by quadrilateral regions within a psychrometric chart [55], and also falls into the category of a fixed constrained region. Another example of simplified comfort model is defined as upper and lower bounds on perceivable building temperatures, and this comfort index is used by authors in [90, 95, 76, 19, 102, 105].

Model

Model-based control relies on building models to predict the building states, energy models to predict the electricity energy consumptions, and load models to predict the future load induced by occupancy, solar radiation as well as electrical devices. The choice of models affects the performance and computational complexity of the resulting model predictive controller. The models reported in literature can be categorized into three groups: 1) high fidelity models that involve the numerical solution of differential algebraic equations (DAE) describing thermal balance of buildings, 2) simplified physical based models based on RC analogy, and 3) grey/black box models that do not rely on physical building models.

High fidelity models are usually developed for simulation purpose. Existing building energy simulation programs are EnergyPlus [44], TRNSYS [77], ESP-r [40], DOE-2 [139], IDA [118], and BuildOpt [136, 138]. A number of studies have focused on designing optimal controllers of the type (1.1) using high fidelity models as prediction models in (1.1b).

Henze et al. in [68] demonstrated model predictive control of active and passive building thermal storage inventory in a test facility in real time using time-varying electricity prices without demand charges. In their study, the building was modeled by TRNSYS while Matlab and its optimization toolbox were interfaced with the building simulation program.

Brian Coffey et al. in [41] proposed a software framework for model predictive control based on an optimization package called GenOpt [137]. His approach allows the use of any text-based simulation tools and simplifies the setup process for simulation-based studies or real-time implementations. His work also introduces the use of optimization starting points and dynamic search-space constraints based on the results of previous time-steps and on heuristic rules.

Despite the aforementioned success, optimal control with high fidelity prediction models has the following drawbacks: 1) the identification and validation of high fidelity models are nontrivial, and require excessive parameters tuning and simulation, 2) the complexity and size of the prediction models quickly lead to a computationally intractable optimization problem, and 3) high fidelity simulation models prevent the optimization solvers from exploring the sparse structure of the resulting optimization problem. These issues are the major obstacles for their online implementations.

The issues related to high fidelity prediction models can be mitigated by using simplified physical-based models or lumped parameter physical models. In particular, in this thesis we present low order RC based bilinear building models based on thermal energy conservation law. The bilinear terms arise from the multiplication of air flow rate and supply air temperature to compute the thermal energy delivered to zones. This simplified building model approach has appeared in [43], and then extensively adopted in model-based optimal control for building temperature regulations [19, 90, 76, 105, 5]. RC based load models are proposed in [92, 64] to estimate the thermal load of buildings. In particular, buildings are modeled as interconnections of thermal resistances and thermal capacitances representing walls, windows, ceilings, and furniture. The building thermal load is a function of average room temperature, room temperature set points, cloud coverage, ambient temperature, and time. The model parameters are estimated by fitting historical measurement. In [107], a reduced order RC-based model is developed to predict room temperatures, and the building thermal load is estimated using extended Kalman filter that combines models with temperature sensors. The author in [4] proposed to model room thermal dynamics using a first order linear model subject to additive uncertain loads, and the loads are estimated by semi-parametric regression using only temperature measurements. In the MPC implementation reported in [4], the load prediction is assumed constant over the prediction horizon.

Lastly, a few approaches in the literature use gey/black box models to represent the building dynamics. These models do not require full knowledge of the system or process. They are developed by fitting historical behaviors of the system. The parameters in these models do not reflect physical significance. Typical examples of black/grey box models are polynomial curve fits, ARMAX model, and artificial neural networks (ANNs). Black/grey box models are simple to develop and implement since they require only data. However, black/grey box models cannot ensure reliable prediction for operating points outside the range covered by training data, and thus extensive and adequate data training is needed in order to guarantee prediction accuracy. Black/grey box building dynamics models have been used in MPC design in [55, 73, 131, 141]. Black box building load models have been reported in [90, 93], and the building loads are modeled as a bounded time-varying uncertainty. The disturbance load envelopes can be learned from historical data, shared calendars, and weather predictions.

Implementations

The optimal control actions obtained by solving Problem (1.1) can be implemented in different ways. The implementation approach depends on the problem complexity and available computational power as well as hardware specifications. The two major implementation methods include, 1) online optimization that computes optimal control signal in real-time, and 2) off-line method that pre-computes optimal solutions and looks up the optimal control signal on line. Online model predictive control requires to solve the optimization problem (1.1) at each control sampling time. The optimal command signal is applied to the system only during the current sampling interval. At the next time step a new optimal

control problem based on new measurements of the state is solved over a shifted horizon. Many of the optimal control methods for building energy system reported in literature fall within this online paradigm [19, 90, 76, 105, 5]. However, as the building size increase, the optimization solver might not be able to compute a satisfying solutions within the control sampling time.

In order to reduce the computation time for the optimal control actions, and enable real time implementation, off-line methods attempt to solve an optimal solution set to Problem (1.1) parameterized over initial states off-line. Instead of solving an optimization problem at each time step as online methods, off-line methods merely require to measure the current states and find the corresponding optimal control action within the solution set pre-computed off-line [17, 3]. However, the complexity of this method requires excessive memory to store the optimal solution set, and suffers from the curse of dimension. Several approximation methods are proposed to reduce the memory usage of the off-line methods by simplifying the representation of the optimal solution set. In particular, the authors in [46] make use of standard machine learning algorithms to approximate the optimal solution set from randomly chosen point-wise sample values. The work in [47] deals with learning decision rules from simulation data to mimic the binary decisions of a hybrid MPC building controller as closely as possible while maintaining high performance. A well-known concept AdaBoost from machine learning is employed to combine a number of weak learners into a strong classifier. The resulting controller maps directly to a majority voting system using simple if-then-else rules, a structure that building operators are familiar with. In addition, a complementary statistical technique has been introduced that allows for the extraction of logistic decision models from the optimal control results for windows operation in buildings [97]. The process works best when some time-lagged information is present as a predictor variable to ensure that some process memory is preserved. A generalized linear model (GLM) in the form of a multi-logistic regression was able to mimic the general characteristics of the optimizer results, achieving energy savings, but at a small fraction of the computational expense.

1.3 Thesis Outline

The remainder of the thesis is structured as follow.

- Chapter 2 presents the models used in the control design for building heating and cooling systems. A typical building heating and cooling system has energy conversion and energy distribution components for both water-loop and air-loop systems. At the water-loop system, the energy conversion systems include a set of cooling towers and chillers to produce the chilled water, and a heating pump or boiler to produce hot water. The energy distribution system includes water pumps and an energy storage device. At the air-loop system, subsystems known as air handling units (AHUs) transfer energy from water into localized air flows. These air flows are transported by a supply fan to buildings' spaces, delivering cooling and heating energy where required.

For the water-loop system, a simplified model based on mass and energy conservation law is developed and validated for an energy storage tank, and the thermal load demanded by the buildings are predicted using a solar and internal load predictor and a building thermal load predictor. The solar and internal load predictor computes the thermal load induced by solar radiation, occupancy, and electrical devices in the buildings as a function of weather information, time, and date. The building thermal load predictor computes the heating and cooling energy required by buildings to maintain thermal comfort using a simplified RC model. In addition, polynomial models are developed to predict the energy consumptions for chillers, cooling towers, and pumps.

For the air-loop system, two types of simplified models are presented for the dynamics of zone temperatures. The first one is a two-state bilinear model, where the first state describes the fast dynamic thermal mass (air within the room) and the second state is the slow dynamic thermal mass (furniture, floors, ceilings, and walls). The bilinear term results from the computation of cooling and heating energy delivered to building spaces. The second model is a bilinear ARMAX model. Both models are tested on historical data collected from Bancroft DOE library at UC Berkeley. The thermal room models are subject to uncertainty load disturbance, and the load is calculated as follows. The room model parameters are identified using data from unoccupied hours with minimum internal load, and the load during occupied hours are computed as the difference between the temperature measurement and the temperature predicted by models. The energy consumption models for each component (dampers, heating/cooing coils, and supply fan) are modeled as static polynomial functions of control actions and system states.

- Chapter 3 outlines the model-based predictive control design for both water-loop and air-loop systems. The MPC controllers for water-loop and air-loop systems are denoted as HMPC and LMPC, respectively. First the MPC concept is briefly outlined for a simplified one-state linear building energy model. Simulation results are presented to demonstrate its ability to save energy and reduce peak power consumption compared to baseline controller.

  A detailed case study for HMPC for the cooling system on the campus of UC Merced is reported. To the best of the author's knowledge, this is the first real-time implementation of model predictive control on large scale building energy systems. In particular, a nonlinear MPC is designed to optimize the set points and schedule of the central chilling plant and the charge level of the energy storage tank. The MPC problem is solved using sequential quadratic programming. In addition, move blocking strategy is applied to reduce the problem size, and a periodic robust invariant set is used as terminal constraints so that there is always enough chilled water in the tank to meet all possible realization of the uncertain demands. Experimental results suggested that MPC is able to reduce the energy consumptions and increase the thermal efficiency of the chillers and cooling towers.

Chapter 3 also presents the details for LMPC. Extensive simulations are carried out to show that MPC is able to maintain thermal comfort while consuming least energy consumptions by air-loop systems. Furthermore, time varying electricity price can be incorporated in MPC design with minimum modifications to reduce the electricity bills associated with HVAC operations.

- Chapter 4 proposes a distributed model predictive control for building temperature regulations. The size of the centralized predictive control problem rapidly grows when a realistic number of rooms together with a meaningful control horizon are considered. Therefore the real-time implementation of a MPC scheme is a challenge for the low-cost embedded platforms currently used for HVAC control algorithms. The techniques presented in Chapter 4 enable the implementation of a MPC algorithm by distributing the computational load on a set of VAV box embedded controllers coordinated by the embedded controllers on the AHU system. This is achieved by adopting sequential quadratic programming (SQP), proximal minimization, and dual decomposition to handle the system nonlinearities and the decentralization, respectively. The effectiveness and computational efficiency of the proposed distributed algorithm is demonstrated by simulations results.

- Chapter 5 proposes a stochastic MPC for building HVAC system so that the expected energy consumption is minimized while guaranteeing low probability of comfort constraint violations. In this framework, the uncertainties of model disturbances are learned from historical data, and are modeled as finitely supported probability distribution functions. The uncertainty descriptions are explicitly taken into account in the design stage by propagating the statistics of system states along the horizon, and bounding the chances of constraints violations. Compared to existing stochastic MPC approaches for building energy system [105], the proposed method is able to directly handle bilinear systems subject to non Gaussian distributions. In particular, feedback linearization is used to linearize the bilinear system, and the probability density function of the uncertainties are propagated using discrete method and sample based method. Simulative and experimental studies are carried out to demonstrate the effectiveness of the method.

Chapter 2

System Description

2.1 Introduction

In this chapter, we describe a building cooling and heating system with energy storage elements. A typical building heating and cooling system has energy conversion and energy distribution systems for both water-loop and air-loop systems. In this thesis, we focus on the building cooling system. The results in following chapters can be adapted to heating systems with minor modifications.

The water-loop system, in general, consists of a series of chillers/cooling towers for energy conversion, an electrical pump as well as thermal storage elements for energy distribution. The chillers and cooling towers are responsible for the chilled water generation. The electrical pump is operated to maintain enough static duct pressure so that a desired mass flow rate of chilled water is supplied to the down-streaming systems. The thermal storage element is installed to enable load shifting. The concept of load shifting allows the chillers and cooling towers to operate when the energy price is lower and when the ambient conditions allow to operate the chilling system more efficiently. The chilled water in the energy storage element is used to satisfy the buildings thermal demands.

The air-loop system includes air handling units and variable air volume boxes. An air handling unit transfers energy from distributed chilled water into localized air flows, and these air flows are transported to buildings space, delivering cooling and heating energy where required.

Simulative models for both water-loop and air-loop systems have been developed and widely used in building industry for design and control purpose [138]. However, these models are too complex to be included in the design of model-based control. Therefore it is necessary to develop simplified yet descriptive models for both water-loop and air-loop systems. In the following sections, control-oriented models for buildings are developed and validated. In particular, for each level of the system, we developed a set of difference (differential) equations describing the system dynamics, nonlinear static equations computing the energy consumption for each building component, and empirical equations estimating the future

load.

For the water-loop system, a simplified model is developed and validated for a class of energy storage tank. The model is based on mass and energy conservation laws, and the building thermal load is predicted by solar and internal load predictor and the building thermal load predictor. The solar and internal load predictor computes the thermal load induced by solar radiation, occupancy, and electrical devices in the buildings as a function of weather information, time, and date. The building thermal load predictor computes the heating and cooling energy required by buildings to maintain occupants' thermal comfort using a simplified RC model. In addition, polynomial curve fitting models are developed to predict the energy consumptions for chillers, cooling towers, and pumps.

For the air-loop system, two simplified models are presented for thermal zone dynamics. The first model is a two-state bilinear system, where the first state describes the fast responding thermal mass (air within the room) and the second state is the slow responding thermal mass (furniture, floors, ceilings, and walls). The bilinear term results from the computation of cooling and heating energy delivered to zones. The second model is a bilinear ARMAX model. Both models are tested on historical data collected from Bancroft library at UC Berkeley. The thermal zone models are subject to uncertain load, and the load is calculated as the difference between the temperature measurement and the temperature predicted by the zone models. The energy consumption for each component (dampers, heating/cooing coils, and supply fan) is modeled as a static polynomial function of control signals and system states.

2.2 Cooling Control Systems

This section describes a building HVAC system by dividing it into two loops, namely water-loop and air-loop systems. At the water-loop system, a centralized chilled water generation system produces the required cooling energy. It is assumed that an energy storage device is available. This thesis focuses on thermal energy stored in a stratified water tank. At the air-loop system, subsystems known as air handling units (AHU) transfer energy from distributed chilled water into localized air flows. These air flows are transported to building spaces, delivering cooling energy where required. The next sections describe the water-loop chilled water generation, the storage element, and the air-loop system.

Water-Loop: Components Outside a Building

Figure 2.1 depicts the main components of a cooling system based on chilled water conversion, storage, and distribution. The system can serve either a single building or multiple buildings. Chillers and cooling towers are responsible for generating chilled water. The chillers remove heat from the chilled water loop by means of a vapor-compression or absorption refrigeration cycle. Cooling towers reject heat from the chiller cycle to the ambient environment by evaporation and forced convection using electrical fans.

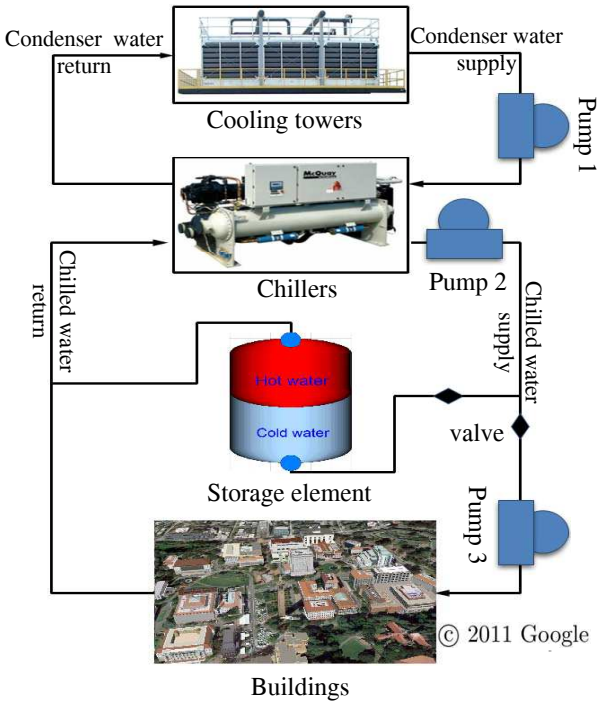


Figure 2.1: Schematic of a cooling system.

The chilled water storage element in Figure 2.1 can shift the peak cooling thermal energy load so that the chillers and cooling towers can run only when it is most efficient to do so. The logic behind the load shifting depends on various factors including time varying utility prices, availability of renewable energy, lower ambient temperature, and load shedding signals received from the utility grid. The chilled water is distributed through insulated piping using hydraulic pumps. The valve in Figure 2.1 controls the chilled water flow to either fill the storage element, or serve the buildings with a desired mass flow rate of chilled water.

Several thermal storage elements are available in the building industry. This thesis focuses on a stratified chilled water tank that can store chilled water, such as the one installed on the campus of University of California at Merced [94, 64]. Other storage devices are available using ice balls or concrete slabs to store thermal energy.

Air-Loop: Components Inside a Building

The main components used to produce and distribute cool air in a building are depicted in Figure 2.2. They are air handling units (AHUs) and variable air volume (VAV) boxes. The AHU recirculates return air from building spaces, and mixes it with fresh outside air. The ratio of return air flow to outside air flow is controlled by dampers located inside the AHU. The mixed air is cooled by a cooling coil that transfers cooling energy from the chilled water that is generated or stored by the water-loop system.

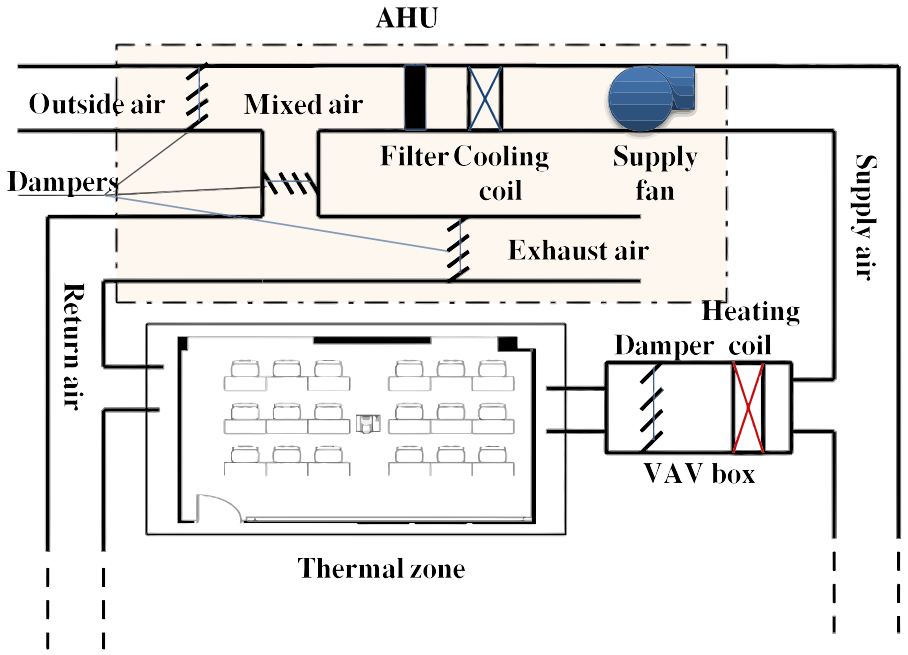


Figure 2.2: Schematic of the air distribution system.

The air temperature downstream of the cooling coil depends on the mass flow rate of chilled water through the cooling coil, the temperature of the chilled water, the temperature of mixed air entering the cooling coil, the mass flow rate of the mixed air, and the physical characteristics as well as thermal effectiveness of the cooling coil. Cool air is delivered to the building spaces by electric fans. Before reaching the building spaces, the air goes though variable air volume (VAV) boxes. At each VAV box the air flow rate supplied to the space is regulated by a damper. In addition, air temperature can be increased using a reheat coil installed in the VAV box when needed. A space served by one VAV box is referred to as a *thermal zone*. The delivered air enters a thermal zone through diffusers that are designed to fully mix the incoming air with the air in the thermal zone.

2.3 Hierarchical Levels and System Modeling

We follow the structure of the previous section and present reduced order models for both the water-loop and air-loop of building cooling and heating systems. For each loop we describe the associated dynamics, components that use energy, and thermal energy load. When a multitude of buildings being controlled, the relevant model dynamics at the water-loop are the dynamics of thermal energy storage devices. The buildings' cooling or heating thermal energy loads are modeled as a lumped heat flux, and the main components consuming energy are chillers, cooling towers, and pumps. At the air-loop the relevant model dynamics are

the dynamics of thermal zones. The thermal energy load combines the loads generated by occupants, solar radiation, and electrical devices. The components consuming energy are a supply fan, a cooling coil, and heating coils.

Water-Loop: Modeling a Building from the Outside

Storage Dynamics

A model of a water tank used for actively storing chilled water is presented and validated with data collected from UC Merced campus. The water in the tank is subject to negligible mixing, thus the tank can be modeled as a stratified system with layers of warmer water at the top and cooler water at the bottom. Figure 2.3 depicts the temperature of water measured inside the tank at different heights at 8:30 AM on November 29, 2007. A thin layer of water, known as a thermocline, that has a steep temperature gradient over the height of the tank can be observed. Warmer water above the thermocline and cooler water below the thermocline are lumped up to obtain a four state system describing the height and internal energy of the warmer and cooler water, respectively.

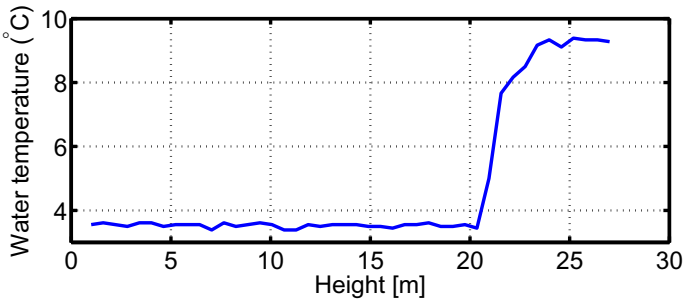


Figure 2.3: Temperature of water layers in the tank.

The tank is assumed to be part of a closed hydraulic loop, thus the mass flow rate entering the tank is equal to the mass flow rate exiting the tank. Subsequently, the total height of water in the tank z_{tank} is the sum of the height of warm water z_a and the height of cool water z_b in the tank. The tank dynamics are governed by mass and internal energy conservation laws,

$$\dot{z}_b = \frac{\dot{m}_{CHWS} - \dot{m}_{cmp}}{\rho \pi r_{tank}^2}, \quad \dot{z}_a + \dot{z}_b = 0, \tag{2.1a}$$

$$\dot{U}_a = \dot{H}_a + \dot{Q}_{b>a} + \dot{Q}_{oa>a}, \quad \dot{U}_b = \dot{H}_b + \dot{Q}_{a>b} + \dot{Q}_{oa>b}, \tag{2.1b}$$

where ρ is the density of the water, r_{tank} is the inner radius of the tank, $\dot{m}_{CHWS}$ is the mass flow rate of water supplied to the buildings, and $\dot{m}_{cmp}$ is the mass flow rate of water returning from the buildings. $U_* = \rho \pi r_{tank}^2 z_* c_p T_*$ is the internal energy of the water in the tank, where $* = a$ denotes variables for warmer water, and $* = b$ denotes variables for cooler

water. $\dot{Q}_{oa>a}$ is the heat transferred from ambient to the warmer water in the tank, $\dot{Q}_{oa>b}$ is the heat transferred from ambient to the cooler water in the tank,

$$\dot{Q}_{oa>a} = (T_{oa} - T_a)(2\pi r_{tank} z_a)k_1, \tag{2.2a}$$
$$\dot{Q}_{oa>b} = (T_{oa} - T_b)(2\pi r_{tank} z_b)k_1, \tag{2.2b}$$

where $\dot{Q}_{a>b}$ is the heat conducted from warmer water to cooler water in the tank and $\dot{Q}_{a>b}$ is the heat conducted from cooler water to warmer water in the tank,

$$\dot{Q}_{a>b} = (T_a - T_b)(\pi r_{tank}^2)k_2, \tag{2.3a}$$
$$\dot{Q}_{b>a} = -\dot{Q}_{a>b}, \tag{2.3b}$$

where k_1 and k_2 are heat transfer coefficients.

The thermal storage tank can operate in two modes. When the chilled water flow $\dot{m}_{CHWS}$ is greater than the water flow demanded by the buildings $\dot{m}_{cmp}$, the excess flow fills the tank. Hence the water flow enthalpy rates are calculated as

$$\dot{H}_a = -(\dot{m}_{CHWS} - \dot{m}_{cmp})c_p T_a, \tag{2.4a}$$
$$\dot{H}_b = (\dot{m}_{CHWS} - \dot{m}_{cmp})c_p T_{CHWS}, \tag{2.4b}$$

where $\dot{H}_a$ is the enthalpy rate for the warm water in the tank contributed by water flow, and similarly $\dot{H}_b$ for the cool water.

When the chilled water flow $\dot{m}_{CHWS}$ is less than the water flow demanded by the buildings, the water in the tank compensates for the insufficient chilled water supply. Hence the water flow enthalpy rates are calculated as,

$$\dot{H}_a = -(\dot{m}_{CHWS} - \dot{m}_{cmp})c_p T_{cmp,r}, \tag{2.5a}$$
$$\dot{H}_b = (\dot{m}_{CHWS} - \dot{m}_{cmp})c_p T_b, \tag{2.5b}$$

where $T_{cmp,r}$ is the temperature of return water from buildings.

The simplified model (2.1)–(2.5) is validated using measured data collected from May 22 to May 29, 2007. The measured inputs $[\dot{m}_{CHWS}, T_{CHWS}, \dot{m}_{cmp}, T_{cmp,r}]$ are applied to the tank model, and the output of the model $[z_a, z_b, T_a, T_b]$ is compared with the measurements. Figure 2.4 shows the tank water temperature validation results. The solid lines are the temperature measurements of top layer water T_a and bottom layer water T_b in the tank, and the dotted lines show the predicted cooler and warmer water temperature. The tank model captures the temperature dynamics of the top and bottom layers of the tank water as well as the dynamics of the cool water height (Figure 2.5). However, the second peak of the top water temperature during the day is not captured due to the formation of a second thermocline (note in Figure 2.4 the bumps above 15.2 °C every afternoon). A higher order model overcomes this limitation. We preferred to not increase the model complexity in order to avoid real-time implementation issues.

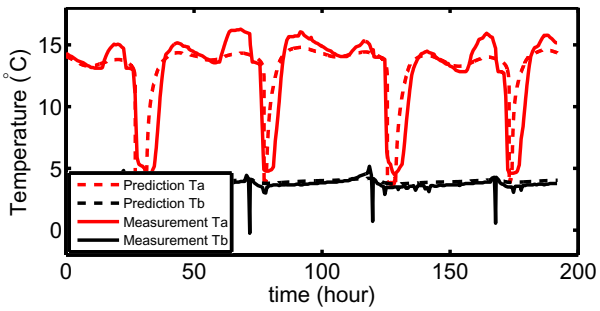


Figure 2.4: Water temperature validation.

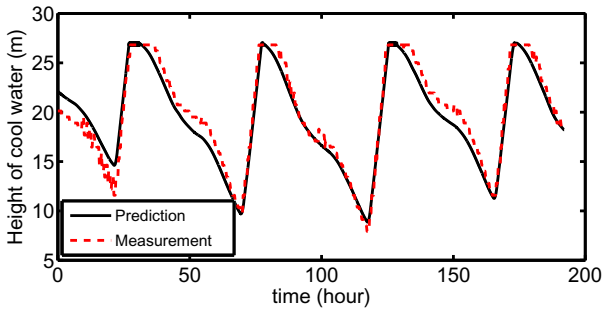


Figure 2.5: Water height validation.

Load Modeling

At the water-loop, each building can be modeled as a *load demand* element. A lumped load model predicts the total energy requested by a building based on date, time, occupancy, and weather. In the approach studied in[64, 92], the building load model has two subcomponents, namely the solar and internal load predictor and the building thermal load predictor. Both components are depicted in Figure 2.6.

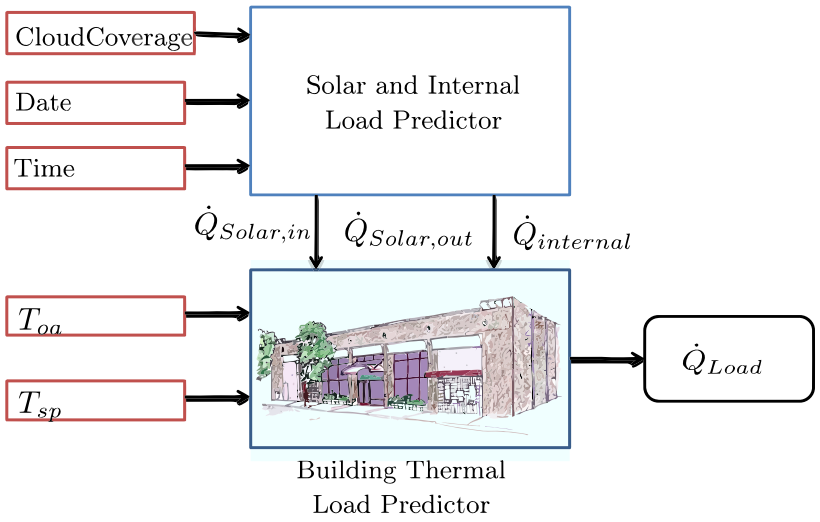


Figure 2.6: The building thermal load predictor.

The solar and internal load predictor uses time (t_{td} [sec], a number ranging from 0 to 86400), date (t_{dy}, a number between 1 to 365 (366); t_{dw}, a number between 1 to 7) and cloud coverage (β_{cloud}, a continuous value from 0 to 1, reflecting the impacts of cloud on the solar energy) as its inputs and calculates inside and outside solar loads and internal load. The outside solar load reflects the solar radiation on the outer surface of the building, while the inside solar load is the solar radiation into the building (e.g. sunshine through the windows into rooms). The internal load includes the heat from occupants, lights, and equipment. The

outside and inside solar loads are calculated as follows:

$$S_{Hour} = (t_{td} - 12) \times 15^\circ, \tag{2.6a}$$

$$S_{Dec} = -23.45^\circ \times \cos(360^\circ \frac{t_{dy} + 10}{365}), \tag{2.6b}$$

$$E_{Solar} = \max\left(0, \cos(S_{Hour})\cos(S_{Dec})\cos(S_{Lat}) + \sin(S_{Dec})\sin(S_{Lat})\right), \tag{2.6c}$$

$$\dot{Q}_{Solar,in} = \beta_{cloud} E_{Solar} \theta_{in}, \tag{2.6d}$$

$$\dot{Q}_{Solar,out} = \beta_{cloud} E_{Solar} \theta_{out}, \tag{2.6e}$$

where S_{Hour} and S_{Dec} are variables calculated based on time and date to indicate the solar projection angle, S_{Lat} is the latitude specifying the geographic coordinate of the building, E_{Solar} is the extraterrestrial horizontal radiation, θ_{in} (θ_{out}) is the solar load on the inner (outer) wall mass per unit of clear-sky extraterrestrial horizontal radiation and it depends on the building geometry. The internal load is assumed to be a piecewise constant signal with a period of one week, and is evaluated as:

$$\gamma = \begin{cases} 1 & \text{if } t_{start} \leq t_{td} \leq t_{end} \\ 0 & \text{Otherwise} \end{cases} \tag{2.7a}$$

$$\dot{Q}_{internal} = \begin{cases} \gamma \dot{Q}_{Saturday} \\ \text{if } t_{dw} = 6 \\ \gamma \dot{Q}_{Sunday} \\ \text{if } t_{dw} = 7 \\ \gamma \dot{Q}_{Weekday} \\ \text{Otherwise} \end{cases} \tag{2.7b}$$

where t_{start} and t_{end} indicate the time interval when the internal load is different from zero. The parameters $\dot{Q}_{Saturday}$, $\dot{Q}_{Sunday}$ and, $\dot{Q}_{Weekday}$ are constant internal load values during different days of the week.

The building thermal load predictor predicts the cooling load of buildings. The buildings are conventionally modeled by RC circuits [61, 87]. The building thermal load model is sketched in Figure 2.7, where R_1 represents the thermal resistance of windows. The walls are separated into two layers, where C_{in} and C_{out} capture the thermal capacitance of the wall when influenced by outside and inside solar load, respectively. The thermal resistance between C_{in} and C_{out} is modeled by R_3, while R_2 and R_4 capture the thermal resistance associated with heat convection. The interconnection of the thermal components is shown in Figure 2.7. The model inputs are outside air temperature T_{oa}, outside solar load $\dot{Q}_{Solar,out}$, inside solar load $\dot{Q}_{Solar,in}$, internal load $\dot{Q}_{internal}$, and the indoor temperature setpoint T_{sp}. The model internal states are the temperatures of the thermal masses T_{in} and T_{out}. The model output is the cooling load demand $\dot{Q}_{Load}$.

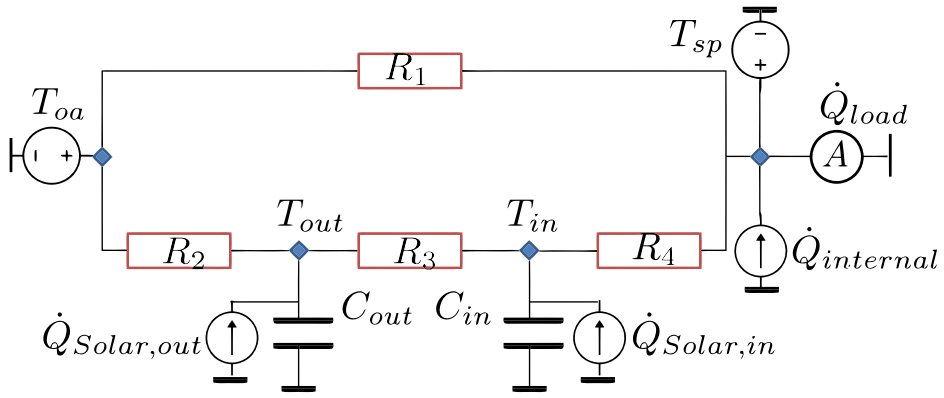


Figure 2.7: Building thermal load model.

The detailed equations describing the model in Figure 2.7 are

$$\dot{T}_{in} = \frac{\dot{Q}_{Solar,in} + \frac{T_{out}-T_{in}}{R_3} + \frac{T_{sp}-T_{in}}{R_4}}{C_{in}}, \tag{2.8a}$$

$$\dot{T}_{out} = \frac{\dot{Q}_{Solar,out} + \frac{T_{oa}-T_{out}}{R_2} + \frac{T_{in}-T_{out}}{R_3}}{C_{out}}, \tag{2.8b}$$

$$\dot{Q}_{Load} = \max(0, \dot{Q}_{internal} + \frac{T_{in}-T_{sp}}{R_4} + \frac{T_{oa}-T_{sp}}{R_1}). \tag{2.8c}$$

The load model (2.8a)–(2.8c) is used to model each building on the UC Merced campus where the parameters R_1, R_2, R_3, R_4, C_{in}, C_{out} are estimated using historical data. A nonlinear regression problem is solved to minimize the error between the actual cooling load demand and the predicted load demand. The predicted load demand is obtained by simulating the model (2.8a)–(2.8c) with measured inputs. The load model with estimated parameters is validated using load measurements from June 1 to June 5, 2009 at UC Merced. Figure 2.8 presents the validation results for one of the buildings at UC Merced. The measured building load is depicted as the dotted line, and the solid line shows the building load predicted by model (2.8a)–(2.8c). When the prediction mismatch exceeds desired tolerances, the system parameters need to be re-identified based on a new set of measured data. For other types of building load models see [96, 20].

Main Components Modeling

At the water-loop, the main components consuming energy are chillers, cooling towers, and pumps. It is assumed that the set points sent to these components are tracked instantly without errors.

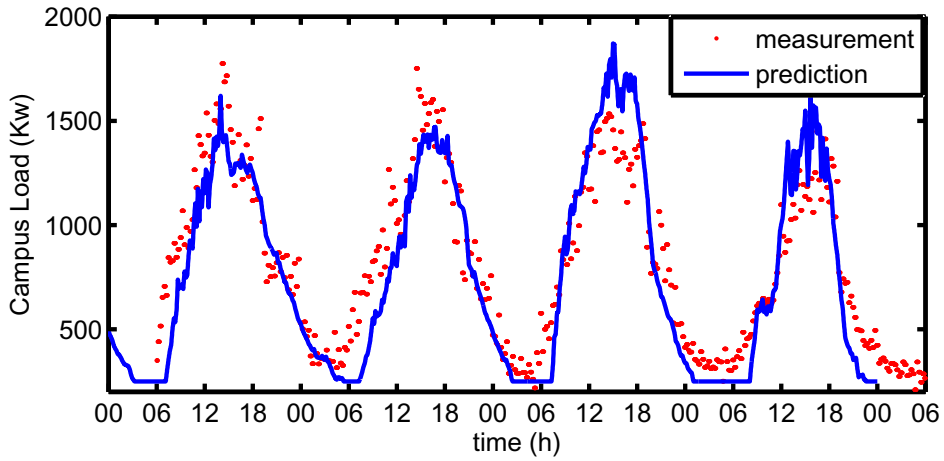


Figure 2.8: Campus load validation.

Chillers High fidelity models for chillers are based on thermodynamic refrigeration cycles. The refrigerant through the evaporator absorbs energy $\dot{Q}_{evap}$ from chiller water, and the vaporized refrigerant is transformed to liquid by the compressor with external work. The refrigerant through the condenser then loses its energy $\dot{Q}_{cond}$ to the condenser water. The refrigeration cycle enables the chillers to generate chilled water by ejecting heat to the condenser water loop. In order to develop control oriented model for chillers to enable real time optimization, several empirical functions are used to relate the boundary conditions and cooling load of the chillers to the power consumption [44]. It is assumed that chillers follow the energy conservation law. The energy balances across the evaporator and the condenser heat exchangers are respectively represented as,

$$\dot{Q}_{evap} = \dot{m}_{CHWS} c_p (T_{CHWR} - T_{CHWS}), \tag{2.9a}$$

$$\dot{Q}_{cond} = \dot{m}_{CWS} c_p (T_{CWR} - T_{CWS}), \tag{2.9b}$$

where $\dot{Q}_{evap}$ is the thermal energy change across the evaporator, and $\dot{Q}_{cond}$ is the thermal energy change across the condenser.

The cooling capacity of the chiller is predicted by the biquadratic function

$$\dot{Q}_{cap} = \dot{Q}_0 (a_0 + a_1 T_{CHWS} + a_2 T_{CHWS}^2 + a_3 T_{CWR} + a_4 T_{CWR}^2 + a_5 T_{CWR} T_{CHWS}), \tag{2.10}$$

where $\dot{Q}_0$ is the nominal capacity of chillers to be identified.

The part load ratio (PLR) is the fraction of the total cooling capacity used by the chillers.

$$PLR = \max(PLR_{min}, \dot{Q}_{evap}/\dot{Q}_{cap}), \tag{2.11}$$

where PLR_{min} is the minimal part load ratio allowed to start up the chillers.

Two energy input ratio (EIR) functions are defined to capture the power consumption of the chillers operating in various conditions, namely the energy input ratio temperature curve $EIRF_T(T_{CHWS}, T_{CWS})$, and energy input ratio PLR curve $EIRF_{PLR}(PLR)$,

$$EIRF_T = b_0 + b_1 T_{CHWS} + b_2 T_{CHWS}^2 + b_3 T_{CWR} + b_4 T_{CWR}^2 + b_5 T_{CHWS} T_{CWR}, \quad (2.12)$$

$$EIRF_{PLR} = c_0 + c_1 PLR + c_2 PLR^2 + c_2 PLR^3. \quad (2.13)$$

The chiller power consumption is calculated as

$$P_{chiller} = \dot{Q}_{cap} EIRF_T(T_{CHWS}, T_{CWS}) EIRF_{PLR}(PLR)/COP_0, \quad (2.14)$$

where COP_0 is the nominal coefficient of performance of the chiller when the chiller is operating under designed nominal conditions.

There are total 16 parameters to be identified in this simplified chiller model, namely the the nominal capacity of chillers $\dot{Q}_0$, the polynomial coefficients in (2.10) $[a_0,\ a_1, \ldots,\ a_5]$, the polynomial coefficients in (2.12) $[b_0,\ b_1, \ldots,\ b_5]$, and the polynomial coefficients in (2.13) $[c_0,\ c_1,\ c_2]$.

Historical data are used to populate this map or improve maps provided by chiller vendors. Model (2.10)-(2.14) is used to describe the chillers performance at UC Merced. Figure 2.9 presents the validation results based on the data collected from June 1 to June 3, 2009. The predicted chillers' power in solid line successfully captures the measurements in dotted line.

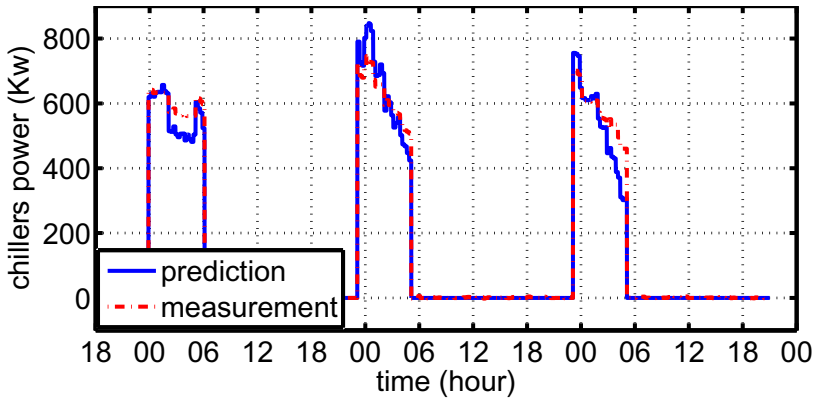


Figure 2.9: Chiller model validation.

Cooling Towers The cooling tower uses variable speed fans to track a set point for the condenser water supply temperature by rejecting the condenser water heat to the ambient environment through evaporation and conduction. In this thesis, the cooling towers are modeled based on the cooling tower approach temperature [64]. The approach temperature is the difference in temperature between the condenser water supply temperature and the entering air wet bulb temperature $T_{app} = T_{CWS} - T_{wb}$. Since the cooling towers are based on

the principles of evaporative cooling, the maximum cooling tower efficiency depends on the wet bulb temperature of the air. The approach temperature model implemented in Dymola is estimated by a polynomial regression model of the temperature range $T_R = T_{CWR} - T_{CWS}$, the wet bulb temperature T_{wb}, and the air to water ratio $R_{aw} = \dot{m}_{CWS}/\omega_{fan}$. In particular, the approach temperature is fitted using polynomial functions,

$$\begin{aligned} T_{app} &= APP(T_{wb}, T_{CWS}, T_{CWR}, \dot{m}_{CWS}, \omega_{fan}) \\ &+d_0 + d_1 T_{wb} + d_2 T_{wb}^2 + d_3 T_R + d_4 T_{wb} T_R + d_5 T_{wb}^2 T_R + d_6 T_R^2 + d_7 T_{wb} T_R^2 + d_8 T_{wb}^2 T_R^2 \\ &+d_9 R_{aw} + d_{10} T_{wb} R_{aw} + d_{11} T_{wb}^2 R_{aw} + d_{12} T_R R_{aw} + d_{13} T_{wb} T_R R_{aw} + d_{14} T_{wb}^2 T_R T_{aw} \\ &+d_{15} T_R^2 R_{aw} + d_{16} T_{wb} T_R^2 R_{aw} + d_{17} T_{wb}^2 T_R^2 R_{aw} + d_{18} R_{aw}^2 + d_{19} T_{wb} R_{aw}^2 + d_{20} T_R R_{aw}^2 \\ &+d_{21} T_{wb} T_R R_{aw}^2 + d_{22} T_{wb}^2 T_R R_{aw}^2 + d_{23} T_R^2 R_{aw}^2 + d_{24} T_{wb} T_R^2 R_{aw}^2 + d_{25} T_{wb}^2 T_R^2 R_{aw}^2. \end{aligned} \quad (2.15)$$

The approach temperature model (2.15) has 26 parameters to be identified $[d_0,\ d_1, \ldots,\ d_{25}]$.

Given a desired condenser water supply temperature T_{CWS}, the fan speed ω_{fan} is computed by solving the implicit equation (2.15) with the information of T_{wb}, T_{CWR}, and $\dot{m}_{CWS}$. The cooling tower power consumption is approximated as a cubic function of the fan speed

$$P_{ct} = c(\omega_{fan})^3. \quad (2.16)$$

This model (2.15)–(2.16) is applied to model the cooling towers at UC Merced, and the validation results are reported in Figure 2.10. The dotted line depicts the measured electric power consumption of cooling towers located at UC Merced, and the solid line shows the power consumption of the cooling towers predicted by model (2.15)–(2.16). Data was collected from June 1 to June 3, 2009.

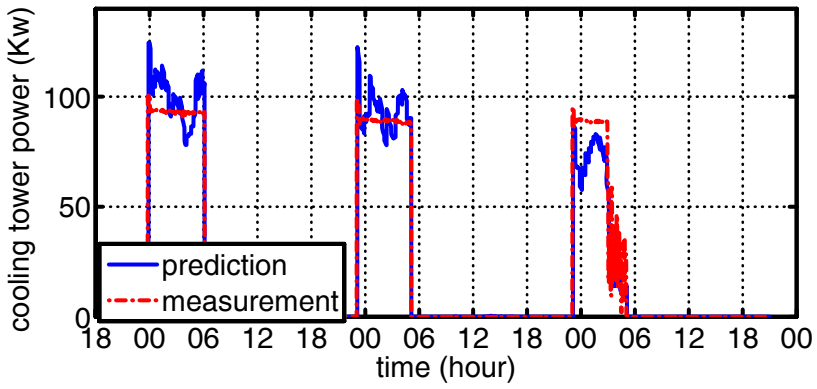


Figure 2.10: Cooling tower model validation.

Pumps When modeling the energy consumption of the pumps, it is assumed that the enthalpy change of water through the pumps are constant. It is also assumed that the pump

volumetric water flow q_{pump}, the pump speed ω_{pump}, the pressure difference Δp_{pump} generated by the pump, and their corresponding nominal values denoted with the superscript ^{0} satisfy the equations

$$\frac{q_{pump}}{q^0_{pump}} = \frac{\omega_{pump}}{\omega^0_{pump}}, \ \frac{\Delta p_{pump}}{\Delta p^0_{pump}} = \left(\frac{\omega_{pump}}{\omega^0_{pump}}\right)^2. \tag{2.17}$$

The pressure difference $\Delta p^0_{pump}(q^0_{pump})$ and efficiency $\eta^0_{pump}(q^0_{pump})$ under nominal operation conditions are approximated as polynomial functions of the nominal volumetric water flow q^0_{pump},

$$\Delta p^0_{pump} = c_0 + c_1 q^0_{pump} + c_2 (q^0_{pump})^2, \tag{2.18}$$
$$\eta^0_{pump} = d_0 + d_1 q^0_{pump} + d_2 (q^0_{pump})^2, \tag{2.19}$$

and the corresponding coefficients are obtained by fitting historical pump performance data. For a given fan speed ω_{pump} and the volumetric water flow rate q_{pump}, the pressure difference Δp_{pump}, and efficiency η_{pump} are computed as,

$$\begin{aligned}\Delta p_{pump} &= \left(\frac{\omega_{pump}}{\omega^0_{pump}}\right)^2 \Delta p^0_{pump} \\ &= \left(\frac{\omega_{pump}}{\omega^0_{pump}}\right)^2 \left(c_0 + c_1 \left(\frac{\omega^0_{pump}}{\omega_{pump}} q_{pump}\right) + c_2 \left(\frac{\omega^0_{pump}}{\omega_{pump}} q_{pump}\right)^2\right),\end{aligned} \tag{2.20}$$

$$\eta_{pump} = d_0 + d_1 \left(\frac{\omega^0_{pump}}{\omega_{pump}} q_{pump}\right) + d_2 \left(\frac{\omega^0_{pump}}{\omega_{pump}} q_{pump}\right)^2. \tag{2.21}$$

The pump power then is calculated as

$$P_{pump} = \Delta p_{pump} q_{pump} / \eta_{pump}. \tag{2.22}$$

The pump model (2.21)–(2.22) is validated by using measured data from the chilled water supply pump at UC Merced. The validation results in Figure 2.11 suggest that the model correctly predicts the pump power consumption. The dotted line is the measured electric power consumption of the hydraulic pump 2 in Figure 2.1 located at UC Merced. The solid line is the pump power consumption predicted by model (2.21)–(2.22). Data was collected from June 1 to June 3, 2009.

Air-Loop: Modeling a Building from the Inside

Thermal Zones Temperature Dynamics

We use an undirected graph structure to represent the room network and their dynamic couplings in the following way. We associate the i-th room with the i-th node of a graph,

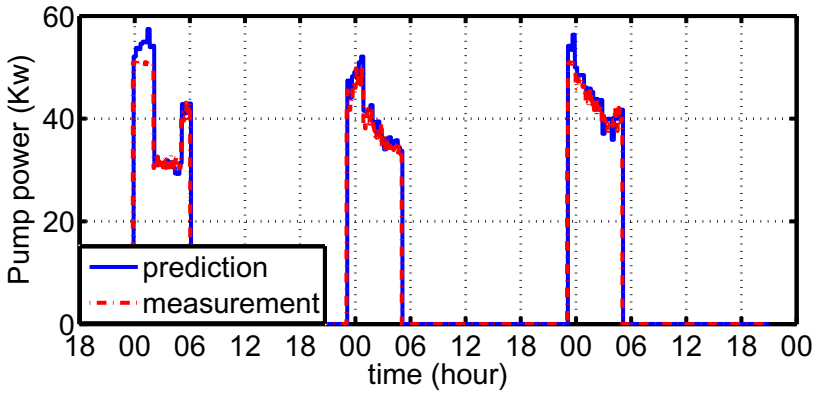


Figure 2.11: Pump model validation.

and if an edge (i, j) connecting the i-th and j-th node is present, the rooms i and j are subject to direct heat transfer. The graph $\mathcal{G}$ will be defined as

$$\mathcal{G} = (\mathcal{V}, \mathcal{A}), \tag{2.23}$$

where $\mathcal{V}$ is the set of nodes (or vertices) $\mathcal{V} = \{1, \ldots, N_v\}$ and $\mathcal{A} \subseteq \mathcal{V} \times \mathcal{V}$ the set of edges (i, j) with $i \in \mathcal{V},\ j \in \mathcal{V}$. We denote $\mathcal{N}^i$ the set of neighboring nodes of i, i.e., $j \in \mathcal{N}^i$ if and only if $(i, j) \in \mathcal{A}$.

Now consider a single room $j \in \mathcal{V}$. The air enters the room j with a mass flow rate $\dot{m}_s^j$, and supply air temperature T_s^j, It is assumed that in the AHU, the outside air fully mixes with the return air without delay, and the mixing proportion δ between the return air and outside air is controlled by the damper configurations in the AHU system to obtain:

$$T_m = \delta T_r + (1 - \delta) T_{oa}, \tag{2.24}$$

where T_{oa} is the outside air temperature, and T_m is the temperature of the mixed air. T_r is the return air temperature calculated as weighted average temperature of return air from each room

$$T_r = \sum_{i \in \mathcal{V}} \dot{m}_s^i T^i / \sum_{i \in \mathcal{V}} \dot{m}_s^i. \tag{2.25}$$

The return air is not recirculated when $\delta = 0$, and no outside fresh air is used when $\delta = 1$. δ can be used to save energy through recirculation but it has to be strictly less than one to guarantee a minimal outdoor fresh air delivered to the rooms.

The supply air temperature then is computed as

$$T_s^j = T_m - \Delta T_c + \Delta T_h^j, \forall j \in \mathcal{V}, \tag{2.26}$$

where ΔT_c represents the supply air temperature difference across the cooling coil in the AHU, and ΔT_h^j is the air temperature difference across the heating coil in the j-th VAV box.

We present two classes of models, and both are validated using historical data from Bancroft library at UC Berkeley.

RC Model We model the room as a two-mass system. C_1^j is the fast-dynamic mass that has lower thermal capacitance (e.g. air around VAV diffusers), and C_2^j represents the slow-dynamic mass that has higher thermal capacitance (e.g. solid parts which include floor, walls and furniture). We remark that the phenomenon of fast and slow dynamics has been observed in [73]. The thermal dynamic model of a room is:

$$C_1^j \frac{dT_1^j}{dt} = \dot{m}_s^j c_p (T_s^j - T_1^j) + (T_2^j - T_1^j)/R^j + (T_{oa} - T_1^j)/R_{oa}^j + \sum_{i \in \mathcal{N}^j} (T_1^i - T_1^j)/R_{ij} + P_d^j, \tag{2.27a}$$

$$C_2^j \frac{dT_2^j}{dt} = (T_1^j - T_2^j)/R^j, \ \forall j \in \mathcal{V}, \tag{2.27b}$$

where T_1^j and T_2^j are system states representing the temperatures of the lumped masses C_1^j and C_2^j, respectively. T^j is the perceived temperature of room j, which is assumed to be equal to the temperature of the fast-dynamic mass C_1^j. R_{oa}^j is the thermal resistance between room j and outside air, and c_p is the specific heat capacity of room air. R^j models the heat resistance between C_1^j and C_2^j, $R_{ij} = R_{ji}$ models thermal resistances between room i and the adjacent room j, and P_d^j is an unmeasured load induced by external factors such as occupancy, equipment, and solar radiation.

The model (2.27) is used to capture the temperature dynamics of a thermal zone in the Bancroft library located on the campus of University of California at Berkeley. By using historical data we have identified the model parameters for each thermal zones and validated the resulting model. The dimension of the conference room is $5 \times 4 \times 3$ m, and it has one door and no windows. As a result, the effect of solar radiation is negligible. The major source of load derives from occupants and electronic equipment. The conference room has one neighboring office room ($\mathcal{N}^1 = \{2\}$).

The model parameters ($[C_1^1,\ C_2^1,\ R^1,\ R_{12},\ R_{oa}^1]$) are identified by fitting measured data collected on January 20, 2012 over 24 hours using a nonlinear regression algorithm. This corresponds to a Sunday when the conference room has no occupants ($P_d^1 = 0$). Measurements of room temperature (T^1), supply air temperature (T_s^1), mass flow rate of the supply air ($\dot{m}_s^1$), the neighboring room temperature (T^2), and outside air temperature T_{oa} are used for the identification. The identified parameters values are reported in Table 2.1.

The identification results plotted in Figure 2.12 show that the proposed model successfully captures the thermal dynamics of the conference room without occupants. In Figure 2.12 the solid line depicts the measured room temperature trend and the dashed line is the room temperature predicted by model (2.27) when driven by the measured inputs.

The proposed model (2.27) with the identified parameters in Table 2.1 is validated against measurements during other unoccupied hours. Figure 2.13 plots the validation results for January 21, 2012. One can observe that the predictions match well the experimental data.

Table 2.1: Identification results for conference room model on January 20, 2012.

Parameter	Value	Parameter	Value
C_1^1	9.163×10^3 kJ/°C	R_{12}	2.000 °C/kW
C_2^1	1.694×10^5 kJ/°C	R_{oa}^1	57 °C/kW
R^1	1.700 °C/kW		

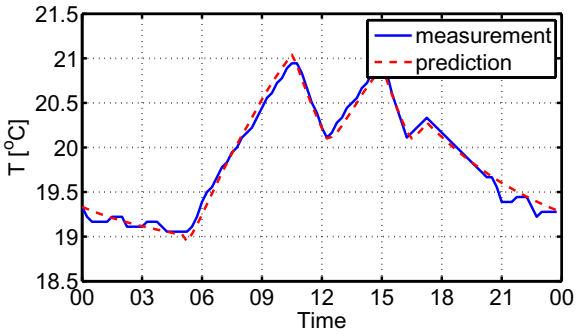


Figure 2.12: Identification results of the thermal zone model (2.27).

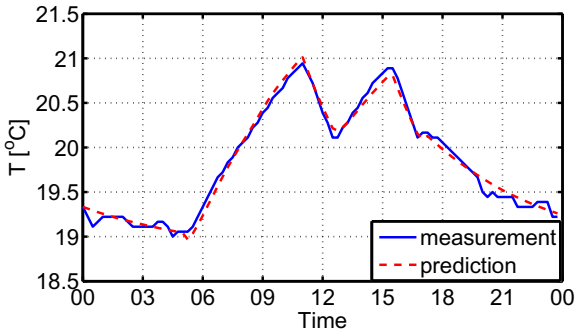


Figure 2.13: Validation results of simplified room model (2.27).

Bilinear Regression Model It has been shown that the physical based models are able to capture to dynamics of the thermal zones. Also the physical based models are capable of extrapolating the behavior of the thermal zones outside the training region. However, the identification problem for model (2.27) is non-convex, and it quickly becomes computationally demanding with increasing number of thermal zones. In order to reduce the computational complexity of the identification problem, the temperature dynamic of a network of zones can be modeled as a bilinear regression model as in [73, 131]. Each thermal zone $i \in \mathcal{V}$ is simplified as an ARMAX model, and the exogenous signals consist of the thermal energy input $(T_s^i - T^i)\,\dot{m}_s^i$, internal thermal load P_d, and measured disturbances including ambient temperature T_{oa} and solar radiation intensity I^i.

$$T^i(t+1) = \sum_{q=0}^{q=q_d} \left(p_{1,q}^i T_{oa}(t-q) + p_{2,q}^i I^i(t-q)\right) + p_3^i \left(T_s^i(t) - T^i(t)\right) \dot{m}_s^i(t) + \sum_{q=0}^{q=q_x} \left(p_{4,q}^i T^i(t-q) + \sum_{j \in \mathcal{N}^i} p_{5,q,j}^i T^j(t-j) \right) + p_6 + P_d^i(t), \tag{2.28}$$

where q_x and q_d are the autoregressive order and moving average order, respectively, and the selection of q_x and q_d affects the complexity and accuracy of the model. The bilinear regression model (2.28) is identified using data collected from unoccupied hours when the internal load P_d is minimal. Given historical measurements of zone temperatures T, ambient

temperature T_{oa}, solar radiation intensity I, and recorded control inputs ($\dot{m}_s$ and T_s), the model parameters can be computed by solving a linear regression problem.

The bilinear zone model (2.28) with $q_x = q_d = 2$ is used to model the conference room in Bancroft library at UC Berkeley. The floor plan of the Bancroft library is depicted in Figure 2.14. Results for the classroom (labeled VAV C-2-15 in Figure 2.14) are reported next.

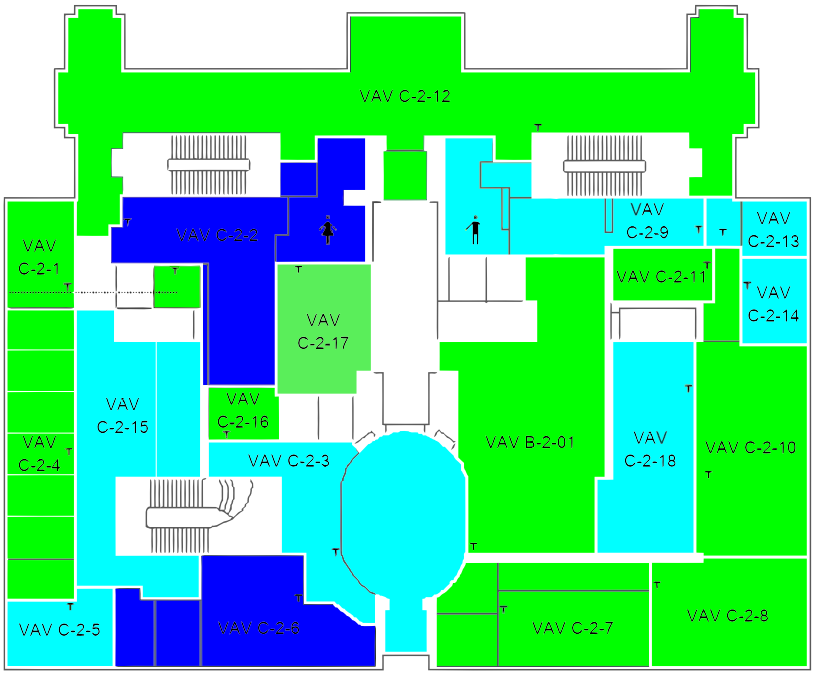


Figure 2.14: Bancroft library floor plan.

The identification results are reported in Figure 2.15 for the weekend of January 20, 2012, and the conference room is unoccupied during the weekends, i.e. $P_d^{i\in\mathcal{V}} = 0$. The identified model then is validated for the weekend of January 21, 2012, and the results are plotted in Figure 2.16.

Load

The load prediction $P_d^j(t)$ for each thermal zone j is important for designing predictive feedback controllers and assessing potential energy savings. Various approaches are available to estimate occupancy load. For instance, the authors of [86] develop an agent-based model to simulate the occupants' behavior in a building, and the work in [52, 100] focuses on occupants' behavior and mobility patterns using a wireless camera sensor network.

Time varying bounds on the disturbance load $P_d(t)$ can be computed from the mismatch between a nominal model and historical data, and correlating the load bounds with shared calendars, weather predictions as well as predicted cloud coverage. This concept can be illustrated using the conference room discussed previously. The conference room calendar

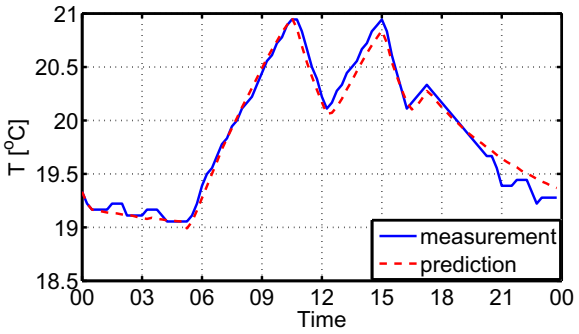


Figure 2.15: Identification results of the thermal zone model (2.28).

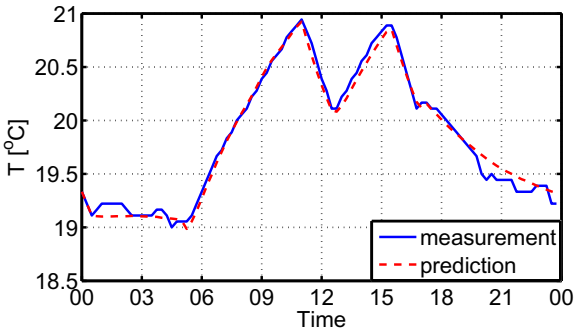


Figure 2.16: Validation results of simplified room model (2.28).

contained two regularly scheduled group meetings at 10:00 AM and 02:00 PM every Wednesday. The same meetings can be identified by inspecting the model mismatch between nominal model predictions and historical data. Figure 2.17 depicts the envelope-bounded disturbance load during all Wednesdays in July, 2012. The envelope is computed as the pointwise minimum and maximum difference between measured data and the nominal model described by (2.27) with $P_d^1 = 0$. The two peaks in the disturbance load envelope in Figure 2.17 correspond to two regularly scheduled group meetings. The off-peak prediction errors can be attributed to unmodeled dynamics and external disturbances.

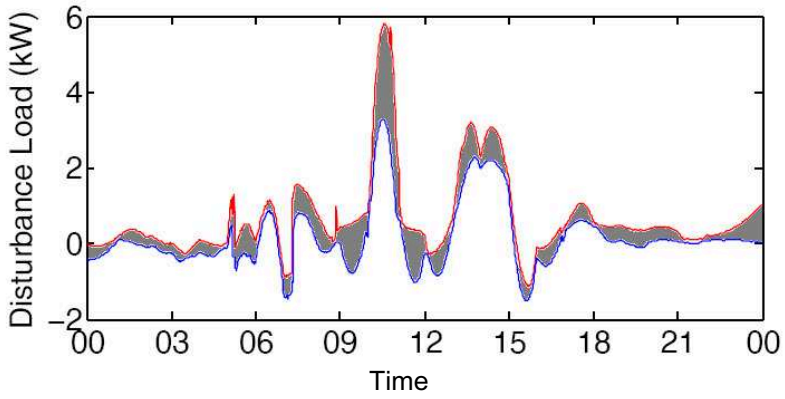


Figure 2.17: Envelope-bounded disturbance load during all Wednesdays in July 2012 in the conference room of the Bancroft Library at UC Berkeley.

Main Components

The components at the air-loop of the architecture that use energy include dampers, supply fans, heating coils, and cooling coils as shown in Figure 2.2. The supply fan needs electrical

power to drive the system while the heating and cooling coils consume the energy of the chilled and hot water. It is assumed that the power to drive the dampers is negligible. A simple energy consumption model for each component is presented next.

Fan Power The fan power can be approximated as a second order polynomial function of the total supply air mass flow rate $\dot{m}_{fan}$ driven by the fan, and the supply air mass flow rate by the fan is equal to the summation of air flow to each room $\dot{m}_{fan} = \sum_{j \in \mathcal{V}} \dot{m}_s^j$.

$$P_f = c_0 + c_1 \dot{m}_{fan} + c_2 \dot{m}_{fan}^2, \tag{2.29}$$

where c_0, c_1, c_2 are parameters to be identified by fitting recorded data. The simplified fan

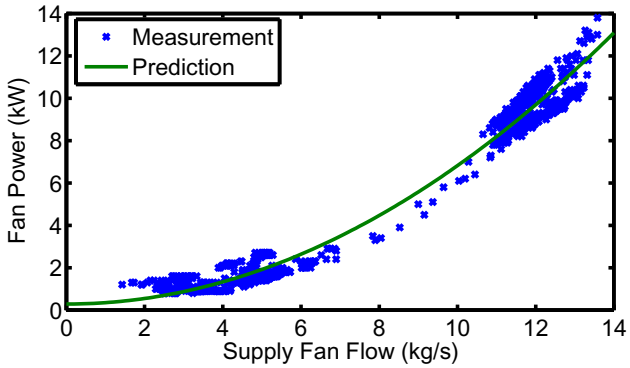


Figure 2.18: Identification results for the simplified fan power consumption model (2.29).

model (2.29) is tested on the recorded data from the UC Berkeley Bancroft Library from October 1st to October 10th 2010. The identification results plotted in Figure 2.18 suggest that the polynomial function successfully predicts the electricity consumption of the fan.

Cooling and Heating Coils Cooling coils and heating coils are air-water heat exchangers. There has been extensive studies to develop simplified yet descriptive models of coil units [116, 37, 134]. The authors in [116] developed simple empirical equations with four parameters by using a finite difference method to capture the transient response of counter-flow heat exchangers. In [37], the authors presented an improved simulation model-based on ASHRAE Secondary HVAC Toolkit, and in [134] a simplified control oriented cooling coil unit model is presented based on energy and mass conservation laws.

In this thesis, we use a simple coil model with constant efficiency (η_c for the cooling coils and and η_h for the heating coils). With this simplification the energy consumption model is a static function of the load on the air-side

$$P_c = \frac{\sum_{j \in \mathcal{V}} \dot{m}_s^j c_p \Delta T_c}{\eta_c \, \mathrm{COP}_c}, \;\; P_h^j = \frac{c_p \dot{m}_s^j \Delta T_h^j}{\eta_h \, \mathrm{COP}_h}, \tag{2.30}$$

where P_c is the electrical power consumption related to the generation of chilled water consumed by the cooling coils in AHU and P_h^j is the power used to generate the hot water consumed by the heating coils in the VAV box connected with room j. ΔT_c is the temperature difference through the cooling coils, ΔT_h^j the temperature difference through the heating coils in the VAV box j, COP_c is the chilling coefficient of performance, and COP_h is the heating coefficient of performance. The coefficient of performance (COP) is defined as

$$\text{COP} = \frac{E_{thermal}}{E_{input}}. \tag{2.31}$$

COP captures the efficiency of the exchange system, i.e., the amount of thermal energy $E_{thermal}$ (J) generated by the system with one Joule of energy consumed. The input energy E_{input} can be from different resources such as electricity, fuel, and gas for different systems. Model (2.30) is oversimplified as compared to the aforementioned literature. However, the model is adequate to capture the energy consumption if coils are operating in a narrow performance range.

2.4 Summary

In this chapter, the building cooling and heating systems with energy storage elements are presented. The overall system is divided into water-loop system and air-loop system. Simplified yet descriptive control oriented models are derived and discussed for both water-loop and air-loop systems. The models are used in the design of model-based predictive control in the following chapters.

Chapter 3

Deterministic Predictive Control Design

3.1 Introduction

This chapter describes the design of model predictive control for building cooling and heating systems. The chapter starts with an illustrative MPC design for a one-state building model. The example is used to demonstrate the basic principles of active thermal storage.

In the second part, a MPC scheme is presented to optimize the operation of water-loop in building cooling and heating systems. The algorithm is implemented for the control of the cooling system on UC Merced campus. A nonlinear MPC is designed to optimize the operation of the central chilling plant, and the charge level of a energy storage tank. The MPC problem is solved using sequential quadratic programming. In addition, move blocking strategy is applied to reduce the problem size, and a periodic robust invariant set is used as terminal constraints so that there is always enough chill water in the tank to meet all possible realization of the uncertain demands. Experimental results suggested that MPC is able to reduce the energy consumptions and increase the thermal efficiency of the chillers and cooling towers.

In the third part, we present the design of MPC logic for air-loop of building cooling and heating systems. Extensive simulations show that MPC is able to maintain thermal comfort while reducing energy consumptions. Furthermore, time varying electricity price can be incorporated in MPC design with minimum modifications to reduce the electricity bills.

The control design of this chapter uses nominal models and expected predictions. However, prediction uncertainties will degrade the performance of MPC. We conclude with a simple numerical example to demonstrated that simple control can possibly outperform nominal MPC design if prediction uncertainty is high. We will address MPC design with prediction uncertainty using stochastic MPC in Chapter 5.

3.2 Model Predictive Control and Thermal Storage: a Simple Example

In order to make people feel comfortable, energy is converted and delivered to building spaces. In this thesis occupants' comfort is measured by the perceivable air temperature of a given space. The objective of this section is to use a simple thermal mass model to show the basic principles of active thermal storage, to demonstrate active storage that naturally emerges as closed-loop behavior of MPC scheme, and to discuss the main tradeoffs that arise in active thermal storage.

The temperature dynamics of a given space can be modeled using resistance-capacitance (RC) circuit analogy

$$C\dot{T} = u + P_d + (T_{oa} - T)/R, \tag{3.1}$$

where T is the temperature of the space, P_d is the external disturbance load generated by occupants, direct sunlight, and electrical devices, T_{oa} is the temperature of outside air, and u is the heating and cooling power input to the space. The space is cooled when $u \leq 0$ and heated when $u \geq 0$.

In water-loop problem, the simplified thermal mass model (3.1) can be viewed as the abstraction of an entire building and the temperature T is an average temperature of all building spaces. In air-loop problem, the simplified thermal mass model (3.1) can be viewed as the abstraction of one zone within a building, and T is the temperature measurement of the zone. In this case, the lumped parameter R describes the thermal resistance of walls and windows that isolate the zone from the outside environment, and the parameter C lumps up the thermal capacitance of the zone. The active storage mechanism explained next shares the same properties for both control problems.

The representation of system (3.1) in discrete time is obtained using Euler discretization with a sampling time of Δt,

$$T(k+1) = AT(k) + Bu(k) + d(k), \tag{3.2}$$

where $A = 1 - \dfrac{\Delta t}{RC}$, $B = \dfrac{\Delta t}{C}$, $d = \dfrac{P_d \Delta t}{C} + \dfrac{T_{oa}\Delta t}{RC}$.

A simple model predictive control problem is formulated with the objective of minimizing total heating and cooling energy consumption, minimizing the peak power consumption, and maintaining zones within a desired temperature range despite the predicted load changes.

At each time step, the predictive controller solves the following problem

$$\min_{\mathbf{U}_t,\ \underline{\varepsilon},\ \overline{\varepsilon}} \sum_{k=0}^{N-1} |u_{t+k|t}|\Delta t + \kappa \max\{|u_{t|t}|, \ldots, |u_{t+N-1|t}|\} + \rho \sum_{k=1}^{N} \left(|\overline{\varepsilon}_{t+k|t}| + |\underline{\varepsilon}_{t+k|t}|\right) \tag{3.3}$$

subject to

$$T_{t+k+1|t} = AT_{t+k|t} + Bu_{t+k|t} + d_{t+k|t}, \tag{3.4}$$

$$\underline{T} - \underline{\varepsilon}_{t+k|t} \leq T_{t+k|t} \leq \overline{T} + \overline{\varepsilon}_{t+k|t}, \tag{3.5}$$

$$\underline{\varepsilon}_{t+k|t},\ \overline{\varepsilon}_{t+k|t} \geq 0, \tag{3.6}$$

where the symbol $v_{t+k|t}$ is read as "the variable v at time $t+k$ predicted at time t." For instance, $T_{3|1}$ represents the predicted temperature at time 3 when the prediction is made at time $t = 1$ starting from the current temperature $T(1)$. $T_{3|1}$ is in general different from $T_{3|2}$, which is the predicted temperature at time 3 when the prediction is made at time $t = 2$ starting from the current temperature $T(2)$. With this notation in place, $\mathbf{U}_t = \left[u_{t|t},\ u_{t+1|t}, \ldots, u_{t+N-1|t}\right]$ is the vector of energy control inputs, $\underline{\varepsilon} = \left[\underline{\varepsilon}_{t+1|t}, \ldots, \underline{\varepsilon}_{t+N|t}\right]$ is the vector of temperature violations below the lower bound, $\overline{\varepsilon}$ collects the temperature violation above the upper bound, $T_{t+k|t}$ is the thermal zone temperature, and $d_{t+k|t}$ is the load prediction. $\underline{T}$ and $\overline{T}$ are the lower and upper bounds on the zone temperature, respectively. ρ is the penalty on the comfort constraint violation, κ is the penalty on peak power consumption, and N is the length of the prediction horizon.

Let $\mathbf{U}_t^* = \{u_{t|t}^*, \ldots, u_{t+N-1|t}^*\}$ be the optimal solution to Problem (3.3)–(3.6) at time t. The first element of $\mathbf{U}_t^*$ is applied to system (3.2)

$$u(t) = u_{t|t}^*. \tag{3.7}$$

The optimization problem (3.3)–(3.6) is repeated at the next time step $t+1$ based on the new measured temperature $T_{t+1|t+1} = T(t+1)$, yielding a moving or receding horizon control strategy.

The following parameters are used in the simulations presented next. Thermal capacitance $C = 9.2 \times 10^3$ kJ/°C, thermal resistance $R = 50$ °C/kW, sampling time $\Delta t = 1$ hour, prediction horizon $N = 24$ hours, and thermal comfort interval $[\underline{T}, \overline{T}] = [21,\ 26]$ °C. Note that the plant model (3.2) and the model used in the MPC scheme (3.4) are the same. It is assumed that weather and load are periodic with a period of one day, and that their predictions are perfect without mismatch between predictions and actual measurements. The outside air temperature profile $T_{oa}(t)$ used in (3.1) is depicted in Figure 3.1, and the disturbance load profile $P_d(t)$ used in (3.1) is depicted in Figure 3.2. The load $P_d(t)$ resembles a 3 kW thermal load of a conference room with a daily meeting scheduled from 9:00 AM to 11:00 AM.

Two controllers are considered. Controller C1 is a proportional controller designed to reject the load without predictive information. Controller C1 inputs zero power when the

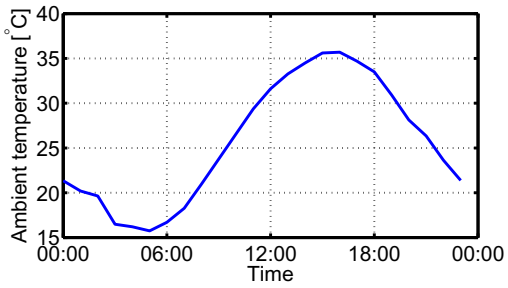


Figure 3.1: Ambient temperature (°C).

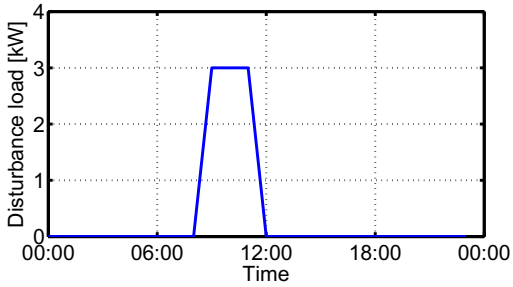


Figure 3.2: Disturbance load (kW).

space temperature is within the comfort range, otherwise the proportional control law

$$u(t) = \begin{cases} K(\overline{T} - T(t)) & T(t) \geq \overline{T} \\ 0 & \underline{T} < T(t) < \overline{T} \\ K(\underline{T} - T(t)) & T(t) \leq \underline{T} \end{cases} \tag{3.8}$$

is applied to system (3.2). Controller C2 implements the MPC Problem (3.3)–(3.7).

Simulations of system (3.2) in closed-loop with C1 and C2 are performed until the system settles to steady-periodic behavior. The performance of the controllers is measured by the closed-loop total energy consumption

$$J^u = \sum_{k=0}^{N-1} |u^*(k)| \Delta t, \tag{3.9}$$

the peak power consumption

$$J^p = \max\{|u^*(0)|, \ldots, |u^*(N-1)|\}, \tag{3.10}$$

and the total comfort violation

$$J^\varepsilon = \sum_{k=0}^{N} \left(|\overline{\varepsilon}^*(k)| + |\underline{\varepsilon}^*(k)|\right) \Delta t. \tag{3.11}$$

Closed-loop simulations are performed with various gains K of the controller C1 and various tuning ρ of MPC with no penalty on peak power consumption, that is $\kappa = 0$. Figure 3.3 shows the tradeoff between comfort violation and total energy consumption. It is observed that C1 and C2 use the same energy for the same amount of constraint violation. As expected, increased comfort violation corresponds to a lower energy use for both controllers. Note that the results above are valid for the model (3.2) and the performance indices defined by (3.9)-(3.11).

When $\kappa \neq 0$, a different behavior is observed. Simulation results for C1 with $K = 400$ and C2 with $\rho = 1000$ and $\kappa = 2$ are plotted in Figure 3.4-3.5.

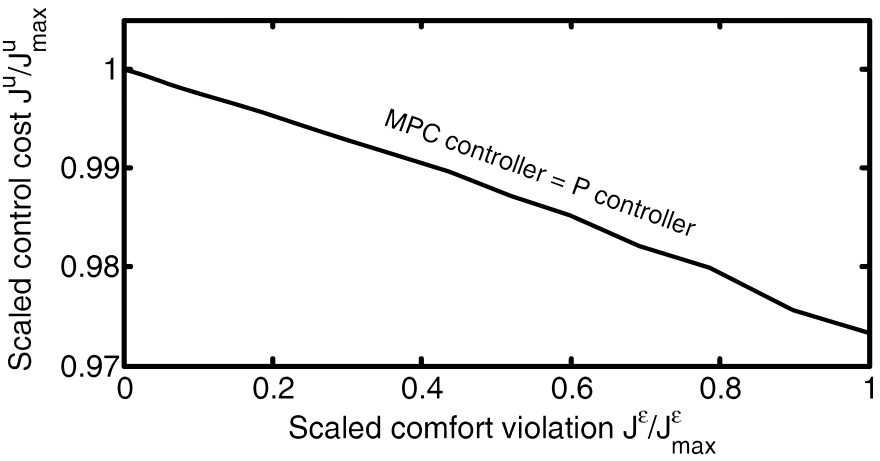


Figure 3.3: Comparing MPC (C2) and proportional controller (C1). Observe that C1 and C2 use the same energy $J^u_{C1} = J^u_{C2}$ for the same amount of constraint violation $J^\varepsilon_{C1} = J^\varepsilon_{C2}$. As expected, increased comfort violation corresponds to a lower energy use for both controllers.

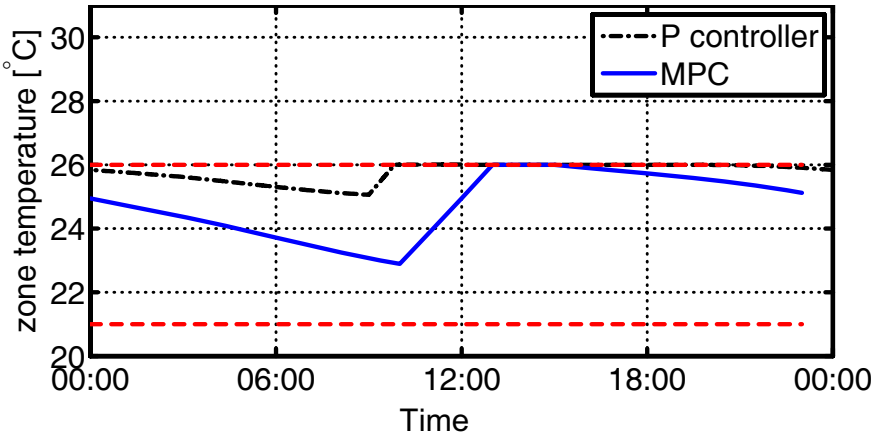


Figure 3.4: Zone temperature (°C).

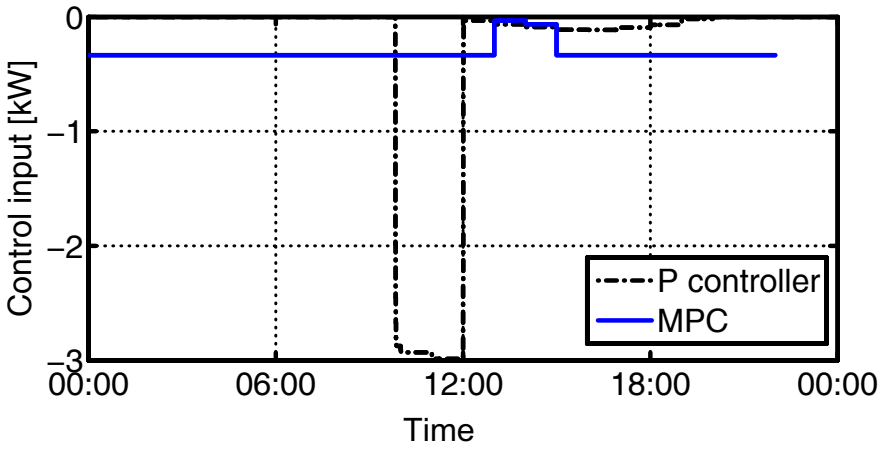


Figure 3.5: Cooling input (kW).

The peak power consumption J^p is reduced by 89% relative to the proportional controller C1 when MPC C2 is used. For both controllers the comfort violation index J^ϵ is zero. This behavior is obtained by taking advantage of the predictive information of the disturbance and using the space thermal storage. In fact MPC precools the space temperature to 22.8 °C before the occupancy load begins. This *precooling* behavior reduces the peak power consumptions of the system and flattens the control profile.

In the previous simulation, the total energy consumption J^p of MPC is increased by 6.3% compared to the proportional controller. The increase in energy consumption is due to energy losses through the resistance R while precooling. The tradeoff between the total energy losses $\frac{J^u_{C2}-J^u_{C1}}{J^u_{C1}}$ and the peak power reduction $\frac{J^p_{C1}-J^p_{C2}}{J^p_{C1}}$ is further explored in Figure 3.6. The tradeoff

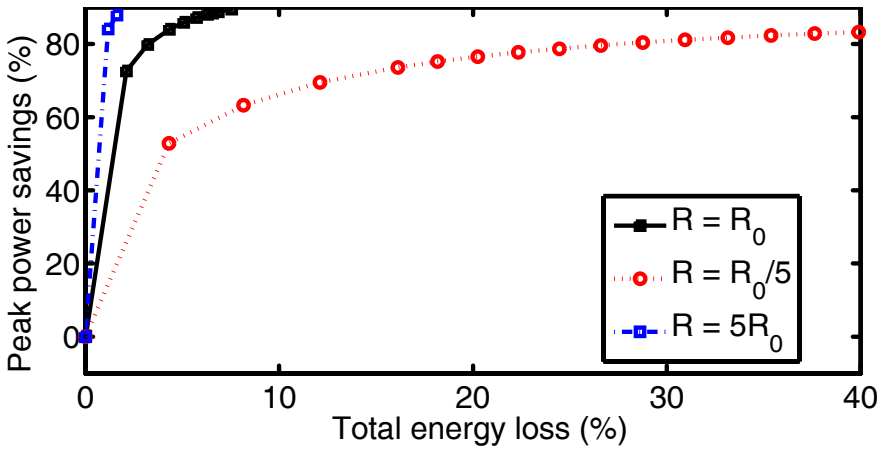


Figure 3.6: Comparison between MPC and proportional control in terms of closed-loop total energy losses ($\frac{J^u_{C2}-J^u_{C1}}{J^u_{C1}}$) and peak power reduction ($\frac{J^p_{C1}-J^p_{C2}}{J^p_{C1}}$).

lines are generated for different tunings of κ taken from the interval $[0, 5]$ and three different thermal resistances with values $R_0 = 50$ °C/kW, $R_0/5$, and $5R_0$. MPC achieves lower peak power consumption at the cost of higher energy consumption. MPC energy losses relative to the proportional controller decrease as the thermal resistance R in (3.1) increases. Thorough studies on load shifting have appeared in [19, 102].

The above example shows the benefits and tradeoffs resulting from the use of MPC. However, two important elements are not captured by the example. The first element is the complexity of the real problem. Buildings are more complicated than simple RC systems. It is therefore necessary to adopt descriptive yet simple enough models for real-time optimization in MPC and bring substantial savings to the real world. Also, the cost function of a MPC scheme in practice includes more detailed energy consumption functions for system components and external signals such as time varying utility price, availability of renewable energy, and load shedding signals received from the utility grid. The resulting MPC logic combines load shifting with additional features such as peak electrical power reduction and free cooling. The second element not captured by the example is the uncertainty in predictions. In reality the model and predictions of load and weather are uncertain.

In this chapter, we address the first issue by incorporating more descriptive dynamic models and energy models introduced in Chapter 2. The second issue will be addressed in Chapter 5 using stochastic MPC.

For the sake of brevity, we concentrate on the cooling side. Despite of different central equipment for heating generation, the heating architecture is similar to the cooling one. As a result, the abstraction levels and control methodology in this thesis can be extended to heating systems with minimal effort.

The MPC architecture presented in this chapter is depicted in Figure 3.7. A high-level MPC (HMPC) is deployed to optimize the operation and schedule of the water-loop in cooling systems with active thermal storage and a low-level MPC (LMPC) controls the air-loop including VAV boxes and the air handling units in each building to guarantee occupant thermal comfort. At both levels various predictions can be included in the constraints and in the cost function to control the system in an efficient and effective way. These predictions include building loads, load shedding signals from the power grid, utility prices, weather, occupancy, and solar radiation. In addition, HMPC and LMPC can exchange information to achieve better performance. For example, the occupancy load prediction from LMPC can help HMPC achieve better accuracy of building load predictions. Also, the chilled and hot water temperature predictions from HMPC impose constraints for the LMPC on achievable supply air temperature downstream of the cooling and heating coils, respectively.

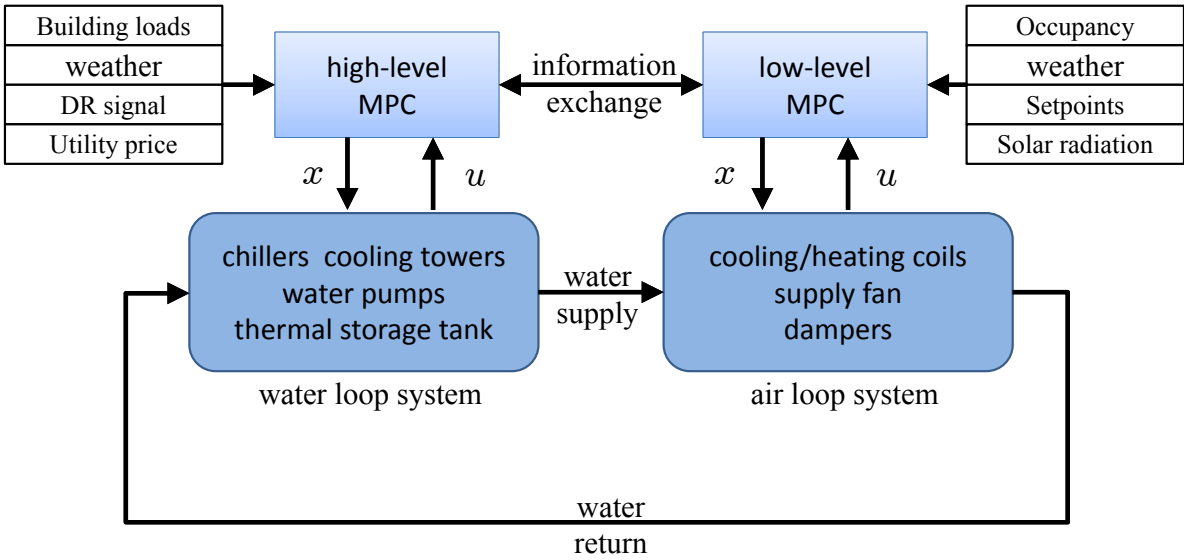


Figure 3.7: Hierarchical MPC structure for a building control system.

The following optimization problem is used to describe both the HMPC and LMPC

$$J^{\star}(x(t),t) = \min_{u_{t|t},\ldots,u_{t+N-1|t}} \sum_{k=0}^{N-1} J(x_{t+k|t}, u_{t+k-1|t}, k) + J_N(x_{t+N|t}) \tag{3.12}$$

subject to

$$x_{t+k+1|t} = f(x_{t+k|t}, u_{t+k|t}, d_{t+k|t}, k), \forall k = 0, 1, \ldots, N-1, \tag{3.13}$$
$$y_{t+k|t} = g(x_{t+k|t}, u_{t+k-1|t}, d_{t+k|t}, k), \forall k = 0, 2, \ldots, N, \tag{3.14}$$
$$y_{t+k|t} \in \mathbb{Y}, \forall k = 1, 2, \ldots, N, \tag{3.15}$$
$$u_{t+k|t} \in \mathbb{U}, \forall k = 0, 1, \ldots, N-1, \tag{3.16}$$
$$d_{t+k|t} \in \mathbb{D}(t+k), \forall k = 1, 2, \ldots, N, \tag{3.17}$$
$$x_{t|t} = x(t), \tag{3.18}$$

where $\mathbb{Y}$ is set of feasible system outputs y, $\mathbb{U}$ is the feasible set of control inputs u, $J_N(x)$ is the terminal cost function, $f(x,u,d,k)$ is the state update equation, d is the disturbance, and $\mathbb{D}$ is the set of possible disturbances realizations. Disturbances d_k can be predicted by a dynamic model such as building load model (2.8a)–(2.8c). An alternative approach is to obtain the future admissible set of disturbances $\mathbb{D}(t+k)$ by external modules such as the occupancy model in Figure 2.17.

Let $U^{\star}_{t\to t+N-1|t} = \{u^{\star}_{t|t}, \ldots, u^{\star}_{t+N-1|t}\}$ be the optimal solution to problem (3.12)–(3.18) at time t. Then the first element of $U^{\star}_{t\to t+N-1|t}$ is applied to the system, $u(t) = u^{\star}_{t|t}$. The optimization problem (3.12)–(3.18) is repeated at $t+\Delta t$, with the updated state $x_{t+\Delta t|t+\Delta t} = x(t+\Delta t)$, yielding a moving or receding horizon control strategy.

The cost function, model dynamics, constraints, and disturbances depend on the abstraction level and the specific problem of interest. Two detailed implementations of HMPC and LMPC schemes are presented in the following sections. We remark that when a nominal model of the disturbances d_k is replaced by a set valued model, that is $d_k \in \mathbb{D}(k)$ with a given probability distribution function, then robust or stochastic MPC formulations [112, 7, 30, 130, 42] need to be used in place of (3.12)–(3.18). More details on stochastic MPC are presented in Chapter 5.

3.3 Predictive Control for Water-Loop System

The objective of HMPC is to minimize the electrical energy consumption while generating enough chilled water. A typical cost function of the HMPC in (3.12) penalizes total electricity bill and the deviation from the building thermal energy demand satisfaction. The cost can be further extended to include the peak load requests, time varying utility prices, and time varying availability of renewable energy. The control variables to be optimized by MPC include the chilled water supply temperature T_{CHWS}, condenser water supply temperature T_{CWS}, chilled water supply flow rate $\dot{m}_{CHWS}$, chilling system start time t_s, and chilling system end time t_{end}. The dynamic system $f(x, u, d, k)$ includes the storage dynamics in (2.1a)–(2.1b), and the disturbance d includes weather and building load demand.

The following section reports a real time implementation of HMPC for the cooling system installed on the UC Merced campus.

3.4 University of California at Merced Experimental Testing

The HMPC controller for the UC Merced campus computes the set points for cooling towers, chillers, and the thermal storage tank at the central plant. The MPC algorithm is implemented in Matlab® and runs in real time on a Pentium 4 Intel processor. The MPC algorithm receives and sends data to the campus through the building automation system "Automated Logic Web Control." In the following sections, the MPC problem formulation and implementation details are presented.

Control Variables

The control variables of the HMPC implemented at UC Merced are listed as follow,

1. $T_{CWS,ref}(t)$: Reference temperature of water exiting cooling towers. The sampling rate is 1 hour.

2. $\dot{m}_{CHWS,ref}(t)$: Mass flow rate of the chiller water supply. It is a disconnected set. The mass flow rate is 0 when chillers are off, and $[148, 235]$ kg/s while the chillers are operating. The sampling rate is 1 hour.

3. $T_{CHWS,ref}(t)$: Reference temperature of water supplied by chillers. The sampling rate is 1 hour.

4. t_s (t_f): Start-up (Shut-down) time of chillers and cooling towers. The sampling rate is 1 day.

Measured Variables

The following variables are measured to initialize prediction of the HMPC controller (3.18),

1. T_a: Temperature of the cool water in the tank.

2. T_b: Temperature of the warm water in the tank.

3. z_a: Height of the warm water in the tank above the thermocline.

4. z_b: Height of the cool water in the tank below the thermocline.

The following variables are predictions downloaded from the national weather service.

1. T_{amb}: Ambient temperature.

2. T_{wb}: Ambient temperature.

3. β_{cloud}: Cloud coverage.

Operation Constraints

The following constraints avoid the malfunction of the system components.

1. $T_{CWS,ref} \in [288, 295]K$.

2. $\dot{m}_{CHWS,ref} \in \{0\} \bigcup [148, 235] kg/s$.

3. $T_{CHWS,ref} \in [276.5, 280.4]K$.

4. $T_{CHWR} \in [283, 295]K$.

5. $z_b \in [0.3, 1] z_{tank}$.

Model Summary

The dynamic equations of the water-loop system include the energy storage tank (2.1a)–(2.5), the load model (2.6)–(2.8) that estimates the total energy requested by buildings, and the chilling system including the chillers (2.14), cooling towers (2.16), and pumps (2.22). The dynamic system is discretied with sampling time (Δt) of 1 hour:

$$x(t+\Delta t) = f_{wl}(x(t), u(t), \Phi(t), t), \tag{3.19a}$$
$$y(t) = g_{wl}(x(t), u(t), \Phi(t), t), \tag{3.19b}$$

where

$$f_{wl} = \begin{cases} f_1(x(t), u(t), \Phi(t), t); & \text{if } \dot{m}_{CHWS} \leq \dot{m}_{cmp} \\ f_2(x(t), u(t), \Phi(t), t); & \text{if } \dot{m}_{CHWS} > \dot{m}_{cmp} \end{cases}$$
$$u(t) = [T_{CWS,ref}; \dot{m}_{CHWS}; T_{CHWS,ref}] \in \mathbb{U}$$
$$x(t) = [U_a; U_b; z_a; z_b; T_{in}; T_{out}]$$
$$y(t) = [T_{CHWR}; z_b] \in \mathbb{Y}.$$

$\mathbb{U}$ and $\mathbb{Y}$ are the feasible control input set and feasible output set defined in Section 3.4, respectively.

Energy Price

UC Merced is currently enrolled in a special plan, electric schedule E-20. The unit price for the service under Schedule E-20 varies depending on the period of time. Table 3.1 shows the definition of the time periods.

Table 3.1: Definition of time periods.

SUMMER	Period A	(May 1st though Oct. 31st)
Peak	12:00–18:00	except holidays
Partial-peak	8:30–12:00	except holidays
	AND 18:00–9:30	
Off-peak	21:30–8:30	Mon. through Fri.
	ALL DAY	Sat., Sun, and holidays
WINTER	Period B	(Nov. 1st though Apr. 30st)
Partial-peak	8:30–21:30	except holidays
Off-peak	21:30–8:30	Mon. through Fri.
	ALL DAY	Sat., Sun, and holidays

We denote the unit electricity price for energy charge defined by Table 3.2 as the function C_e. The customer pays for energy by the kilowatt hour (kWh).

Table 3.2: Total energy rates.

Total Energy Rates ($ per kWh)	
Peak Summer	$ 0.13593
Part-Peak Summer	$ 0.09204
Off-Peak Summer	$ 0.07392
Part-Peak Winter	$ 0.08155
Off-Peak Winter	$ 0.07118

The total bill then is computed as

$$\text{Bill}_{wl}(t) = \sum_{k=0}^{k=N} C_e(t + k\Delta t)\text{Power}_{wl}(t + k\Delta t)\Delta t, \tag{3.20}$$

where Power_{wl} is the function adding up the power consumption of chillers (2.14), cooling towers (2.15)-(2.16), and pumps (2.21)-(2.22).

HMPC Problem Formulation

This section presents the design of a MPC controller whose objective is to find the optimal control sequence that satisfies the required cooling load and minimizes electricity usage. Consider the following optimization problem: Consider the following optimization problem:

$$\begin{aligned}
J^\star(x(t), t) = \min_{\hat{u}_{t|t},\ldots,\hat{u}_{t+M-1|t},t_s,t_f} & \text{Bill}_{wl,t|t} && \text{(3.21a)}\\
s.t.\ & y_{t+k|t} \in \mathbb{Y}, \forall k = 1, 2, \ldots, N-1 && \text{(3.21b)}\\
& u_{t+k|t} \in \mathbb{U}, \forall k = 0, 1, \ldots, N-1 && \text{(3.21c)}\\
& y_{t+N|t} \in \mathbb{Y}_f(t) && \text{(3.21d)}\\
& [u'_{t|t}, \ldots, u'_{t+(N-1)|t}]' = B \otimes I_m[\hat{u}'_{t|t}, \ldots, \hat{u}'_{t+M-1|t}] && \text{(3.21e)}\\
& x_{t+k+1|t} = f_{wl}(x_{t+k|t}, u_{t+k|t}, \Phi_{t+k|t}, k) && \text{(3.21f)}\\
& \forall k = 0, 1, \ldots, N-1,\\
& y_{t+k|t} = g_{wl}(x_{t+k|t}, u_{t+(k-1)|t}, \Phi_{t+k|t}, k)\\
& \forall k = 1, 2, \ldots, N && \text{(3.21g)}
\end{aligned}$$

where $\mathbb{Y}_f(t)$ is the terminal constraint set; $\text{Bill}_{wl}(\cdot)$ is the electricity energy bill defined in (3.20).

In (3.21) $x_{t+k|t}$ denotes the state vector at time $t + k\Delta t$ predicted at time t obtained by starting from the current state $x_{t|t} = x(t)$ and applying the input sequence $u_{t|t}, \ldots, u_{t+N-1|t}$ to the system model (3.21f).

Let $U^\star_{t\to t+N-1|t} = \{u^\star_{t|t}, \ldots, u^\star_{t+N-1|t}\}$ be the optimal solution of problem (3.21) at time t, and $J^\star_t(x(t))$ the corresponding value function. Then, the first element of $U^\star_{t\to t+N-1|t}$ is

implemented to the system (3.19): $u(t) = u^{\star}_{t|t}$. The optimization problem (3.21) is repeated at $t + \Delta t$, with the updated state $x_{t+\Delta t|t+\Delta t} = x(t + \Delta t)$, yielding a *moving or receding horizon control* strategy. The proposed MPC controller uses a move blocking strategy (3.21e) to reduce the computational time required for its real time implementation. Details are discussed in the following section.

Move Blocking Strategy

The prediction horizon of the proposed MPC controller is 24 hours, and the control sampling time is one hour. As a result, there would be total 72 optimization variables as the control input dimension is 3. It is common practice to reduce the degrees of freedom by fixing the input or its derivatives to be constant over several time steps [110]. In this thesis, we are using the moving window blocking approach proposed in [21]. For the sake of completeness of the thesis, the algorithm to compute the blocking matrix B in (3.21e) is briefly outlined. We first need the following definitions before providing the algorithm used.

Definition 1 (Admissible Blocking Matrix) *A matrix $B \in \{0,1\}^{N\times M}$ is an admissible blocking matrix if $M < N$, and one entry in each row of B is equal to 1, the elements of the matrix are arranged in an "upper staircase" form, i.e. if the column in which a 1 occurs in the i'th row is*

$$j^{\star}(i) := \{j | B_{i,j} = 1\},$$

then $j^{\star}(i+1) \geq j^{\star}(i)$ for all $i \in \{1, 2, \cdots, N-1\}$. $B_{i,j}$ denotes the element of i'th row and j'th column of matrix B.

Definition 2 (Blocking Length Vector) *Given an admissible blocking matrix B, the blocking length vector $L(B)$ is defined as the columnwise summation of the matrix B. An admissible block vector corresponds to a unique blocking length matrix.*

The following algorithm is in the proposed MPC.

Algorithm 1 (Moving window blocking)

Initial: *Select an initial blocking length matrix L_0, and let $i = 0$.*

Step 1: *if $L_i(1) > 1$,*

$$\begin{aligned} L_{i+1} &:= L_i; \\ L_{i+1}(1) &:= L_i(1) - 1; \\ L_{i+1}(end) &:= L_i(end) + 1. \end{aligned}$$

if $L_i(1) = 1$,

$$L_{i+1} := [L_i(2 : end), L_i(1), 0].$$

Step 2: *if $L_i = L_0$, stop. Otherwise, go to next step.*

Step 3: *let $i := i + 1$, and go to step 1.*

Where $L_i(end)$ is the last element of L_i.
By following Algorithm 1, we can get a series of blocking length matrices for each step. In this thesis, we choose $L_0 = [2,\ 2,\ 18,\ 1,\ 1,\ 0]$, and Algorithm 1 will give

$$\begin{aligned} L_1 &= [1,\ 2,\ 18,\ 1,\ 1,\ 1], \\ L_2 &= [2,\ 18,\ 1,\ 1,\ 2,\ 0], \\ L_3 &= [1,\ 18,\ 1,\ 1,\ 2,\ 1], \\ L_4 &= [18,\ 1,\ 1,\ 2,\ 2,\ 0], \\ &\vdots \\ L_{24} &= [2,\ 2,\ 18,\ 1,\ 1,\ 0]. \end{aligned}$$

More details can be found in [21].

Terminal Constraints

It is well known that stability and feasibility are not ensured by the MPC law without terminal cost or terminal constraints. Usually the problem is augmented with a terminal cost and a terminal constraint set $\mathbb{Y}_f$. Typically $\mathbb{Y}_f$ is a robust control invariant set which guarantees that if Problem (3.21) is feasible for a given x_0, then it is always feasible for $t \geq 0$.

A formal definition of robust control invariant sets follows [98, 8].

Definition 3 (Robust Control Invariant Set) *A set $\mathcal{C} \subseteq \mathbb{X}$ is said to be a robust control invariant set for system (3.19) if*

$$\forall x(k) \in \mathcal{C},\ \exists u(k) \in \mathbb{U} |\ f(x(k), u(k), d(k)) \subseteq \mathcal{C} \quad \forall d(k) \in \mathbb{D}.$$

The set $\mathcal{C}_\infty$ is said to be the maximal *robust control invariant if it is robust control invariant and contains all robust control invariant set contained in $\mathbb{X}$. Where $f(x(k), u(k), d(k))$ is defined in Equation (3.19).*

We use historical data of $T_{cmp,s}$, $T_{cmp,r}$ and $\dot{m}_{cmp}$ in order to compute the possible range of $\dot{Q}_{Load}$. Figure 3.8 plots historical daily campus load during September 2008, and we observe that the load has a period of one day. It is reasonable to model the admissible campus load as a periodic disturbance with periodic envelope constraints (the bounds are represented with thicker lines in Figure 3.8).

Since the disturbance is periodic, the idea proposed by F. Blanchini and W. Ukovich in [13] can be applied to the proposed MPC controller. The invariant sets, if it exists, will be time variant and periodic with the same period as the disturbances. In order to guarantee that the tank has enough cold water to satisfy the demand, we use the algorithm proposed in [13] to calculate the CPI (Controlled Periodic Invariant) set for the system described in (2.1a). The system for calculating the CPI set is a simple buffer plant subject to constraints in Section 3.4 and periodic disturbance modeled in Figure 3.8.

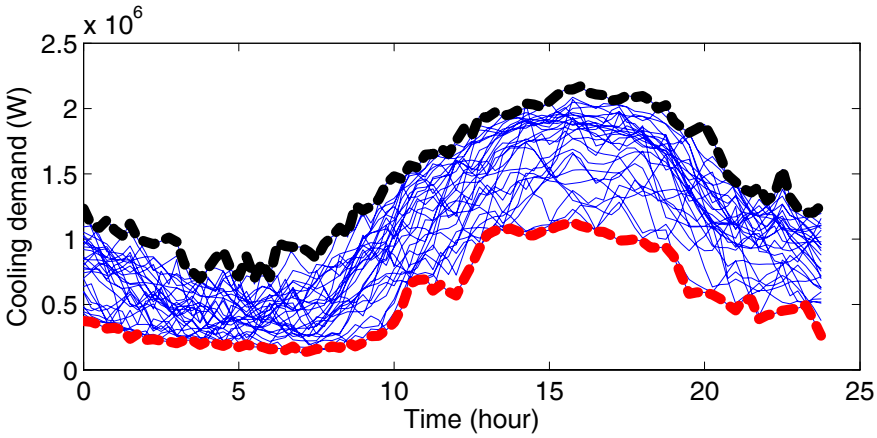


Figure 3.8: Campus load.

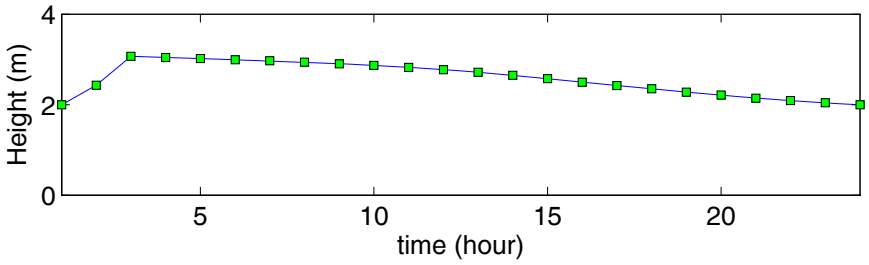


Figure 3.9: Lower bound $b(t)$ of the controlled periodic invariant set $\mathbb{Y}_f(t)$ in (3.22).

We implemented the algorithm proposed in [13] and Figure 3.9 plots the lower bound $b(t)$ of the computed periodic set $\mathbb{Y}_f(t)$:

$$\mathbb{Y}_f(t) = \{z_b(t) : b(t) \leq z_b(t) \leq z_{tank}\}. \tag{3.22}$$

If the height of the cool water in the tank is greater than the lower bounds $b(t)$, there exists a feedback control law that will satisfy any disturbance belonging to the envelope in Figure 3.8 without violating the states and inputs constraint.

Experimental Setup

The MPC controller computes the set points for cooling towers, chillers and the thermal storage tank at the central plant. Because of lower level control loops, the closed loop system indirectly affects all the components of the campus including the pumps and fan coils of the distribution system.

The MPC algorithm is implemented in Matlab® and running in real-time on a Pentium 4 Intel® processor. The average computational time for solving an optimization problem

was 20 minutes which ensured real-time implementation with the chosen one hour sampling time. The MPC algorithm receives and sends data to the campus through the Automated Logics Web Control (ALC) system. ALC is a building automation system, offering a user interface and some control features. ALC enables one to access all building management functions including (1) set and change schedules; (2) adjust set points and other control parameters; (3) trend building conditions; (4) view and acknowledge alarms and events; (5) run preconfigured and customized reports on energy usage, occupant overrides, tenant billing.

The implementation of the experiment involves three steps:

1. Download data from ALC system and pre-processing the data to get the information required by MPC;
2. Run MPC to get the control set points;
3. Update the control set points onto the ALC system, thus controlling the plant.

Experiment

Four scenarios have been studied in order to evaluate the performance of the MPC controller:

[S1] Scenario 1 is the baseline performance. The plant is operated manually by using the policy defined by the plant managers. There is no optimal control algorithm involved. Rather, the control policy is based on the operators' experience. The data for experiment S1 are collected from May 27 to May 31, 2009.

[S2] Scenario 2 implements the MPC control (3.21) with the additional constraint that start time and stop time (t_s and t_f) can only be multiple of the sampling time (1 hour) [94]. The data for experiment S2 are collected from June 2 to June 6 2009.

[S3] In Scenario 3 the plant is operated manually by using a modified policy defined by the plant managers. The modifications are extracted by observing the policy used by the MPC controller in S2. The data for experiment S3 are collected from June 8 to June 12, 2009.

[S4] Scenario 4 implements the MPC controller (3.21). The data for experiment S4 are collected from October 6 to October 10, 2009.

In all four scenarios, the quantity of chilled water stored in the tank at the end of the experiment is forced to be equal to the one available at the beginning of the experiment. Despite the difference in time, the weather conditions during experiments S1 and S4 are similar. This allows us to fairly compare the MPC performance to the one obtained with the baseline control logic.

Comparison Metrics

Two comparison metrics are defined to evaluate the performance of MPC: the electricity bills and the coefficient of performance.

Electricity Bill The electricity bill function is described by Equation (3.20), and by comparing the electricity bill we can quantify the cost savings generated by the MPC controller.

Coefficient of Performance The Coefficient of Performance (COP)

$$COP = E^{Thermal}_{Generated}/E^{Electrical}_{plant} \tag{3.23}$$

captures the efficiency of the central plant, i.e. the amount of thermal energy (J) generated by the central plant with 1 J of electricity power. $E^{Electrical}_{plant}$ is the electricity power consumed by the central plant, and $E^{Thermal}_{Generated}$ is the thermal energy generated by the central plant defined as

$$E^{Thermal}_{Generated} = \int_{t_0}^{t_f} \dot{m}_{CHWS}(T_{CHWR} - T_{CHWS})dt \tag{3.24}$$

By comparing the COP between the 3 scenarios S1, S2 and S3, we can better understand if MPC improves the efficiency of the central plant.

Discussion of Experimental Results

Next we compare the four experiments S1, S2, S3 and S4 by analyzing the performance of the central plant and the corresponding control profiles.

Performance Comparison The performance of the central plant will be compared by using the metrics defined in Section 3.4.

Table 3.3: Central plant performance comparison (all quantities correspond to daily average).

	S1	S2	S3	S4
Energy Consumption [KJ]	8.63e6	4.25e6	4.40e6	3.58e6
Energy Generated [KJ]	4.05e7	2.01e7	2.31e7	2.01e7
COP	4.70	4.77	5.26	5.60
Bill [dollar]	1.68e3	4.18e2	4.75e2	4.00e2

Table 3.3 lists the electrical energy consumption, thermal energy generated, COP and the electricity bill for experiments S1, S2, S3, and S4. We can observe that

- Comparing S1 with S2. The MPC controller has significantly reduced the daily electricity bill in experiment S2 by \$1265 compared to experiment S1. Meantime, the efficiency of central plant, COP, is also improved by 1.5%.

- Comparing S3 with S1. The electricity bill reduction is \$1205 and COP is increased by 11.9%.
- Comparing S4 with S3 and S1. The COP of the central plant reaches 5.60 in experiment S4, increased by 19.1% over baseline (S1). The daily electricity bill is reduced by \$75 when compared to S3 and by \$1280 when compared to S1.

The performance improvement is further discussed by looking at the implemented control profiles in the rest of the section.

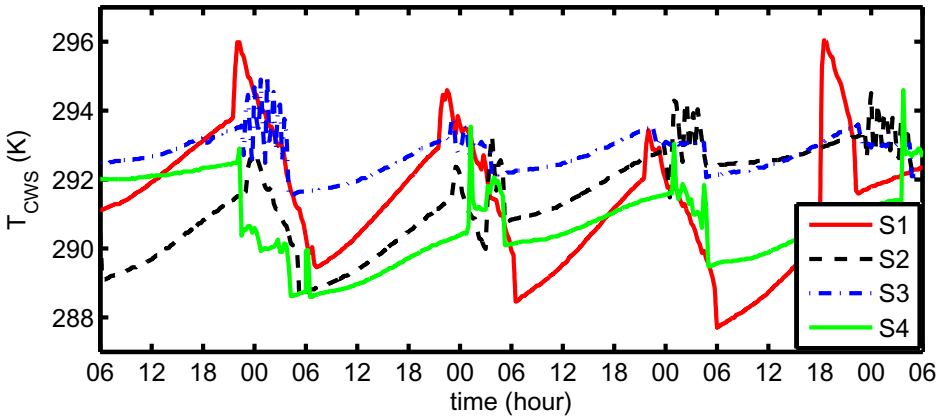


Figure 3.10: Control input set points $T_{CWS,ref}$.

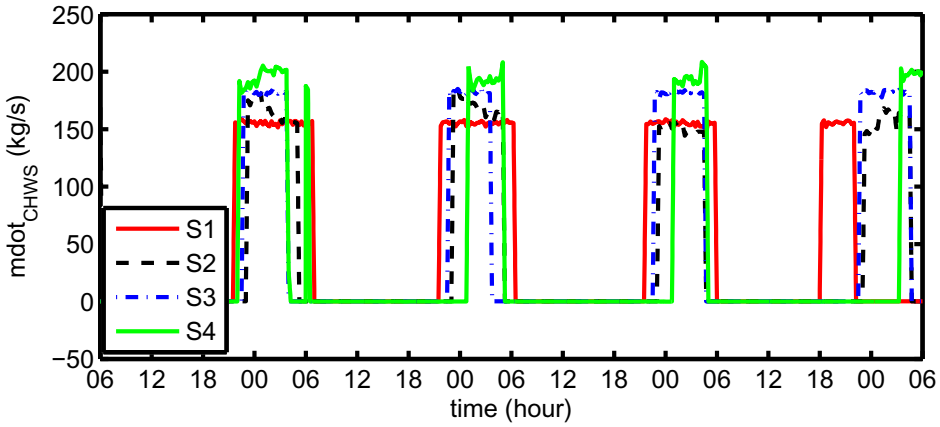


Figure 3.11: Control input set points $\dot{m}_{CHWS,ref}$.

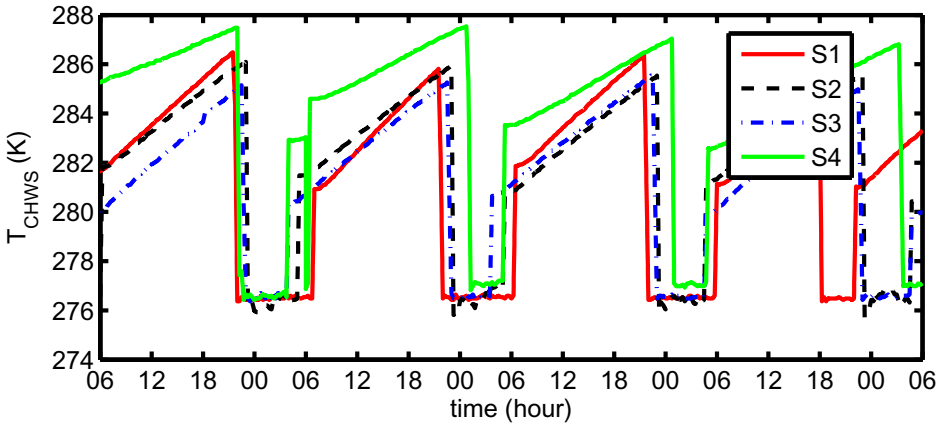


Figure 3.12: Control input set points $T_{CHWS,ref}$.

Table 3.4: Average values of central plant flows and temperatures during charging time.

	S1	S2	S3	S4
T_{CWS} [K]	289.0	292.3	293.2	290.72
$\dot{m}_{CHWS}$ [Kg/s]	152.6	158.2	178.0	187.0
T_{CHWS} [K]	276.7	276.4	276.9	278.6

Control Profile Figure 3.10-3.12 shows the control profiles for experiments S1, S2, S3 and S4 respectively. Table 3.4 lists the average values of the control set points during the charging time. Based on these information the following remarks can be drawn:

- The baseline control logics in S1 works as follow: condensed water supply temperature ($T_{CWS,ref}$) is set as low as possible so that the cooling towers always work at full load, chilled water supply temperature set point $T_{CHWS,ref}$ is fixed to $276.5K$, and the average mass flow rate $\dot{m}_{CHWS,ref}$ is set to $150kg/s$. The operation schedule starts at 10 pm and ends when the tank is fully charged.

- The MPC controller in S2 applies higher condensed water supply temperature (T_{CWS}) for cooling towers than experiment S1. In the baseline control (experiment S1), the operators set the T_{CWS} as low as possible. This overloads the cooling towers, and a higher T_{CWS} can help balance the tradeoff between power consumed by the chilling system while meeting cooling loads.

- The MPC in S4, applies a desired condensed water supply temperature T_{CWS} of 293.1 K. However due to a lower level controller malfunctioning, T_{CWS} did not track its reference but was as low as $290.72K$ in the first three days.

- During experiments S2 and S4, the central plant is working with shorter charging windows, and the average mass flow rate $\dot{m}_{CHWS}$ is greater than the one used by the operators in S1.
- The set points of chilled water supply temperature $T_{CHWS,ref}$ for S1, S2, S3 and S4 are reported in Figure 3.12, and for S1, S2 and S3 scenarios, there is no noticeable difference.

We notice that experiment S3 improves COP over experiment S2 (with MPC in the loop). The reason is that the MPC in S2 assumes that start time and stop time (t_s and t_f in (3.21)) can only be multiple of the sampling time (one hour). Because of such coarse sampling time, a constant and high mass flow rate would overcharge the tank. As it can be observed in Figure 3.11, the mass flow rate ($\dot{m}_{CHWS,ref}$) in experiment S2 is high only at the beginning of the charging window. Then, it decreases in order to satisfy the load demand. Since for the specific scenario and chillers performance curves, a high COP is always obtained for higher mass flow rates ($\dot{m}_{CHWS,ref}$), the decrease in $\dot{m}_{CHWS,ref}$ erodes the efficiency of the central plant. This problem is fixed in experiment S4 where chillers start time and stop time (t_s and t_f) are allowed to assume any continuous value in the optimization problem (3.21). As a result, in scenario S4 a high flow $\dot{m}_{CHWS,ref}$ is maintained over the charging period (Figure 3.11).

After experiment S2 the operators observed the behavior of the MPC and decided to apply maximum chilled water supply mass flow rate and set the condensed water supply temperature around 293.7K. These two modification are used in scenario S3. As observed from Table 3.4, the performance of the central plant, in terms of COP, is improved by 11.9% compared to their original baseline control S1.

Weather Dependence

The MPC performance is affected by the weather patterns. In order to better understand the potential improvement under a variety of weather conditions, an extensive simulation study over six months was performed. The proposed MPC in Section 3.4 was simulated in closed loop with the tank dynamic model in Section 2.3. The campus load is estimated by using model presented in Section 2.3.

We performed extensive simulations from December 1, 2008 to July 1, 2009 by using the weather conditions at UC Merced. Figure 3.14 shows that the simulations cover a daily average ambient temperature from $278K$ in winter to $300K$ in summer. We note that under such a wide range of weather conditions, the COP with the MPC proposed in Section 3.4 constantly outperforms the COP of the baseline control (see Figure 3.13). The missing points in Figure 3.13-3.14 corresponds to missing data in the (corrupted) weather database.

Figure 3.15 plots the correlation between the absolute COP improvement over baseline and the average ambient temperature during the charging time. The dashed line shows the upper bound of the COP improvement and the solid line is the lower bound. The MPC controller can achieve better COP improvement with average ambient temperature ranging

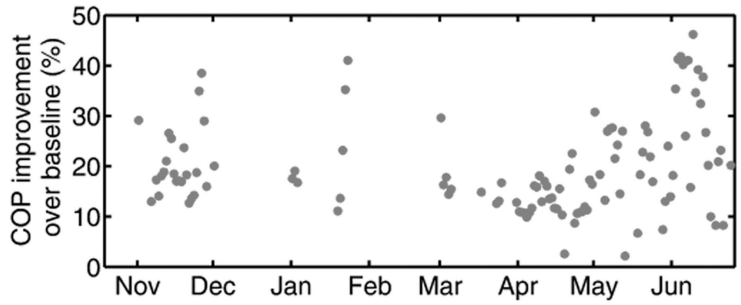


Figure 3.13: COP improvement of MPC over baseline.

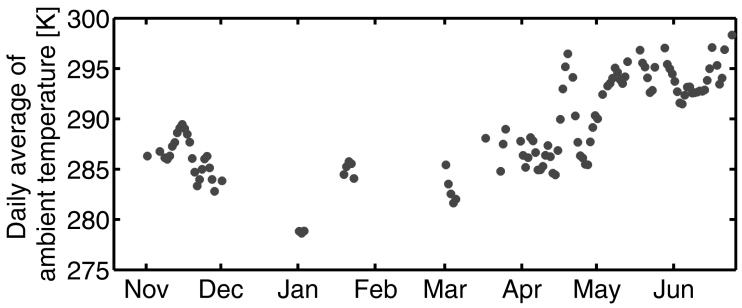


Figure 3.14: Daily average of ambient temperature.

from $285K$ to $291K$. This can be explained as follows. Low ambient temperatures limits the achievable condensed water temperature (T_{CWS}) for cooling towers and, as pointed out in Section 3.4, higher condensed water temperature provides higher COP. On the other hand, higher ambient temperatures reduce the maximum COP achievable $\eta T_{CHWS}/(T_{amb}-T_{CHWS})$, where η is the efficiency of the system, and $T_{CHWS}/(T_{amb}-T_{CHWS})$ is the COP of an ideal Carnot compression refrigeration cycle [2].

3.5 Predictive Control for Air-Loop System

The objective of LMPC is to minimize the energy consumption in the form of cold water, hot water, and electricity while maintaining the thermal zones within the comfort range.

Control Variables

The control variables of LMPC are listed as follow,

1. $\dot{m}^i_{s,ref}(t)$: Reference for supply air mass flow rate to ith zone.

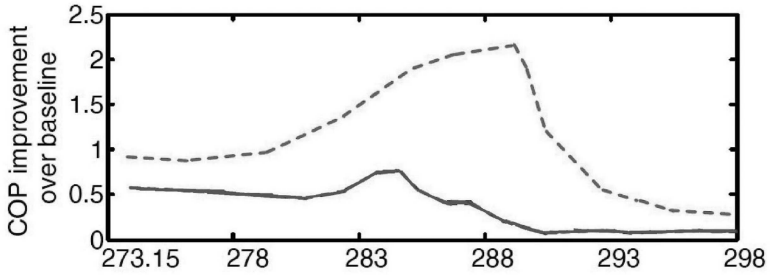


Figure 3.15: Max (dash line) and min (solid line) COP improvement as a function of average ambient temperature (K) during charging time.

2. $\Delta T_{c,ref}(t)$: Reference for temperature difference of supply air through the cooling coil in AHU.

3. $\Delta T^i_{h,ref}(t)$: Reference temperature difference of supply air through the reheating coil in VAV box connected to ith zone.

4. δ: The mixing proportion of return air and outside air.

Measured Variables

The following variables are measured to initialize prediction of the HMPC controller (3.18),

1. T^i: Average air temperature of ith zone.

The following variables are load predictions.

1. T_{oa}: Outside air temperature.

2. I^i: Solar radiation intensity of ith zone.

3. P^i_d: Internal thermal load for ith zone.

Operation Constraints

The following constraints avoid the malfunction of the system components and guarantee thermal comfort of occupants.

1. $T^i \in [\underline{T}^i, \overline{T}^i]$: The thermal comfort is defined as a box constraints on perceivable zone air temperatures.

2. $\dot{m}^i_{s,ref} \in [\underline{m}^i_s, \overline{m}^i_s]$: The supply air flow rate to each zone is limited by the capacity of VAV boxes.

3. $\sum_i \dot{m}^i_{s,ref} \in [\underline{m}^i_s, \overline{m}^i_s]$: The total supply air flow rate delivered by the fan in AHU is constrained by the fan size.

4. $\Delta T_{c,ref} \in [0, \overline{\Delta T}_c]$: The supply air temperature difference through the cooing coil is upper bounded due to the capacity of cooling coil in AHU.

5. $\Delta T^i_{h,ref} \in [0, \overline{\Delta T}^i_h]K$: The supply air temperature difference through the reheating coil is upper bounded due to the capacity of reheating coil in VAV box.

6. $\delta \in [0, \overline{\delta}]$. The mixing ratio of return air and outside air is upper bounded by $\overline{\delta} < 1$ to guarantee minimum ventilation requirement.

Model Summary

The main dynamics of the air-loop system are thermal zones, and in this section the RC model (2.27) is implemented. The dynamic system is discretied with sampling time (Δt) of 15 minutes:

$$x(t + \Delta t) = f_{al}(x(t), u(t), w(t), t), \tag{3.25a}$$
$$y(t) = g_{al}(x(t)), \tag{3.25b}$$

where

$$\begin{aligned} u &= [\dot{m}^{i\in\mathcal{V}}_{s,ref}; \Delta T^{i\in\mathcal{V}}_{h,ref}; \Delta T_{c,ref}; \delta] \in \mathbb{U}, \\ x &= [T_1^{i\in\mathcal{V}}; T_2^{i\in\mathcal{V}}], \\ w &= [T_{oa}; I^{i\in\mathcal{V}}; P_d^{i\in\mathcal{V}}], \\ y &= [T_1^{i\in\mathcal{V}}] \in \mathbb{Y}. \end{aligned}$$

$\mathbb{U}$ and $\mathbb{Y}$ are the feasible control input set and feasible output set defined in Section 3.5, respectively.

Energy Price

The energy price for air-loop systems is defined as Table 3.2, and the total bill is computed as

$$\text{Bill}_{al}(t) = \sum_{k=0}^{k=N} C_e(t + k\Delta t)\text{Power}_{al}(t + k\Delta t)\Delta t, \tag{3.26}$$

where Power_{al} is the function adding up the power consumption of fan (2.29) and cooling and heating coils (2.30).

LMPC Problem Formulation

This section presents the design of a MPC controller whose objective is to find the optimal control sequence that satisfies the required occupants' comfort and minimizes electricity usage. Consider the following optimization problem:

$$J^{\star}(x(t),t) = \min_{\hat{u}_{t|t},\ldots,\hat{u}_{t+N-1|t},t_s,t_f} \text{Bill}_{al,t|t} \tag{3.27a}$$

$$s.t.\ y_{t+k|t} \in \mathbb{Y}, \forall k = 1,2,\ldots,N \tag{3.27b}$$

$$u_{t+k|t} \in \mathbb{U}, \forall k = 0,1,\ldots,N-1 \tag{3.27c}$$

$$x_{t+k+1|t} = f_{al}(x_{t+k|t}, u_{t+k|t}, w_{t+k|t}, k) \tag{3.27d}$$
$$\forall k = 0,1,\ldots,N-1,$$
$$y_{t+k|t} = g_{al}(x_{t+k|t})$$
$$\forall k = 1,2,\ldots,N \tag{3.27e}$$

where $\text{Bill}_{al}(\cdot)$ is the electricity energy bill defined in (3.26). The LMPC optimization problem (3.27) is solved by Ipopt, a software package for large-scale nonlinear optimization developed by [135].

Simulation Results

In this section[1], we construct a simple 5-zone building model with input thermal loads as shown in Figure 3.16 to demonstrate LMPC (3.27). The heat transfer between the five zones is neglected. The first four zones have equal and negative load that requires heating except briefly in the afternoon. Zone 5 has high positive load that requires cooling during occupied hours, with a small negative load in unoccupied hours. The nominal LMPC results in Figure 3.17 and 3.18 show the tradeoff between supply temperature and mass flow rate. Between 6:30 and 10:00, we can see economizer and temperature reset-like operation where cooling of zone 5 is performed using outside air, warmer supply temperatures, and high mass flow rates.

This control scheme saves energy because the rest of the zones are in heating mode during this period, and any cooler supply temperature would require reheat to keep those zone temperatures above their lower bounds. Once all of the zones are in cooling mode, controlling the cooling coil to the minimum feasible supply temperature and using lower flow rates becomes a more efficient strategy. We see a brief supply temperature reset behavior again near the end of the occupied hours at 18:00. Anticipating less cooling demand for the unoccupied period, the LMPC controller starts increasing the supply temperature early.

[1]This section is extracted from Anthony Kelman's contribution to the paper titled "Model Predictive Control for Energy Efficient Buildings with Thermal Storage: Modeling, Simulation, and Experiments" in IEEE Control System Magazine, vol. 32, no. 1, pp. 44-64, February 2012.

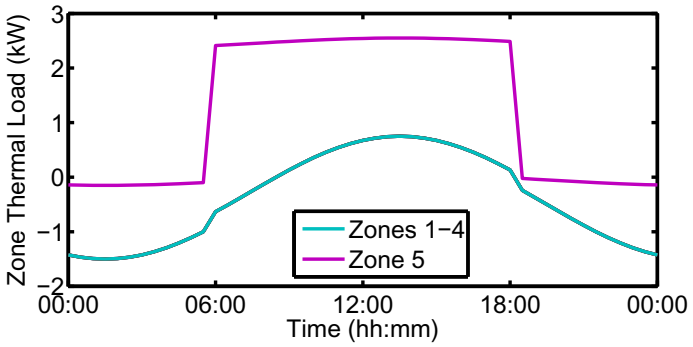


Figure 3.16: Zone thermal loads P_d^i.

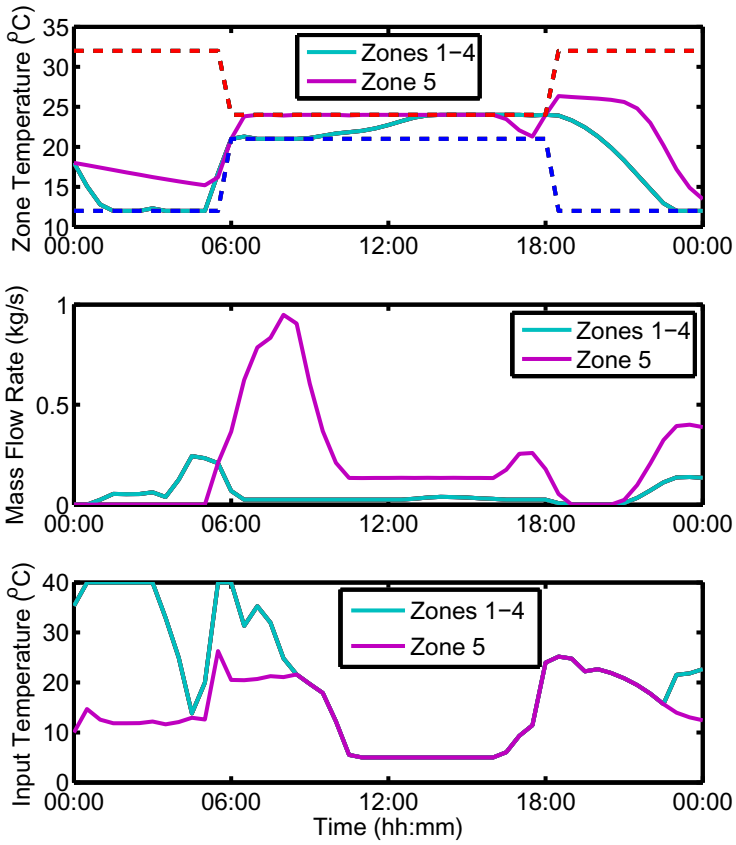


Figure 3.17: Nominal case zone results.

To compare against the nominal case LMPC results in Figure 3.17 and Figure 3.18, we repeat the calculations with modified versions of the bill function. First, in Figure 3.19, we

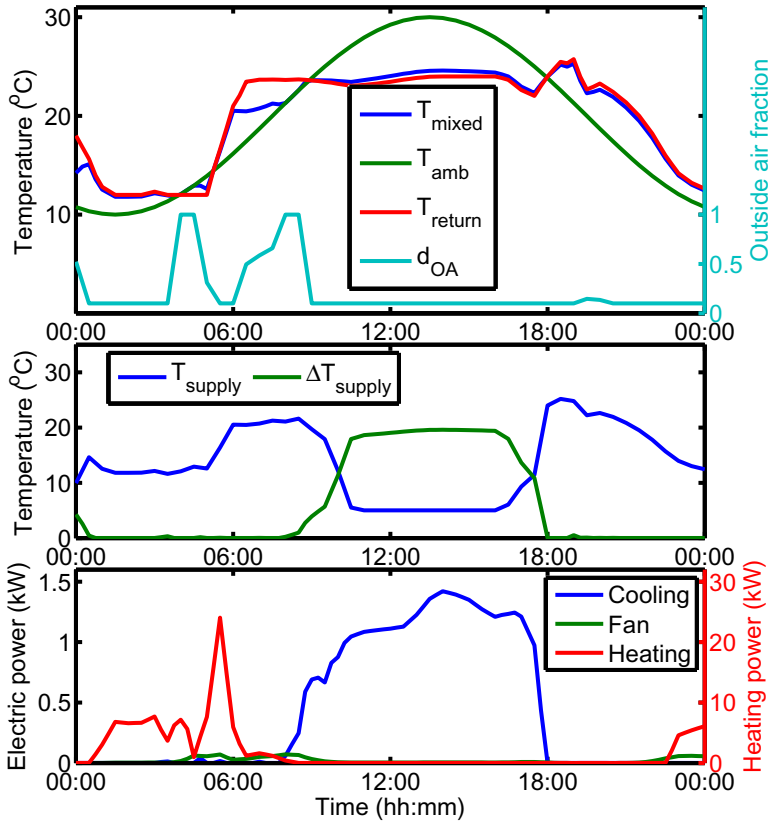


Figure 3.18: Nominal case AHU results.

modify the electric utility rate $C_e(k)$ to have a higher value between 12:00 and 16:30. In Figure 3.20 the utility rates are constant throughout, but we add an additional penalty term to the cost function to minimize the peak electric power over the entire day. The modified bill function is

$$\text{Bill}_{al}^{\text{mod}}(t) = \left(\sum_{k=0}^{k=N} C_e(t + k\Delta t)\text{Power}_{al}(t + k\Delta t)\Delta t \right) + \varphi \max_{k=0}^{N} \text{Power}_{al}(t + k\Delta t), \qquad (3.28)$$

where φ is a penalty weighting factor in dollars per unit power.

Both these modified cases in Figures 3.19 and 3.20 demonstrate precooling of zone 5 and lengthened cooling of zones 1-4, but with different timing and intent. In Figure 3.19, the peak electric power is not penalized in the cost function but the electric utility rate has a higher value between 12:00 and 16:30. As a consequence, precooling is only performed immediately before noon, with a corresponding spike in cooling power, so that less cooling energy is used between 12:00 and 16:30. In Figure 3.20, the peak electric power is included

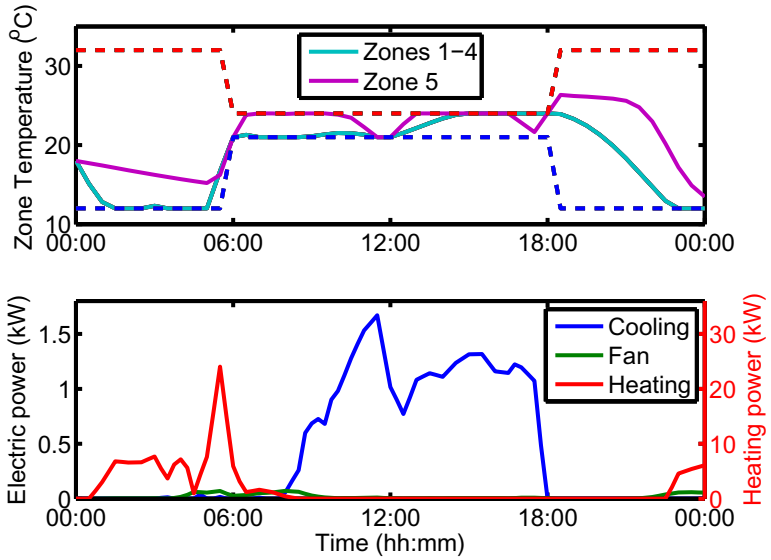


Figure 3.19: Variable utility rate case. Note the precooling and spike in cooling power immediately before noon.

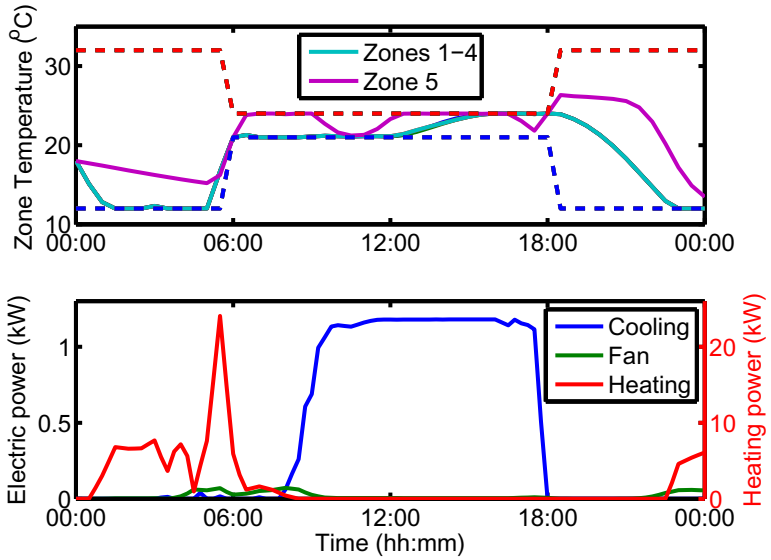


Figure 3.20: Peak power limiting case. Note the timing of the precooling and the intentional plateau in cooling power.

in the cost function so zone 5 is precooled beginning earlier in the morning. As a result, cooling power is increased at a time when it would normally be low, shifting electric power use away from the times it would normally be at maximum.

3.6 Summary

In this chapter, a simple thermal mass model has been used to show the basic mechanism of active thermal storage and how this mechanism naturally emerges in a predictive control scheme. The model is also used to demonstrate a fundamental tradeoff involving savings, losses, and uncertainty in load shifting. We also provided the main ingredients of a predictive control framework implementable in a building equipped with thermal storage. MPC formulations are developed for both air-loop and water-loop systems in building cooling and heating systems. Simulations and experimental results have shown the effectiveness of the proposed control scheme. However, the appealing advantages of MPC shown though simulations and experiments in the previous sections do not come without a price. Several issues have to be considered while designing and implementing MPC for buildings. In the remainder of this section we conclude with some design and implementation considerations.

Design Considerations

Stability and Feasibility

Stability and feasibility of MPC are well-studied issues [98]. In particular, it has been shown that stability and feasibility are not ensured by the MPC law without terminal cost and terminal constraints[98]. Typically the terminal constaint is a robust control invariant set so that the persistent feasibility of the MPC strategy is guaranteed. Persistent feasibility ensures that if the MPC (3.12)–(3.18) is feasible for a given initial state $x(0)$, then it is feasible for all $t \geq 0$. Definitions and properties of invariant sets can be found in [12, 98]. In the specific context of the MPC considered in this thesis, the terminal set ensures that enough energy is actively stored in thermal storage elements to counteract a bounded unpredicted change in demand. A treatment of sufficient conditions guaranteeing persistent feasibility of MPC problems has been demonstrated for the specific case of the UC Merced study in Section 3.4.

Prediction Uncertainty

The example in the first section of this chapter showed benefits of MPC under the assumption that MPC has perfect knowledge of predicted disturbances and system dynamics. This section tries to highlight potential issues associated with this assumption. We focus on total energy consumption using the simple MPC problem (3.3)–(3.6) with $\kappa = 0$. The control design assumes that weather prediction in Figure 3.1 is perfect and occupancy load prediction in Figure 3.2 is perfect. This time we assume that in reality the occupancy load differs from what predicted. Two scenarios are considered. In scenario S1 the future occupancy load is exactly the same as predicted in Figure 3.2 with probability $P(S1)$ equal to α. In scenario S2 the occupancy load is zero over the entire day with probability $P(S2)$ equal to $1 - \alpha$. In short, the controller is designed based on S1 but the probability of S1 happening is α.

The expected value of the control input cost $\mathbb{E}[J^u]$ and constraint violation $\mathbb{E}[J^\varepsilon]$ for MPC C2 and proportional controller C1 are computed in closed-loop. The closed-loop simulations use different occupancy load profiles depending on the chosen probability α.

Simulation results for various values of α and various tunings for MPC C2 and proportional controller C1 are summarized in Figure 3.21. When the prediction is perfect with $\alpha = 1$, the performance of MPC C2 is the same as the proportional controller C1 in terms of total energy consumption and constraint violation. However, the MPC performance deteriorates as α decreases. In fact, MPC fails to keep the zone temperature within the comfort constraints due to the misleading predictions. MPC consumes more energy than the proportional controller for $\alpha = 0.5$ and $\alpha = 0$ because MPC is performing precooling even if occupants do not enter the space.

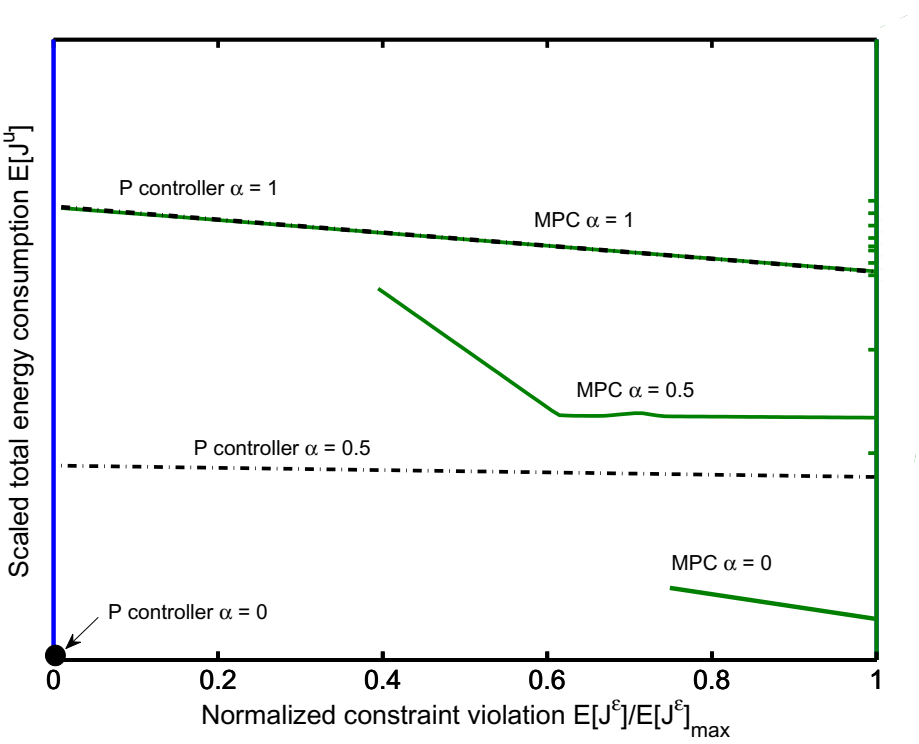


Figure 3.21: Comparison results.

Stochastic MPC [105, 31] might be a better approach to address this issue when probability distribution functions of the loads are available. In this case, we would minimize expected costs and satisfy constraints with a given probability. Further investigations on prediction uncertainty are presented in Chapter 5.

Implementation Considerations

Convergence to Suboptimal Solutions

The product between air temperatures and mass flow rates in the thermodynamic energy balance equations (2.27) leads to a non-convex MPC problem which might have distinct locally optimal solutions. Fast computational techniques for solving non-convex optimization such as sequential quadratic programming (SQP) can only provide certificates of local optimality. These locally optimal solutions might be less efficient than those obtained with simpler control design. We are currently analyzing different types of local optima and their physical interpretation. The analysis can be used to derive branch and bound rules which allow a SQP solver to converge to globally optimal control sequences.

Computational Complexity of Model Predictive Control

As the complexity of the building model increases, centralized MPC might become computationally intractable due to the limited computational resources available on current building control platforms. This limitation is critical at the low-level of the control architecture where distributed inexpensive computing platforms are common.

One approach to address the computational complexity of MPC is to precompute the control action for a set of initial states and external parameters. A lookup table can be generated by gridding the space of parameters and states and solving the optimization problem offline for each grid point. For linear and switched linear systems the gridding can be avoided using multiparametric optimization [14]. In the specific context of MPC for buildings the authors of [97] present a rule extraction approach. Rather than running an online MPC in real time, many simulations of the MPC are executed offline. The simulation results are then used to generate simplified rule-based controllers as functions of operating conditions.

Alternatively, the limitation might be overcome by efficient numerical solvers tailored to the specific hardware or with the use of distributed model predictive control [114, 90]. In distributed MPC, the centralized problem is decomposed into a set of smaller problems which can be associated with different subsystems such as VAV boxes and AHUs. Each subsystem solves local small MPC problems with information from local and neighboring subsystems. The local MPC control modules communicate with each other to converge to an optimal solution [90]. More details will be presented in Chapter 4.

Equipment Retrofitting

Model predictive control requires sensor data from a building in order to initialize simulations and make predictions. Additionally, there must be some way to communicate the computed optimal control inputs either to lower level controllers (for setpoint tracking) or directly to the control actuators. Modern digital building automation systems satisfy these requirements, but are only present in new buildings. In order to apply MPC to the existing stock of older buildings, HVAC equipment must be retrofitted for digital control and additional sensors

need to be added or existing sensors replaced with digital versions. This can be prohibitively expensive, and must be offset by the operational energy cost savings of MPC versus the baseline control.

Chapter 4

Distributed Model Predictive Control

4.1 Introduction

The size of the centralized predictive control problem rapidly grows when a realistic number of rooms together with a meaningful control horizon are considered. Therefore the real-time implementation of an MPC scheme is a challenge for the low-cost embedded platforms currently used for HVAC control algorithms. In this chapter, we present a distributed model-based predictive control (DMPC) for regulating heating and cooling in order to minimize energy consumption while satisfying comfort constraints. We make use of the nonlinear models presented in Section 2.3 to model the thermal behavior of buildings. The techniques presented in this chapter enable the implementation of an MPC algorithm by distributing the computational load on a set of VAV box embedded controllers coordinated by the embedded controllers on the AHU system. Compared to existing DMPC schemes [16, 114, 127, 15], the proposed method is tailored to the specific class of problems considered in this thesis. In particular, it makes use of sequential quadratic programming (SQP) [104, 62], proximal minimization [9], and dual decomposition [83] to handle the system nonlinearities and the decentralization, respectively. The SQP and proximal minimization methods are used to derive a strictly convex Quadratic Program (QP) from the original nonlinear optimization problem. The dual decomposition scheme takes advantage of the separability of the dual Lagrangian QP problem. By doing so, the dual QP is solved iteratively by updating dual and primal variables in a distributed fashion. In this thesis we show that if the centralized MPC problem is properly formulated, the resulting primal and dual update laws can be easily derived. Simulation results show good performance and computational tractability of the resulting scheme.

4.2 Model Summary

The thermal dynamic model (2.27) in Chapter 2 is compactly rewritten as

$$\frac{dx^j}{dt} = f_c(x^j, u^j, u^a, T_m, w^j) + \sum_{i\in\mathcal{N}^j} E_i^j x^i,\ \forall j \in \mathcal{V}, \tag{4.1a}$$

$$g_c(x^{j\in\mathcal{V}}, u^{j\in\mathcal{V}}, u^a, w^{j\in\mathcal{V}}, T_r, T_m) = 0, \tag{4.1b}$$

where $x^j = (T_1^j,\ T_2^j)$ is the state of the j-th room, $u^j = (\dot{m}_s^j,\ \Delta T_h^j)$ are the control inputs to the j-th VAV box, and $u^a = (\delta,\ \Delta T_c)$ collects the AHU control inputs. The vector $w^j = (P_d^j,\ T_{oa})$ is the disturbance assumed to be perfectly known. Equation (4.1b) lumps up the algebraic equations that describe the static model for return air temperature and mixed air temperature, respectively. The set $x^{j\in\mathcal{V}}$ is defined as $\{x^1, x^2, \ldots, x^{\mathcal{N}_v}\}$. Note that the room dynamics in the network are coupled through states (the second term in (4.1a)) and inputs (δ and ΔT_c are common to all rooms).

The continuous time system (4.1) is discretized as follow. We consider a control law with uniform sampling time Δt,

$$u^j(t) = u_k^j,\ \forall j \in \mathcal{V}, \tag{4.2a}$$

$$u^a(t) = u_k^a, \tag{4.2b}$$

$$t \in [k\Delta t,\ (k+1)\Delta t).$$

With the control law defined by (4.2), the system equations (4.1) are discretized over $t \in [k\Delta t,\ (k+1)\Delta t)$ using the trapezoidal method to obtain:

$$\frac{x_{k+1}^j - x_k^j}{\Delta t} = \frac{1}{2} f_c(x_k^j, u_k^j, u_k^a, T_{mk}, w_k^j) + \frac{1}{2} f_c(x_{k+1}^j, u_k^j, u_k^a, T_{mk}^+, w_{k+1}^j) + \sum_{i\in\mathcal{N}^j} E_i^j \frac{x_k^i + x_{k+1}^i}{2},\ \forall j \in \mathcal{V}, \tag{4.3a}$$

$$g_c(x_k^{j\in\mathcal{V}}, u_k^{j\in\mathcal{V}}, u_k^a, w_k^{j\in\mathcal{V}}, T_{rk}, T_{mk}) = 0, \tag{4.3b}$$

$$g_c(x_{k+1}^{j\in\mathcal{V}}, u_k^{j\in\mathcal{V}}, u_k^a, w_{k+1}^{j\in\mathcal{V}}, T_{r,k}^+, T_{m,k}^+) = 0, \tag{4.3c}$$

where $T_{r,k}^+$ and $T_{m,k}^+$ are the temperature of return air and mixed air when time approaches $(k+1)\Delta t$, respectively. Note that $T_{r,k}^+$ ($T_{m,k}^+$) is generally different from $T_{r,k+1}$ ($T_{m,k+1}$) due to the discontinuity introduced by the control law (4.2).

If $u_k^c = (u_k^a,\ T_{r,k},\ T_{m,k},\ T_{r,k}^+,\ T_{m,k}^+)$, then the discretized room model (4.3) can be compactly rewritten as

$$f(x_{k+1}^j, x_k^j, u_k^j, u_k^c, w_k^j, w_{k+1}^j) + \sum_{i \in \mathcal{N}^j} E_i^j \frac{x_k^i + x_{k+1}^i}{2} = 0, \ \forall j \in \mathcal{V}, \tag{4.4a}$$

$$g(x_{k+1}^{j\in\mathcal{V}}, x_k^{j\in\mathcal{V}}, u_k^{j\in\mathcal{V}}, u_k^c, w_k^{j\in\mathcal{V}}, w_{k+1}^{j\in\mathcal{V}}) = 0, \tag{4.4b}$$

$$u_k^c \in \mathcal{U}^c; \ x_k^j \in \mathcal{X}, \ u_k^j \in \mathcal{U}^j, \ \forall j \in \mathcal{V}. \tag{4.4c}$$

In the next section we will also use the following linearized version of model (4.4) around the trajectory of states and control inputs $(\bar{x}_k^1, \ldots, \bar{x}_k^{N_v}, \ \bar{u}_k^1, \ldots, \bar{u}_k^{N_v}, \ \bar{u}_k^c), \ k = 0, 1, \ldots, N-1$:

$$T_k^j dx_{k+1}^j + A_k^j dx_k^j + B_k^j du_k^j + B_k^c du_k^c + f_k^{e,j} + \sum_{i \in \mathcal{N}^j} E_i^j \frac{dx_{k+1}^i + dx_k^i}{2} = 0, \tag{4.5a}$$

$$\forall j \in \mathcal{V}, \tag{4.5b}$$

$$G_k^c du_k^c + \sum_{j \in \mathcal{V}} G_{j,k}^+ dx_{k+1}^j + \sum_{j \in \mathcal{V}} G_{j,k}^x dx_k^j + \sum_{j \in \mathcal{V}} G_{j,k}^u du_k^j + g_k^e = 0, \tag{4.5c}$$

$$\bar{u}_k^c + du_k^c \in \mathcal{U}^c, \ \bar{x}_k^j + dx_k^j \in \mathcal{X}, \ \ \bar{u}_k^j + du_k^j \in \mathcal{U}^j, \ \forall j \in \mathcal{V}, \tag{4.5d}$$

$$T_k^j = \left.\frac{\partial f}{\partial x_{k+1}^j}\right|_{\bar{x}_{k+1}^j}, \ A_k^j = \left.\frac{\partial f}{\partial x_k^j}\right|_{\bar{x}_k^j}, \ B_k^j = \left.\frac{\partial f}{\partial u_k^j}\right|_{\bar{u}_k^j}, \ B_k^c = \left.\frac{\partial f}{\partial u_k^c}\right|_{\bar{u}_k^c}, \tag{4.5e}$$

$$G_{j,k}^+ = \left.\frac{\partial g}{\partial x_{k+1}^j}\right|_{\bar{x}_{k+1}^j}, \ G_{j,k}^x = \left.\frac{\partial g}{\partial x_k^j}\right|_{\bar{x}_k^j}, \ G_{j,k}^u = \left.\frac{\partial g}{\partial u_k^j}\right|_{\bar{u}_k^j}, \ G_k^c = \left.\frac{\partial g}{\partial u_k^c}\right|_{\bar{u}_k^c}, \tag{4.5f}$$

$$f_k^{e,j} = f(\bar{x}_{k+1}^j, \bar{x}_k^j, \bar{u}_k^j, \bar{u}_k^c, w_k^j, w_{k+1}^j) + \sum_{i \in \mathcal{N}^j} E_i^j \frac{\bar{x}_{k+1}^i + \bar{x}_k^i}{2}, \ \forall j \in \mathcal{V}, \tag{4.5g}$$

$$g_k^e = g(\bar{x}_{k+1}^{j\in\mathcal{V}}, \bar{x}_k^{j\in\mathcal{V}}, \bar{u}_k^{j\in\mathcal{V}}, \bar{u}_k^c, w_k^{j\in\mathcal{V}}, w_{k+1}^{j\in\mathcal{V}}), \tag{4.5h}$$

where dx_k^j, du_k^j, and du_k^c are the deviations of states and control inputs around the trajectory. f_k^e and g_k^e are residuals of nonlinear equality constraints (4.4a) and (4.4b).

4.3 Distributed Model Predictive Control

In this section we formalize the MPC control problem and provide details on the distributed MPC (DMPC) design. We are interested in solving at each time step t the following opti-

mization problem:

$$\min_{\mathbf{U},\mathbf{X}} \quad J(\mathbf{U},\mathbf{X}) = \sum_{k=0}^{N-1} \left\{ P_c + P_{\text{fan}} + \sum_{j\in\mathcal{V}} P_h^j \right\} \Delta t \tag{4.6a}$$

subj. to:

$$f(x^j_{t+k+1|t}, x^j_{t+k|t}, u^j_{k|t}, u^c_{t+k|t}, w^j_{t+k|t}, w^j_{t+k+1|t}) + \sum_{i\in\mathcal{N}^j} E^j_i (x^i_{t+k+1|t} + x^i_{t+k|t})/2 = 0,$$

$$\forall j \in \mathcal{V},\ k = 0, 1, \ldots, N-1, \tag{4.6b}$$

$$g(x^{j\in\mathcal{V}}_{t+k+1|t}, x^{j\in\mathcal{V}}_{t+k|t}, u^{j\in\mathcal{V}}_{t+k|t}, u^c_{t+k|t}, w^{j\in\mathcal{V}}_{t+k|t}, w^{j\in\mathcal{V}}_{t+k+1|t}) = 0,$$

$$k = 0, 1, \ldots, N-1, \tag{4.6c}$$

$$x^j_{t+k|t} \in \mathcal{X}^j,\ \forall j \in \mathcal{V},\ k = 1, \ldots, N, \tag{4.6d}$$

$$u^j_{t+k|t} \in \mathcal{U}^j,\ u^c_{t+k|t} \in \mathcal{U}^c,\ \forall j \in \mathcal{V},\ k = 0, \ldots, N-1, \tag{4.6e}$$

$$x^j_{t|t} = x^j(t),\ \forall j \in \mathcal{V}, \tag{4.6f}$$

where $\mathbf{U} = (u^1_{t|t}, \ldots, u^1_{t+N-1|t}, \ldots, u^{N_v}_{t|t}, \ldots, u^{N_v}_{t+N-1|t}, u^c_{t|t}, \ldots, u^c_{t+N-1|t})$ lumps up u^c and $u^{j\in\mathcal{V}}$ over the prediction horizon, and $\mathbf{X} = (x^1_{t|t}, \ldots, x^1_{t+N|t}, \ldots x^{N_v}_{t|t}, \ldots, x^{N_v}_{t+N|t})$ is the vector of system state predictions over the prediction horizon. Let $\mathcal{T} = (\mathbf{X}, \mathbf{U})$ be the vector collecting all optimization variables. The cost function in (4.6) is the total energy consumed by all VAV boxes and the AHU system over the prediction horizon.

In (4.6) $x^j_{t+k|t}$ denotes the state of room j at time $t + k\Delta t$ predicted at time t starting from the current state $x^j_{t|t} = x^j(t),\ \forall j \in \mathcal{V}$.

Let $(\mathbf{U}^\star, \mathbf{X}^\star)$ be the optimal solution of problem (4.6). Then, only the first element of every control sequence is implemented to the system, i.e. $u^j(t) = u^{j\star}_{t|t}$, $u^a(t) = u^{a\star}_{t|t}$.

The optimization (4.6) is repeated at time $t + \Delta t$, with the updated state $x_{t+\Delta t|t+\Delta t} = x(t + \Delta t)$, yielding a *moving or receding horizon control* strategy.

The MPC problem (4.6) has a non-convex cost (4.6a) which includes bilinear terms for the energy consumption of cooling and heating coils, bilinear equality constraints (4.6b) and (4.6c), and box constraints on system states and control inputs. The size of the nonlinear optimization problem rapidly grows when a realistic number of rooms and a meaningful horizon length N are considered. In order to solve the MPC problem (4.6) in a distributed fashion, we apply sequential quadratic programming (SQP), proximal minimization, and dual decomposition. Next we show the main idea of these techniques and implementation details for the specific class of problems considered in this thesis.

The SQP procedure is an efficient method to solve nonlinear programming problems [62, 104]. The basic idea is to linearize the nonlinear constraints around a candidate solution and replace the objective with a quadratic function around this guess. The solution to the resulting QP is then used to update the candidate solution. The iterations are repeated until convergence is achieved [62]. Note that the hessian of the cost (4.6a) is, in general, indefinite since the cost is nonconvex.

The proximal minimization algorithm is adopted to solve the indefinite QP problem obtained from the SQP procedure by iteratively solving a set of subproblems whose cost functions have extra convex quadratic terms [9]. By doing so, the primal cost function of the subproblem is strictly convex, which implies differentiability of the dual cost function. This enables faster convergence in the procedure of dual decomposition [9].

The concept of dual decomposition traces back to the 70's [83], and it has been extensively studied since then [9, 113]. The QP derived from the proximal minimization algorithm is a separable convex optimization problems and therefore the gradient of the dual problem can be calculated in a distributed fashion, and the dual variables can be optimized separately by using gradient or subgradient based approaches. The primal optimal solution then can be reconstructed from the dual variables.

To summarize, the optimal solution to Problem (4.6) is obtained through three nested iterative algorithms. The outer iteration solves the original nonlinear optimization problem (4.6) by solving a sequence of QPs, the second and third levels of iteration solve the QP in a distributed fashion by using proximal minimization and dual decomposition, respectively.

The details of the proposed algorithms are described next. Problem (4.6) is time-variant because of the disturbance load profile w. With abuse of notation and for the sake of simplicity, in the rest of the chapter the lower indices "$t+k|t$" is denoted as "k".

Level 1: Modified Sequential Quadratic Programming

At the SQP iteration n_s, problem (4.6) is linearized by replacing the nonlinear system dynamics (4.6b) with the linearized ones (4.5) at a candidate solution $\mathcal{T}^{n_s} = (\mathbf{U}^{n_s}, \mathbf{X}^{n_s})$. The cost function (4.6a) is approximated by a quadratic function around $\mathcal{T}^{n_s}$ while neglecting the off-diagonal terms. The resultant optimization problem (4.7) is convex.

$$\min_{\mathbf{dU},\mathbf{dX}} dJ(\mathbf{dU},\mathbf{dX}) = \sum_{k=0}^{N-1}\left\{\sum_{j\in\mathcal{V}}\left(\frac{1}{2}du_k^{jT}Q_k^j du_k^j + c_k^{jT}du_k^j\right) + \frac{1}{2}du_k^{cT}Q_k^c du_k^c + c_k^{cT}du_k^c\right\} \tag{4.7a}$$

subject to:

$$T_k^j dx_{k+1}^j + A_k^j dx_k^j + B_k^j du_k^j + B_k^c du_k^c + f_k^{e,j} + \sum_{i\in\mathcal{N}^j} E_i^j \frac{dx_{k+1}^i + dx_k^i}{2} = 0,\ \forall j\in\mathcal{V},\ k=0,1,\ldots,N-1, \tag{4.7b}$$

$$G_k^c du_k^c + \sum_{j\in\mathcal{V}} G_{j,k}^+ dx_{k+1}^j + \sum_{j\in\mathcal{V}} G_{j,k}^x dx_k^j + \sum_{j\in\mathcal{V}} G_{j,k}^u du_k^j + g_k^e = 0,\ k=0,1,\ldots,N-1, \tag{4.7c}$$

$$dx_k^j \in d\mathcal{X}^j,\ \forall j\in\mathcal{V},\ k=1,\ldots,N, \tag{4.7d}$$

$$du_k^j \in d\mathcal{U}^j,\ du_k^c \in d\mathcal{U}^c,\ \forall j\in\mathcal{V},\ k=0,\ldots,N-1, \tag{4.7e}$$

where Q_k^c and Q_k^j are diagonal matrices obtained by removing the off-diagonal terms of the hessian of (4.6a) at candidate solution $\mathcal{T}^{n_s}$. The vector $\mathbf{dU} = (du_0^1, \ldots, du_{N-1}^1, \ldots, du_0^{N_v}, \ldots, du_{N-1}^{N_v}, du_0^c, \ldots, du_{N-1}^c)$ collects control input difference from the candidate solution $\mathbf{U}^{n_s}$, and the vector $\mathbf{dX} = (dx_1^1, \ldots, dx_N^1, \ldots, dx_1^{N_v}, \ldots, dx_N^{N_v})$ collects system state deviations from the trajectory $\mathbf{X}^{n_s}$. The constraint sets $d\mathcal{X}^j = \mathcal{X}^j - x_k^{j\,n_s}$, $d\mathcal{U}^j = \mathcal{U}^j - u_k^{j\,n_s}$, and $d\mathcal{U}^c = \mathcal{U}^c - u_k^{c\,n_s}$ define the feasible state variations $\mathbf{dX}$ and control input variations $\mathbf{dU}$, respectively.

The optimal solution $d\mathcal{T}^\star = (\mathbf{dX}^\star,\ \mathbf{dU}^\star)$ to problem (4.7) is computed by the iterative algorithms at level 2 and level 3 described next. The vector $d\mathcal{T}^\star$ is used to update the candidate solution as

$$\mathcal{T}^{n_s+1} = \mathcal{T}^{n_s} + \alpha \cdot d\mathcal{T}^\star.$$

In this thesis, a constant step length α is applied. At the SQP iteration $n_s + 1$, the process of linearizing problem (4.6) and solving problem (4.7) is repeated. The SQP algorithm is terminated if

$$\|d\mathcal{T}^\star\| \leq \kappa, \tag{4.8}$$

where κ is a predefined convergence tolerance.

Remark 1 *The convergence of the proposed SQP algorithm for general non-convex programs is not guaranteed. Extensive numerical tests have failed in finding an instance of the problem considered in this chapter where the proposed algorithm would not converge. The convergence rate of SQP procedure can be improved when an adaptive step length α is selected using an advanced line search algorithm [104].*

Level 2: Proximal Minimization

The iterative algorithm at the second level solves problem (4.7) by using the proximal minimization algorithm proposed in [9]. The quadratic objective in Problem (4.7) is positive semi-definite. Its solution is obtained by optimizing a sequence of subproblems obtained by adding a quadratic term to the original cost. At the proximal minimization iteration n_p, we

consider the subproblem

$$\min_{\mathbf{d\widehat{U}},\mathbf{d\widehat{X}}} d\widehat{J}^{n_p}(\mathbf{d\widehat{U}},\mathbf{d\widehat{X}}) = \sum_{k=0}^{N-1}\left\{\sum_{j\in\mathcal{V}}\left(\frac{1}{2}d\widehat{u}_k^{j\,T}Q_k^j d\widehat{u}_k^j + c_k^{j\,T} d\widehat{u}_k^j\right) + \frac{1}{2}d\widehat{u}_k^{c\,T}Q_k^c d\widehat{u}_k^c + c_k^{c\,T} d\widehat{u}_k^c\right\} + \frac{\rho}{2}\left(\|\mathbf{d\widehat{U}} - \mathbf{d\widehat{U}}^{n_p-1}\|_2^2 + \|\mathbf{d\widehat{X}} - \mathbf{d\widehat{X}}^{n_p-1}\|_2^2\right) \tag{4.9a}$$

subj. to:

$$T_k^j d\widehat{x}_{k+1}^j + A_k^j d\widehat{x}_k^j + B_k^j d\widehat{u}_k^j + B_k^c d\widehat{u}_k^c + f_k^{e,j} + \sum_{i\in\mathcal{N}^j} E_i^j \frac{d\widehat{x}_{k+1}^i + d\widehat{x}_k^i}{2} = 0,\ \forall j\in\mathcal{V},\ k=0,1,\ldots,N-1, \tag{4.9b}$$

$$G_k^c d\widehat{u}_k^c + \sum_{j\in\mathcal{V}} G_{j,k}^+ d\widehat{x}_{k+1}^j + \sum_{j\in\mathcal{V}} G_{j,k}^x d\widehat{x}_k^j + \sum_{j\in\mathcal{V}} G_{j,k}^u d\widehat{u}_k^j + g_k^e = 0,\ k=0,1,\ldots,N-1, \tag{4.9c}$$

$$d\widehat{x}_k^j \in d\mathcal{X}^j,\ \forall j\in\mathcal{V},\ k=1,\ldots,N, \tag{4.9d}$$

$$d\widehat{u}_k^j \in d\mathcal{U}^j,\ d\widehat{u}_k^c \in d\mathcal{U}^c,\ \forall j\in\mathcal{V},\ k=0,\ldots,N-1, \tag{4.9e}$$

where $\rho > 0$ is a strictly positive constant predetermined such that the resulting cost (4.9a) is positive definite. $\left(\mathbf{d\widehat{X}}^{n_p-1},\ \mathbf{d\widehat{U}}^{n_p-1}\right)$ is the optimal solution to Problem (4.9) at iteration $n_p - 1$. When $n_p = 1$, let $\mathbf{d\widehat{X}}^0 = \mathbf{0},\ \mathbf{d\widehat{U}}^0 = \mathbf{0}$.

The optimal solution $(\mathbf{d\widehat{X}}^\star, \mathbf{d\widehat{U}}^\star)$ to problem (4.9) is computed by the algorithm at level 3. We set $\mathbf{d\widehat{X}}^{n_p} = \mathbf{d\widehat{X}}^\star$ and $\mathbf{d\widehat{U}}^{n_p} = \mathbf{d\widehat{U}}^\star$, and terminate the iterative algorithm if

$$\|\widehat{d\mathbf{U}}^{n_p} - \widehat{d\mathbf{U}}^{n_p-1}\|_2 \le \kappa,\ \|\widehat{d\mathbf{X}}^{n_p} - \widehat{d\mathbf{X}}^{n_p-1}\|_2 \le \kappa. \tag{4.10}$$

It has been proved in [9] that if the optimization problem (4.7) is convex, then the vector $(\mathbf{d\widehat{X}}^{n_p},\ \mathbf{d\widehat{U}}^{n_p})$ converges to an optimum $(\mathbf{dX}^\star,\ \mathbf{dU}^\star)$ of problem (4.7).

Level 3: Dual Decomposition and Fast Gradient Method

The dual decomposition algorithm is used to solve problem (4.9) by solving its dual problem. The dual problem of the QP (4.9) is formulated by assigning dual variables λ_k^j and μ_k to the constraints (4.9b) and (4.9c), respectively. The dual problem can be formulated as follows:

$$\max_{\lambda,\mu\ \text{free}} \min_{\mathbf{d\widehat{U}},\mathbf{d\widehat{X}}} \quad d\widehat{J}^{n_p} + L^c + L^f \tag{4.11a}$$

subj. to

$$d\widehat{x}_k^j \in d\mathcal{X}^j,\ \forall j\in\mathcal{V},\ k=1,\ldots,N, \tag{4.11b}$$

$$d\widehat{u}_k^j \in d\mathcal{U}^j,\ \forall j\in\mathcal{V},\ k=0,\ldots,N-1, \tag{4.11c}$$

$$d\widehat{u}_k^c \in d\mathcal{U}^c,\ \forall k=0,\ldots,N-1, \tag{4.11d}$$

where $\widehat{dJ}^{n_p}$ is the cost defined in (4.9a). The term

$$L^c = \sum_{k=1}^{N} \mu_k^T \left(G_k^c d\widehat{u}_k^c + \sum_{j\in\mathcal{V}} G_{j,k}^+ d\widehat{x}_{k+1}^j + \sum_{j\in\mathcal{V}} G_{j,k}^x d\widehat{x}_k^j + \sum_{j\in\mathcal{V}} G_{j,k}^u d\widehat{u}_k^j + g_k^e \right)$$

is the dual term corresponding to constraint (4.9c). The term

$$L^f = \sum_{j\in\mathcal{V}} \sum_{k=1}^{N} {\lambda_k^j}^T \left(T_k^j d\widehat{x}_{k+1}^j + A_k^j d\widehat{x}_k^j + B_k^j d\widehat{u}_k^j + B_k^c d\widehat{u}_k^c + f_k^{e,j} + \sum_{i\in\mathcal{N}^j} E_i^j \frac{d\widehat{x}_{k+1}^i + d\widehat{x}_k^i}{2} \right)$$

is the dual term for constraint (4.9b).

In [90], the dual problem is solved by a projected subgradient method with a constant step size, which suffers from relatively slow convergence. The algorithm convergence speed can be improved by applying the fast gradient method since the dual function becomes Lipschitz-smooth when a proximal terms is added [103, 115].

We note that the cost function (4.11a) and the constraints of the inner minimization problem (4.11) are separable. This special structure allows us to solve problem (4.11) using the fast gradient method in a distributed way as described next.

Three sets of variables are updated in the fast gradient method [115], namely the dual variables λ and μ, the auxiliary variables $\bar{\lambda}$ and $\bar{\mu}$ of the same dimension as the dual variables, and a parameter γ controlling the step size β. They are initialized as $\lambda^0 = \lambda^{\star n_p-1}$, $\mu^0 = \mu^{\star n_p-1}$, $\bar{\lambda}^0 = \lambda^0$, $\bar{\mu}^0 = \mu^0$, and $\gamma^0 = \frac{\sqrt{5}-1}{2}$. Note that for $n_p = 1$, we set $\lambda^{\star n_p-1} = \mathbf{0}$ and $\mu^{\star n_p-1} = \mathbf{0}$.

The optimal solution $\left(\lambda^{\star n_p-1},\ \mu^{\star n_p-1}\right)$ to the dual problem at the second level iteration $n_p - 1$ is used as a warm start for faster convergence.

At the third level iteration n_d, the dual and auxiliary variables are updated by using the fast gradient method as follows [115]:

$$\lambda_k^{j,n_d} = \bar{\lambda}_k^{j,n_d-1} + \frac{1}{L} h_{\lambda_k^j}(\bar{\lambda}_k^{j,n_d-1}),\ \forall j \in \mathcal{V},\quad \forall k = 1, 2, \ldots, N, \tag{4.12a}$$

$${\mu_k}^{n_d} = {\bar{\mu}_k}^{n_d-1} + \frac{1}{L} h_{\mu_k}(\bar{\mu}_k^{n_d-1}),\ \forall k = 0, 1, \ldots, N-1, \tag{4.12b}$$

$$\gamma^{n_d} = \frac{\gamma^{n_d-1}}{2} \left(\sqrt{(\gamma^{n_d-1})^2 + 4} - \gamma^{n_d-1} \right), \tag{4.12c}$$

$$\beta = \frac{\gamma^{n_d-1}(1-\gamma^{n_d-1})}{(\gamma^{n_d-1})^2 + \gamma^{n_d}}, \tag{4.12d}$$

$$\bar{\lambda}_k^{j,n_d} = {\lambda_k^j}^{n_d} + \beta({\lambda_k^j}^{n_d} - {\lambda_k^j}^{n_d-1}),\ \forall j \in \mathcal{V},\quad \forall k = 1, 2, \ldots, N, \tag{4.12e}$$

$${\bar{\mu}_k}^{n_d} = {\mu_k}^{n_d} + \beta({\mu_k}^{n_d} - {\mu_k}^{n_d-1}),\ \forall k = 0, 1, \ldots, N-1, \tag{4.12f}$$

where L is the Lipschitz constant of the dual gradient calculated as in [115]. $h_{\lambda_k^j}(\bar{\lambda}_k^{j,n_d-1})$ and $h_{\mu_k}(\bar{\mu}_k^{n_d-1})$ are the gradients of the dual cost function (4.11a) at ${\bar{\lambda}_k^j}^{n_d-1}$ and ${\bar{\mu}_k}^{n_d-1}$,

respectively, The fast gradient method algorithm is terminated if

$$\|h_{\lambda_k^j}(\bar{\lambda}_k^{j,n_d})\|_2 \leq \kappa, \ \forall j \in \mathcal{V}, \ k = 1, 2, \ldots, N, \tag{4.13a}$$

$$\|h_{\mu_k}(\bar{\mu}_k^{n_d})\|_2 \leq \kappa, \ k = 1, 2, \ldots, N. \tag{4.13b}$$

The computation of the gradients $h_{\lambda_k^j}(\bar{\lambda}_k^{j,n_d})$ and $h_{\mu_k}(\bar{\mu}_k^{n_d})$ is the most time consuming step of the algorithm. The proposed algorithm uses the following approach.

First, we compute the primal variables in (4.11) associated to the set of dual variables $\left(\bar{\lambda}^{n_d}, \bar{\mu}^{n_d}\right)$:

$$d\widehat{x}_k^{\star\, j\, n_d} = \Pi_{d\mathcal{X}^j}\left(d\widehat{x}_k^{j,n_p-1} + \frac{1}{\rho}(-G_{j,k}^{+\,T}\bar{\mu}_{k-1}^{n_d} - G_{j,k}^{x\ T}\bar{\mu}_k^{\ n_d} - A_k^{j\,T}\bar{\lambda}_k^{j\,n_d} - T_k^{j^T}\bar{\lambda}_{k-1}^{j,n_d}\right.$$

$$\left. - \sum_{i\in\mathcal{N}^j} E_j^{i\,T}\frac{\bar{\lambda}_{k-1}^{i,n_d} + \bar{\lambda}_k^{i\ n_d}}{2})\right), \ \forall j \in \mathcal{V}, \quad \forall k = 1, 2, \ldots, N, \tag{4.14a}$$

$$d\widehat{u}_k^{\star\, j\, n_d} = \Pi_{d\mathcal{U}^j}\left((\rho I + Q_k^j)^{-1}(\rho d\widehat{u}_k^{j\,n_p-1} - c_k^j - G_{j,k}^{u\ T}\bar{\mu}_k^{\ n_d} - B_k^{j^T}\bar{\lambda}_k^{j\,n_d})\right),$$

$$\forall j \in \mathcal{V}, \quad \forall k = 0, 1, \ldots, N-1, \tag{4.14b}$$

$$d\widehat{u}_k^{\star\, c\, n_d} = \Pi_{d\mathcal{U}^c}\left((\rho I + Q_k^c)^{-1}(\rho d\widehat{u}_k^{c\,n_p-1} - c_k^c - G_{c,k}^{u\ T}\bar{\mu}_k^{\ n_d} - \sum_{j\in\mathcal{V}} B_k^{c^T}\bar{\lambda}_k^{j\,n_d})\right),$$

$$\forall k = 0, 1, \ldots, N-1, \tag{4.14c}$$

where $\Pi_{\mathcal{S}}(\star)$ is the operation of projecting $\star$ onto the convex set $\mathcal{S}$.[1] Then, the gradients of the dual cost function (4.11a) at $(\bar{\mu}^{n_d}, \ \bar{\lambda}^{n_d})$ can be computed as follows [115]:

$$h_{\lambda_k^j}(\bar{\lambda}_k^{j,n_d}) = T_k^j d\widehat{x}_{k+1}^{\star\, j\ n_d} + A_k^j d\widehat{x}_k^{\star\, j\, n_d} + B_k^j d\widehat{u}_k^{\star\, j\, n_d} + B_k^c d\widehat{u}_k^{\star\, c\, n_d} + f_k^{e,j}$$

$$+ \sum_{i\in\mathcal{N}^j} E_i^j \frac{d\widehat{x}_{k+1}^{\star\, i\ \ n_d} + d\widehat{x}_k^{\star\, i\, n_d}}{2}, \tag{4.15a}$$

$$\forall j \in \mathcal{V}, \quad \forall k = 0, 1, \ldots, N-1,$$

$$h_{\mu_k}(\bar{\mu}_k^{n_d}) = G_k^c d\widehat{u}_k^{\star\, c\, n_d} + \sum_{j\in\mathcal{V}} G_{j,k}^{+} d\widehat{x}_{k+1}^{\star\, j\, n_d} + \sum_{j\in\mathcal{V}} G_{j,k}^{x} d\widehat{x}_k^{\star\, j\, n_d}$$

$$+ \sum_{j\in\mathcal{V}} G_{j,k}^{u} d\widehat{u}_k^{\star\, j\, n_d} + g_k^e, \tag{4.15b}$$

$$\forall k = 0, 1, \ldots, N-1.$$

In summary, the proposed MPC problem is solved locally by three nested levels of iterations: the outer iteration solves the original nonlinear optimization problem (4.6) by solving a sequence of QPs (4.7). The second and third levels of iteration solve the QP (4.7) in a

[1] Note that as Q_k^j and Q_k^c are diagonal matrices, the matrix inversion in (4.14b) and (4.14c) can be easily evaluated. The projection is easy to calculate as $d\mathcal{X}^j$, $d\mathcal{U}^j$, and $d\mathcal{U}^c$ are box constraints.

distributed fashion by using proximal minimization and dual decomposition. Algorithm 2 summarizes the main steps of the proposed distributed model predictive control (DMPC) scheme.

Algorithm 2 (DMPC for building control)

Initial:	*Let* $\mathbf{U}^1$, $\mathbf{X}^1$ *be initial guesses for primal variables for Problem (4.6). Set the SQP iteration index to* $n_s = 1$.
Step 1:	*Let* n_s *be the current SQP iteration index. Linearize the system model (4.4) at* $(\mathbf{U}^{n_s}, \mathbf{X}^{n_s})$ *to obtain the coefficients in (4.7). Let* $n_p = 1$, *and set* $\widehat{\mathbf{dU}}^0 = \mathbf{0}$ *and* $\widehat{\mathbf{dX}}^0 = \mathbf{0}$ *in (4.9). Also let* $\lambda^{\star 0} = \mathbf{0}$ *and* $\mu^{\star 0} = \mathbf{0}$.
Step 2:	*Let* n_p *be the current proximal minimization iteration index, and set* $n_d = 1$. *Initialize* $\widehat{\lambda}^1 = \lambda^1 = \lambda^{\star\, n_p - 1}$ *and* $\widehat{\mu}^1 = \mu^1 = \mu^{\star\, n_p - 1}$
Step 3:	*Let* n_d *be the current dual decomposition iteration index. Update primal variables* $(\widehat{d\mathbf{U}}^\star, \widehat{d\mathbf{X}}^\star)$ *as in the distributed algorithm (4.14).*
Step 4:	*Exchange the updated primal variables and calculate the gradients (4.15) for the dual function.*
Step 5:	*Update the dual sequences* λ^{n_d+1}, μ^{n_d+1}, $\widehat{\lambda}^{n_d+1}$, $\widehat{\lambda}^{n_d+1}$, *and* γ^{n_d+1} *as in the distributed algorithm (4.12).*
Step 6:	*If condition (4.13) is satisfied, go to the next step. Otherwise, set* $n_d = n_d + 1$ *and go to Step 3.*
Step 7:	*Update* $\widehat{d\mathbf{U}}^{n_p} = \widehat{d\mathbf{U}}^{\star\, n_d}$ *and* $\widehat{d\mathbf{X}}^{n_p} = \widehat{d\mathbf{X}}^{\star\, n_d}$. *If condition (4.10) is satisfied, go to the next step. Otherwise, set* $n_p = n_p + 1$ *and go to Step 2.*
Step 8	*Set* $d\mathbf{U}^\star = \widehat{d\mathbf{U}}^{n_p}$ *and* $d\mathbf{X}^\star = \widehat{d\mathbf{X}}^{n_p}$. *If condition (4.8) is satisfied, terminate and the optimal control sequence is* $\mathbf{U}^{n_s}$. *Otherwise, update* $\mathbf{U}^{n_s+1} = \mathbf{U}^{n_s} + \alpha \cdot d\mathbf{U}^\star$ *and* $\mathbf{X}^{n_s+1} = \mathbf{X}^{n_s} + \alpha \cdot d\mathbf{X}^\star$, *set* $n_s = n_s + 1$, *and go to Step 1.*

4.4 Simulation Results

This section presents a numerical example to show the effectiveness of the proposed controller design methodology.

We compare the proposed control methodology with a baseline control logic (BC), which is a simplified version of a production control logic. The BC works as follows. When all the room temperatures are within the comfort range, the mass flow rate of the supply air ($\dot{m}_s^j$) is set to its minimum and the valves of cooling and heating coils are closed. When a room temperature hits the lower bound, the air mass flow rate to the room is maintained at its minimum, and the supply air temperature will be adjusted by the heating coil in the corresponding VAV box so that the room temperature stays at the lower bound value. When a room temperature violates the upper constraints, the AHU supply air temperature is set to its minimum, and the mass flow rate of the supply air is controlled so that the room temperature is within the comfort range.

Table 4.1: Parameters for the numerical example.

param	value	param	value	param	value	param	value
$\underline{\dot{m}}$	0.005 kg/s	$\overline{\dot{m}}$	5 kg/s	η_c	0.7	η_h	0.8
$\underline{\Delta T_c}$	0 K	$\overline{\Delta T_c}$	8 K	P_f	0.08	α	0.25
$\underline{\Delta T_h}$	0 K	$\overline{\Delta T_h}$	8 K	$\underline{\delta}$	0	$\overline{\delta}$	0.8
$\underline{COP_c}$	5	COP_h	0.9	N	48	κ	1×10^{-3}

The numerical example considers a network of 10 rooms. All the rooms have the same model parameters as in Table 2.1. The undirected graph describing the topology of the room network is $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, where $\mathcal{V} = \{1, 2, \ldots, 10\}$, and $\mathcal{A} = \{(1,2), (2,3), \ldots, (9,10)\}$. The weather information is downloaded from July 2nd to July 3rd, 2009 at UC Berkeley, and the temperature profile is plotted in Figure 4.1. Because of the warm weather, only cooling is critical in the considered scenario. Figure 4.2 depicts the nominal internal load profile

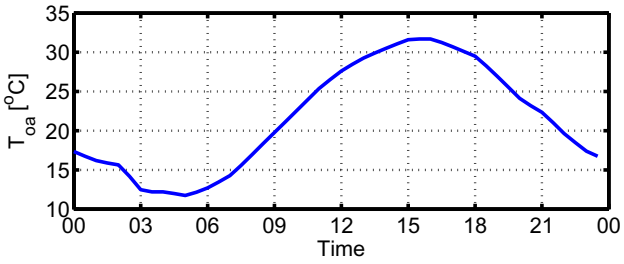


Figure 4.1: Ambient temperature (T_{oa}).

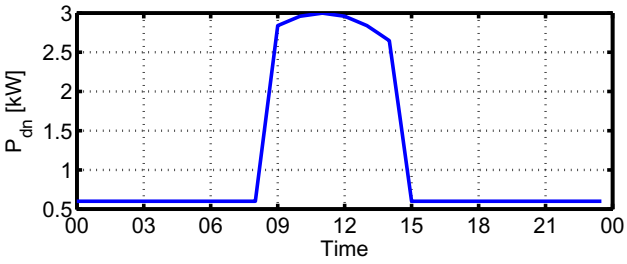


Figure 4.2: Internal load profile (P_{dn}).

(P_{dn}) in our simulations. We assume that during 15:00 and 8:00 the next day, the rooms are empty without occupancy, leaving a minimum internal load of 0.01 kW due to lighting or other electrical devices. We use an internal load profile different for each room. In particular, we compute the internal load for room j as

$$P_d{}^j = (1 + 0.2j)P_{dn}, \quad j = 1, 2, \ldots, 10.$$

In our simulations, the parameters for controllers in Section 4.3 are listed in Table 4.1. The sampling time Δt is chosen to be thirty minutes, and the prediction horizon is one day ($N = 48$). The comfort constraints are defined as [21, 25]°C from 6:00 to 18:00 when there is occupancy in the buildings, and the comfort set is relaxed to [19, 27]°C when offices are unoccupied.

Figure 4.3 shows the simulation results for the network of rooms controlled by the baseline controller. Figure 4.3(a) shows that all the room temperatures are within the comfort range defined by the dotted lines. During early morning till 04:00, all zone temperatures are within the comfort range. As a result, the supply fan only maintains the minimum required air mass flow rate to each zone, and the valve of the cooling coils in the AHU is fully closed. The

occupancy load at noon results in a peak total air mass flow rate of 7.2 kg/s. The cooling coils are operating at maximum capacity as soon as one of the zone temperatures hits the upper constraints so that the thermal comfort can be guaranteed. The return air damper position is fully closed to take advantage of free cooling when the ambient temperature is lower than the zone temperature.

The performance of the proposed DMPC controller is reported in Figure 4.4. It cools down the room temperature to the lower bounds of the comfort range during the early morning (Figure 4.4(a)) while the baseline controller remains inactive until the room temperature hits the upper bounds around 4:00 (Figure 4.3(a)). This *precooling* saves energy, since during the early morning the lower ambient temperature enables free cooling. The free cooling is illustrated in Figure 4.4(d). The MPC algorithm decides to open the cooling coil valve from 8:00, which is three hours later than the schedule proposed by baseline control logics in Figure 4.3(d).

Moreover, it is noted that instead of cooling all zones simultaneously, MPC cools down zones consecutively as Figure 4.4(a) illustrates. This feature significantly reduces the peak total air flow rate from 7.2 kg/s of BC to 5.8 kg/s (Figure 4.4(b)), and thus saves fan energy consumption (note that we use a quadratic penalty of total supply air mass flow rate in (4.6a)). The simulation results suggested that the pre-cooling and consecutive cooling strategies induced by DMPC enable a 43.5% energy saving compared to the baseline.

DMPC Algorithm Complexity

The proposed DMPC Algorithm 1.1 can be implemented in a network of embedded processors with low computational capacity since Step 3 and Step 4 of Algorithm 1.1 require only a few algebraic operations and simple projections. Figure 4.5 shows that a large number of iterations is required. This imposes a requirement for high network communication speed.

The DMPC Algorithm 1.1 was coded in Matlab® and runs on a single PC with Intel Core Duo CPU 3.00GHz. The runtime of the DMPC algorithm is estimated based on the assumption that the computation of Step 3 and Step 4 in Algorithm 1.1 are executed in parallel on $N_v + 1$ units including the controller units equipped on each VAV box and the AHU unit, and that the communication time is neglected. The results are reported in Figure 4.6 for different numbers of thermal zones considered. The dashed line shows the runtime of DMPC algorithm *when implemented on* N_v+1 *CPUs in parallel*, and the solid line depicts the time required to solve Problem (4.6) by the Interior Point OPTimizer (IPOPT) interfaced via AMPL *on one CPU*. Both the DMPC algorithm and the IPOPT start from the same initial guess of optimal solutions. One can notice that when the number of zones is less than three, IPOPT is faster than DMPC on a single PC. As the number of zones and the size of problem (4.6) increase, one can notice that DMPC could be implemented with a faster control sampling rate than IPOPT.

Figure 4.7 plots the variations of the optimal cost $J^{\star}_{DMPC}$ obtained by Algorithm 1.1 relative to the optimal cost $J^{\star}_{IPOPT}$ by IPOPT when solving MPC problem (4.6) with different

numbers of zones N_v. It is noted that DMPC results in slightly higher optimal costs than IPOPT. The reasons for this include the selection of convergence tolerance κ in Table 4.1.

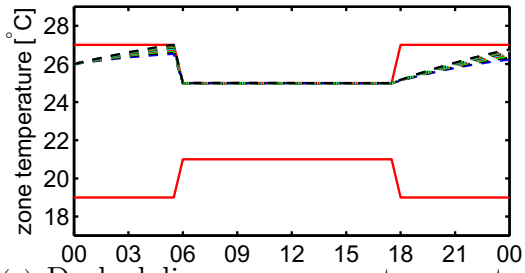
Fast Gradient Method Improvement

The advantage of applying the fast gradient method instead of the projected subgradient method in level 3 of the DMPC scheme is illustrated in Figure 4.8. We focus on the MPC problem (4.6) at $t = 0$. The modified SQP algorithm converged in 301 iterations. The dashed line in Figure 4.8 depicts the total number of iterations $\sum_{i=1}^{n_p} n_d^i$ required to solve the subproblems generated from the SQP algorithm when the fast gradient method in Section 4.3 is applied. The solid line depicts the number of iterations for the projected subgradient method with constant step size.

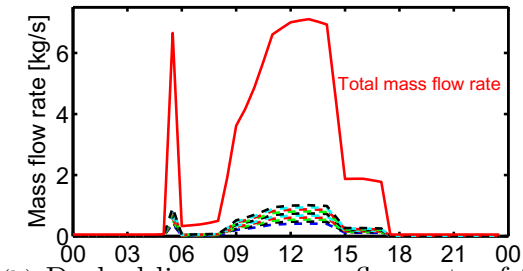
In this thesis, n_d^i is the number of iterations to solve the i-th subproblem (4.9) obtained in the proximal minimization level. If the fast gradient method is used, the number of iterations required to satisfy the stopping criterion is, on average, about four times less than for the projected subgradient method.

4.5 Summary

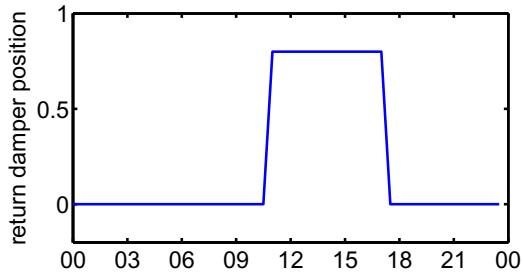
In this chapter, a distributed model predictive control is designed to regulate thermal comfort while minimizing energy consumption. A three-level iterative algorithm is presented for the solution of the nonlinear MPC problem. The key advance of this approach is the ability to solve the nonlinear MPC problem in a distributed manner and also in parallel. The resulting scheme is suitable for being implemented on a set of distributed low-cost processors. Simulation results show interesting behavior and short computation time. Future research will focus on analyzing stability issues of the resultant nonlinear MPC controller and improving the convergence rate of the proposed algorithm using adaptive step length and parameters.



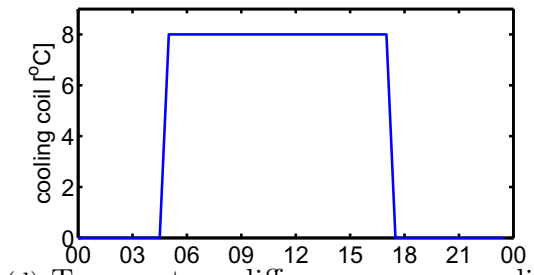
(a) Dashed lines are room temperatures ($T^{j\in\mathcal{V}}$), and solid lines define the comfort sets.



(b) Dashed lines are mass flow rate of the supply air to each zone ($\dot{m}_s^{j\in\mathcal{V}}$), and the solid line is the total mass flow rate $\dot{m}_{\text{fan}}$.

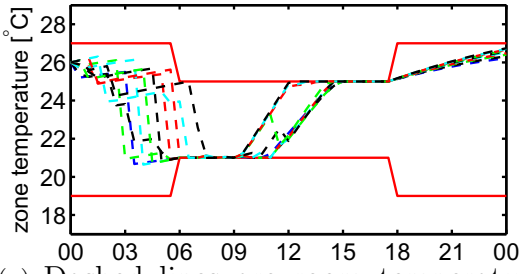


(c) Return air damper position (δ)

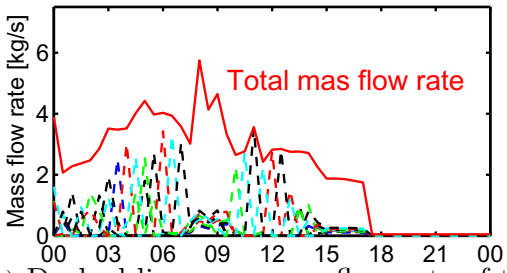


(d) Temperature difference across cooling coil (ΔT_h)

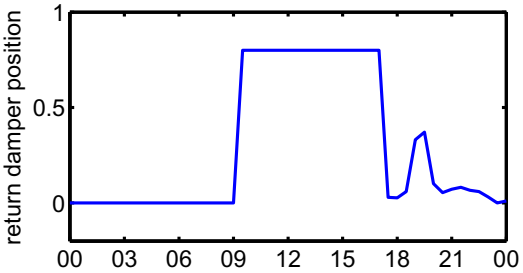
Figure 4.3: System behavior for simplified baseline control logics.



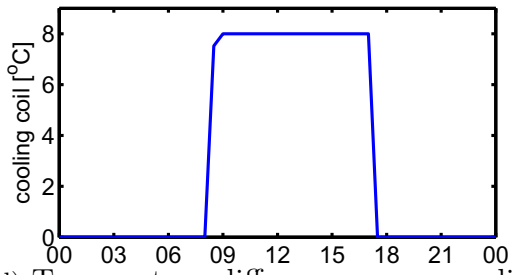
(a) Dashed lines are room temperatures ($T^{j\in\mathcal{V}}$), and solid lines define the comfort sets.



(b) Dashed lines are mass flow rate of the supply air to each zone ($\dot{m}_s^{j\in\mathcal{V}}$), and the solid line is the total mass flow rate $\dot{m}_{\text{fan}}$.



(c) Return air damper position (δ).



(d) Temperature difference across cooling coil (ΔT_h).

Figure 4.4: System behavior for distributed model predictive control.

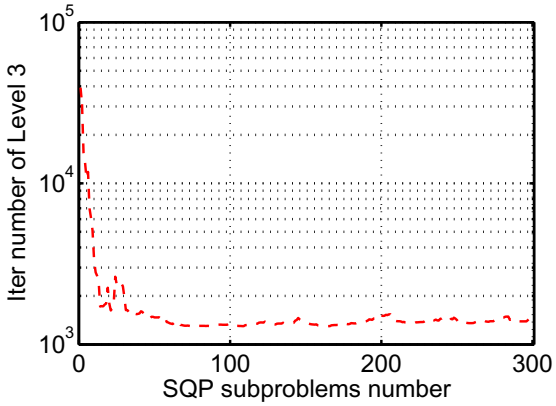


Figure 4.5: Number of dual decomposition iterations at time $t = 0$.

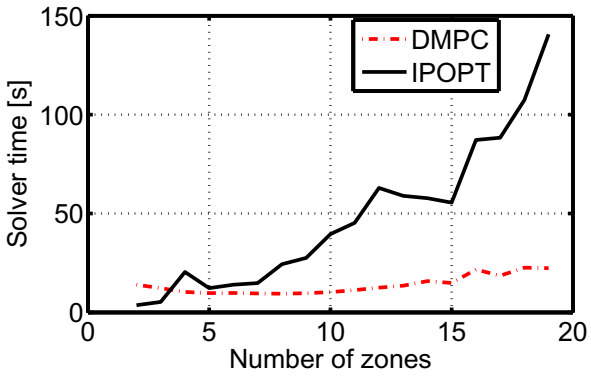


Figure 4.6: Comparison between CPU time for DMPC and IPOPT.

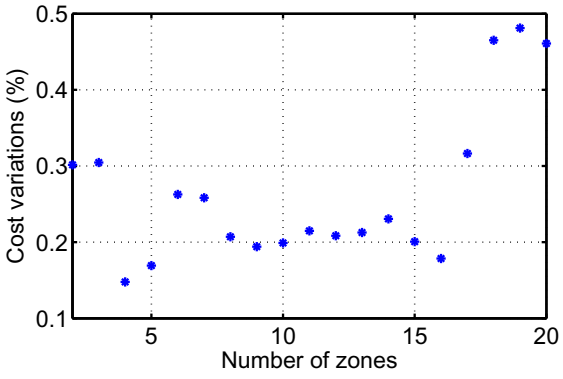


Figure 4.7: Cost variations $(J^{\star}_{DMPC} - J^{\star}_{IPOPT})/J^{\star}_{IPOPT}$.

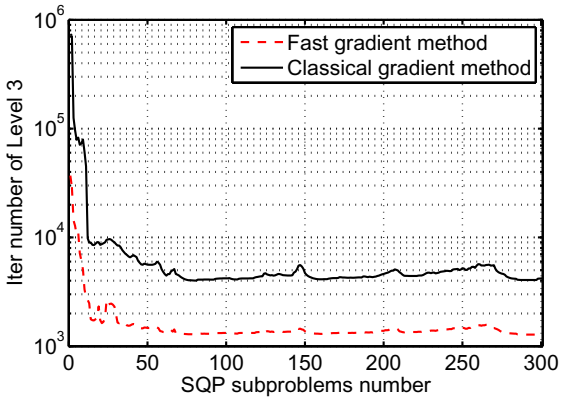


Figure 4.8: Number of dual decomposition iterations at time $t = 0$.

Chapter 5

Stochastic Model Predictive Control

5.1 Introduction

In the past years there has been a renewed interest in modeling and predictive control for energy conversion, storage, and distribution in commercial buildings [94, 64, 92, 93, 95, 74, 75, 76, 106, 19, 102, 5]. The majority of existing model predictive control (MPC) schemes has the objective of minimizing energy consumption while satisfying occupant thermal comfort using predictive knowledge of weather and occupancy (the building "load"). Simulative and experimental results have shown that closed-loop energy savings and comfort depend on the load predictions that are often different from actual realizations [95].

In this chapter, a stochastic predictive control (SMPC) design is proposed for the aforementioned problem. In particular, we make use of stochastic information of weather and load learned from historical data, and minimize average energy consumption while bounding the probability of comfort violations. For commercial buildings of medium size, the simplest models reach the complexity of hundreds of states and control inputs. The complexity of the resulting SMPC problem motivates our research.

The following finite time stochastic optimization problem will be used to present the SMPC design.

$$\min_{\pi_{t|t},\ldots,\pi_{t+T-1|t}} \mathbb{E}\left(\sum_{k=0}^{k=T-1} \text{Energy}(x_{t+k|t}, u_{t+k|t}, d_{t+k|t})\right) \tag{5.1a}$$

$$\text{subj. to} \quad x_{t+k+1|t} = f(x_{t+k|t}, u_{t+k|t}, d_{t+k|t}),\ \forall k \in \mathbb{N}_{T-1}, \tag{5.1b}$$

$$u_{t+k|t} = \kappa_{t+k|t}(x_{t+k|t}),\ \forall k \in \mathbb{N}_{T-1}, \tag{5.1c}$$

$$u_{t+k|t} \in \mathcal{U},\ \forall k \in \mathbb{N}_{T-1}, \tag{5.1d}$$

$$\mathbb{P}\left(x_{t+k|t} \in \mathcal{X}\right) \geq 1-\epsilon,\ \forall k \in \mathbb{N}_{T}, \tag{5.1e}$$

$$x_{t|t} = x(t), \tag{5.1f}$$

where the symbol $v_{t+k|t}$ is the random variable v at time $t+k$ predicted at time t, $\mathbb{E}(v)$ denotes the expected value of v, $\mathbb{P}\left(v \in \mathcal{V}\right)$ is the probability of the event $v \in \mathcal{V}$, $\mathbb{N}_T$ is the set

of integers $\{0, 1, \ldots, T\}$, x is the system state, u is the control input, and d is the uncertain load. The functions Energy(x, u, d) and $f(x, u, d)$ define the energy consumption and system dynamics, respectively. Problem (5.1) seeks a set of control feedback laws $\kappa_{t+k|t} \in \mathcal{U}$ that minimizes the expected energy consumption (5.1a), and that has small probability of state constraint violation $x \notin \mathcal{X}$.

The solution to the stochastic MPC problem (5.1) requires four main steps: *(i)* the translation of the optimization over *control polices* $\kappa(\cdot)$ into a finite-dimensional optimization problem, *(ii)* the propagation of the stochastic system states over the prediction horizon, *(iii)* the translation of probabilistic constraints into deterministic constraints, and *(iv)* the solution of the resulting mathematical program.

With the exception of linear systems and special classes of distributions (e.g., normal distributions), steps *(i)-(iii)* are non-trivial and affect the complexity of the mathematical problem at step (*iv*) and the conservatism of the resulting control policy [84, 130, 109, 28, 71]. Step *(i)* has different solutions [6]. Open-loop prediction schemes are conservative since they look for one optimal open-loop control sequence that has to cope with all possible future disturbance realization, without taking future measurements into account. Closed-loop formulations overcome this issue but they can quickly lead to intractable problems. A compromise consists in fixing the control structure (e.g. affine state-feedback policies or affine disturbance feedback), parameterizing the control sequence in the feedback gains, and optimizing over these parameters [79, 80].

The chance constraints (5.1e) in step *(iii)* are translated into deterministic ones by enforcing tightened constraints on expected values of states and inputs. The tightening offset is computed based on the tails of the disturbance probability distributions [70, 123]. In practice, the distribution of the ambient temperature and occupancy load in buildings are finitely-supported and non-Gaussian. In addition, the simplified control oriented models for buildings are bilinear systems [95, 73, 106, 58]. The bilinear terms arise from the multiplication of supply air mass flow rate and temperature to compute the cooling and heating energy delivered to thermal zones. Exact solutions to stochastic MPC problem (5.1) for nonlinear systems subject to non-Gaussian disturbances is, in general, computationally intractable for real-time implementation in HVAC systems. Obtaining computationally tractable approximation to this problem is crucial for the real-time implementation of stochastic MPC. The authors in [106] proposed to compute the nonlinear SMPC problem by solving a sequence of tractable subproblems. For each subproblem, the nonlinear dynamics are linearized around the simulated trajectory, and the tightening offsets for chance constraints are computed by bounding the tails of Gaussian distributions. The resulting problems are second order cone programs if the linear disturbance feedback gain is optimized, and the problems become quadratic programs if the controller feedback gain is determined off-line.

Non-Gaussian and finitely-supported disturbances have been studied in [32, 29, 80, 79, 23, 27] in the context of SMPC. The authors in [32, 29] use a tube-based method to translate the chance constraints. The work in [80, 79] suggests to translate the chance constraints to deterministic ones by using tightening offsets. The offsets are computed off-line using numerical approximations of convolution integrals. Sample-based methods provide an alternative

approach to transform the chance constraints for non-Gaussian random variables [23, 22, 27]. The approach consists in transforming the chance constraints (5.1e) into deterministic counterparts by evaluating them at a large number of disturbance samples.

In this thesis, we build on the work in [80, 79, 27] to design SMPC algorithms for networks of bilinear systems subject to non-Gaussian disturbances while retaining computational tractability. In particular, the bilinear system (5.1b) is linearized by using feedback linearization. The chance constraints (5.1e) then are transformed to deterministic ones by using two techniques: discrete convolution integrals and sample-based method. The complexity of both approaches is studied as a function of the problem size. The resulting nonlinear program is solved using Ipopt, a software package for large-scale nonlinear optimization problems [135].

The state feedback linearization proposed in this paper has several advantages. In particular, it allows us to easily optimize over feedback policy. Also, the transformation of the chance constraints into deterministic ones does not depend on the linearized systems dynamics as in [106].

The proposed SMPC design is carefully analyzed and compared with existing approaches. In particular, in Section 5.4 we discuss the tradeoff between performance, conservatism, and complexity. We will also try to shed some lights on the following questions.

(i) Does one need a stochastic MPC formulation, or nominal MPC expected forecasts provides "good" results for HVAC systems?

(ii) Is there value in using nonlinear probability distribution functions or Gaussian approximations work well and what is the price one has to pay for it?

(iii) Can the proposed approach be implemented on large scale buildings?

(iv) Should one transform the chance constraints by using convolution integrals or sample-based methods?

The effectiveness of the proposed approach will be demonstrated by using simulation and experiments. In both simulative and experimental studies, models and probability distribution functions of prediction uncertainties are generated by using measured historical data.

5.2 System Model

HVAC System

The main equipments used to produce and distribute cold or hot air in a building are depicted in Figure 2.2. Detailed model descriptions have been presented in Chapter 2. In order to improve its readability and to facilitate the presentation of the control design approach later in this chapter, we briefly include main relevant findings in Chapter 2.

The air handling unit (AHU) recirculates return air from building spaces, and mixes it with fresh outside air. The proportion of return air to outside air is controlled by damper positions in the AHU. The mixed air is cooled by the cooling coils that extract the cooling energy from chilled water produced by chillers. The air temperature after the cooling coil depends on the cooling coil valve position, the temperature of the chilled water, the temperature of mixed air entering the cooling coil, the mass flow rate of the mixed air, and the physical characteristics as well as thermal effectiveness of the cooling coil. Cool air is delivered to building spaces by electrical fans. Before reaching a given space, the air goes through variable air volume (VAV) boxes. At each VAV box the mass flow rate of the air supplied to the space is adjusted by a damper position. In addition, air temperature can be increased using reheat coils installed in the VAV box when needed. Space served by one VAV box is denoted by a thermal zone. The delivered air enters a zone through diffusers that are designed to fully mix the incoming air with the air in the thermal zone.

In this chapter, a network of thermal zones is described by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, and the thermal dynamic of a thermal zone is described by a bilinear regression model presented in Chapter 2,

$$\begin{aligned} T^i_{t+1} = &\sum_{q=0}^{q=q_d} \left(p^i_{1,q} T_{oa,t-q} + p^i_{2,q} I^i_{t-q}\right) + p^i_3 \left(T^i_{s,t} - T^i_t\right) \dot{m}^i_{s,t} + \\ &+ \sum_{q=0}^{q=q_x} \left(p^i_{4,q} T^i_{t-q} + \sum_{j \in \mathcal{N}^i} p^i_{5,q,j} T^j_{t-j} \right) + p^i_6 + P^i_{d,k}, \ \forall i \in \mathcal{V}, \end{aligned} \tag{5.2}$$

where T^i_k is the temperature of zone i at time k. For zone i at time k, $\dot{m}^i_{s,k}$ is the supply air mass flow rate, $T^i_{s,k}$ is the supply air temperature, $P^i_{d,k}$ is the internal thermal load, $T_{oa,k}$ is the ambient temperature, I^i_k is the solar radiation intensity. The supply air mass flow rates and temperatures $\dot{m}^i_s$ and T^i_s are control inputs. In (5.2) T^i are system states, and P^i_d, T_{oa} and I^i are exogenous signals. In (5.2) q_x and q_d are the autoregressive order and moving average order, respectively. The selection of q_x and q_d affects the complexity and the accuracy of the model. The bilinear term $(T^i_s(t) - T^i(t))\dot{m}^i_s(t)$ in (5.2) is introduced to match first-order energy-balance equations. The bilinear regression model (5.2) is identified using data collected from unoccupied hours when the internal load P_d is minimal. Given historical measurements of zone temperatures, ambient temperature, solar radiation intensity, and recorded control inputs, the model parameters can be computed by solving a linear regression problem.

We remark that the SMPC framework presented later in this chapter can be easily adopted to RC-based bilinear building models (2.27) with minor modifications.

Next we provide more details on the generation of forecasts for thermal load P^i_d, ambient temperature T_{oa}, and solar radiation intensity I^i in (5.2).

Load Models

The internal thermal loads are calculated as follows. The zone model (5.2) is first identified using data from unoccupied hours. By comparing occupied hour data to the model predictions, we can capture thermal load by occupants. We define a load profile as the difference in temperature between a local model prediction and the measurements:

$$P^i_{d,t-1} = \left(T^i_{\text{meas},t} - T^i_t\right), \quad \forall i \in \mathcal{V}, \tag{5.3}$$

where P^i_d is the load for zone i, $T^i_{\text{meas},t}$ is the measured temperature of zone i at time t, and T^i_t is the zone i temperature prediction with the identified model (5.2) using data at time-step $t-1$. To illustrate the effectiveness of the approach, next we report the results for a classroom of the Bancroft library at UC Berkeley (labeled as VAV C-2-15 in Figure 2.14). Figure 5.1 reports the estimated occupancy load profile from May 2011 to February 2012. Figure 5.1 overlays all the estimated load (dotted lines) by day-of-the-week. The solid line is the weekly mean occupancy load over the whole period (May 2011 to February 2012).

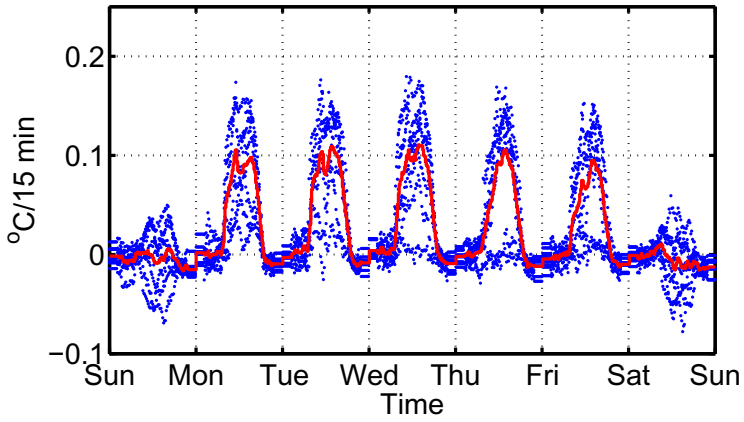


Figure 5.1: Occupancy load.

In Figure 5.1 we observe that during weekdays the average occupancy load profile has a peak load of 0.15°C, and there is no significant occupancy load during weekends. This phenomenon matches the occupancy schedule of the classroom. The model (5.3) is used to generate uncertainty map for an occupancy load predictor. The occupancy loads are predicted using a look up table with period of one week,

$$P^i_{d,k+t|t} = \hat{P}^i_{d,k+t|t} + \tilde{P}^i_{d,k+t|t}, \forall i \in \mathcal{V}, \tag{5.4}$$

where $P^i_{d,k+t|t}$ is the predicted occupancy load at time $t+k$ when the prediction starts from t, $\hat{P}^i_{d,k+t|t}$ is the mean of the internal load prediction at time $t+k$ plotted as the solid line in Figure 5.1, and $\tilde{P}^i_{d,k+t|t}$ is the prediction uncertainty described by a probability density

function (PDF) with finite support. The finitely-supported probability density function can be computed from samples of P_d^i estimated through (5.3) by using simple piecewise constant approximation or kernel density estimation [122].

The solar load for each zone is computed as the sun radiation intensity projected onto the normal vector to the outside wall. Figure 5.2 reports the computed daily sun load for an office room (labeled as VAV C-2-5 in Figure 2.14) from December 01, 2011 to January 31, 2012 in DOE library at UC Berkeley. In Figure 5.2, the dots are the computed daily sun load samples, and the solid line denotes the mean value of the computed daily solar load. We

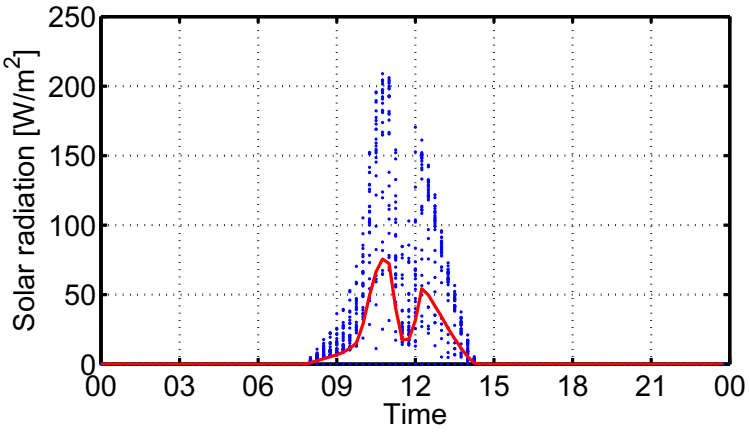


Figure 5.2: Solar load for a classroom labeled as VAV C-2-5 in Figure 2.14.

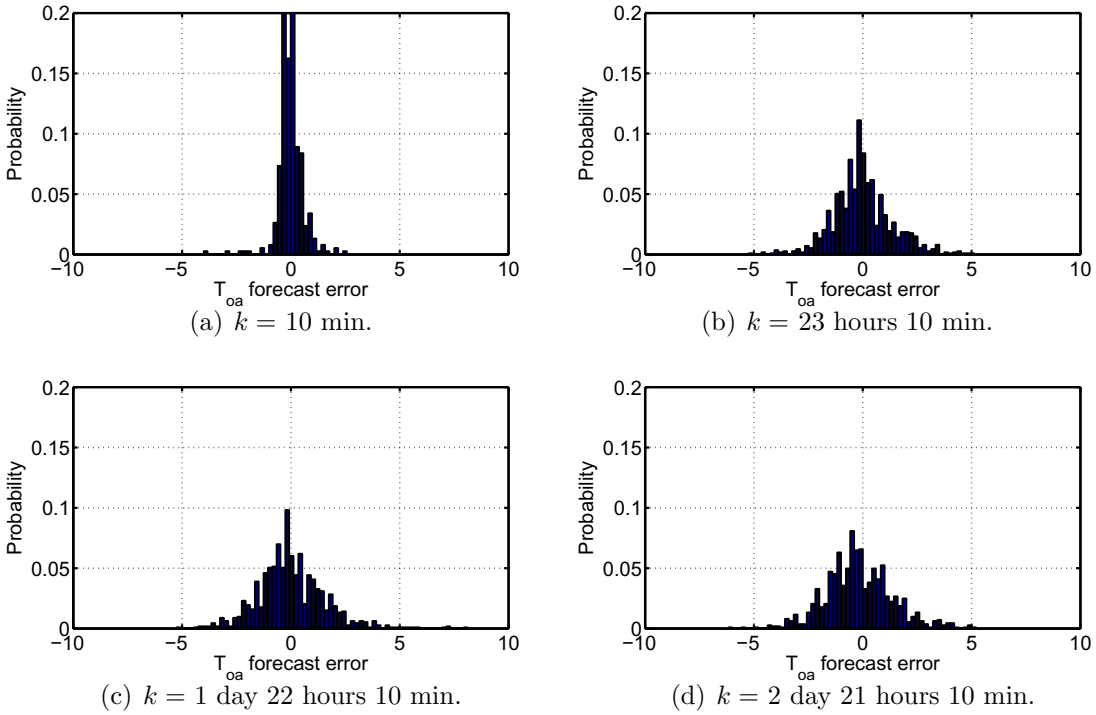
use a periodic solar load $I^i_{k+t|t}$ in zone i prediction model with period of one day. The mean $\hat{I}^i_{k+t|t}$ and the probability density function of the prediction uncertainty $\tilde{I}^i_{k+t|t}$ are calculated from historical data.

The ambient temperature at time $t+k$ predicted at t is

$$T_{oa,k+t|t} = \hat{T}_{oa,k+t|t} + \tilde{T}_{oa,k+t|t} \tag{5.5}$$

where $\hat{T}_{oa,k+t|t}$ is weather forecast obtained from the local weather station, and $\tilde{T}_{oa,k+t|t}$ is weather forecast uncertainty. The weather forecast error is defined as the difference between historical weather measurements and archived weather forecast data, and the probability density function of the ambient temperature prediction uncertainty $\tilde{T}_{oa}$ is modeled from historical weather forecast errors. In our study, the prediction uncertainty $\tilde{T}_{oa}$ is learned from archived weather prediction and measurements from May 20, 2012 to July 07, 2012. Figure 5.3 shows the probability density functions of the weather forecast uncertainty for four selected prediction times k. As expected, the measured ambient temperature forecast uncertainty increases with longer prediction horizon k.

Remark 2 *The approach presented in this section to forecast the building loads and to estimate their uncertainty might sound simplistic. The control design methodology presented*



(a) $k = 10$ min. (b) $k = 23$ hours 10 min.

(c) $k = 1$ day 22 hours 10 min. (d) $k = 2$ day 21 hours 10 min.

Figure 5.3: Ambient temperature prediction error.

in the following section does not depend on the specifics of the forecast algorithms. It only requires probability distribution function of forecasted loads. Different methods could be used to obtain more accurate predictions.

Energy Model

The AHU and VAV components that use energy are dampers, supply fans, and heating coils. The supply fan needs electrical power to drive the system, the heating coils consume the energy of hot water, and the power to drive the dampers is assumed to be negligible. The power models for supply fan and coils have been reported in Chapter 2. In particular, the fan power is approximated as a second order polynomial function of the total supply air mass flow rate ($\dot{m} = \sum_{i\in\mathcal{V}} \dot{m}_s^i$) driven by the fan.

$$P_{fan} = c_0 + c_1\dot{m} + c_2\dot{m}^2, \tag{5.6}$$

where c_0, c_1, c_2 are parameters to be identified by fitting recorded data. Heating and cooling coils are air-water heat exchangers, and the power consumption of the coils are derived from the energy conservation law,

$$P_c = \frac{\sum_{j\in\mathcal{V}} \dot{m}_s^j c_p (T_m - T_c)}{\eta_c \,\mathrm{COP}_c}, \;\; P_h^j = \frac{c_p \dot{m}_s^j (T_s^j - T_c)}{\eta_h \,\mathrm{COP}_h}, \tag{5.7}$$

where P_c (P_h) is the power used by the cooling (heating) coils to deliver supply air with temperature T_s^i, COP_* and η_* are model parameters, T_c is the air temperature after the AHU cooling coil, and T_m is the mixed air temperature before the cooling coil (see Figure 2.2). The mixed air temperature is computed as

$$T_m = \delta T_{oa} + (1-\delta)\frac{\sum_{i\in\mathcal{V}} \dot{m}_s^i T^i}{\sum_{i\in\mathcal{V}} \dot{m}_s^i}, \tag{5.8}$$

where δ is the mixing ratio between the outside air and return air. It is assumed that the return air temperature is a weighted sum of the zone temperatures with weights being the mass flow rates supplied to the corresponding zones. Both δ and T_c can be controlled through the AHU cooling coil and return damper. The total electricity power consumption of the HVAC system at time t then is calculated as

$$P_{tot} = (P_h + P_c + P_{fan}). \tag{5.9}$$

Constraints

The HVAC system is subject to thermal comfort constraints and operational constraints defined next.

C1- $\mathbb{P}\left\{T_k^i \geq \underline{T}_k^i\right\} > 1-\epsilon,\ \mathbb{P}\left\{T_k^i \leq \overline{T}_k^i\right\} > 1-\epsilon,\ \forall i \in \mathcal{V}$. The probability that zone temperatures at time step k are within the comfort bounds is greater than $1-\epsilon$. Comfort bounds $\underline{T}_k^i$, $\overline{T}_k^i$ and allowed violation probability ϵ are design parameters.

C2- $T_{s,k}^i \in \left[\underline{T}_{s,k}^i,\ \overline{T}_{s,k}^i\right], \forall i \in \mathcal{V}$. The supply air temperature is limited by the chilled water temperature through the coils and the physical characteristics of coils.

C3- $\dot{m}_{s,k}^i \in \left[\underline{\dot{m}}_{s,k}^i,\ \overline{\dot{m}}_{s,k}^i\right],\ \forall i \in \mathcal{V}$. Allowed mass flow rate for the supplied air. The lower bounds $\underline{\dot{m}}_{s,k}^i$ are strictly positive to meet minimum ventilation requirement.

C4- $\sum_{i\in\mathcal{V}} \dot{m}_s^i T_c \leq \sum_{i\in\mathcal{V}} \dot{m}_s^i T_m$. The air temperature after the cooling coil T_c cannot be warmer than the mixed air temperature T_m.

C5- $\dot{m}_s^i T_c \leq \dot{m}_s^i T_s^i,\ \forall i \in \mathcal{V}$. The air temperatures across the heating coil can only increase.

Model Summary

The bilinear ARMAX model (5.2) for all zones and the energy model (5.9) can be compactly rewritten as

$$x_{k+1} = f(x_k, u_k^{nl}, w_k), w_k \in \mathcal{W}_k, \tag{5.10a}$$
$$T_k^i = C^i x_k, \forall i \in \mathcal{V}, \tag{5.10b}$$
$$P_{tot,k} = P_{\text{tot}}(x_k, u_k^{nl}, w_k), \tag{5.10c}$$

where $x_k = [T^1_k, T^1_{k-1}, T^1_{k-2}, \ldots, T^{N_v}_k, T^{N_v}_{k-1}, T^{N_v}_{k-2}] \in \mathbb{R}^{n_x \times 1}$ is the system state, $u^{nl}_k = [\dot{m}^1_{s,k}, T^1_{s,k}, \ldots, \dot{m}^{N_v}_{s,k}, T^{N_v}_{s,k}, T_{c,k}, \delta_k] \in \mathbb{R}^{n_u \times 1}$ is the control input, and $w_k = [P^1_{d,k}, \ldots, P^{N_v}_{d,k}, I^1_k, I^1_{k-1}, I^1_{k-2}, \ldots, I^{N_v}_{k-1}, I^{N_v}_{k-2}, T_{oa,k}, T_{oa,k-1}, T_{oa,k-2}] \in \mathbb{R}^{n_d \times 1}$ is the system load. The forecasts of w_k are obtained as discussed in Section 5.2. The state vector x_k collects the state of the ARMAX model (5.2) at every zone of the building. The input vector u_k collects the supply air mass flow rate $\dot{m}_{s,k}$ and the supply air temperature $T_{s,k}$ to each zone, as well as the air temperature after the AHU cooling coil and the mixing ratio between the outside air and return air.

The constraints C1–C5 are compactly written as

$$\mathbb{P}\left\{h^{j\,T} x_k \leq g^j_k\right\} > 1 - \epsilon, \ \forall j \in \mathbb{N}_{N_c} \tag{5.11a}$$

$$G(x_k, u^{nl}_k, w_k) \leq 0, \forall w_k \in \mathcal{W}_k, \tag{5.11b}$$

where $N_c = 2N_v$ is the number of state inequality constraints, $\mathbb{N}_{N_c}$ is the index set $\{0, 1, \ldots, N_c\}$, and v^T is the transpose of a vector v.

5.3 Stochastic Model Predictive Control

Consider the building model (5.10) and its constraints (5.11). We formulate the following stochastic optimization problem with chance constraints for comfort constraints:

$$\min_{X_t, U_t} \sum_{k \in \mathbb{N}_{T-1}} P_{tot}(\hat{x}_{k|t}, \hat{u}^{nl}_{k|t}, \hat{w}_{k|t})\Delta t, \tag{5.12a}$$

subj. to:

$$x_{k+1|t} = f(x_{k|t}, u^{nl}_{k|t}, w_{k|t}), \forall k \in \mathbb{N}_T, \tag{5.12b}$$

$$\mathbb{P}\left\{h^{j\,T} x_{k|t} \leq g^j_k\right\} > 1 - \epsilon, \forall j \in \mathbb{N}_{N_c}, \forall k \in \mathbb{N}_T, \tag{5.12c}$$

$$G(x_{k|t}, u^{nl}_{k|t}, w_{k|t}) \leq 0, \forall w_{k|t} \in \mathcal{W}_{k|t}, \forall k \in \mathbb{N}_{T-1}, \tag{5.12d}$$

$$x_0 = x(t), \tag{5.12e}$$

where $X_t = \{x_{0|t}, \ldots, x_{T|t}\}$ and $U_t = \{u^{nl}_{0|t}, \ldots, u^{nl}_{T-1|t}\}$ are the optimization variables and Δt is the sampling time. We use the notation $v_{t+k|t}$ to denote the value of the variable v at time $t + k$ predicted at time t. Similarly, $\hat{v}_{t+k|t}$ is the expected the value of the random variable v at time $t + k$ predicted at time t.

A stochastic model predictive controller solves problem (5.12) at each time step t. In particular, let $X^\star_t = \{x^\star_{0|t}, \ldots, x^\star_{T|t}\}$ and $U^\star_t = \{u^{nl\,\star}_{0|t}, \ldots, u^{nl\,\star}_{T-1|t}\}$ be the optimal solution of problem (5.12) at time t. Then, the first element of optimal control sequence is implemented to the system (5.10a),

$$u^{nl}_t = u^{nl\,\star}_{0|t}.$$

The optimization problem (5.12) is repeated at $t = t + \Delta t$, with the updated state $x_0 = x_{t+1}$.

The formulation (5.12) has the following features; *(i)* the power consumption at expected values of states, inputs, and disturbances is minimized in (5.12a), *(ii)* chance constraints (5.12c) are used for temperature bounds, *(iii)* constraint C2-C5 are robustly enforced for all admissible disturbances realizations in (5.12d). As it will be clear from the results presented later in this chapter, this formulation will enable complexity reduction (by avoiding the computation of nonlinear probability distribution functions) while reducing the conservatism typically associated to the use of simple linear models, gaussian uncertainty, and robust temperature bound satisfaction.

The next section presents the main contributions of this chapter.

1. Propose a computationally tractable approach to solve stochastic MPC problem while reducing conservatism. We first implement state feedback linearization to linearize the bilinear system model (5.12b), and this enables the computation of the state error distribution using numerical methods. The chance constraints then can be reformulated as deterministic constraints by bounding the distribution tails. The resulting nonconvex optimization problem is solved using Ipopt, a software package for large-scale nonlinear optimization developed by [135].

2. Analyze the complexity and conservatism of discretization method and sampled-based method when computing the tightening offset for the chance constraints (5.12c).

3. Carry out extensive simulation tests to demonstrate the effectiveness of the proposed SMPC scheme compared with alternative MPC designs. The proposed SMPC method is compared with existing SMPC approach that assumes Gussian disturbances [106], and the simulation results show that the proposed SMPC method is less conservative than existing SMPC reported in [106].

Remark 3 *With abuse of notation and for the sake of simplicity, in the rest of the chapter, the notation $x_{k+t|t}$ will be replaced with x_k.*

System Linearization

The bilinear dynamic model (5.12b) is linearized by introducing deterministic virtual inputs u_s^i and u_z^i for each zone i,

$$u_{s,k}^i = \dot{m}_{s,k}^i T_{s,k}^i, \forall i \in \mathcal{V}, \tag{5.13a}$$
$$u_{z,k}^i = \dot{m}_{s,k}^i T_k^i, \forall i \in \mathcal{V}, \tag{5.13b}$$

to obtain the linear model

$$x_{k+1} = Ax_k + Bu_k + Dw_k,\ w_k \in \mathcal{W}_k, \tag{5.14}$$
$$T_k^i = C^i x_k, \forall i \in \mathcal{V}, \tag{5.15}$$

where the new control input vector u_k is $u_k = \left[u^1_{s,k}, u^1_{z,k}, \ldots, u^{N_v}_{s,k}, u^{N_v}_{z,k}, T_{c,k}, \delta_k\right]$. For a given virtual input vector u_k, the input signals u^{nl}_k can be uniquely determined if $T^i \neq 0,\ \dot{m}^i_s \neq 0,\ \forall i \in \mathcal{V}$. Any feasible system trajectory satisfies these two constraints (see Section 5.2).

In model (5.14) w_k is a random variable with bounded support $\mathcal{W}_k$. Model (5.14) implies that the dynamics of state mean $\hat{x}$ and state error $\tilde{x} = x - \hat{x}$ are

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + D\hat{w}_k, \tag{5.16a}$$
$$\hat{T}^i_k = C^i\hat{x}_k, \tag{5.16b}$$
$$\tilde{x}_{k+1} = A\tilde{x}_k + D\tilde{w}_k, \tag{5.16c}$$
$$\tilde{T}^i_k = C^i\tilde{x}_k, \tag{5.16d}$$
$$\tilde{x}_0 = 0. \tag{5.16e}$$

The linear system (5.16) can be used in the optimization problem (5.12) instead of the nonlinear dynamics (5.12b). The new control variables $u^i_{z,k}$ in (5.13) can be interpreted as a deterministic state-feedback gain linking mass air flow rate $\dot{m}^i_s$ and supply air temperature T^i_s to room temperature T^i:

$$\dot{m}^i_{s,k} = u^i_{z,k}\frac{1}{T^i_k}, \forall i \in \mathcal{V}. \tag{5.17a}$$
$$T^i_{s,k} = u^i_{s,k}\frac{1}{\dot{m}^i_{s,k}} = \frac{u^i_{s,k}}{u^i_{z,k}}T^i_k, \forall i \in \mathcal{V}. \tag{5.17b}$$

Therefore, while the optimization in the virtual inputs u^i_s and u^i_z is over open-loop policies [6], the variable substitution (5.13) provides state-feedback policies. These advantages come with a price. In fact, the energy models in Section 5.2 and constraints defined in Section 5.2 have to be rewritten as a function of the virtual inputs u_k. By using the change of control variables (5.13) the energy models for supply fan (5.6), cooling and heating coils (5.7) are rewritten as

$$P_{fan} = c_0 + c_1\sum_{i\in\mathcal{V}}\frac{u^i_z}{T^i} + c_2\left(\sum_{i\in\mathcal{V}}\frac{u^i_z}{T^i}\right)^2, \tag{5.18}$$
$$P_h = p_h\sum_{i\in\mathcal{V}}\left(u^i_s - \frac{u^i_z}{T^i}T_c\right), \tag{5.19}$$
$$P_c = p_c\sum_{i\in\mathcal{V}}\left(\frac{u^i_z}{T^i}\delta T_{oa} + (1-\delta)u^i_s - \frac{u^i_z}{T^i}T_c\right), \tag{5.20}$$
$$P_{tot} = (P_{fan} + P_c + P_h) = P_{tot}(x,u,w). \tag{5.21}$$

The constraints on supply air temperature C2 are rewritten as the following bilinear

constraints,

$$\max_{\tilde{w}_0,\ldots,\tilde{w}_{k-1}} u^i_{s,k}T^i_k = u^i_{s,k}C^i\hat{x}_k + u^i_{s,k}\tilde{T}^i_{k,\max} \leq \overline{T}^i_s u^i_z,\ \forall i \in \mathcal{V} \tag{5.22}$$

$$\min_{\tilde{w}_0,\ldots,\tilde{w}_{k-1}} u^i_{s,k}T^i_k = u^i_{s,k}C^i\hat{x}_k + u^i_{s,k}\tilde{T}^i_{k,\min} \geq \underline{T}^i_s u^i_z,\ \forall i \in \mathcal{V}. \tag{5.23}$$

The constraints on supply air mass flow rate C3 are rewritten as the following robust constraints

$$u^i_{z,k} \leq \min_{\tilde{w}_{k\in\mathbb{N}_{k-1}}} \overline{\dot{m}}^i_{s,k}T^i_k = \overline{\dot{m}}^i_{s,k}C^i\hat{x}_k + \overline{\dot{m}}^i_{s,k}\tilde{T}^i_{k,\min},\ \forall i \in \mathcal{V}, \tag{5.24}$$

$$u^i_{z,k} \geq \max_{\tilde{w}_{k\in\mathbb{N}_{k-1}}} \underline{\dot{m}}^i_{s,k}T^i_k = \underline{\dot{m}}^i_{s,k}C^i\hat{x}_k + \underline{\dot{m}}^i_{s,k}\tilde{T}^i_{k,\max},\ \forall i \in \mathcal{V}, \tag{5.25}$$

where

$$\tilde{T}^i_{k,\min} = \min_{\tilde{w}_0,\ldots,\tilde{w}_{k-1}} C^i\left(\tilde{x}_0 + \sum_{i=0}^{k-1} A^k D\tilde{w}_{k-1-i}\right), \tag{5.26}$$

$$\tilde{T}^i_{k,\max} = \max_{\tilde{w}_0,\ldots,\tilde{w}_{k-1}} C^i\left(\tilde{x}_0 + \sum_{i=0}^{k-1} A^k D\tilde{w}_{k-1-i}\right). \tag{5.27}$$

The robust constraints C4 on the air temperature after the cooling coil in AHU T_c can be conservatively robustified as follow

$$\begin{aligned}
&\max_{\tilde{w}_0,\ldots,\tilde{w}_{k-1}} \left(\sum_{i\in\mathcal{V}}\left[\frac{u^i_{z,k}}{T^i_k}(T_{c,k} - \delta_k T_{oa,k}) + (\delta_k - 1)u^i_{s,k}\right]\right) \\
&\leq \sum_{i\in\mathcal{V}}\left[\frac{u^i_{z,k}}{C^i\hat{x}_k + T^i_{k,\min}}T_{c,k} - \delta_k \underline{T}_{oa,k}\frac{u^i_{z,k}}{C^i\hat{x}_k + T^i_{k,\max}} + (\delta_k - 1)u^i_{s,k}\right] \\
&\leq 0,
\end{aligned} \tag{5.28}$$

where $\underline{T}_{oa,k}$ is the lower bound of $T_{oa,k}$. The constraints C5 on the air temperature after the cooling coil in AHU T_c become

$$u^i_z T_c \leq \min_{\tilde{w}_0,\ldots,\tilde{w}_{k-1}} u^i_s T^i = u^i_s C^i\hat{x}_k + u^i_s\tilde{T}^i_{k,\min},\ \forall i \in \mathcal{V}. \tag{5.29}$$

In summary, the constraints (5.24)–(5.29) can be compactly rewritten as the following deterministic nonlinear constraint

$$G^{FL}(\hat{x}_k, u_k, \hat{w}_k) \leq 0, \forall k \in \mathbb{N}_T. \tag{5.30}$$

Solving Chance Constraints

In this section, we show the approach we used to transform the chance constraints (5.12c) to convex deterministic ones with tightening offsets [36]. We denote by π_k^j the univariate random variable ${h^j}^T x_k$. The chance constraints (5.12c) then can be reformulated as deterministic ones [39]

$$\hat{\pi}_k^j \leq g_k^j - \alpha_k^j, \ \forall j \in \mathbb{N}_{N_c}, \ \forall k \in \mathbb{N}_T, \tag{5.31}$$

where $\hat{\pi}_k^j$ is the expected value of π_k^j, and the offset α_k^j is calculated as follows

$$\int_{\alpha_k^j}^{\infty} \mathrm{pdf}_{\tilde{\pi}_k^j}(y)dy = 1 - \mathrm{cdf}_{\tilde{\pi}_k^j}(\alpha_k^j) = \epsilon, \ \forall j \in \mathbb{N}_{N_c}, \ \forall k \in \mathbb{N}_T, \tag{5.32}$$

where $\mathrm{pdf}_{\tilde{\pi}_k^j}(y)$ is the probability density function of $\tilde{\pi}_k^j = \pi_k^j - \hat{\pi}_k^j$ evaluated at y, and $\mathrm{cdf}_{\tilde{\pi}_k^j}(\alpha_k^j) = \mathbb{P}\left\{\tilde{\pi}_k^j \leq \alpha_k^j\right\}$ is the cumulative density function of $\tilde{\pi}_k^j$ evaluated at α_k^j. A solution to (5.31) is also a solution to (5.12c).

The univariate random variable $\tilde{\pi}_k^j$ can be expressed as a linear function of previous disturbances

$$\tilde{\pi}_k^j = {h^j}^T \tilde{x}_k = {h^j}^T \tilde{x}_0 + \sum_{i=0}^{k-1} {h^j}^T A^k D \tilde{w}_{k-1-i}. \tag{5.33}$$

If the measurement of the state x_0 is exact, i.e. $\tilde{x}_0 = 0$, then equation (5.33) is compactly written as

$$\tilde{\pi}_k^j = {\phi_k^j}^T \tilde{W}_{k-1}, \ \forall j \in \mathbb{N}_{N_c}, \ \forall k \in \mathbb{N}_T, \tag{5.34}$$

where $\tilde{W}_{k-1} = \left[\tilde{w}_0^T, \tilde{w}_1^T; \ldots; \tilde{w}_{k-1}^T\right]^T$ is a column vector of length $N_w = n_d\, k$. The disturbances error $\tilde{W}_{k-1}$ is finitely-supported with lower bound $\underline{\tilde{W}}_{k-1}$ and upper bound $\overline{\tilde{W}}_{k-1}$.

Next we present and compare two numerical methods to compute the offset α_k^j computed by (5.32) given the probability density function for the disturbances error $\tilde{w}$.

Discrete Convolution Integral Method

If the disturbance errors $\tilde{w}$ are statistically independent, the probability density function of $\tilde{\pi}_k^j$ in (5.34) can be computed recursively as follow

$$\mathrm{pdf}_{\tilde{\pi}_k^j}^{i+1} = \mathrm{pdf}_{\tilde{\pi}_k^j}^{i} * \mathrm{pdf}_{w_{(i+1)}}, \tag{5.35}$$

$$\mathrm{pdf}_{\tilde{\pi}_k^j}^{0} = \mathrm{pdf}_{w_{(0)}}, \tag{5.36}$$

$$\mathrm{pdf}_{\tilde{\pi}_k^j} = \mathrm{pdf}_{\tilde{\pi}_k^j}^{N_w}, \tag{5.37}$$

where $\mathrm{pdf}_{w_{(n)}}$ is the probability density function for n^{th} element of $\mathrm{diag}(\phi_k^j)\tilde{W}_{k-1}$ in (5.34). The operator $*$ denotes the convolution integral,

In general, there is no analytical solution for convolution integrals (5.35). Numerical algorithms to approximate the convolution integrals (5.35) can be found in [79, 80, 33]. We used the approach in [79]. The probability density function $\text{pdf}_{w_{(n)}}$ in (5.35) is first discretized over N_s evenly distributed samples from $[\underline{\pi},\ \overline{\pi}]$ with discretization interval $\Delta\pi = \frac{\overline{\pi}-\underline{\pi}}{N_s}$.

The upper bound $\overline{\pi}$ and lower bound $\underline{\pi}$ of the discretization range is computed as

$$\overline{\pi}_T^j = \max_{\underline{\tilde{W}}_{T-1} \le \tilde{W}_{T-1} \le \overline{\tilde{W}}_{T-1}} (\phi_T^j)^T \tilde{W}_{T-1},\ \forall j \in \mathbb{N}_{N_c}, \tag{5.38a}$$

$$\underline{\pi}_T^j = \min_{\underline{\tilde{W}}_{T-1} \le \tilde{W}_{T-1} \le \overline{\tilde{W}}_{T-1}} (\phi_k^j)^T \tilde{W}_{T-1},\ \forall j \in \mathbb{N}_{N_c}, \tag{5.38b}$$

$$\overline{\pi} = \max_{j \in \mathbb{N}_{N_c}} \overline{\pi}_T^j,\ \underline{\pi} = \min_{j \in \mathbb{N}_{N_c}} \underline{\pi}_T^j. \tag{5.38c}$$

Within the discretization interval, the probability density function $\text{pdf}_{w_{(n)}}$ is kept constant

$$\text{pdf}_{w_{(n)}}[k] = \text{pdf}_{w_{(n)}}(\underline{\pi}_k^j + k\ \Delta\pi),\ \forall k \in \mathbb{N}_{N_s},\ \forall n \in \mathbb{N}_{N_w}. \tag{5.39}$$

The convolution integral (5.35) is then approximated by discrete convolution,

$$\left(\text{pdf}^i_{\tilde{\pi}_k^j} * \text{pdf}_{w_{(i+1)}}\right)[k] = \sum_{l \in \mathbb{N}_{Ns}} \text{pdf}^i_{\tilde{\pi}_k^j}[l]\, \text{pdf}_{w_{(i+1)}}[k-l]\Delta\pi, \forall k \in \mathbb{N}_{N_s}. \tag{5.40}$$

The probability density function $\text{pdf}_{\tilde{\pi}_k^j}$ in (5.35) is then obtained by approximating each convolution integral in (5.35) using discrete convolution (5.40). Once $\text{pdf}_{\tilde{\pi}_k^j}$ in (5.35) is computed, the tightening offsets α_k^j in (5.32) are computed as

$$\alpha_k^j = \underline{\pi}_k^j + i_{\min}\Delta\pi,\ i_{\min} = \min\{i \in \mathbb{N}_{N_s} \mid \sum_{n=0}^{n=i} \text{pdf}_{\tilde{\pi}_k^j}[n] \ge 1-\epsilon\}. \tag{5.41}$$

The discrete convolution integral method to compute the tightening offsets α_k^j in (5.31) is summarized as follow.

Algorithm 3 (Discrete method)

Initial:	*Compute the upper and lower bounds of discretization range as in (5.38),and select the sample number N_s.*
Step 1:	*Discretize the probability density function $w_{(n)},\ \forall n \in \mathbb{N}_{N_w}$ as in (5.39).*
Step 2:	*Approximate the probability density function $pdf_{\tilde{\pi}_k^j}$ by discrete convolutions as in (5.40).*
Step 3:	*Compute the tightening offsets α_k^j as in (5.41).*

Sample-based Method

The sampling-based approach[1] to single-stage chance constrained problems (CCP) has been studied in [23], [22], [27], while recently it has been extended to multiple-stage chance constrained problems [120]. It provides an alternative solution to the computation of the offset α_k^j in (5.32). The sampling-based approach transforms the chance constraints (5.12c) into a deterministic ones by generating for each individual constraint a large number $N_{s,k}^j$ of independent identically distributed disturbance samples $\tilde{w}_k^{j,1}, \tilde{w}_k^{j,2}, ..., \tilde{w}_k^{j,N_{s,k}^j}$ (usually referred to as scenarios) according to the probability density function of the disturbance. We use (5.33) and replace the chance constraints (5.32) with

$$h^{j^T}\hat{x}_k + \sum_{i=0}^{k-1} h^{j^T} A^k D\tilde{w}_{k-1-i}^{j,l} \leq g_k^j, \ \forall j \in \mathbb{N}_{N_c}, \forall k \in \mathbb{N}_T, , \forall l \in \mathbb{N}_{N_{s,k}^j}. \tag{5.42}$$

For a sufficiently large $N_{s,k}^j$, the satisfaction of (5.42) guarantees that each chance constraint in (5.12c) is satisfied with the high probability $1 - \beta_k^j$.

Such an approach introduces significant conservatism, since a solution satisfying the constraints for "many" disturbance realizations is close to a worst-case solution (i.e. ϵ in (5.12c) is close to zero). In order to reduce the conservatism of the solution, a certain number of generated samples $N_{r,k}^j$ can be removed from the set of samples without a significant loss of reliability (1-β_k^j) of the solution. The link between the number of original samples $N_{s,k}^j$, the removed ones $N_{r,k}^j$, the allowable violation probability ϵ and the reliability parameter β_k^j has been studied in [23], [22], [27], [120]. Here we use the inequality [120]:

$$\binom{N_{r,k}^j + \zeta_k^j - 1}{N_{r,k}^j} \sum_{l=0}^{N_{r,k}^j + \zeta_k^j - 1} \binom{N_{s,k}^j}{l} \epsilon^l (1-\epsilon)^{N_{s,k}^j - l} \leq \beta_k^j, \tag{5.43}$$

where ζ_k^j is the support rank of each individual constraint in (5.12c). The support rank of a chance constraint is the dimension of the vector space spanned by the constraint. In summary the chance constraint (5.12c) is transformed into

$$h^{j^T}\hat{x}_k \leq g_k^j - \alpha_k^{j,l}, \ \forall l \in \mathbb{I}^{N_{s,k}^j} \setminus \mathbb{I}^{N_{r,k}^j}, \forall k \in \mathbb{N}_T, \text{ and } \forall j \in \mathbb{N}_{N_c}, \tag{5.44}$$

with offsets $\alpha_k^{j,l}$ calculated as:

$$\alpha_k^{j,l} = \sum_{i=0}^{k-1} h^{j^T} A_{cl}^i D\tilde{w}_{k-1-i}^l, \tag{5.45}$$

[1]This section is partially extracted from Jadranko Matuško's contribution to the paper titled "Scenario-Based Approach to Stochastic Linear Predictive Control" in IEEE Conference on Decision and Control, December 2012.

where $\mathbb{I}^{N^j_{s,k}} = \{1, ..., N^j_{s,k}\}$ and $\mathbb{I}^{N^j_{r,k}} \subseteq \mathbb{I}^{N^j_{s,k}}$ is a set containing the indices of $N^j_{r,k}$ removed samples.

Since the support rank of the constraints (5.44) is $\zeta^j_k = 1$, inequality (5.43) reduces to:

$$\sum_{l=0}^{N^j_{r,k}} \binom{N^j_{s,k}}{l} \epsilon^l (1-\epsilon)^{N^j_{s,k}-l} \leq \beta^j_k. \tag{5.46}$$

In order to ensure that the solution to the overall optimization problem is a feasible solution to the original optimization problem with high probability $1-\beta$, the β^j_k for the j-th constraint and the k time index have to will be selected to satisfy

$$\sum_{j \in \mathbb{N}_{N_c}} \sum_{k \in \mathbb{N}_T} \beta^j_k < \beta. \tag{5.47}$$

From the equation (5.44), one can easily notice that the optimal sample removal strategy corresponds to the removal of the $N^j_{r,k}$ largest offsets $\alpha^{j,l}_k$ (notice that for fixed $j \in \mathbb{N}_{N_c}$ and $k \in \mathbb{N}_T$ all constraints in (5.44) are parallel), resulting in the least conservative approximation of the corresponding individual chance constraint in (5.12c). Finally, the offsets α^j_k in (5.32) are calculated as:

$$\forall k \in \mathbb{N}_T, \text{ and } \forall j \in \mathbb{N}_{N_c}, \alpha^j_k = \max(\alpha^{j,l}_k),\ l \in \mathbb{I}^{N^j_{s,k}} \setminus \mathbb{I}^{N^j_{r,k}} \tag{5.48}$$

The algorithm to compute $\mathbb{I}^{N^j_{s,k}} \setminus \mathbb{I}^{N^j_{r,k}}$ in (5.44) and the corresponding offsets α^j_k is summarized next.

Algorithm 4 (Sampling-based method)

Initial: *$\forall k \in \mathbb{N}_T$ and $\forall j \in \mathbb{N}_{N_c}$ compute the number of samples to generate $N^j_{s,k}$ and to remove $N^j_{r,k}$ that satisfy inequalities (5.46) and (5.47).*

Step 1: *$\forall k \in \mathbb{N}_T$, $\forall j \in \mathbb{N}_{N_c}$ and $\forall l \in \mathbb{N}^j_{s,k}$* ***do***

S1.1 Generate the set of samples $\tilde{\mathcal{W}}^j_k = \{\tilde{w}^{j,l}_k\}$.

S1.2 Calculate the set of the offsets $\{\alpha^{j,l}_k\}$ using (5.45).

S1.3 Sort the set of the offsets to obtain ordered set of the offsets $\{\bar{\alpha}^{j,l}_k\}$

Step 2: *$\forall k \in \mathbb{N}_T$ and $\forall j \in \mathbb{N}_{N_c}$* ***do***

S2.1 Remove $N^j_{r,k}$ largest offsets from the set $\{\bar{\alpha}^{j,l}_k\}$.

S2.2 Calculate the offsets α^j_k using (5.48).

Optimization Problem

With the change of variables presented in Section 5.3, and the transformation of chance constraints in Section 5.3, the stochastic MPC problem (5.12) is transformed into the deterministic nonlinear optimization problem

$$\min_{\hat{x}_{k\in\mathbb{N}_T}, u_{k\in\mathbb{N}_{T-1}}} \sum_{k\in\mathbb{N}_{T-1}} P_{\text{tot}}(\hat{x}_k, u_k, \hat{w}_k)\Delta t, \tag{5.49a}$$

subj. to:

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + D\hat{w}_k, \forall k \in \mathbb{N}_T, \tag{5.49b}$$

$$h^{j^T}\hat{x}_k \leq g_k^j - \alpha_k^j,\ \forall j \in \mathbb{N}_{N_c}, \forall k \in \mathbb{N}_T, \tag{5.49c}$$

$$G^{FL}(\hat{x}_k, u_k, \hat{w}_k) \leq 0, \forall k \in \mathbb{N}_{T-1}, \tag{5.49d}$$

$$\hat{x}_0 = x(0). \tag{5.49e}$$

We compute the optimal solution $u^*_{k\in\mathbb{N}_T}$ to Problem (5.49) by using Ipopt [135]. We remark that the steps presented in the previous section are crucial for obtaining a resulting problem (5.49) which is computationally tractable and whose solution is not too conservative. In the next section we will provide metrics from conservatism and computational tractability. We will also present results which confirm the effectiveness of our approach.

5.4 Results

The stochastic MPC algorithm solving (5.49) at every time step k is analyzed in this section. For the class of building HVAC systems studied in this thesis, we will investigate the following questions:

(i) Does one need a stochastic MPC formulation or nominal MPC with expected forecasts provides "good" results?

(ii) Is there value in using nonlinear probability distribution functions or Gaussian approximations work well and what is the price one has to pay for it?

(iii) Can the proposed approach be implemented on large scale buildings?

(iv) Should one transform the chance constraints by using convolution integrals (Section 5.3) or sample-based methods (Section 5.3)?

Although the answers depend on the HVAC system and the level of uncertainty, the next sections will shed some light on the aforementioned questions by simulation and experiments. In both simulation and experiments study, models and probability distribution functions of forecasts uncertainty are generated by using measured historical data.

Bancroft Library Simulations

Simulation Setup

We present closed loop simulation of several MPC algorithms for the second floor of the Bancroft library at UC Berkeley. The second floor plan of the Bancroft library is depicted in Figure 2.14 and consists of classrooms, offices, and conference rooms. The HVAC system has a single duct configuration shown in Figure 2.2, and it is controlled and monitored using WebCTRL® developed by Automated Logic Corporation.

The original stochastic nonlinear control problem (5.12) is not real-time feasible for this system by using current desktop capabilities. The goal of this section is to study the compromise between complexity, performance, and conservatism of different approximations to (5.12), including the one proposed in this chapter.

We call the control logic implemented by WebCTRL® the baseline control (BC). The BC has been fine-tuned by professional and implements a "trim and respond" control algorithm. Details on the BC algorithm can be found in [78]. The main idea is to control the heating coil valve position command and the airflow set point as a function of the difference between zone temperature and the bounds on comfort level. When the zone temperature is within the comfort range, the VAV box maintains the minimum ventilation level. If a zone violates the thermal comfort upper-bound, a cooling request is triggered. If a zone violates the thermal comfort lower-bound, a heating request is triggered. The total number of cooling requests and heating requests is then used to control the AHU unit.

We compare the measured BC performance with the following MPC algorithms.

L1 Perfect MPC control (PMPC). A model predictive controller with perfect knowledge of disturbance prediction, i.e. the disturbance prediction error $\tilde{w}_t = 0$, and the mean value of the predicted disturbance over the horizon $\hat{w}_t$ is equal to the future disturbance realization. This is not physically implementable but will be used as reference for what could be potentially achieved without uncertainty.

L2 Certainty equivalent MPC (CMPC). A nominal model predictive controller which uses the mean of the predicted disturbance $\hat{w}_t$ (modeled in Section 5.2). Clearly, in closed-loop simulations the nominal disturbance prediction might be different from the actual disturbance realization.

L3 Stochastic MPC (ESMPC) solving (5.49) with chance constraints approximated using the discrete convolution method in Section 5.3.

L4 Stochastic MPC (SSMPC) solving (5.49) with chance constraints approximated using the sample-based method in Section 5.3.

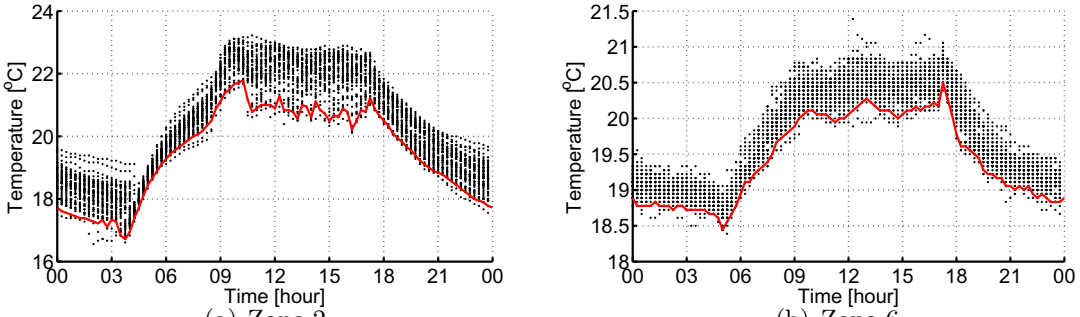
L5 Stochastic MPC (GSMPC) with disturbances probability density function approximated as Gaussian distributions [106].

We focus on eight zones in the second floor of Bancroft library from December 1, 2011 to February 1, 2012. The eight zones are subject to negligible thermal interaction between each other. The zone model parameters are identified by linear regression using historical weekend data from December 2011 to February 2012.

In our simulations, the control parameters for the SMPC controllers (5.12) are set as follow. The control sampling time Δt is 15 minutes, the comfort constraint is allowed to violate with a chance of $\varepsilon = 5\%$, the prediction horizon is $T = 20$ (5 hours).

The thermal comfort constraints ($\overline{T}_t^i$ and $\underline{T}_t^i$) are zone dependent and have a period of one day. Rather than using the original system comfort constraints, we will use what the BC actually achieved. This approach allows us to properly compare the MPC and BC control performances. In particular, for each zone, the comfort constraints are defined as the 95% envelop of the zone temperatures controlled by BC from December 2011 to February 2012. In other words, the chosen thermal comfort lower bounds are violated by the BC controller with probability of 5% at each time instant. Our simulation focuses HVAC systems (Figure 2.2) operating on heating seasons, thus only the lower comfort bounds are of interest, and the upper bounds are always satisfied for well-controlled HVAC systems. We also remark that by enlarging these constraints, any of the five MPC controllers will improve its performance compared to the BC.

Figure 5.4(a) and Figure 5.4(b) illustrate the comfort bounds for two of the eight zones in Bancroft library at UC Berkeley. The dots in Figure 5.4 represents the daily historical zone temperatures, and the solid lines are the computed comfort bounds (5% of the dots are outside the comfort sets at each time step).



(a) Zone 2 (b) Zone 6

Figure 5.4: Thermal comfort bounds.

The constraints on supply air mass flow rate and supply air temperature are also learned from historical data from December 2011 to February 2012. The upper bounds and lower bounds of supply air mass flow rate are computed as the point-wise max and min of the historical profiles. The lower bounds of airflow rates guarantee minimum ventilation levels during occupied hours. The maximum achievable supply air temperature is limited by the hot water temperature through heating coils and the physical characteristics of coils. It is assumed to be constant for each VAV box.

Comparison Metrics

The performance of the model predictive controllers L1–L5 is evaluated by closed-loop simulation with system model (5.10a) for fifty five days. In the simulation, the disturbance realizations are recorded measurements from December 2011 to February 2012. Three metrics are proposed to compare L1–L5 with the baseline control logic.

- Closed-loop energy savings compared to baseline control

$$S^{\diamond} = \frac{E_{\text{tot}}^{\text{BC}} - E_{\text{tot}}^{\diamond}}{E_{\text{tot}}^{\text{BC}}}, \tag{5.50}$$

  where $S^{\diamond}$ is the energy saving for controller $\diamond$, and E_{tot} is the total energy consumption

$$E_{tot} = \sum_{t_s}^{t_f} P_{\text{tot}}\left(x_t, u_t^*, w_t\right) \Delta t, \tag{5.51}$$

  where t_s is the simulation start time, t_f is the simulation end time, u_t^* is the implemented control input at time t, and w_t is the disturbance realization at time t in closed-loop simulation.

- Total comfort improvement compared to baseline control

$$\Delta^{\diamond} = \frac{V_{\text{tot}}^{\text{BC}} - V_{\text{tot}}^{\diamond}}{V_{\text{tot}}^{\text{BC}}}, \tag{5.52}$$

  where $\Delta^{\diamond}$ is the comfort improvement for controller $\diamond$, and V_{tot} is the total comfort violation

$$V_{tot} = \sum_{t_s}^{t_f} \sum_{i \in \mathcal{V}} \max(\underline{T}^i(t) - T^i(t), 0) \Delta t. \tag{5.53}$$

- Thermal efficiency of the HVAC system η

$$\eta = \frac{E_{tot}^{thermal}}{E_{tot}}, \tag{5.54}$$

  where $E_{tot}^{thermal}$ is the total thermal energy delivered by HVAC system, defined as

$$E_{tot}^{thermal} = \sum_{t_s}^{t_f} \left(P_c \eta_c \text{COP}_c + P_h \eta_h \text{COP}_h\right) \Delta t. \tag{5.55}$$

Table 5.1: Simulation results.

Controller	PMPC	CMPC	GSMPC	ESMPC	SSMPC
Energy savings S (%)	22.54	22.43	20.39	20.33	19.93
comfort improvement Δ (%)	96.25	-121.91	17.26	88.07	93.97

Simulation Results

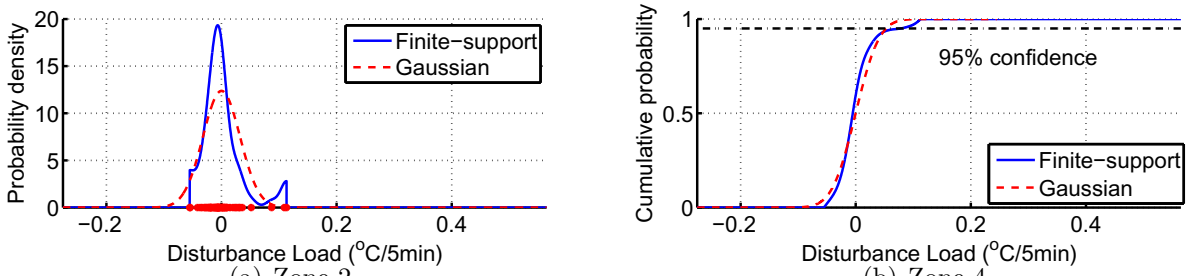
Table 5.1 summarizes the results of our tests. Table 5.1 compares the comfort improvement Δ and the total energy savings S. The tradeoff between performance, conservatism, and complexity is clear. With perfect knowledge of future predictions, PMPC presents the highest comfort improvement while achieving maximum energy savings of 22.54%. The simplest MPC controller to implement, CMPC, achieves comparable energy savings of 22.43%. However, the zone temperature regulated by CMPC violates the comfort constraints 121.17% more than the baseline control. CMPC ignores the disturbance uncertainty at the design stage and this lead to the incapability of the CMPC to maintain the comfort constraints.

The performance of the ESMPC and SSMPC proposed in this paper is not too far from the PMPC. This confirms the effectiveness of our approach. We notice that the ESMPC consumes less energy than SSMPC at the cost of more comfort violations. This is due to a more conservative approximation of chance constraints in the SSMPC approach presented in Algorithm 3 compared to sample-based algorithm 4.

GSMPC is computationally simpler than ESMPC and SSMPC. Compared to SSMPC and ESMPC proposed in Section 5.3, GSMPC achieves similar energy savings while violating more comfort constraints. The coarse Gaussian approximations of disturbance is the reason for this. The approximation error of load probability distribution function is illustrated in Figure 5.5. In particular, Figure 5.5(a) depicts the load Gaussian model (dashed lines) compared to load model which uses finitely-supported density approximation (solid lines) using Gaussian kernels [122]. Figure 5.5(b) shows the cumulative density function of the PDF in Figure 5.5(a). The horizontal dash-dotted line Figure 5.5(b) indicates the 95% confidence level. With a 95% confidence level, the cumulative density function of the Gaussian model underestimates the tail. This leads to an underestimation of the tightening offsets α in (5.32) and thus a higher probability of comfort violations than that specified by chance constraints (5.12c).

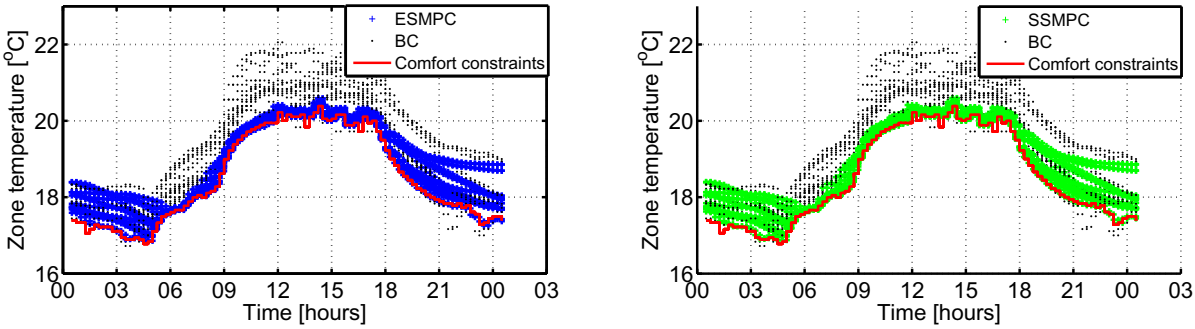
In ESMPC and SSMPC controllers the energy savings compared to baseline control can be explained as follows. The main reason is that the zone temperatures controlled by ESMPC and SSMPC are closer to the lower bounds (this can be observed in Figure 5.6). The dots in Figure 5.6(a) and 5.6(b) are the zone temperatures controlled by BC, the cross markers in Figure 5.6(a) and 5.6(b) represent the zone temperatures controlled by ESMPC and SSMPC, respectively.

The ESMPC and SSMPC controllers also optimize the combination of supply airflow rate and supply air temperature so that the required heating energy is delivered with minimum energy consumption. Table 5.2 listed the thermal efficiency of the HVAC system controlled



(a) Zone 2. (b) Zone 4.

Figure 5.5: Gaussian approximations for disturbances.



(a) Explicit stochastic MPC. (b) Sample based stochastic MPC.

Figure 5.6: Daily temperature for Zone 2.

by L1–L5. It can be noticed that the coordination aspects of the energy-savings in MPC is minimal, and the BC controller has similar performance.

Table 5.2: Thermal efficiency of HVAC systems.

BC	PMPC	CMPC	ESMPC	SSMPC	GSMPC
80.08%	80.38%	80.30%	80.37%	80.42%	80.45%

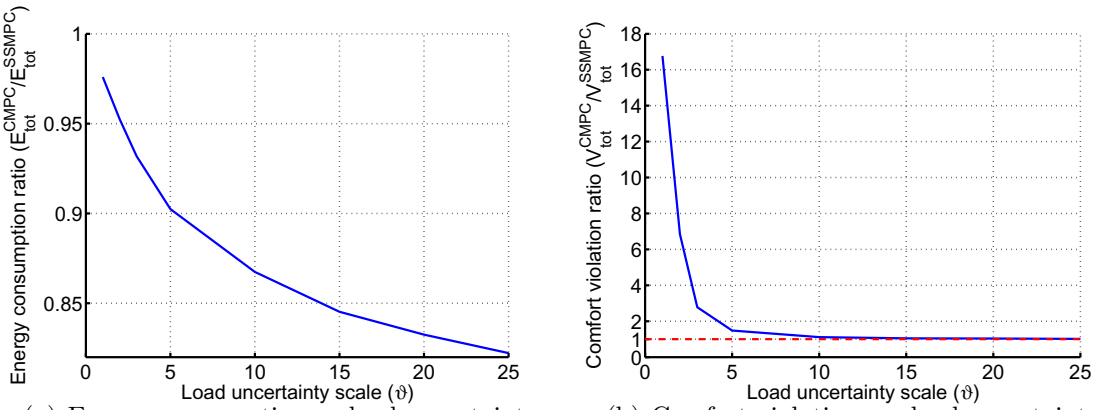
The results in Table 5.1 show the advantages of stochastic MPC formulation compared to nominal MPC with expected forecasts. Next we show that this depends on the uncertainty level. We compare the energy consumption and comfort violation of CMPC and SSMPC with increasing level of load uncertainty. At the design stage, CMPC ignores the uncertainty of the load prediction, and only consider the mean value of predictions. On the other hand, SSMPC seeks control signals to guarantee the level of comfort satisfaction while respecting the uncertain load predictions. In this section, the load uncertainty level is scaled by the parameter ϑ as follow. For SSMPC, we scale the independent disturbance samples $\{\tilde{w}_k^{j,1}, \tilde{w}_k^{j,2}, ..., \tilde{w}_k^{j,N_{s,k}^j}\}$ extracted from the probability density function of the load uncertainty as $\{\vartheta\tilde{w}_k^{j,1}, \vartheta\tilde{w}_k^{j,2}, ..., \vartheta\tilde{w}_k^{j,N_{s,k}^j}\}$. The scaled set of samples is used to compute the tightening

offsets in Algorithm 4. For CMPC, the resulting control policy is independent from the scaling of load uncertainty as CMPC only takes into account the mean value of load prediction at design stage. The load realizations in close-loop simulation have to be scaled accordingly,

$$w(t)_\vartheta = \hat{w}(t) + \vartheta(w(t) - \hat{w}(t)), \tag{5.56}$$

where $w(t)_\vartheta$ is the scaled load realization, $w(t)$ is the original load realization, and $\hat{w}(t)$ is the weekly mean load computed as Figure 5.1 in Section 5.2.

Figure 5.7 reports the simulation results of CMPC and SSMPC for eight zones in the second floor of Bancroft library from December 1, 2011 to February 1, 2012. Figure 5.7(a) shows the energy consumption ration $E_{tot}^{CMPC}/E_{tot}^{SSMPC}$ with increasing level of load uncertainty, and Figure 5.7(b) shows the comfort violation ration $V_{tot}^{CMPC}/V_{tot}^{SSMPC}$ with increasing level of load uncertainty. It is observed that with low level of load uncertainty ($\vartheta < 5$), SSMPC is able to improve the comfort satisfaction compared to CMPC while consuming comparable energy. However, for load uncertainty scaling factor $\vartheta > 5$, SMPC loses its advantage, and CMPC shows comparable performance to SSMPC despite its simplicity at design stage.



(a) Energy consumption vs load uncertainty (b) Comfort violations vs load uncertainty

Figure 5.7: Performance of CMPC and ESMPC with varying load uncertainty.

Complexity Analysis for SSMPC and ESMPC

This section focuses on the complexity of SSMPC and ESMPC. In particular we focus on the complexity of the transformation of the chance constraints (5.12c) into the corresponding deterministic ones (5.31). Computational demands of Algorithms 3 and 4 are analyzed with respect to the three problem size parameters: number of thermal zones N_v, prediction horizon length T and number of samples N_s.

The discrete convolutions in Step 2 are the most computational demanding operations in Algorithm 3. The complexity of discrete convolution is $\mathcal{O}(N_s^2)$, where N_s is the number of samples used to discretize the probability density functions in (5.40). The number of discrete

convolutions needed to approximate the probability density function $\text{pdf}_{\tilde{\pi}_k^j}$ is a linear function of $N_v T$, where N_v is the number of zones and T is the number of prediction steps. Algorithm 3 is repeated for $N_c = 2N_v T$ number of chance constraints in (5.12c), thus the complexity of the discrete method to compute the offsets in (5.12) is of the order $\mathcal{O}(N_s^2 N_v^2 T^2)$.

If we assume that the N_v thermal zones are, in the worst case, fully coupled, Algorithm 4 has complexity $\mathcal{O}(N_v^2)$ (see Step S1.2 of Algorithm 4). Similarly, the algorithm complexity as a function of the prediction horizon T is $\mathcal{O}(T^2)$. With respect to the number of samples N_s, the most computationally demanding step of Algorithm 4 is sorting the offsets $\alpha_k^{j,l}$ in (5.45). This has complexity of $\mathcal{O}(N_s \log N_s)$, if a quick-sort algorithm is used. Taking all above mentioned into account we can conclude that overall complexity of Algorithm 4 is $\mathcal{O}(N_s \log N_s N_v^2 T^2)$ while the complexity of Algorithm 4 is $\mathcal{O}(N_s^2 N_v^2 T^2)$.

Figures 5.8 shows the execution times associated with the formulation of ESMPC and SSMPC as functions of the prediction horizon length T. Similar results can be obtained if we increase the number of zones N_v. In Figure 5.8, the solid line represents the average computational time for discrete method (Algorithm 3), and the dashed line represents the average computational time for sample-based method (Algorithm 4).

While the ESMPC behaves as expected, the SSMPC shows a super-linear relationship between the SSMPC execution time and the prediction horizon length. The reason for such the discrepancy between theoretical and simulation results is thought to be linked to the sparsity of the matrices involved in calculation of the constraint offsets in Step S2.2 of Algorithm 4. On the other hand, a discrete convolution operation in Step 3 of Algorithm 3 involves a dense matrix-vector multiplication and thus, the complexity of this operation is $\mathcal{O}(T^2)$, regardless of the potential system sparsity. We remark that ESMPC and SSMPC solve the same optimization problem (5.49) with the exception of the tightening offset α_k^j. For the size and complexity of problems considered in this thesis, the average time to solve Problem (5.49) with Ipopt is on the order of seconds.

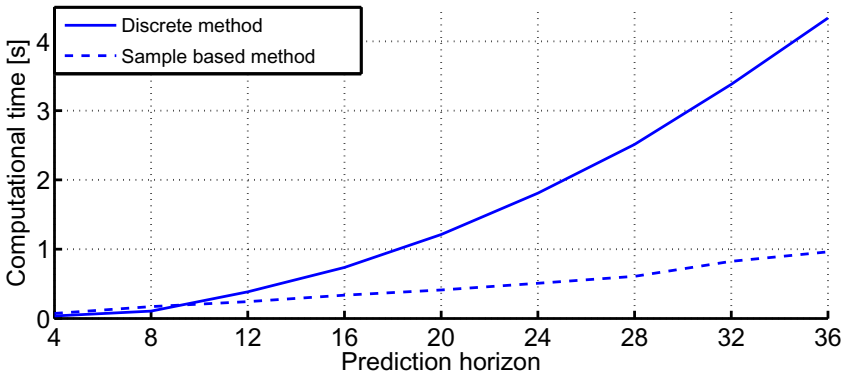


Figure 5.8: Computational time for ESMPC and SSMPC.

Experimental Results

The proposed ESMPC controller has been implemented to control the MPC lab at UC Berkeley. The MPC lab resides on the underground floor of Etcheverry building, it has no windows, and thus is subject to negligible solar radiation. The dimension of the lab is 9m $\times$ 9m $\times$ 3.5m, and it hosts fourteen students with fourteen desktop PCs running during weekdays. The VAV box serving the lab is depicted in Figure 5.9. The inlet air is the cool or warm air supplied by the central air handling unit in the Etcheverry building, and we do not control its temperature. The VAV box in the lab consists of an inlet air damper to control the inlet air flow rate, a set of cooling/heating coils to cool down or warm up the supply air, a supply fan to maintain the static pressure in the duct, a supply air damper to regulate the supply air flow rate, and a return air damper to balance the air pressure in the lab.

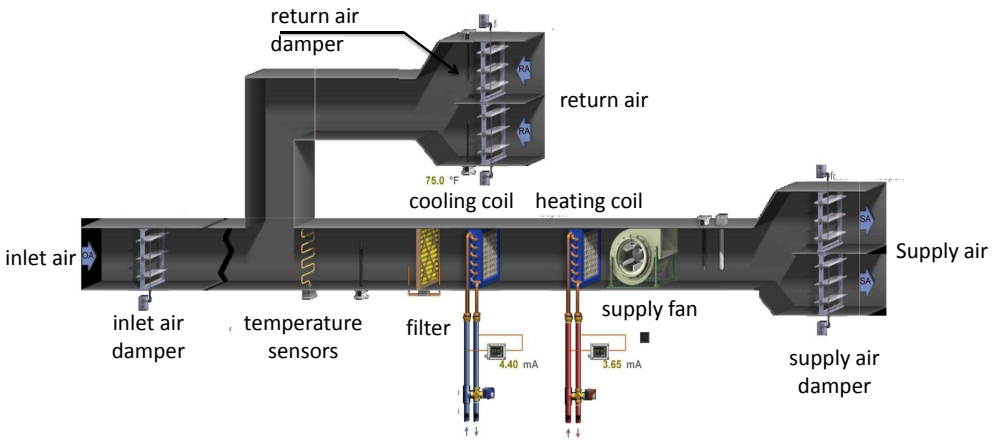


Figure 5.9: Scheme plot for HVAC system in the lab.

The lab is equipped with a modern digital control panels shown in Figure 5.10 which allows us to monitor and analyze the performance of the HVAC systems remotely, and implement advanced control logics. In the following sections, we present the system identification results for the lab, the implementation details for the ESMPC controller, and some experimental results.

System Identification

The thermal dynamic of the lab is modeled by the bilinear regression model (5.2) with $q_x = q_d = 2$. The regression parameters are fitted using historical data from weekend hours on September 27 and September 28, 2012, when there is negligible occupancy load during weekends in the lab. Figure 5.11 depicts the identification results, where the dash dot line is the predicted lab temperature and the solid line represents the measured lab temperature. The identified model parameters are reported in Table 5.3. The load uncertainty model is

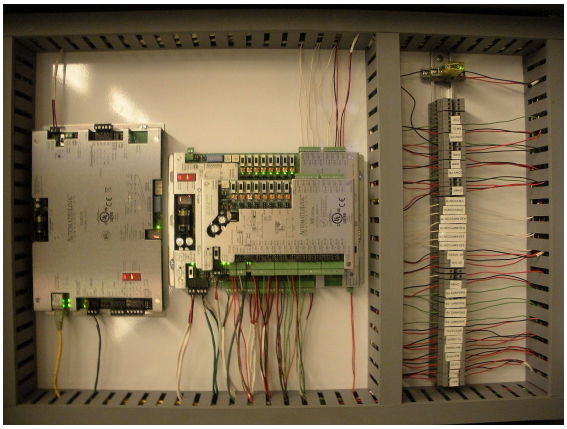


Figure 5.10: ALC control panel in the lab.

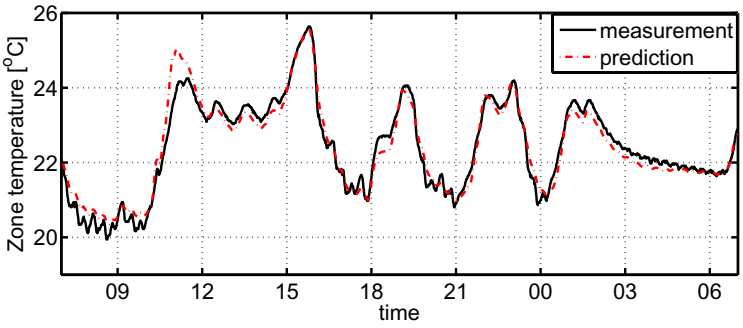


Figure 5.11: Identification results for the lab model.

Table 5.3: Model parameters for the lab.

$p_{1,0}$	$p_{1,1}$	$p_{1,2}$	p_3	$p_{4,0}$	$p_{4,1}$	$p_{4,2}$	p_6
0.0362	-0.0628	0.0331	0.0485	3.9244	0.9459	-0.1275	0.0150

computed as explained in Section 5.2. Figure 5.12 shows the estimated weekly disturbance loads from April 1 to July 21, 2012. In Figure 5.12, the dots are the realizations of the loads while the solid lines is the mean value of the weekly disturbance load. The samples in Figure 5.12 are used to learn the statistics of the load model.

We use the ambient temperature uncertainty prediction model (5.5) where the probability density function is depicted in Figure 5.3 and the mean value of the ambient temperature prediction is downloaded from services provided by National Oceanic and Atmospheric Administration (NOAA).

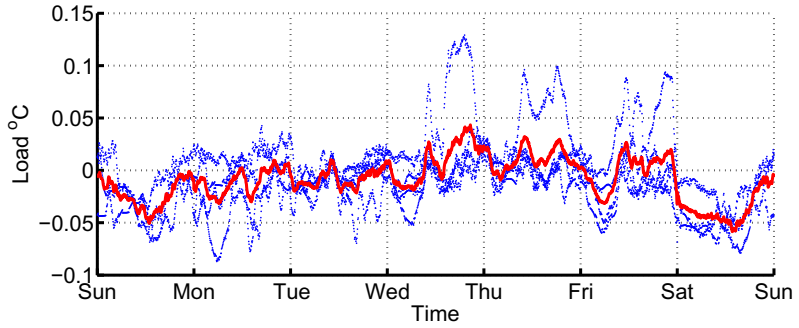


Figure 5.12: Lab thermal load model.

ESMPC Implementation

We use a sampling of 5 minutes, and a prediction horizon of 5 hours. At time t, the ESMPC is implemented as follows

1. Obtain the lab temperature sensor readings.
2. Downloads the weather predictions from the weather forecast services (NOAA).
3. Obtain the uncertain load predictions $P^i_{d,k}$. The uncertain load prediction is modeled as a look-up table learned from samples in Figure 5.12. The load uncertainty model is updated when new measurements become available.
4. Solve the stochastic MPC problem (5.49) to obtain the optimal set point for the supply air mass flow rate and supply air temperature. The optimization problem is solved using Ipopt on a single PC with Intel Core Duo CPU 3.00GHz.
5. Send the optimal set point to low-level PID controllers which regulates the coil valve positions and the fan speed to track the set points.

Experimental Results

Figure 5.13 reports the lab temperature controlled by ESMPC (solid line) from August 07 to August 20, 2012. In Figure 5.13 the dashed lines are the upper bound and lower bound of the comfort region.

In Figure 5.14, the solid line is the supply air temperature delivered to the lab, the dash-dot line is the inlet air temperature delivered by the central AHU in Etcheverry building, and the dashed line is the lower-bound on supply air temperature. It is observed that the supply air temperature satisfies the lower bound, and is kept lower than the inlet air temperature, which is required by constraints C4 in Section 5.2.

The supply air mass flow rate is depicted in Figure 5.15. The lab requires a minimum of 800 cfm supply air flow rate to meet the minimum ventilation requirement.

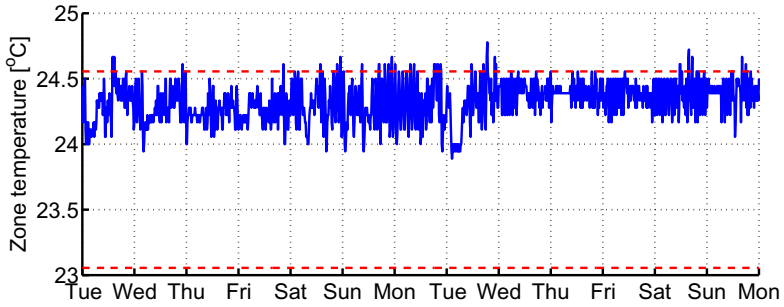


Figure 5.13: Lab temperature.

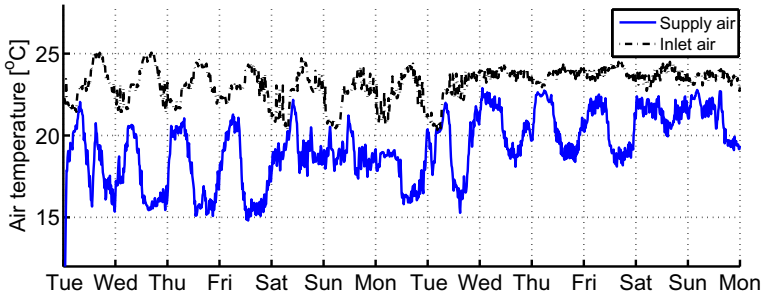


Figure 5.14: Lab supply air temperature.

Figure 5.16 reports the daily electricity energy consumptions for the operation of VAV box in the lab. In Figure 5.16 the blue bars are the energy consumed by cooling coils, and the red bars are the energy consumed by the supply fan.

5.5 Summary

In this chapter, a stochastic model predictive control (SMPC) design methodology is presented for HVAC systems. The SMPC uses uncertain prediction of weather conditions and buildings loads to minimize the expected energy consumption, satisfy the robust operational constraints, and provide guarantees on the probability of comfort violations. We have shown how to modeled the building thermal zones as a network of bilinear systems. The uncertainty models for occupancy load and weather predictions have been modeled as finitely-supported probability distribution functions learned from historical data. In order to reduce the conservatism of the stochastic MPC scheme while retaining the computational tractability, we proposed a feedback linearization scheme. The chance constraints then are transformed to deterministic ones by using two techniques: discrete convolution integrals and sample-based

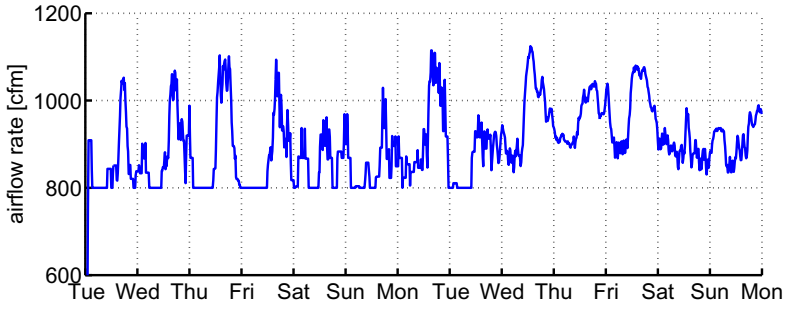


Figure 5.15: Lab supply air mass flow rate.

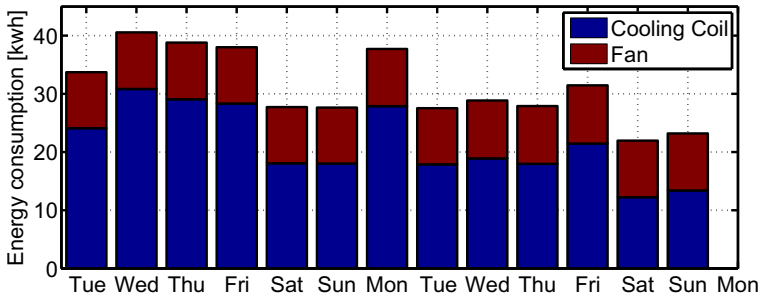


Figure 5.16: Electricity consumption of VAV box in the lab [kwh].

method.

Simulation and experiments have shown the effectiveness of the proposed approach. In particular, we have highlighted the tradeoff between performance, conservatism, and complexity of model predictive control algorithms for HVAC systems.

Bibliography

[1] *2003 ASHRAE Handbook - HVAC Applications*. Tech. rep. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 2003. URL: http://www.knovel.com/web/portal/browse/display?_EXT_KNOVEL_DISPLAY_bookid=2554.

[2] *2009 ASHRAE Handbook - Fundamentals (I-P Edition)*. Tech. rep. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., June 2009. URL: http://www.knovel.com/web/portal/browse/display?_EXT_KNOVEL_DISPLAY_bookid=2554.

[3] A. Alessio and A. Bemporad. "A Survey on Explicit Model Predictive Control". In: *Nonlinear Model Predictive Control*. Ed. by Lalo Magni, Davide Raimondo, and Frank Allgöwer. Vol. 384. Lecture Notes in Control and Information Sciences. Springer Berlin / Heidelberg, 2009. URL: http://dx.doi.org/10.1007/978-3-642-01094-1_29.

[4] A. Aswani et al. *Identifying Models of HVAC Systems Using Semiparametric Regression*. Tech. rep. University of California at Berkeley, 2012.

[5] A. Aswani et al. "Reducing Transient and Steady State Electricity Consumption in HVAC Using Learning-Based Model Predictive Control". In: *Proceedings of the IEEE* 100.1 (Jan. 2012), pp. 240–253.

[6] A. Bemporad, F. Borrelli, and M. Morari. "Min-max Control of Constrained Uncertain Discrete-Time Linear Systems". In: *IEEE Transactions on Automatic Control* 48.9 (Sept. 2003), pp. 1600–1606.

[7] A. Bemporad, F. Borrelli, and M. Morari. "Robust Model Predictive Control: Piecewise Linear Explicit Solution". In: *European Control Conference*. Porto, Portugal, Sept. 2001, pp. 939–944.

[8] A. Bemporad and M. Morari. "Robust Model Predictive Control: A Survey". In: *Robustness in Identification and Control*. Ed. by A. Garulli, A. Tesi, and A. Vicino. Lecture Notes in Control and Information Sciences 245. Springer-Verlag, 1999, pp. 207–226.

[9] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation*. Vol. 290. Englewood Cliffs, NJ: Springer-Verlag, 1989.

[10] L. Blackmore, H. Li, and B. Williams. "A Probabilistic Approach to Optimal Robust Path Planning with Obstacles". In: *Proceedings of the American Control Conference*. 2006.

[11] L. Blackmore and M. Ono. "Convex Chance Constrained Predictive Control without Sampling". In: *Jet Propulsion* 4800.1 (2009), pp. 0–6. URL: `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.157.9876&rep=rep1&type=pdf`.

[12] F. Blanchini. "Set Invariance in Control–A Survey". In: *Automatica* 35.11 (Nov. 1999), pp. 1747–1768.

[13] F. Blanchini and W. Ukovich. "Linear Programming Approach to the Control of Discrete-time eriodic sySstems with Uncertain Inputs". In: *J. Optim. Theory Appl.* 78.3 (1993), pp. 523–539.

[14] F. Borrelli. *Constrained Optimal Control of Linear and Hybrid Systems*. Vol. 290. Springer-Verlag, 2003.

[15] F. Borrelli and T. Keviczky. "Distributed LQR Design for Identical Dynamically Decoupled Systems". In: *IEEE Transactions on Automatic Control* 53.8 (Sept. 2008), pp. 1901–1912.

[16] F. Borrelli, T. Keviczky, and G.E. Stewart. "Decentralized Constrained Optimal Control Approach to Distributed Paper Machine Control". In: *44th IEEE Conference on Decision and Control and European Control Conference. CDC-ECC '05*. Dec. 2005, pp. 3037–3042.

[17] F. Borrelli et al. "On the Computation of Linear Model Predictive Control Laws". In: *Proceedings of the 48th IEEE Conference on Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference*. Dec. 2009, pp. 7375–7380.

[18] L. Boullart, A. Krijgsman, and R.A. Vingerhoeds. *Application of Artificial Intelligence in Process Control*. Vol. 290. June 1993.

[19] J. Braun. "Reducing Energy Costs and Peak Electrical Demand Through Optimal Control of Building Thermal Storage". In: *ASHRAE Transactions* 96.2 (1990), pp. 876–887.

[20] J. Braun and N. Chaturvedi. "Model for Transient Building Load Prediction". In: *HVAC&R Research* 8.1 (2002), pp. 73–99.

[21] R. Cagienard et al. "Move Blocking Strategies in Receding Horizon Control". In: *Journal of Process Control* 17.6 (2007), pp. 563–570. URL: `http://www.sciencedirect.com/science/article/B6V4N-4N44076-1/2/c47500954eb2cef2b1542b8a59ecdf63`.

[22] G.C. Calafiore. "Random Convex Programs". In: *SIAM Journal on Optimization* 20.6 (2010), pp. 3427–3464.

[23] G.C. Calafiore and M.C. Campi. "The Scenario Approach to Robust Control Design". In: *IEEE Transactions on Automatic Control* 51.5 (May 2006), pp. 742–753.

[24] G.C. Calafiore and L. Fagiano. "Robust Model Predictive Control: The Random Convex Programming Approach". In: *2011 IEEE International Symposium on Computer-Aided Control System Design (CACSD)*. Sept. 2011, pp. 222–227.

[25] G.C. Calafiore and L. Fagiano. "Robust Model Predictive Control via Random Convex Programming". In: *50th IEEE Conference on Decision and Control and European Control Conference*. Dec. 2011, pp. 1910–1915.

[26] E. Camacho and C. Bordons. "Nonlinear Model Predictive Control: An Introductory Review". In: *Assessment and Future Directions of Nonlinear Model Predictive Control*. Ed. by Rolf Findeisen, Frank Allgöwer, and Lorenz Biegler. Vol. 358. Lecture Notes in Control and Information Sciences. Springer Berlin / Heidelberg, 2007, pp. 1–16.

[27] M.C. Campi and S. Garatti. "A Sampling-and-discarding Approach to Chance Constrained Optimization: Feasibility and Optimality". In: *Journal of Optimization Theory and Applications* 148.2 (2011), pp. 257–280.

[28] M. Cannon, P. Couchman, and B. Kouvaritakis. "MPC for Stochastic Systems". In: *Assessment and Future Directions of Nonlinear Model Predictive Control*. Ed. by Rolf Findeisen, Frank Allgöwer, and Lorenz Biegler. Vol. 358. Lecture Notes in Control and Information Sciences. Springer Berlin / Heidelberg, 2007, pp. 255–268.

[29] M. Cannon, B. Kouvaritakis, and D. Ng. "Probabilistic Tubes in Linear Stochastic Model Predictive Control". In: *Systems & Control Letters* 58.10-11 (2009), pp. 747–753.

[30] M. Cannon, B. Kouvaritakis, and X. Wu. "Probabilistic Constrained MPC for Multiplicative and Additive Stochastic Uncertainty". In: *IEEE Transactions on Automatic Control* 54.7 (2009), pp. 1626–1632.

[31] M. Cannon, B. Kouvaritakis, and X. Wu. "Probabilistic Constrained MPC for Multiplicative and Additive Stochastic Uncertainty". In: *IEEE Transactions on Automatic Control* 54.7 (2009), pp. 1626–1632.

[32] M. Cannon, D. Ng, and B. Kouvaritakis. "Successive Linearization NMPC for a Class of Stochastic Nonlinear Systems". In: *Nonlinear Model Predictive Control*. Ed. by Lalo Magni, Davide Raimondo, and Frank Allgöwer. Vol. 384. Lecture Notes in Control and Information Sciences. Springer Berlin / Heidelberg, 2009, pp. 249–262. URL: `http://dx.doi.org/10.1007/978-3-642-01094-1_20`.

[33] M. Cannon et al. "Stochastic Tubes in Model Predictive Control With Probabilistic Constraints". In: *IEEE Transactions on Automatic Control* 56.1 (Jan. 2011), pp. 194–200.

[34] M. Cannon et al. "Stochastic Tubes in Model Predictive Control With Probabilistic Constraints". In: *Automatic Control, IEEE Transactions on* 56.1 (Jan. 2011), pp. 194–200.

[35] M. Castilla et al. "A comparison of thermal comfort predictive control strategies". In: *Energy and Buildings* 43.10 (Oct. 2011), pp. 2737–2746. URL: http://www.sciencedirect.com/science/article/pii/S0378778811002799.

[36] A. Charnes and W.W. Cooper. "Chance-Constrained Programming". In: *Management Science* 6.1 (1959), pp. 73–79. URL: http://www.jstor.org/stable/2627476.

[37] R.J. Chillar and R.J. Liesen. "Improvement of the ASHRAE Secondary HVAC Toolkit Simple Cooling Coil Model for Simulation". In: *Proceedings of the SimBuild 2004 Conference, Boulder, Colorado* (Aug. 2004).

[38] J. Cigler et al. "Optimization of Predicted Mean Vote index within Model Predictive Control framework: Computationally tractable solution". In: *Energy and Buildings* 52.0 (2012), pp. 39–49. URL: http://www.sciencedirect.com/science/article/pii/S0378778812002770.

[39] E. Cinquemani et al. *On Convex Problems in Chance-constrain3ed Stochastic Model Predictive Control*. Tech. rep. Swiss Federal Institute of Technology Zürich, 2011. URL: http://control.ee.ethz.ch/index.cgi?page=publications;action=details;id=3317.

[40] J.A. Clarke. *Energy Simulation in Building Design, Second Edition*. Butterworth-Heinemann, Oxford, United Kingdom, 2001.

[41] B. Coffey et al. "A Software Framework for Model Predictive Control with GenOpt". In: *Energy and Buildings* 42.7 (2010), pp. 1084–1092.

[42] P.D. Couchman, M. Cannon, and B. Kouvaritakis. "Stochastic MPC with Inequality Stability Constraints". In: *Automatica* 42.12 (2006), pp. 2169–2174.

[43] J.A. Crabb, N. Murdoch, and J.M. Penman. "A Simplified Thermal Response Model". In: *Building Services Engineering Research & Technology* 8 (1987), pp. 13–19.

[44] D.B. Crawley et al. "EnergyPlus: Creating a New-generation Building Energy Simulation Program". In: *Energy and Buildings* 33.4 (2001), pp. 443–457.

[45] M. Dahleh and J.J. Pearson. "L1-optimal Feedback Controllers for MIMO Discrete-time Systems". In: *IEEE Transactions on Automatic Control* 32.4 (Apr. 1987), pp. 314–322.

[46] A. Domahidi et al. "Learning a Feasible and Stabilizing Explicit Model Predictive Control Law by Robust Optimization". In: *IEEE Conference on Decision and Control*. Orlando, FL, USA, Dec. 2011, pp. 513–519.

[47] A. Domahidi et al. "Learning Near-optimal Decision Rules for Energy Efficient Building Control". In: *IEEE Conference on Decision and Control*. Grand Wailea Maui, HI, USA, Dec. 2012.

[48] E. Donaisky et al. "PMV-Based Predictive Algorithms for Controlling Thermal Comfort in Building Plants". In: *IEEE International Conference on Control Applications*. Oct. 2007, pp. 182–187.

[49] Ltd D&R International. *2011 Building Energy Data Book*. Mar. 2012. Chap. 3. URL: http://buildingsdatabook.eren.doe.gov/docs/DataBooks/2011_BEDB.pdf.

[50] M.M. Eftekhari and L.D. Marjanovic. "Application of Fuzzy Control in Naturally Ventilated Buildings for Summer Conditions". In: *Energy and Buildings* 35.7 (2003), pp. 645–655. URL: http://www.sciencedirect.com/science/article/pii/S0378778802002220.

[51] P.G. Ellis et al. "Automated Multivariate Optimization Tool for Energy Analysis". In: *IBPSA SimBuild 2006 Conference* (Aug. 2006). URL: http://apps1.eere.energy.gov/buildings/publications/pdfs/commercial_initiative/automated_opt_tools.pdf.

[52] V.L. Erickson et al. "Energy Efficient Building Environment Control Strategies Using Real-Time Occupancy Measurements". In: *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. Berkeley, California, 2009, pp. 19–24. URL: http://doi.acm.org/10.1145/1810279.1810284.

[53] P.O. Fanger. *Thermal Comfort, Analysis and Applications in Environmental Engineering*. McGraw-Hill, 1972.

[54] B.A. Flake. "Parameter Estimation and Optimal Supervisory Control of Chilled Water Plants". PhD thesis. University of Wisconsin-Madison., 1998. URL: http://minds.wisconsin.edu/bitstream/handle/1793/7702/disserta.pdf?sequence=1.

[55] R.Z. Freire, G.H.C. Oliveira, and N. Mendes. "Predictive Controllers for Thermal Comfort Optimization and Energy Savings". In: *Energy and Buildings* 40.7 (2008), pp. 1353–1365. URL: http://www.sciencedirect.com/science/article/pii/S0378778808000029.

[56] M. Gilli and P. Winker. *A Review of Heuristic Optimization Methods in Econometrics*. Swiss Finance Institute Research Paper Series 08-12. Swiss Finance Institute, June 2008. URL: http://ideas.repec.org/p/chf/rpseri/rp0812.html.

[57] J. Glover and F. Schweppe. "Control of Linear Dynamic Systems with Set Constrained Disturbances". In: *Automatic Control, IEEE Transactions on* 16.5 (Oct. 1971), pp. 411–423.

[58] S. Goyal and P. Barooah. "A Method for Model-Reduction of Nonlinear Building Thermal Dynamics". In: *Energy and Buildings* 47 (Apr. 2011), pp. 332–340. URL: http://dx.doi.org/10.1016/j.enbuild.2011.12.005.

[59] R.E. Griffith and R.A. Stewart. "A Nonlinear Programming Technique for the Optimization of Continuous Processing Systems". In: *Management Science* 7.4 (July 1961), pp. 379–392. URL: http://www.jstor.org/stable/2627058.

[60] P. Gutman and M. Cwikel. "Admissible Sets and Feedback Control for Discrete-time Linear Dynamical Systems with Bounded Controls and States". In: *The 23rd IEEE Conference on Decision and Control* 23.1 (Dec. 1984), pp. 1727–1731.

[61] M. Gwerder et al. "Control of Thermally-Activated Building Systems (TABS)". In: *Applied Energy* 85.7 (2008), pp. 565–581.

[62] S.P. Han. "A Globally Convergent Method for Nonlinear Programming". In: *Journal of Optimization Theory and Applications* 22.3 (July 1977), pp. 297–309. URL: http://www.springerlink.com/content/q34q4j1550705425/fulltext.pdf.

[63] T. Hartman. "Designing Efficient Systems with the Equal Marginal Performance Principle". In: *ASHRAE Journal* 47.7 (July 2005), pp. 64–70.

[64] P. Haves et al. "Model Predictive Control of HVAC Systems: Implementation and Testing at the University of California, Merced". In: *Final Report to US DOE and CEC PIER* (2010). URL: http://www.escholarship.org/uc/item/3pt2d32h.pdf.

[65] G.P. Henze, C. Felsmann, and G. Knabe. "Evaluation of Optimal Control for Active and Passive Building Thermal Storage". In: *International Journal of Thermal Sciences* 43.2 (2004), pp. 173–183. URL: http://www.sciencedirect.com/science/article/B6VT1-4B6TWX9-1/2/8930c1a2f0a11accd5a089727edc1c65.

[66] G.P. Henze, M. Krarti, and M.J. Brandemuehl. "Guidelines for Improved Performance of Ice Storage Systems". In: *Energy and Buildings* 35.2 (2003), pp. 111–127.

[67] G.P. Henze and J. Schoenmann. "Evaluation of Reinforcement Learning Control for Thermal Energy Storage Systems". In: *HVAC&R Research* 9.3 (2003), pp. 259–275. URL: http://www.tandfonline.com/doi/abs/10.1080/10789669.2003.10391069.

[68] G.P. Henze et al. "Experimental Analysis of Model-based Predictive Optimal Control for Active and Passive Building Thermal Storage Inventory". In: *HVAC&R Research* 11.2 (2005), pp. 189–214.

[69] G.P. Henze et al. "Impact of Adaptive Comfort Criteria and Heat Waves on Optimal Building Thermal Mass Control". In: *Energy and Buildings* 39.2 (2007), pp. 221–235.

[70] D.H. van Hessem and O.H. Bosgra. "A Conic Reformulation of Model Predictive Control including Bounded and Stochastic Disturbances under State and Input Constraints". In: *Proceedings of the 41st IEEE Conference on Decision and Control*. Vol. 4. Dec. 2002, pp. 4643–4648.

[71] P. Hokayem et al. "Stochastic Receding Horizon Control with Output Feedback and Bounded Control Inputs". In: *IEEE Conference on Decision and Control*. Dec. 2010, pp. 6095–6100.

[72] R. Horst and H. Tuy. *Global Optimization: Deterministic Approaches*. Springer, Berlin, German, 1990.

[73] J. Jang. "System Design and Dynamic Signature Identification for Intelligent Energy Management in Residential Buildings". PhD thesis. University of California at Berkeley, 2008. URL: http://escholarship.org/uc/item/0v83w3kw.pdf.

[74] A. Kelman and F. Borrelli. "Bilinear Model Predictive Control of a HVAC System Using Sequential Quadratic Programming". In: *Proceedings of the 18th IFAC World Congress*. Vol. 18. 2011.

[75] A. Kelman, Y. Ma, and F. Borrelli. "Analysis of Local Optima in Predictive Control for Energy Efficient Buildings". In: *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*. Dec. 2011, pp. 5125–5130.

[76] A. Kelman, Y. Ma, and F. Borrelli. "Analysis of Local Optima in Predictive Control for Energy Efficient Buildings". In: *Journal of Building Performance Simulation* (2012), pp. 1–20. URL: http://www.tandfonline.com/doi/abs/10.1080/19401493.2012.671959.

[77] S.A. Klein, J.A. Duffie, and W.A. Beckman. "TRNSYS - A Transient Simulation Program". In: *ASHRAE Transactions* 82.1 (1976), pp. 623–633.

[78] S. Koehler and F. Borrelli. *Building Temperature Distributed Control via Explicit MPC and "Trim and Respond" Methods*. Tech. rep. University of California at Berkeley, 2012. URL: http://www.mpc.berkeley.edu/people/sarah-koehler/files/ECCTrimRespondv10.pdf.

[79] B. Kouvaritakis, M. Cannon, and R. Yadlin. "On the Centres and Scalings of Robust and Stochastic Tubes". In: *United Kingdom Automatic Control Council (UKACC) International Conference on Control* (2011).

[80] B. Kouvaritakis et al. "Explicit Use of Probabilistic Distributions in Linear Predictive Control". In: *IEEE Transactions on Automatic Control* 46.10 (2010), pp. 1719–1724.

[81] M. Krarti. "An Overview of Artificial Intelligence-based Methods for Building Energy Systems". In: *Journal of Solar Energy Engineering* 25 (2003), pp. 331–342.

[82] M. Krarti. *Artificial Neural Networks Applied to Loanstar Data, Final Report*. Tech. rep. report prepared for Texas Energy Experiment Station. University of Colorado at Boulder, 1993.

[83] L.S. Lasdon. "Duality and Decomposition in Mathematical Programming". In: *IEEE Transactions on Systems Science and Cybernetics* 4.2 (July 1968).

[84] J.H. Lee and B.L. Cooley. "Optimal Feedback Control Strategies for State-space Systems with Stochastic Parameters". In: *IEEE Transactions on Automatic Control* 43.10 (Oct. 1998), pp. 1469–1475. ISSN: 0018-9286.

[85] J.H. Lee, M. Morari, and C.E. Garcia. "State-Space Interpretation of Model Predictive Control". In: *Automatica* 30.4 (1994), pp. 707–717.

[86] C. Liao and P. Barooah. "An Integrated Approach to Occupancy Modeling and Estimation in Commercial Buildings". In: *2010 American Control Conference*. June 2010, pp. 3130–3135.

[87] J.H. Lienhard. *A Heat Transfer Textbook*. Cambridge, Massachusetts: Phlogiston Press, 2002, pp. 62–77.

[88] S. Liu and G.P. Henze. "Experimental Analysis of Simulated Reinforcement Learning Control for Active and Passive Building Thermal Storage Inventory: Part 1. Theoretical Foundation". In: *Energy and Buildings* 38.2 (2006), pp. 142–147. ISSN: 0378-7788.

[89] J. Löfberg. "Approximations of closed-loop MPC". In: *IEEE Conference on Decision and Control*. Maui, Hawaii, Dec. 2003, pp. 1438–1442.

[90] Y. Ma, G. Anderson, and F. Borrelli. "A Distributed Predictive Control Approach to Building Temperature Regulation". In: *2011 American Control Conference*. June 2011.

[91] Y. Ma, S. Vichik, and F. Borrelli. *Fast Stochastic MPC with Optimal Risk Allocation Applied to Building Control Systems*. Tech. rep. University of California at Berkeley, 2012. URL: `http://www.mpc.berkeley.edu/people/yudong-ma/files/YudongMa_CDC12_ver1.pdf`.

[92] Y. Ma et al. "Model Predictive Control for the Operation of Building Cooling Systems". In: *Proceedings on American Control Conference, 2010*. June 2010, pp. 5106–5111.

[93] Y. Ma et al. "Model Predictive Control for the Operation of Building Cooling Systems". In: *IEEE Transactions on Control Systems Technology* 20.3 (May 2012), pp. 796–803.

[94] Y. Ma et al. "Model Predictive Control of Thermal Energy Storage in Building Cooling Systems". In: *Proceedings of the 48th IEEE Conference on CDC/CCC* (Dec. 2009), pp. 392–397.

[95] Y. Ma et al. "Predictive Control for Energy Efficient Buildings with Thermal Storage: modeling, Simulation, and Experiments". In: *Control Systems Magazine, IEEE* 32.1 (Feb. 2012), pp. 44–64.

[96] K. Marik, Z. Schindler, and P. Stluka. "Decision Support Tools for Advanced Energy Management". In: *17th International Congress of Chemical and process Engineering* 33.6 (June 2008), pp. 858–873. ISSN: 0196-8904.

[97] P. May-Ostendorp et al. "Model Predictive Control of Mixed-Mode Buildings with Rule Extraction". In: *Building and Environment* 46.2 (2011), pp. 428–437. URL: `http://www.sciencedirect.com/science/article/pii/S0360132310002477`.

[98] D.Q. Mayne et al. "Constrained Model Predictive Control: Stability and Optimality". In: *Automatica* 36.6 (June 2000), pp. 789–814.

[99] S. Mehrotra. "On the Implementation of a Primal-Dual Interior Point Method". In: 2.4 (1992), pp. 575–601.

[100] S. Meyn et al. "A Sensor-Utility-Network Method for Estimation of Occupancy in Buildings". In: *Proceedings of the 48th IEEE Conference on CDC/CCC*. Dec. 2009, pp. 1494–1500.

[101] J.W. Moon et al. "Comparative Study of Artificial Intelligence-based Building Thermal Control Methods–Application of Fuzzy, Adaptive Neuro-fuzzy Inference System, and Artificial Neural Network". In: *Applied Thermal Engineering* 31.14-15 (2011), pp. 2422–2429. URL: http://www.sciencedirect.com/science/article/pii/S1359431111001943.

[102] T. Nagai. "Optimization Method for Minimizing Annual Energy, Peak Energy Demand, and Annual Energy Cost Through Use of Building Thermal Storage". In: *ASHRAE Transactions* 108.1 (2002), pp. 976–887.

[103] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2004.

[104] J. Nocedal and S. J. Wright. "Numerical Optimization". In: Springer-Verlag, 1999. Chap. 18.

[105] F. Oldewurtel et al. "Energy Efficient Building Climate Control using Stochastic Model Predictive Control and Weather Predictions". In: *2010 American Control Conference*. Baltimore, USA, June 2010.

[106] F. Oldewurtel et al. "Energy Efficient Building Climate Control using Stochastic Model Predictive Control and Weather Predictions". In: *American Control Conference (ACC), 2010*. July 2010, pp. 5100–5105.

[107] Z. ONeill, S. Narayanan, and R. Brahme. "Model-Based thermal load estimation in Buildings". In: *Fourth National Conference of IBPSA-USA, SimBuild 2010* (Aug. 2010). URL: http://www.ibpsa.us/pub/simbuild2010/papers/SB10-DOC-TS09B-02-ONeill.pdf.

[108] M. Ono and B.C. Williams. "Iterative Risk Allocation: A New Approac to Robust Model Predictive Control with a Joint Chance Constraint". In: *47th IEEE Conference on Decision and Control*. Dec. 2008, pp. 3427–3432.

[109] J.A. Primbs. "A Soft Constraint Approach to Stochastic Receding Horizon Control". In: *46th IEEE Conference on Decision and Control*. Dec. 2007, pp. 4797–4802.

[110] S.J. Qin and T.A. Badgwell. "An Overview Of Industrial Model Predictive Control Technology". In: *Control Engineering Practice* (1997). URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.52.8909.

[111] S. Rakovic et al. "Parameterized Tube Model Predictive Control". In: *IEEE Transactions on Automatic Control* PP.99 (2012), pp. 1–8.

[112] S. V. Raković et al. "Optimized Robust Control Invariance for Linear Discrete-time Systems: Theoretical Foundations". In: *Automatica* 43.5 (2007), pp. 831–841.

[113] A. Rantzer. "Dynamic Dual Decomposition for Distributed Control". In: *American Control Conference*. June 2009, pp. 884–888.

[114] J.B. Rawlings and B.T. Stewart. "Coordinating Multiple Optimization-based Controllers: New Opportunities and Challenges". In: *Journal of Process Control* 18.9 (2008), pp. 839–845. URL: http://www.sciencedirect.com/science/article/B6V4N-4T8331W-1/2/c2bd50ac8c0b600df5c43cd0a077aa97.

[115] S. Richter, M. Morari, and C.N. Jones. "Towards Computational Complexity Certification for Constrained MPC Based on Lagrange Relaxation and the Fast Gradient Method". In: *Proc. 50th IEEE Conf. on Decision and Control*. Dec. 2011.

[116] F.E. Romie. "Transient Response of the Counterflow Heat Exchanger". In: *Journal of Heat Transfer* 106.3 (1984), pp. 620–626. URL: http://link.aip.org/link/?JHR/106/620/1.

[117] K.W. Roth et al. *Energy Impact of Commercial Building Controls and Performance Diagnostics: Market Characterization, Energy Impact of Building Faults*. Tech. rep. TIAX LLC, Nov. 2005, pp. 25–29. URL: http://envtraining.org/images/udf/file/EES/Commercial_Bldgcontrols_TIAXNov_05.pdf.

[118] P. Sahlin and A. Bring. "IDA Solver–Tool for Building and Energy Systems Simulation". In: *Proceedings of the IBPSA Conference, Nice, France* (1991).

[119] G. Schilbach et al. "Randomized Model Predictive Control for Stochastic Linear Systems". In: *American Control Conference*. 2012.

[120] G. Schildabach, L. Fagiano, and M. Morari. *Randomized Solutions to Convex Programs with Multiple Chance Constraints*. 2012.

[121] D.F. Shanno and R.J. Vanderbei. *Interior-Point Methods For Nonconvex Nonlinear Programming: Orderings And Higher-Order Methods*. Tech. rep. MATH. PROG, 1999.

[122] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics, Applied Probability, London: Chapman, and Hall, Apr. 1986.

[123] J. Skaf and S.P. Boyd. "Design of Affine Controllers via Convex Optimization". In: *IEEE Transactions on Automatic Control* 55.11 (Nov. 2010), pp. 2476–2487.

[124] J. Slotine and W. Li. *Applied Nonlinear Control*. Prentice Hall, Oct. 1991.

[125] J.M. Sousa, R. Babuška, and H.B. Verbruggen. "Fuzzy Predictive Control Applied to an Air-conditioning System". In: *Control Engineering Practice* 5.10 (1997), pp. 1395–1406.

[126] T. Steven. "Increasing Efficiency With VAV System VAV System Static Pressure Static Pressure Setpoint Reset". In: *ASHRAE journal* (June 2007).

[127] B.T. Stewart et al. "Cooperative Distributed Model Predictive Control". In: *Systems & Control Letters* (July 2010). URL: http://dx.doi.org/10.1016/j.sysconle.2010.06.005.

[128] World Business Council for Sustainable Development (WBCSS). *Energy Efficiency in Buildings Facts & Trends - Summary Report*. Tech. rep. 2007. URL: http://www.wbcsd.org/Pages/EDocument/EDocumentDetails.aspx?ID=13559.

[129] *Thermal Environment Conditions for Human Occupancy, Standard 55-1981 ed.* Tech. rep. American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE), Apr. 2004. URL: http://c0131231.cdn.cloudfiles.rackspacecloud.com/ASHRAE_Thermal_Comfort_Standard.pdf.

[130] D.H. Van Hessem, C.W. Scherer, and O.H. Bosgra. "LMI-based Closed-loop Economic Optimization of Stochastic Process Operation under State and Input Constraints". In: *Proceedings of the 40th IEEE Conference on Decision and Control*. Vol. 5. Dec. 2001, pp. 4228–4233.

[131] S. Vichik and F Borrelli. *Identification of Thermal Model of DOE Library*. Tech. rep. University of California at Berkeley, 2012. URL: http://www.mpc.berkeley.edu/people/sergey-vichik/file/SystemID_DOE_Library.pdf.

[132] M.P. Vitus and C.J. Tomlin. "On Feedback Design and Risk Allocation in Chance Constrained Control". In: *Proceedings of the 50th IEEE Conference on Decision and Control*. Orlando, Florida, Dec. 2011.

[133] S. Wang and Z. Ma. "Supervisory and Optimal Control of Building HVAC Systems: A Review". In: *HVAC&R Research* 14.1 (2008), pp. 3–32.

[134] Y.W. Wang et al. "A Simplified Modeling of Cooling Coils for Control and Optimization of HVAC Systems". In: *Energy Conversion and Management* 45.18-19 (2004), pp. 2915–2930. URL: http://www.sciencedirect.com/science/article/B6V2P-4BP3GJS-4/2/b66a7488beb3b9783f1cca027d18fbf1.

[135] A. Wächter and L.T. Biegler. "On the Implementation of an Interior-point Filter Line-search Algorithm for Large-scale Nonlinear Programming". In: *Mathematical Programming* 106 (1 2006), pp. 25–57. ISSN: 0025-5610.

[136] M. Wetter. "BuildOpt-New Building Energy Simulation Program that is Built on Smooth Models". In: *Building and Environment* 40.8 (Aug. 2005), pp. 1085–1092.

[137] M. Wetter. *GenOpt: generic optimization program*. Tech. rep. Lawrence Berkeley National Laboratory. URL: http://gundog.lbl.gov/GO/.

[138] M. Wetter. "Simulation-Based Building Energy Optimization". PhD thesis. University of California at Berkeley, 2004. URL: http://gundog.lbl.gov/GO/download/mwdiss.pdf.

[139] F.C. Winkelmann et al. *DOE-2 supplement, version 2.1E*. Tech. rep. Lawrence Berkeley National Laboratory, 1993. URL: http://doe2.com/download/doe-21e/DOE-2BDLSummaryVersion21E.pdf.

[140] A.J. Wright and V.I. Hanby. "The Formulation, Characteristics, and Solution of HVAC System Optimized Design Problems". In: *ASHRAE Transactions* 93.2 (Jan. 1987), pp. 2133–2145.

[141] J. Xu et al. "An Optimization-based Approach for Facility Energy Management with Uncertainties". In: *Proceedings of IFAC 2002*. 2002.

[142] Z. Yu and A. Dexter. "Hierarchical Fuzzy Control of Low-energy Building Systems". In: *Solar Energy* 84.4 (2010), pp. 538–548. URL: http://www.sciencedirect.com/science/article/pii/S0038092X09000644.