

Stochastic Model Predictive Control for Building HVAC Systems: Complexity and Conservatism

Yudong Ma, Jadranko Matuško, and Francesco Borrelli

Abstract—This paper presents a stochastic model predictive control (SMPC) approach to building heating, ventilation, and air conditioning (HVAC) systems. The building HVAC system is modeled as a network of thermal zones controlled by a central air handling unit and local variable air volume boxes. In the first part of this paper, simplified nonlinear models are presented for thermal zones and HVAC system components. The uncertain load forecast in each thermal zone is modeled by finitely supported probability density functions (pdfs). These pdfs are initialized using historical data and updated as new data becomes available. In the second part of this paper, we present a SMPC design that minimizes expected energy cost and bounds the probability of thermal comfort violations. SMPC uses predictive knowledge of uncertain loads in each zone during the design stage. The complexity of a commercial building requires special handling of system nonlinearities and chance constraints to enable real-time implementation, minimize energy cost, and guarantee thermal comfort. This paper focuses on the tradeoff between computational tractability and conservatism of the resulting SMPC scheme. The proposed SMPC scheme is compared with alternative SMPC designs, and the effectiveness of the proposed approach is demonstrated by simulation and experimental tests.

Index Terms—Building energy system, nonlinear system, stochastic model predictive control (SMPC).

I. 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 [3], [6], [18]–[22], [25]–[27], [32]–[35], [37]–[39]. The majority of existing model predictive control (MPC) schemes has the objective of minimizing energy cost while satisfying occupant thermal comfort using predictive knowledge of weather and occupancy (the building load). Simulation and experimental results have shown that closed-loop energy savings and comfort depend on the load disturbance predictions that are often different from actual realizations [18], [35].

In this paper, we discuss a stochastic MPC (SMPC) design to the aforementioned problem. In particular, we make use of

Manuscript received July 17, 2013; revised March 13, 2014; accepted March 15, 2014. Manuscript received in final form March 23, 2014. This work was supported by the National Science Foundation under Grant 0844456. Recommended by Associate Editor J. Yu.

Y. Ma and F. Borrelli are with the Department of Mechanical Engineering, University of California, Berkeley, CA 94720-1740 USA (e-mail: myd07@berkeley.edu; fborrelli@berkeley.edu).

J. Matuško is with the Faculty of Electrical Engineering and Computing, Department of Electrical Machines, Drives and Automation, Zagreb HR-10000, Croatia (e-mail: jadranko.matusko@fer.hr).

Digital Object Identifier 10.1109/TCST.2014.2313736

stochastic information of weather and load learned from historical data, and minimizes average energy cost 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 our work:

$$\min_{\kappa_{0|t}, \kappa_{1|t}, \dots, \kappa_{T-1|t}} \sum_{k=0}^{k=T-1} R_{k|t} \text{Energy}(\mathbb{E}(x_{k|t}), \mathbb{E}(u_{k|t}), \mathbb{E}(d_{k|t})), \quad (1a)$$

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

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

$$u_{k|t} \in \mathcal{U}, \forall k \in \mathbb{N}_{T-1}, \forall d_{k|t} \in \mathcal{D}, \quad (1d)$$

$$\mathbb{P}(x_{k|t} \in \mathcal{X}) \geq 1 - \epsilon, \forall k \in \mathbb{N}_T, \quad (1e)$$

$$x_{0|t} = x(t), \quad (1f)$$

where the symbol $v_{k|t}$ is the random variable v at time $t + k$ predicted at time t , $\mathbb{E}(v)$ denotes the expected value of v , $\mathbb{P}(v \in \mathcal{V})$ is the probability of the event $v \in \mathcal{V}$, \mathbb{N}_T is the set of integers $\{0, 1, \dots, T\}$, x is the system state, u is the control input, and d is the uncertain load. The energy rate is R [\$/kJ]. The functions $\text{Energy}(x, u, d)$ and $f(x, u, d)$ define the energy consumption and system dynamics, respectively. \mathcal{D} represents the set of all possible realizations of the uncertain load d . Problem (1) seeks a set of feedback laws $\kappa_{k|t}(\cdot)$, which minimize the average energy cost defined as (1a) and has an associated small probability of constraints violation (1e). Constraints (1e) are referred to as chance constraints. With abuse of notation and for the sake of simplicity, in the rest of this paper, the notation $x_{k|t}$ will be replaced with x_k .

The solution to the stochastic MPC problem (1) requires four main steps: 1) the translation of the optimization over control polices $\kappa(\cdot)$ into a finite-dimensional optimization problem; 2) the propagation of the stochastic system states over the prediction horizon; 3) the translation of probabilistic constraints into deterministic constraints; and 4) the solution of the resulting mathematical program.

With the exception of linear systems and special classes of distributions, (e.g., normal distributions), steps 1)–3) are nontrivial. Various methods are reported in the literature to handle steps 1)–3), and they affect the complexity of the mathematical problem at step 4) as well as the conservatism of the resulting control policy [10], [23], [31], [41], [46].

Step 1) has different solutions [5]. 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 realizations, 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 [29], [30].

The chance constraints (1e) in step 3) 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 [44], [45]. In practice, the distribution of the ambient temperature, and occupancy load in buildings have a finite support and are non-Gaussian. In addition, the simplified control oriented models for buildings are bilinear systems [17], [24], [35], [39]. 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 the stochastic MPC problem (1) for nonlinear systems subject to non-Gaussian disturbances are, in general, computationally intractable for real-time implementation in heating, ventilation, and air conditioning (HVAC) systems. Obtaining a computationally tractable approximation to this problem is crucial for the real-time implementation of stochastic MPC. Oldewurtel *et al.* [39] assume Gaussian distribution of the load and 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 [8], [9], [11], [13], [29], and [30] in the context of SMPC. Cannon *et al.* [11], [13] use a tube-based method to translate the chance constraints. The work in [29] and [30] suggests to translate the chance constraints to deterministic ones 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 [7]–[9], [16], [49], [50]. The approach consists in transforming the chance constraints (1e) into deterministic counterparts by evaluating them at a large number of disturbance samples. In particular, Zavala [50] generate \mathcal{N} disturbance scenarios (weather prediction in his paper) and approximate the cost function with the average of a deterministic cost function evaluated at each realization of the disturbance. The scenario-based approach proposed in the second part of this paper builds on the same idea and proposes a technique to remove samples to reduce its conservatism when state constraints are present.

In this paper, we build on the work in [9], [29], and [30], to design SMPC algorithms for networks of bilinear systems subject to non-Gaussian disturbances while retaining computational tractability. In particular, the bilinear system (1b) is linearized using feedback linearization. The proposed feedback linearization makes the SMPC problem computationally tractable and represents the first contribution of this paper. The chance constraints (1e) then are transformed to deterministic ones using two techniques: discrete convolution integrals and sample-based method. The second contribution of this paper is to propose a technique for scenario removal, which reduces the conservatism of the 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 [48].

The state-feedback linearization proposed in this paper has several advantages. In particular, the transformation of the chance constraints into deterministic ones does not depend on the linearized system dynamics as in [39]. This allows effective handling of non-Gaussian disturbances for nonlinear building systems. In the last part of this paper, the proposed SMPC design is carefully analyzed and compared with existing approaches. In Section IV, we discuss the tradeoff between performance, conservatism, and complexity. This represents our third contribution. In particular, we try to shed some lights on the following questions.

- Q1) Does one need a stochastic MPC formulation, or nominal MPC with expected forecasts provides good results for HVAC systems?
- Q2) Is there value in using non-Gaussian probability distribution functions or does Gaussian approximation work well, and what is the price one has to pay for it?
- Q3) Can the proposed approach be implemented in large scale buildings?
- Q4) Should one transform the chance constraints using convolution integrals or sample-based methods?

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

II. SYSTEM MODEL

A. HVAC System

Fig. 1 shows the main equipment used to produce and distribute cold or hot air in a building. 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.

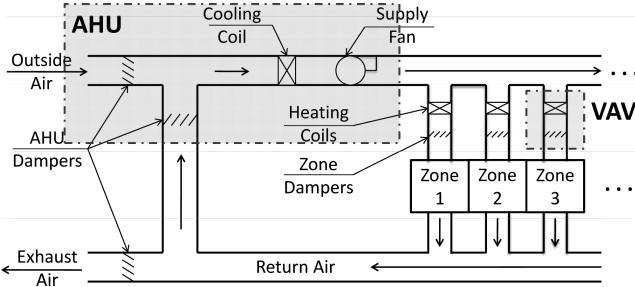


Fig. 1. Schematics diagram of the air distribution system.

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. A space served by one VAV box is denoted as 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 the following sections, we present our approach to AHU and VAV boxes modeling. A combination of reduced order, physics-based models, and static performance maps is employed. Their parameters are learned from measured data. Historical data allows us to quantify load and model uncertainty. The approach has appeared already in [35]. We included the main relevant ideas in this paper to improve its readability and to facilitate the understanding of the control design approach presented later in this paper. Also, the proposed approach can be extended to perform system-wide optimization, for instance by including chillers, cooling towers, and boilers, as proposed and studied in [35] and [50].

B. Thermal Zone Model

A undirected graph structure is adopted to represent the thermal zones and their dynamic couplings in the following way. The i th zone is associated with the i th node of a graph, and when an edge (i, j) connecting the i th and j th node is present, the thermal zones i and j are subjected to direct heat transfer. The graph \mathcal{G} is defined as

$$\mathcal{G} = (\mathcal{V}, \mathcal{A}), \quad (2)$$

where $\mathcal{V} = \{1, \dots, N_v\}$ is the set of nodes or vertices, and $\mathcal{A} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges (i, j) with $i \in \mathcal{V}$, $j \in \mathcal{V}$. \mathcal{N}^i is the set of neighboring nodes of i . We model the temperature dynamics of each thermal zone $i \in \mathcal{V}$ as an autoregressive moving average model with exogenous inputs model (ARMAX model) [24], [47]

$$T_{k+1}^i = \sum_{q=0}^{q=q_d} (p_{1,q}^i T_{oa,k-q} + p_{2,q}^i I_{k-q}^i) + p_3^i (T_{s,k}^i - T_k^i) \dot{m}_{s,k}^i + \sum_{q=0}^{q=q_x} (p_{4,q}^i T_{k-q}^i + \sum_{j \in \mathcal{N}^i} p_{5,q,j}^i T_{k-q}^j) + p_6^i + P_{d,k}^i, \quad (3)$$

where T_k^i is the temperature of zone i at time k . For the zone i at time k , $\dot{m}_{s,k}^i$ is the supply air mass flow rate, $T_{s,k}^i$ is

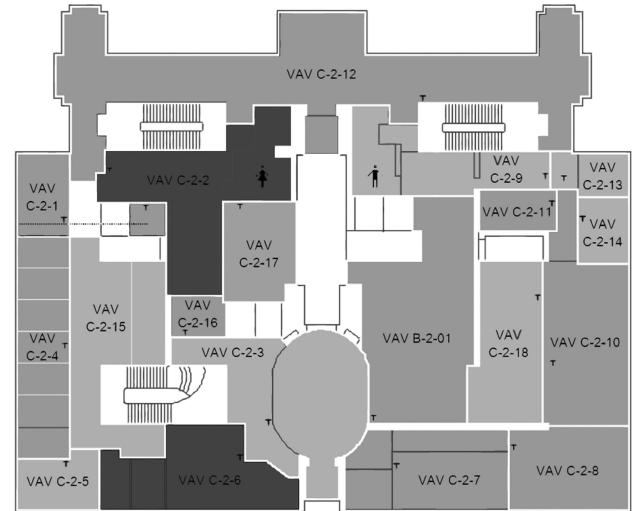


Fig. 2. Bancroft library floor plan.

the supply air temperature, $P_{d,k}^i$ is the internal thermal load, $T_{oa,k}$ is the ambient temperature, and I_k^i is the solar radiation intensity. The supply air mass flow rates and temperatures \dot{m}_s^i and T_s^i are control inputs. In (3), T^i are system states, and P_d^i , T_{oa} and I^i are exogenous signals. In (3), 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 $(T_s^i - T^i) \dot{m}_s^i$ in (3) is introduced to match first-order energy-balance equations. The bilinear regression model (3) is identified using data collected from unoccupied hours when the internal load P_d is minimal. In particular, the historical data for zone temperatures, ambient temperature, solar radiation intensity, and recorded control inputs are first used to propagate model (3) given the initial zone temperature. Then, the parameters p are computed by solving a linear regression problem so that the errors between the measured zone temperatures and predicted zone temperatures are minimized.

We have used model (3) with $q_x = q_d = 2$ to model the second floor of the Bancroft Library at UC Berkeley. These values of q_x and q_d provide a good compromise between model accuracy and problem complexity. The floor plan of the Bancroft library is shown in Fig. 2. Results for the classroom (labeled VAV C-2-15 in Fig. 2) are reported next.

The identification results are shown in Fig. 3 for the weekend of January 20, 2012 when the classroom was unoccupied, 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 Fig. 4.

Next, we provide more details on the generation of forecasts for thermal load P_d^i , ambient temperature T_{oa} , and solar radiation intensity I^i in (3).

C. Load Models

The internal thermal loads are calculated as follows. The zone model (3) is first identified using data from unoccupied hours. By comparing occupied hour data with the model predictions, we can capture thermal load induced by occupants.

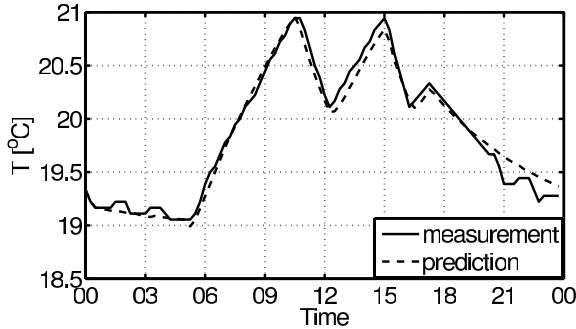


Fig. 3. Identification results of the thermal zone model (3).

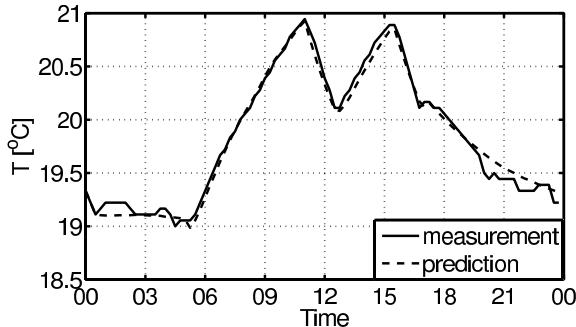


Fig. 4. Validation results of simplified room model (3).

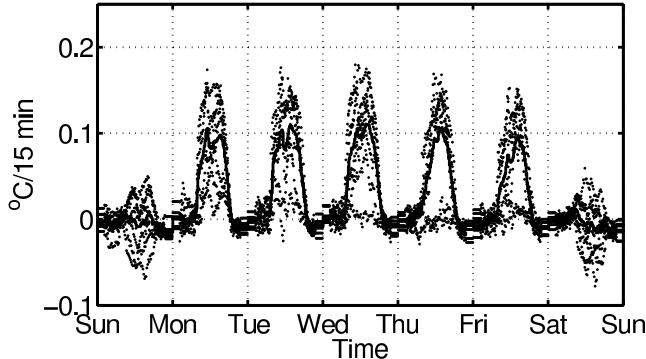


Fig. 5. Occupancy load.

We define a load profile as the difference in temperature between a local model prediction and the measurements

$$P_{d,k}^{i,\text{est}} = T_k^i - T_{\text{meas},k}^i, \quad \forall i \in \mathcal{V}, \quad (4)$$

where $P_d^{i,\text{est}}$ is the load estimated for zone i , T_{meas}^i is the measured temperature of zone i , and T^i is the zone i temperature prediction with the identified model (3) using data at time step $k-1$.

Remark 1: The load estimated by (4) includes model errors. Alternative methods to estimate the loads for building spaces have been reported in [4], [20], and [40], and they can be easily incorporated into our framework with minor modifications.

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 Fig. 2). Fig. 5 shows the weekly occupancy load from May 2011 to February 2012 computed by (4). In Fig. 5, each dot represents the

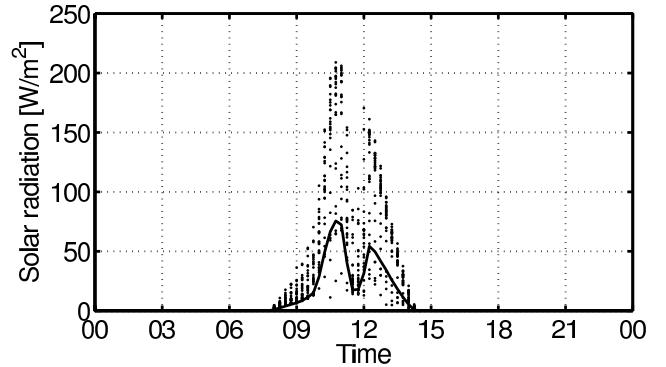


Fig. 6. Solar load for a classroom labeled as VAV C-2-5 in Fig. 2.

estimated load with sampling time of 1 h. The solid line is the weekly mean occupancy load over the whole period (May 2011–February 2012).

In Fig. 5, we observe that during weekdays the average occupancy load profile has a peak load of 0.15 °C over 15 min, and there is no significant occupancy load during weekends. This phenomenon matches the occupancy schedule of the classroom. The model (4) is used to generate uncertainty model for an occupancy load predictor. The occupancy loads are predicted using a lookup table with a period of one week

$$P_{d,k}^i = \hat{P}_{d,k}^i + \tilde{P}_{d,k}^i, \quad \forall i \in \mathcal{V}, \quad (5)$$

where $P_{d,k}^i$ is the predicted occupancy load at time $t+k$, $\hat{P}_{d,k}^i$ is the mean of the internal load prediction at time $t+k$ plotted as the solid line in Fig. 5, and $\tilde{P}_{d,k}^i$ is the prediction uncertainty described by a pdf with finite support. The finitely supported pdf can be computed from load samples of $P_d^{i,\text{est}}$, estimated through (4), using simple piecewise constant approximation or kernel density estimation [43]. In particular, the load samples in Fig. 5 have a bimodal distribution at time instances between Thursdays and Fridays due to occasional meetings in the afternoons.

The solar load for each zone is computed as the sun radiation intensity projected onto the normal vector to the outside wall. Fig. 6 shows the computed daily sun load for an office room (labeled as VAV C-2-5 in Fig. 2) from December 01, 2011 to January 31, 2012 in the Bancroft library at UC Berkeley. In Fig. 6, the dots are the computed daily sun load samples, and the solid line denotes the mean value of the computed daily solar load.

We use a periodic solar load I_k^i in zone i prediction model with period of one day. The mean \hat{I}_k^i and the pdf of the prediction uncertainty \tilde{I}_k^i are calculated from historical data.

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

$$T_{oa,k} = \hat{T}_{oa,k} + \tilde{T}_{oa,k}, \quad (6)$$

where $\hat{T}_{oa,k}$ is weather forecast obtained from the local weather station, and $\tilde{T}_{oa,k}$ is weather forecast uncertainty. The weather forecast error is defined as the difference between historical weather measurements and archived weather forecast data, and the pdf 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

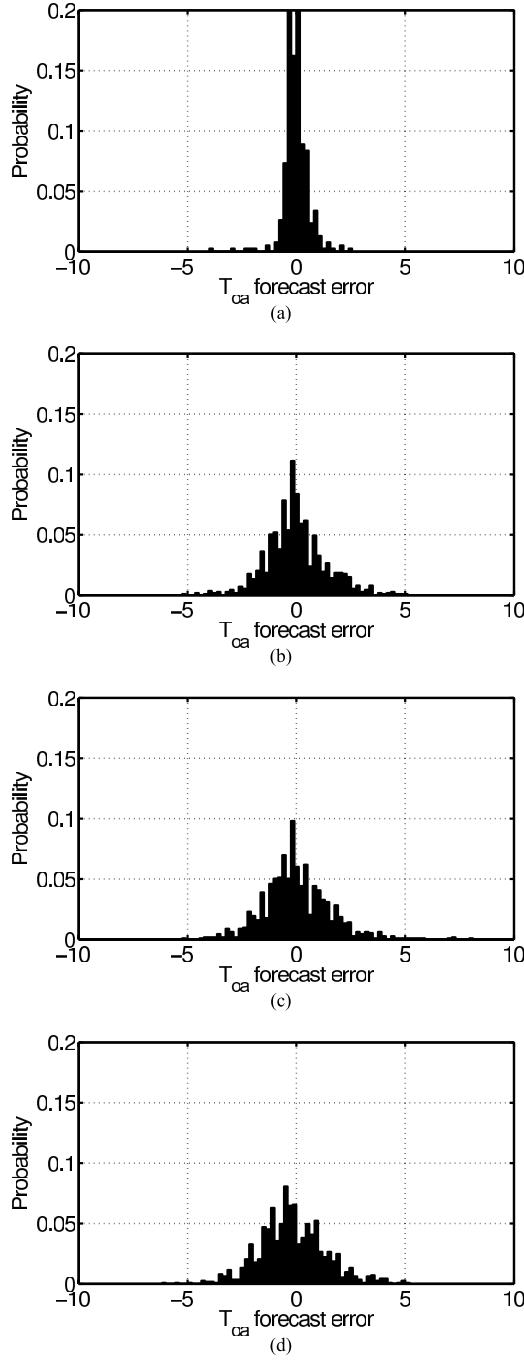


Fig. 7. Ambient temperature prediction error. (a) $k = 10$ min. (b) $k = 1$ day. (c) $k = 2$ day. (d) $k = 3$ day.

from archived weather prediction and measurements from May 20, 2012 to July 07, 2012. Fig. 7 shows the pdfs 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 is simple. The control design methodology presented in the following section does not depend on the specifics of the forecast algorithms. It only requires probability distribution functions of forecasted loads. Different methods could be used to obtain more accurate predictions.

D. Energy Model

The AHU and VAV components that consume 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 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

$$P_{\text{fan}} = c_0 + c_1 \dot{m} + c_2 \dot{m}^2, \quad (7)$$

where c_0 , c_1 , and c_2 are the 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

$$P_h = p_h \sum_{i \in \mathcal{V}} \dot{m}_s^i c_p (T_s^i - T_c), \quad P_c = p_c \sum_{i \in \mathcal{V}} \dot{m}_s^i c_p (T_m - T_c), \quad (8)$$

where P_c (P_h) is the power used by the cooling (heating) coils to deliver supply air with temperature T_s^i , p_c , and p_h are the 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 (Fig. 1). The mixed air temperature is computed as

$$T_m = \delta T_{oa} + (1 - \delta) \frac{\sum_{i \in \mathcal{V}} \dot{m}_s^i T^i}{\sum_{i \in \mathcal{V}} \dot{m}_s^i}, \quad (9)$$

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

$$P_{\text{tot}} = (P_h + P_c + P_{\text{fan}}). \quad (10)$$

E. Constraints

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

- C1) $\mathbb{P}\{T_k^i \geq \underline{T}_k^i\} > 1 - \epsilon$, $\mathbb{P}\{T_k^i \leq \bar{T}_k^i\} > 1 - \epsilon$, $\forall i \in \mathcal{V}$. The probability that zone temperatures are within the comfort bounds is greater than $1 - \epsilon$. Comfort bounds \underline{T}^i , \bar{T}^i and allowed violation probability ϵ are design parameters.
- C2) $T_s^i \in [\underline{T}_s^i, \bar{T}_s^i]$, $\forall i \in \mathcal{V}$. The supply air temperature is limited by the chilled water temperature through the coils and the physical characteristics of coils.
- C3) $\dot{m}_s^i \in [\underline{\dot{m}}_s^i, \bar{\dot{m}}_s^i]$, $\forall i \in \mathcal{V}$. Allowed mass flow rate for the supplied air. The lower bounds $\underline{\dot{m}}_s^i$ 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 computed by (9).
- 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.

Remark 3: Temperature constraint violations in this paper are used as a simple proxy for thermal comfort. The study

of thermal comfort has been the subject of extensive study in [1] and [2]. High fidelity models consider metabolic factors, humidity level, human body thermal radiation, and so on. Including these comfort models renders the problem computationally intractable. However, one could use other type of metrics such as percentage of people dissatisfied.

F. Model Summary

The bilinear ARMAX model (3) for all zones and the energy model (10) can be compactly rewritten as

$$x_{k+1} = f(x_k, u_k^{nl}, w_k), \quad w_k \in \mathcal{W}_k, \quad (11a)$$

$$T_k^i = C^i x_k, \quad \forall i \in \mathcal{V}, \quad (11b)$$

$$P_{\text{tot},k} = P_{\text{tot}}(x_k, u_k^{nl}, w_k), \quad (11c)$$

where $x_k = [T_k^1, T_{k-1}^1, T_{k-2}^1, \dots, T_k^{N_v}, T_{k-1}^{N_v}, T_{k-2}^{N_v}] \in \mathbb{R}^{n_x \times 1}$ is the system state, $u_k^{nl} = [\dot{m}_{s,k}^1, T_{s,k}^1, \dots, \dot{m}_{s,k}^{N_v}, T_{s,k}^{N_v}, T_{c,k}, \delta_k] \in \mathbb{R}^{n_u \times 1}$ is the control input, and $w_k = [P_{d,k}^1, \dots, P_{d,k}^{N_v}, I_k^1, I_{k-1}^1, I_{k-2}^1, \dots, I_{k-1}^{N_v}, I_{k-2}^{N_v}, T_{oa,k}, T_{oa,k-1}, T_{oa,k-2}] \in \mathbb{R}^{n_d \times 1}$ is the system load. The forecasts of w_k are obtained, as discussed in Section II-C. The state vector x_k includes the state of the ARMAX model (3) at every zone of the building. The input vector u_k includes 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

$$\mathbb{P}\{h^{jT} x_k \leq g_k^j\} > 1 - \epsilon, \quad \forall j \in \mathbb{N}_{N_c}, \quad (12a)$$

$$G(x_k, u_k^{nl}, w_k) \leq 0, \quad \forall w_k \in \mathcal{W}_k, \quad (12b)$$

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

III. STOCHASTIC MPC

Consider the building model (11c) and its constraints (12). We formulate the following stochastic optimization problem with chance constraints for comfort constraints:

$$\min_{X_t, U_t} \sum_{k \in \mathbb{N}_{T-1}} R_k P_{\text{tot}}(\hat{x}_k, \hat{u}_k^{nl}, \hat{w}_k) \Delta t, \quad (13a)$$

subj. to:

$$x_{k+1} = f(x_k, u_k^{nl}, w_k), \quad \forall k \in \mathbb{N}_T, \quad (13b)$$

$$\mathbb{P}\{h^{jT} x_k \leq g_k^j\} \geq 1 - \epsilon, \quad \forall j \in \mathbb{N}_{N_c}, \quad \forall k \in \mathbb{N}_T, \quad (13c)$$

$$G(x_k, u_k^{nl}, w_k) \leq 0, \quad \forall w_k \in \mathcal{W}_k, \quad \forall k \in \mathbb{N}_{T-1}, \quad (13d)$$

$$x_0 = x(t), \quad (13e)$$

where $X_t = \{x_0, x_1, \dots, x_T\}$ and $U_t = \{u_0^{nl}, \dots, u_{T-1}^{nl}\}$ are the optimization variables and Δt is the sampling time.

A SMPC solves problem (13) at each time step t . In particular, let $X_t^* = \{x_0^*, x_1^*, \dots, x_T^*\}$ and $U_t^* = \{u_0^{nl*}, u_1^{nl*}, \dots, u_{T-1}^{nl*}\}$ be the optimal solution of problem (13) at time t . Then, the first element of optimal control sequence is implemented to the system (11a)

$$u^{nl}(t) = u_0^{nl*}.$$

The optimization problem (13) is repeated at $t = t + \Delta t$, with the updated state $x_0 = x(t + 1)$.

The formulation (13) has the following features: 1) the power consumption at expected values of states, inputs and disturbances is minimized in (13a); 2) chance constraints (13c) are used for temperature bounds; and 3) constraint C2–C5 are robustly enforced for all admissible disturbances realizations in (13d). As it will be clear from the results presented later in this paper, this formulation will enable complexity reduction (by avoiding the computation of non-Gaussian probability distribution functions) while reducing the conservatism typically associated with the use of simple linear models, Gaussian uncertainty, and robust temperature bound satisfaction.

The following section presents the main contributions of this paper. In particular, we take the following steps.

- 1) Propose a computationally tractable approach to solve the stochastic MPC problem (13). We first make use of a state-feedback linearization to linearize the bilinear system model (13b). The chance constraints then can be reformulated as deterministic constraints by bounding the distribution tails. The resulting nonconvex optimization problem is solved using Ipopt [48].
- 2) Compare the complexity and conservatism of two methods used to reformulate the chance constraints (13c) with deterministic constraints: discrete method and sample-based method.
- 3) Carry out extensive simulation tests to demonstrate the conservatism of the proposed SMPC scheme compared with alternative MPC designs.

A. System Linearization

The bilinear dynamic model (13b) is linearized by introducing deterministic virtual inputs u_s^i and u_z^i for each zone i

$$u_{s,k}^i = \dot{m}_{s,k}^i T_{s,k}^i, \quad \forall i \in \mathcal{V}, \quad (14a)$$

$$u_{z,k}^i = \dot{m}_{s,k}^i T_{z,k}^i, \quad \forall i \in \mathcal{V}, \quad (14b)$$

to obtain the linear model

$$x_{k+1} = Ax_k + Bu_k + Dw_k, \quad w_k \in \mathcal{W}_k, \quad (15)$$

$$T_k^i = C^i x_k, \quad \forall i \in \mathcal{V}, \quad (16)$$

where the virtual control input vector u_k is $u_k = [u_{s,k}^1, u_{z,k}^1, \dots, u_{s,k}^{N_v}, u_{z,k}^{N_v}, T_{c,k}, \delta_k]$. For a given virtual input vector u_k , the input signals u_k^{nl} can be uniquely determined if $T^i \neq 0$, $\dot{m}_s^i \neq 0$, $\forall i \in \mathcal{V}$, (notice that any feasible system trajectory satisfies these two constraints, see Section II-E).

In model (15), w_k is a random variable with bounded support \mathcal{W}_k . Model (15) implies that the dynamics of state mean \hat{x} and state error $\tilde{x} = x - \hat{x}$ are

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + D\hat{w}_k, \quad (17a)$$

$$\hat{T}_k^i = C^i \hat{x}_k, \quad (17b)$$

$$\tilde{x}_{k+1} = A\tilde{x}_k + D\tilde{w}_k, \quad (17c)$$

$$\tilde{T}_k^i = C^i \tilde{x}_k, \quad (17d)$$

$$\tilde{x}_0 = 0. \quad (17e)$$

The linear system (17e) can be used in the optimization problem (13) instead of the nonlinear dynamics (13b). The new control variables $u_{z,k}^i$ in (14) can be interpreted as a deterministic state-feedback gain linking mass air flow rate \dot{m}_s^i and supply air temperature T_s^i to room temperature T^i

$$\dot{m}_{s,k}^i = u_{z,k}^i \frac{1}{T_k^i}, \quad \forall i \in \mathcal{V}, \quad (18a)$$

$$T_{s,k}^i = u_{s,k}^i \frac{1}{\dot{m}_{s,k}^i} = \frac{u_{s,k}^i}{u_{z,k}^i} T_k^i, \quad \forall i \in \mathcal{V}. \quad (18b)$$

Therefore, while the optimization in the virtual inputs u_s^i and u_z^i is over open-loop policies [5], the variable substitution (14) provides state-feedback policies. These advantages come with a price. In fact, the energy models in Section II-C and constraints defined in Section II-E have to be rewritten as a function of the virtual inputs u_k . Using the change of control variables (14), the energy models for supply fan (7), cooling, and heating coils (8) are rewritten as

$$P_{\text{fan}} = c_0 + c_1 \sum_{i \in \mathcal{V}} \frac{u_z^i}{T^i} + c_2 \left(\sum_{i \in \mathcal{V}} \frac{u_z^i}{T^i} \right)^2, \quad (19)$$

$$P_h = p_h c_p \sum_{i \in \mathcal{V}} \left(u_s^i - \frac{u_z^i}{T^i} T_c \right), \quad (20)$$

$$P_c = p_c c_p \sum_{i \in \mathcal{V}} \left(\frac{u_z^i}{T^i} \delta T_{oa} + (1 - \delta) u_s^i - \frac{u_z^i}{T^i} T_c \right), \quad (21)$$

$$P_{\text{tot}} = (P_{\text{fan}} + P_c + P_h) = P_{\text{tot}}(x, u, w). \quad (22)$$

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

$$\max_{\tilde{w}_0, \dots, \tilde{w}_{k-1}} u_{s,k}^i T_k^i = u_{s,k}^i C^i \hat{x}_k + u_{s,k}^i \tilde{T}_{k,\max}^i \leq \bar{T}_s^i u_z^i, \quad (23)$$

$$\min_{\tilde{w}_0, \dots, \tilde{w}_{k-1}} u_{s,k}^i T_k^i = u_{s,k}^i C^i \hat{x}_k + u_{s,k}^i \tilde{T}_{k,\min}^i \geq \underline{T}_s^i u_z^i. \quad (24)$$

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

$$u_{z,k}^i \leq \min_{\tilde{w}_k \in \mathbb{N}_{k-1}} \bar{\dot{m}}_{s,k}^i T_k^i = \bar{\dot{m}}_{s,k}^i C^i \hat{x}_k + \bar{\dot{m}}_{s,k}^i \tilde{T}_{k,\min}^i, \quad (25)$$

$$u_{z,k}^i \geq \max_{\tilde{w}_k \in \mathbb{N}_{k-1}} \underline{\dot{m}}_{s,k}^i T_k^i = \underline{\dot{m}}_{s,k}^i C^i \hat{x}_k + \underline{\dot{m}}_{s,k}^i \tilde{T}_{k,\max}^i, \quad (26)$$

where

$$\tilde{T}_{k,\min}^i = \min_{\tilde{w}_0, \dots, \tilde{w}_{k-1}} C^i \left(\tilde{x}_0 + \sum_{i=0}^{k-1} A^i D \tilde{w}_{k-1-i} \right), \quad (27)$$

$$\tilde{T}_{k,\max}^i = \max_{\tilde{w}_0, \dots, \tilde{w}_{k-1}} C^i \left(\tilde{x}_0 + \sum_{i=0}^{k-1} A^i D \tilde{w}_{k-1-i} \right). \quad (28)$$

Note that $\tilde{T}_{k,\min}^i$ and $\tilde{T}_{k,\max}^i$ can be computed off-line. Therefore, (23) and (24) are bilinear constraints in $u_{z,k}^i$, \hat{x}_k , and $u_{s,k}^i$. Constraints (25) and (26) are linear constraints in $u_{z,k}^i$ and \hat{x}_k .

The constraints C4 on the air temperature after the cooling coil in AHU T_c can be conservatively approximated as follows:

$$\begin{aligned} \max_{\tilde{w}_0, \dots, \tilde{w}_{k-1}} & \left(\sum_{i \in \mathcal{V}} \left[\frac{u_{z,k}^i}{T_k^i} (T_{c,k} - \delta_k T_{oa,k}) + (\delta_k - 1) u_{z,k}^i \right] \right) \\ & \leq \sum_{i \in \mathcal{V}} \left[\frac{u_{z,k}^i T_{c,k}}{C^i \hat{x}_k + \tilde{T}_{k,\min}^i} - \frac{\delta_k \underline{T}_{oa,k} u_{z,k}^i}{C^i \hat{x}_k + \tilde{T}_{k,\max}^i} + (\delta_k - 1) u_{s,k}^i \right] \leq 0, \end{aligned} \quad (29)$$

where $\underline{T}_{oa,k}$ is a lower bound of $T_{oa,k}$.

The constraint C5 on the air temperature after the cooling coil in AHU T_c becomes

$$u_z^i T_c \leq \min_{\tilde{w}_0, \dots, \tilde{w}_{k-1}} u_s^i T^i = u_s^i C^i \hat{x}_k + u_s^i \tilde{T}_{k,\min}^i, \quad \forall i \in \mathcal{V}. \quad (30)$$

Using the feedback linearization (14), we transform the robust nonlinear input constraints C2–C5 to deterministic constraints (23)–(30). The resulting constraints are nonlinear, and can be compactly rewritten as

$$G^{FL}(\hat{x}_k, u_k, \hat{w}_k) \leq 0, \quad \forall k \in \mathbb{N}_T. \quad (31)$$

B. Transforming Chance Constraints

In this section, we show the approach we used to transform the chance constraints (13c) to convex deterministic ones with tightening offsets [14]. We denote by π_k^j the univariate random variable $h^{jT} x_k$. The chance constraints (13c) then can be reformulated as deterministic ones [15]

$$\hat{\pi}_k^j \leq g_k^j - \alpha_k^j, \quad \forall j \in \mathbb{N}_{N_c}, \quad \forall k \in \mathbb{N}_T, \quad (32)$$

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

$$\int_{\alpha_k^j}^{\infty} \text{pdf}_{\tilde{\pi}_k^j}(y) dy = 1 - \text{cdf}_{\tilde{\pi}_k^j}(\alpha_k^j) = \epsilon, \quad \forall j \in \mathbb{N}_{N_c}, \quad \forall k \in \mathbb{N}_T, \quad (33)$$

where $\text{pdf}_{\tilde{\pi}_k^j}(y)$ is the pdf of $\tilde{\pi}_k^j = \pi_k^j - \hat{\pi}_k^j$ evaluated at y , and $\text{cdf}_{\tilde{\pi}_k^j}(\alpha_k^j) = \mathbb{P}\{\tilde{\pi}_k^j \leq \alpha_k^j\}$ is the cumulative density function of $\tilde{\pi}_k^j$ evaluated at α_k^j . A solution to (32) is also a solution to (13c).

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

$$\tilde{\pi}_k^j = h^{jT} \tilde{x}_k = h^{jT} \tilde{x}_0 + \sum_{i=0}^{k-1} h^{jT} A^i D \tilde{w}_{k-1-i}. \quad (34)$$

If the measurement of the state x_0 is exact, i.e., $\tilde{x}_0 = 0$, then (34) is compactly written as

$$\tilde{\pi}_k^j = \phi_k^{jT} \tilde{W}_{k-1}, \quad \forall j \in \mathbb{N}_{N_c}, \quad \forall k \in \mathbb{N}_T, \quad (35)$$

where

$$\tilde{W}_{k-1} = [\tilde{w}_0^T, \tilde{w}_1^T; \dots; \tilde{w}_{k-1}^T]^T,$$

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 \tilde{W}_{k-1} .

Algorithm 1 Discrete Method

-
- S1 Compute the upper and lower bounds of the discretization range, and select the sample number N_s .
 S2 Discretize the pdf $w_{(n)}$, $\forall n \in \mathbb{N}_{N_w}$.
 S3 Approximate the pdf $\text{pdf}_{\tilde{\pi}_k^j}$ by discrete convolution.
 S4 Compute the tightening offsets α_k^j .
-

Next, we present and compare two numerical methods with compute the offset α_k^j computed by (33) given the pdf for the disturbances error \tilde{w} .

1) Discrete Convolution Integral Method: If the disturbance errors \tilde{w} are statistically independent, the pdf $\text{pdf}_{\tilde{\pi}_k^j}$ of $\tilde{\pi}_k^j$ in (35) can be computed recursively as follows:

$$\text{pdf}_{\tilde{\pi}_k^j}^{(1)} = \text{pdf}_{w_{(1)}}, \quad (36)$$

$$\text{pdf}_{\tilde{\pi}_k^j}^{(n+1)} = \text{pdf}_{\tilde{\pi}_k^j}^{(n)} * \text{pdf}_{w_{(n+1)}}, \quad n = 1, 2, \dots, N_w - 1, \quad (37)$$

where

$$\text{pdf}_{w_{(n)}} = \text{pdf}_{\tilde{\pi}_k^j}^{(N_w)}$$

is the pdf for n^{th} element of $\text{diag}(\phi_k^j)\tilde{W}_{k-1}$ in (35). The operator $*$ denotes the convolution integral.

Numerical algorithms to approximate the convolution integrals (37) can be found in [12], [29], and [30]. We used the approach in [30]. In particular, the pdf $\text{pdf}_{w_{(n)}}$ in (37) is first discretized over N_s evenly distributed samples from $[\underline{\pi}, \bar{\pi}]$ with discretization interval $\Delta\pi = \bar{\pi} - \underline{\pi}/N_s$. Then, the pdf $\text{pdf}_{\tilde{\pi}_k^j}$ in (37) are obtained by approximating each convolution integral with discrete convolution. For the sake of completeness, we include the algorithm for the discrete convolution integral method to compute the tightening offsets α_k^j . More details can be found in [36].

2) Sample-Based Method: The sampling-based approach to single-stage chance constrained problems (CCPs) has been studied in [7]–[9]. Recently, it has been extended to multiple stage CCPs [42]. It provides an alternative solution to the computation of the offset α_k^j in (33). The sampling-based approach transforms the chance constraints (13c) into deterministic ones by generating for each individual constraint a large number $N_{s,k}^j$ of independent identically distributed disturbance samples $\tilde{w}_k^{j,1}, \tilde{w}_k^{j,2}, \dots, \tilde{w}_k^{j,N_{s,k}^j}$ (usually referred to as scenarios) according to the pdf of the disturbance. We use (34) and replace the chance constraints (33) with

$$h^{jT}\hat{x}_k + \sum_{i=0}^{k-1} h^{jT} A^k D \tilde{w}_{k-1-i}^{j,l} \leq g_k^j, \quad \forall j \in \mathbb{N}_{N_c}, \quad \forall k \in \mathbb{N}_T, \quad \forall l \in \mathbb{N}_{N_{s,k}^j}. \quad (38)$$

For a sufficiently large $N_{s,k}^j$, the satisfaction of (38) guarantees that each chance constraint in (13c) is satisfied with probability of $1 - \beta_k^j$.

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 (13c) is close to zero). To reduce the conservatism of the solution, a certain number of generated samples $N_{r,k}^j$ have to be removed from the set of samples without a significant loss of reliability ($1 - \beta_k^j$) 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 [7]–[9], and [42]. Here, we use the inequality [42]

$$\binom{N_{r,k}^j + \zeta_k^j - 1}{N_{r,k}^j} \sum_{l=0}^{N_{r,k}^j + \zeta_k^j - 1} \binom{N_{s,k}^j}{l} \epsilon^l (1 - \epsilon)^{N_{s,k}^j - l} \leq \beta_k^j, \quad (39)$$

where ζ_k^j is the support rank of each individual constraint in (13c). The support rank of a chance constraint is the dimension of the vector space spanned by the constraint. In summary, the chance constraint (13c) is transformed into

$$h^{jT}\hat{x}_k \leq g_k^j - \alpha_k^{j,l}, \quad \forall l \in \mathbb{I}_{s,k}^{N_{s,k}^j} \setminus \mathbb{I}_{r,k}^{N_{r,k}^j}, \quad \forall k \in \mathbb{N}_T, \quad \text{and } \forall j \in \mathbb{N}_{N_c}, \quad (40)$$

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

$$\alpha_k^{j,l} = \sum_{i=0}^{k-1} h^{jT} A_{cl}^i D \tilde{w}_{k-1-i}^l, \quad (41)$$

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

Since the support rank of the constraints (40) is $\zeta_k^j = 1$, inequality (39) reduces to

$$\sum_{l=0}^{N_{r,k}^j} \binom{N_{s,k}^j}{l} \epsilon^l (1 - \epsilon)^{N_{s,k}^j - l} \leq \beta_k^j. \quad (42)$$

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_k^j for the j th constraint, and the k time index has to be selected to satisfy

$$\sum_{j \in \mathbb{N}_{N_c}} \sum_{k \in \mathbb{N}_T} \beta_k^j < \beta. \quad (43)$$

The optimal sample removal in general is NP hard problem that has to be solved within the optimization problem. Because of the proposed feedback linearization approach (Section III-A), the set of samples $\{\tilde{w}_k^{j,l}\}$ can be propagated over the prediction horizon independently, prior to solving the optimization problem, and this enables a simple sample removal procedure described next.

In (40), the optimal sample removal strategy corresponds to the removal of the $N_{r,k}^j$ largest offsets $\alpha_k^{j,l}$ (notice that for fixed $j \in \mathbb{N}_{N_c}$ and $k \in \mathbb{N}_T$ all constraints in (40) are parallel), resulting in the least conservative approximation

Algorithm 2 Sampling-Based Method

-
- S1 $\forall k \in \mathbb{N}_T$ and $\forall j \in \mathbb{N}_{N_c}$ compute the number of samples to generate $N_{s,k}^j$ and to remove $N_{r,k}^j$ that satisfy inequalities (42) and (43).
- S2 $\forall k \in \mathbb{N}_T$, $\forall j \in \mathbb{N}_{N_c}$ and $\forall l \in \mathbb{N}_{s,k}^j$ **do**
- S2.1 Generate the set of samples $\tilde{\mathcal{W}}_k^j = \{\tilde{w}_k^{j,l}\}$.
 - S2.2 Calculate the set of the offsets $\{\alpha_k^{j,l}\}$ using (41).
 - S2.3 Sort the set of the offsets to obtain ordered set of the offsets $\{\bar{\alpha}_k^{j,l}\}$
- S3 $\forall k \in \mathbb{N}_T$ and $\forall j \in \mathbb{N}_{N_c}$ **do**
- S3.1 Remove $N_{r,k}^j$ largest offsets from the set $\{\bar{\alpha}_k^{j,l}\}$.
 - S3.2 Calculate the offsets α_k^j using (44).
-

of the corresponding individual chance constraint in (13c). Finally, the offsets α_k^j in (33) are calculated as

$$\forall k \in \mathbb{N}_T, \forall j \in \mathbb{N}_{N_c}, \alpha_k^j = \max(\alpha_k^{j,l}), l \in \mathbb{N}_{s,k}^j \setminus \mathbb{N}_{r,k}^j. \quad (44)$$

The algorithm to compute $\mathbb{N}_{s,k}^j \setminus \mathbb{N}_{r,k}^j$ in (40) and the corresponding offsets α_k^j is summarized next.

Algorithm 2 is repeated at each time step, prior to solving the MPC optimization problem. The samples are generated using the probability distribution function presented in Section II-C, which can be updated at each time step. Note that the approach proposed here works also for more general cost functions such as the expected valued of the energy cost, as proposed in [50].

C. Optimization Problem

With the change of variables presented in Section III-A, and the transformation of chance constraints in Section III-B, the stochastic MPC problem (13) is transformed into the deterministic nonlinear optimization problem

$$\min_{\hat{x}_k \in \mathbb{N}_T, u_k \in \mathbb{N}_{T-1}} \sum_{k \in \mathbb{N}_T} R_k P_{\text{tot}}(\hat{x}_k, u_k, \hat{w}_k) \Delta t, \quad (45a)$$

subj. to:

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + D\hat{w}_k, \forall k \in \mathbb{N}_T, \quad (45b)$$

$$h^j \hat{x}_k \leq g_k^j - \alpha_k^j, \forall j \in \mathbb{N}_{N_c}, \forall k \in \mathbb{N}_T, \quad (45c)$$

$$G^{FL}(\hat{x}_k, u_k, \hat{w}_k) \leq 0, \forall k \in \mathbb{N}_{T-1}, \quad (45d)$$

$$\hat{x}_0 = x(0). \quad (45e)$$

We compute the optimal solution $u_{k \in \mathbb{N}_T}^*$ to Problem (45) using Ipopt [48]. We remark that the steps presented in the previous section are crucial for obtaining a resulting problem (45), which is computationally tractable and whose solution is not too conservative. In the following section, we will provide metrics from conservatism and computational tractability. We will also present the results, which confirm the effectiveness of our approach.

IV. RESULTS

The stochastic MPC algorithm solving (45) with constant energy rate R_k is analyzed in this section. The effectiveness

of proposed SMPC to handle time-varying energy rate R is demonstrated by experiment results in the following section. For the class of building HVAC systems studied in this paper, we will investigate the questions Q1–Q4 raised in Section I.

Although the answers depend on the HVAC system and the level of uncertainty, the following 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 forecast uncertainty are generated using measured historical data.

A. Bancroft Library Simulations

1) 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 shown in Fig. 2 and consists of classrooms, offices, and conference rooms. The HVAC system has a single duct configuration shown in Fig. 1, and it is controlled and monitored using WebCTRL developed by automated logic corporation.

The original stochastic nonlinear control problem (13) is not feasible for real-time implementation for this system using current desktop capabilities. The goal of this section is to study the compromise between complexity, performance, and conservatism of different approximations to (13), including the one proposed in this paper.

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 [28]. The main idea is to control the heating coil valve position command and the airflow setpoint 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 (PMPC). A MPC 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 implementable (since it is a noncausal controller) but will be used as reference for what could be potentially achieved without uncertainty.
- L2) Certainty equivalent MPC (CMPC). A nominal MPC that solves the optimization problem (45) with the chance constraint offsets $\alpha_k^j = 0$, and the predicted disturbance \hat{w}_t set to be the mean value of load profiles modeled in Section II-C, i.e., the solid line in Fig. 5. Clearly, in closed-loop simulations, the nominal disturbance prediction might be different from the actual disturbance realization.

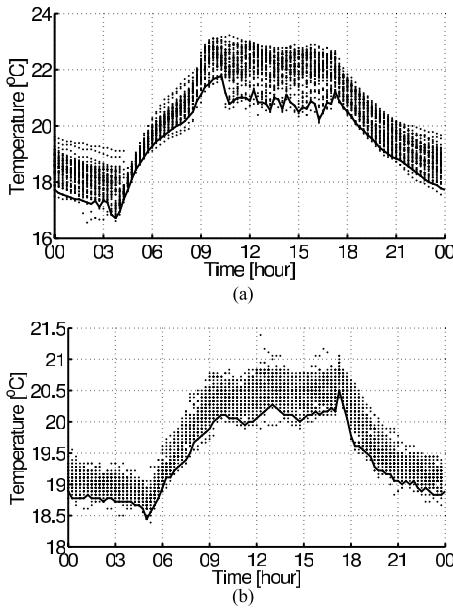


Fig. 8. Thermal comfort bounds. (a) Zone 2. (b) Zone 6.

- L3) SMPC [Explicit Stochastic Model Predictive Control (ESMPC)] solving (45) with chance constraints approximated using the discrete convolution method in Section III-B1.
- L4) SMPC [Sample-based Stochastic Model Predictive Control (SSMPC)] solving (45) with chance constraints approximated using the sample-based method in Section III-B2.
- L5) SMPC [Gaussian Stochastic Model Predictive Control (GSMPC)] solving (45) with disturbances pdf approximated as Gaussian distributions.

We focus on eight zones in the second floor of Bancroft Library from December 1, 2011 to February 1, 2012. The eight zones are subjected 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 SMPC controllers (13) 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 h). The model parameter for heating coils (p_h) and cooling coils (p_c) are 1.2 and 0.25, respectively.

The thermal comfort constraints (\bar{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 on HVAC systems (Fig. 1) 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 reduce its energy consumption compared with the BC.

Fig. 8(a) and (b) shows the comfort bounds for two of the eight zones in Bancroft Library at UC Berkeley. The dots in Fig. 8 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).

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 time invariant for each individual VAV box.

2) *Comparison Metrics*: The performance of the MPCs L1–L5 is evaluated by closed-loop simulation with system model (11a) for 55 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 BC logic.

1) Closed-loop energy savings compared with BC

$$S^\diamond = \frac{E_{\text{tot}}^{\text{BC}} - E_{\text{tot}}^\diamond}{E_{\text{tot}}^{\text{BC}}}, \quad (46)$$

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

$$E_{\text{tot}} = \sum_{t_s}^{t_f} P_{\text{tot}}(x_t, u_t^*, w_t) \Delta t, \quad (47)$$

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.

2) Total comfort improvement compared with BC

$$\Delta^\diamond = \frac{V_{\text{tot}}^{\text{BC}} - V_{\text{tot}}^\diamond}{V_{\text{tot}}^{\text{BC}}}, \quad (48)$$

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

$$V_{\text{tot}} = \sum_{t_s}^{t_f} \sum_{i \in \mathcal{V}} \max(\underline{T}_t^i(t) - T^i(t), 0) \Delta t, \quad (49)$$

3) Thermal efficiency of the HVAC system η

$$\eta = \frac{E_{\text{tot}}^{\text{thermal}}}{E_{\text{tot}}}, \quad (50)$$

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

$$E_{\text{tot}}^{\text{thermal}} = \sum_{t_s}^{t_f} (P_c/p_c + P_h/p_h) \Delta t. \quad (51)$$

TABLE I
SIMULATION RESULTS

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

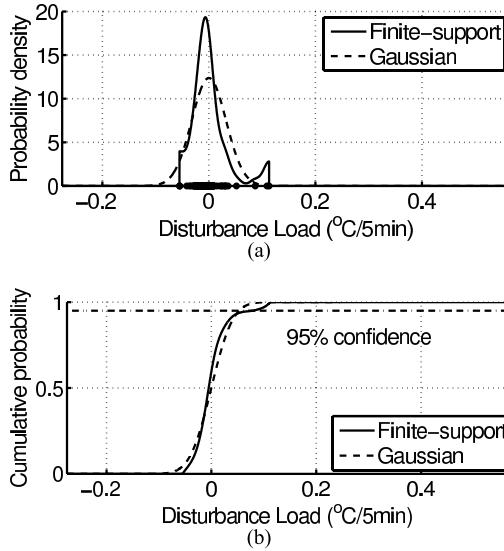


Fig. 9. Gaussian approximations for disturbances. (a) Zone 2. (b) Zone 4.

3) *Simulation Results:* Table I summarizes the results of our tests. Table I 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 BC. CMPC ignores the disturbance uncertainty at the design stage and this lead to the incapability of the CMPC to satisfy 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 the result of a more conservative approximation of the chance constraints in the SSMPC approach compared with discrete convolution.

GSMPC is computationally lighter than ESMPC and SSMPC when computing the chance constraint offset α off-line. Compared with the SSMPC and ESMPC methods proposed in Section III, GSMPC achieves similar energy savings while violating more comfort constraints. The coarse Gaussian approximations of disturbance are the reason for this. The approximation error of the load probability distribution function is shown in Fig. 9. In particular, Fig. 9(a) show the load Gaussian model (dashed lines) compared with a load model, which uses finitely supported density approximation (solid lines) using Gaussian kernels [43]. Fig. 9(b) show the cumulative density function of the pdf in Fig. 9(a).

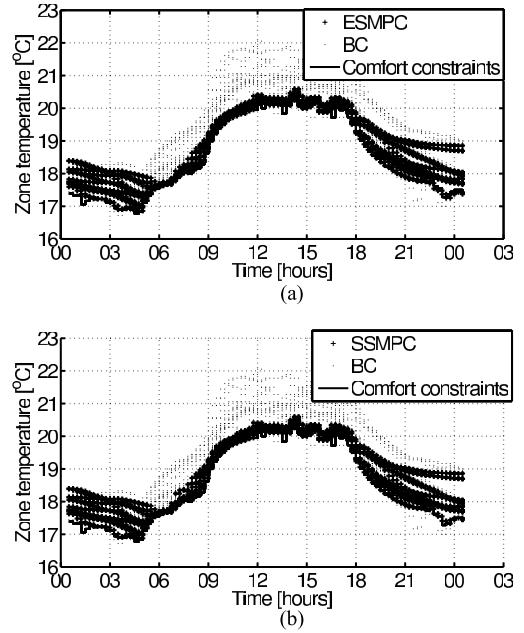


Fig. 10. Daily temperature for Zone 2. (a) Explicit stochastic MPC. (b) Sample based stochastic MPC.

The horizontal dash-dotted line in Fig. 9(b) show 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 (33) and thus a higher probability of comfort violations than that specified by chance constraints (13c).

In the ESMPC and SSMPC controllers the energy savings compared with the BC are the combination of two effects. First, the zone temperatures controlled by ESMPC and SSMPC are closer to the lower bounds (this can be observed in Fig. 10). The dots in Fig. 10(a) and (b) are the zone temperatures controlled by BC, the cross markers in Fig. 10(a) and (b) represent the zone temperatures controlled by ESMPC and SSMPC, respectively. This behavior is the results of using forecasted loads. Second, 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. However, the thermal efficiency of the HVAC system is 80% with any optimal controller L1, L2, ..., L5. This implies that the reason for SMPC uses less energy than BC is due to controlling the zone temperature closer to the lower bounds through the use of forecasts.

The results in Table I show the advantages of stochastic MPC formulation compared with nominal MPC with expected forecasts. Next, we show that this depends on the uncertainty level. We compare the energy consumption and comfort violations of CMPC and SSMPC with increasing levels of load uncertainty. At the design stage, CMPC ignores the uncertainty of the load predictions, and only considers 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 follows.

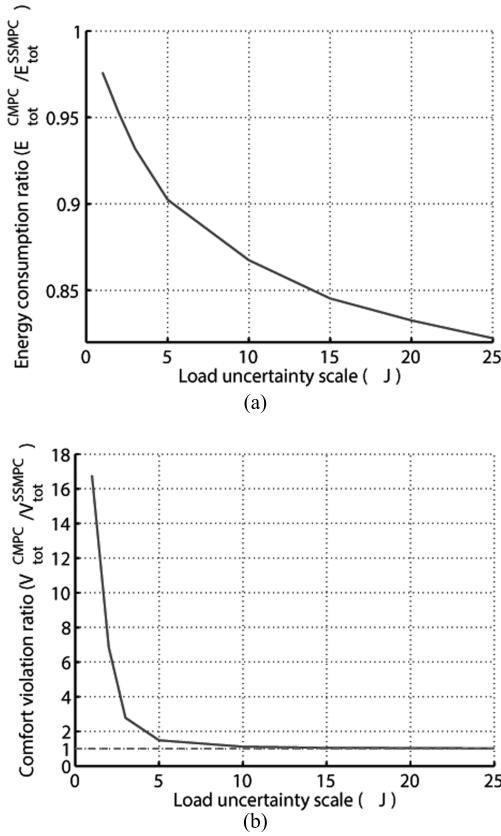


Fig. 11. Performance of CMPC and ESMPC with varying load uncertainty. (a) Energy consumption vs load uncertainty. (b) Comfort violations vs load uncertainty.

For SSMPC, we scale the independent disturbance samples

$$\{\tilde{w}_k^{j,1}, \tilde{w}_k^{j,2}, \dots, \tilde{w}_k^{j,N_{s,k}^j}\},$$

extracted from the pdf of the load uncertainty as

$$\{\vartheta \tilde{w}_k^{j,1}, \vartheta \tilde{w}_k^{j,2}, \dots, \vartheta \tilde{w}_k^{j,N_{s,k}^j}\}.$$

The scaled set of samples is used to compute the tightening offsets in Algorithm 2. For CMPC, the resulting control policy is independent of the scaling of load uncertainty as CMPC only takes into account the mean value of the load prediction at the design stage. The load realizations in closed-loop simulation have to be scaled accordingly

$$w(t)_\vartheta = \hat{w}(t) + \vartheta(w(t) - \hat{w}(t)), \quad (52)$$

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 Fig. 5 in Section II-C.

The higher the scaling factor ϑ the higher the prediction errors is. For example, weather forecast from a cheaper weather station might result in a high value of ϑ .

Fig. 11 reports the simulation results of CMPC and SSMPC for eight zones in the second floor of the Bancroft Library from December 1, 2011 to February 1, 2012. Fig. 11(a) show the energy consumption ratio $E_{\text{tot}}^{\text{CMPC}} / E_{\text{tot}}^{\text{SSMPC}}$ with increasing level of the load uncertainty, and Fig. 11(b) show the comfort violation ratio $V_{\text{tot}}^{\text{CMPC}} / V_{\text{tot}}^{\text{SSMPC}}$ with increasing level of load

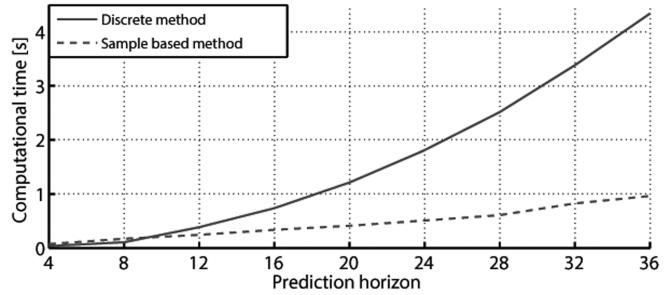


Fig. 12. Computational time for ESMPC and SSMPC.

uncertainty. It is observed that at low levels of load uncertainty ($\vartheta < 5$), SSMPC is able to improve the comfort satisfaction compared with CMPC while consuming comparable energy. However, for load uncertainty scaling factors $\vartheta > 5$, SSMPC loses its advantage, and CMPC shows comparable performance to SSMPC despite its simplicity at design stage.

B. Complexity Analysis for SSMPC and ESMPC

This section focuses on the complexity of SSMPC and ESMPC when transforming the chance constraints (13c) into the corresponding deterministic ones (32). Computational demands of sampled-based method and discrete convolution are analyzed with respect to three parameters: number of thermal zones N_v , prediction horizon length T , and number of samples N_s .

The discrete convolutions in (37) are the most computationally demanding operations in Algorithm 1. The complexity of discrete convolution is $\mathcal{O}(N_s^2)$, where N_s is the number of samples used to discretize the pdfs. The number of discrete convolutions needed to approximate the pdf $\text{pdf}_{\pi_k^j}$ is a linear function of $N_v T$, where N_v is the number of zones and T is the number of prediction steps. The computation of discrete convolution in (37) is repeated for $N_c = 2N_v T$ number of chance constraints in (13c), thus the complexity of the discrete method to compute the offsets in (13) is in 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 2 has complexity $\mathcal{O}(N_v^2)$ (see Step S2.2 of Algorithm 2). 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 2 is sorting the offsets $\alpha_k^{j,l}$ in (41). This has complexity of $\mathcal{O}(N_s \log N_s)$, if a quicksort algorithm is used. Taking all above-mentioned into account, we can conclude that overall complexity of Algorithm 2 is $\mathcal{O}(N_s \log N_s N_v^2 T^2)$ while the complexity of discrete method is $\mathcal{O}(N_s^2 N_v^2 T^2)$.

Fig. 12 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 Fig. 12, the solid line represents the average computational time for discrete method in Algorithm 1, and the dashed line represents the average computational time for sample-based method (Algorithm 2).

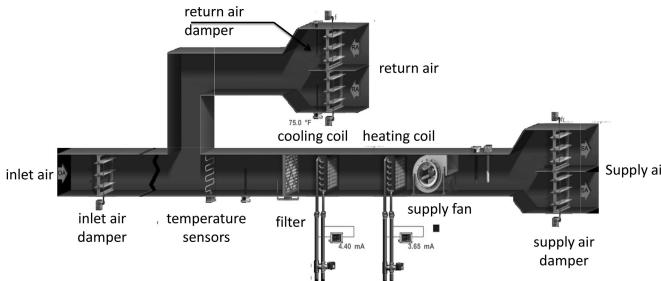


Fig. 13. Scheme plot for HVAC system in the lab.

While the ESMPC behaves as expected, the SSMPC shows a superlinear relationship between the SSMPC execution time and the prediction horizon length. The reason for such the discrepancy between the 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 2. On the other hand, a discrete convolution operation of discrete method in Algorithm 1 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 (45) with the exception of the tightening offset α_k^j . For the size and complexity of problems considered in our study, the average time to solve problem (45) with Ipopt is 5 s.

C. Experimental Results

In this section, 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 subjected to negligible solar radiation. The dimensions of the lab are 9 m × 9 m × 3.5 m, and it hosts 14 students with 14 desktop PCs running during weekdays. The VAV box serving the lab is shown in Fig. 13. The inlet air is the cool or warm air supplied by the central AHU 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.

In the following sections, we present the system identification results for the lab, the implementation details for the ESMPC controller, and some experimental results. The experiment demonstrates the feasibility of implementing stochastic MPC in real time. It also demonstrate how MPC can systematically handle time-varying energy cost in a predictive way. Performance improvement over a baseline controller is not the objective of this experiment. The experimental setup is simple enough (one zone) that a good control design can easily be achieved by an expert control contractor. However, as we have shown in the previous section, the MPC law can be computed in real-time for systems with many zones, multiple

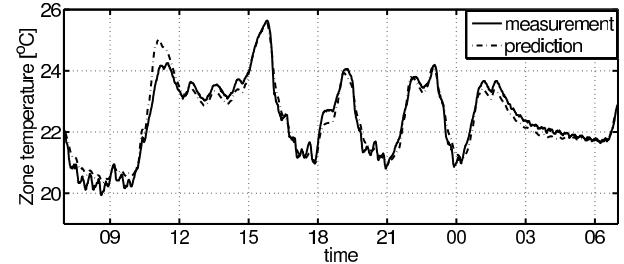


Fig. 14. Identification results for the lab model.

TABLE II
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
0.036	-0.063	0.033	0.049	3.924	0.946	-0.128	0.015

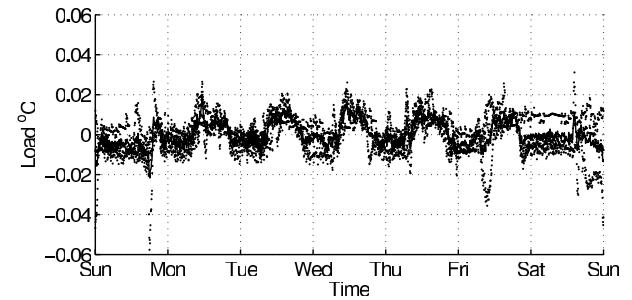


Fig. 15. Lab thermal load model.

components, and and with a complex energy cost structure. Therefore, we expect that energy savings will be observed when the proposed approach will be used for controlling an entire building.

1) *System Identification:* The thermal dynamics of the lab are modeled by the bilinear regression model (3) with $q_x = q_d = 2$. The regression parameters are fitted using historical data from weekend hours on September 27 and 28, 2012, since there is negligible occupancy load during weekends in the lab.

Fig. 14 shows 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 II.

The load uncertainty model is computed, as explained in Section II-C.

Fig. 15 shows the estimated weekly disturbance loads from April 1 to June 21, 2013. In Fig. 15, the dots are the realizations of the loads while the solid lines is the mean value of the weekly disturbance load. The samples in Fig. 15 are used to learn the statistics of the load model.

We use the ambient temperature uncertainty prediction model (6) where the pdf is shown in Fig. 7 and the mean value of the ambient temperature prediction is downloaded from services provided by National Oceanic and Atmospheric Administration (NOAA).

2) *ESMPC Implementation:* We use a sampling of 5 min, and a prediction horizon of 5 h. At time t , the ESMPC is implemented as follows.

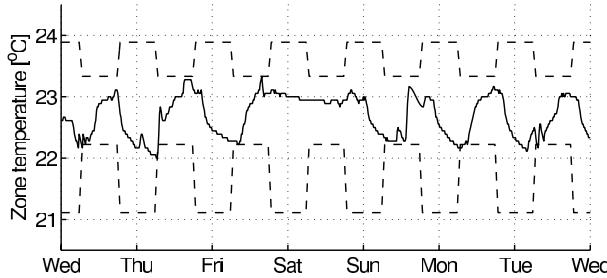


Fig. 16. Lab temperature.

- 1) Obtain the lab temperature sensor readings.
- 2) Downloads the weather predictions from the weather forecast services (NOAA).
- 3) Obtain the uncertain load predictions $P_{d,k}^i$. The uncertain load prediction is modeled as a lookup table learned from samples in Fig. 15. The load uncertainty model is updated when new measurements become available.
- 4) Solve the stochastic MPC problem (45) to obtain the optimal setpoint for the supply air mass flow rate and supply air temperature. The optimization problem is solved using Ipopt on a single Intel Core Duo CPU 3.00 GHz.
- 5) Send the optimal setpoint to low-level PID controllers, which regulate the coil valve positions and the fan speed to track the setpoints.

3) Experimental Results: We present two experiments. Fig. 16 reports the lab temperature of the first experiment from January 23–30, 2013. In Fig. 16, the dashed lines are the lower and upper bounds of the comfort region. The lower (upper) bound of the comfort set is 21.1 (23.9) °C for unoccupied hours from 16:00 to 06:00, and 22.3 (23.3) °C for occupied hours from 06:00 to 16:00 every day. It is observed that ESMPC maintains the zone temperature (solid line) within the comfort region, and the thermal comfort is guaranteed despite the uncertain load predictions in Fig. 15. The lower and upper bounds of the comfort region are read from the existing BMS system. The bounds during unoccupied hours are used by the baseline controller to avoid overcooling or overheating the building. This provides some comfort for people during unoccupied hours and also a smoother transient to occupied hours. This latter objective is easily addressed by the predictive nature of MPC. Therefore one could relax or remove the unoccupied hours constraint when using a MPC controller.

In Fig. 17, the solid line is the supply air temperature delivered to the lab, and the dashed-dotted line is the inlet air temperature delivered by the central AHU in Etcheverry building. The outside air temperature is plotted in Fig. 18.

One can notice four main heating control events on Wednesday, Thursday, Sunday, and Tuesday. They are the result of the MPC controller using forecasted load, occupancy schedule, zone dynamics, energy cost structure, and forecasted outside temperature. The heating events are necessary to keep the zone temperature within the comfort bounds.

The supply air mass flow rate is shown in Fig. 19. The lab requires a minimum of 600 cfm supply air flow rate during the unoccupied hours, and 800 cfm for occupied hours to meet the minimum ventilation requirement.

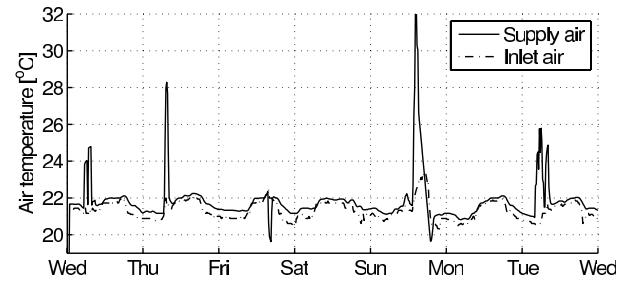


Fig. 17. Lab supply air temperature.

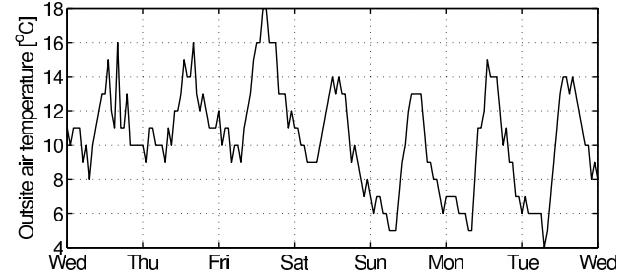


Fig. 18. Lab outside air temperature.

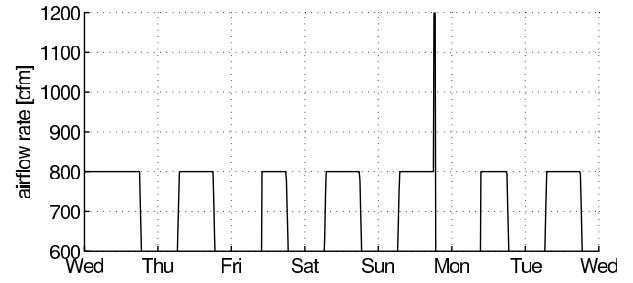


Fig. 19. Lab supply air mass flow rate.

Next, we present the second experiment. We consider a time-varying energy rate R . It is assumed that the energy price between 13:00 and 15:00 is twice higher than the rest of the day. Fig. 20 reports the lab temperature controlled by ESMPC (solid line) on the Friday of July 12th, 2013. In Fig. 20, the dashed lines are the lower and upper bounds of the comfort region. The lower and upper bounds of the comfort set is 21.7 °C and 23.7 °C, respectively. It is observed that ESMPC maintains the zone temperature (solid line) within the comfort region, and the thermal comfort is guaranteed despite the uncertain load predictions in Fig. 15. Additionally, it is noticed that the zone temperature is precooled before 13:00. This is because of the energy price increase between 13:00 and 15:00, and the feature of precooling enables the HVAC system to consume less energy when the energy is expensive.

In Fig. 21, the solid line is the supply air temperature delivered to the lab, and the dash-dot line is the inlet air temperature delivered by the central AHU in Etcheverry building.

The supply air mass flow rate is shown in Fig. 22. The lab requires a minimum of 600 cfm supply air flow rate during the unoccupied hours, and 800 cfm for occupied hours to meet the minimum ventilation requirement. It is

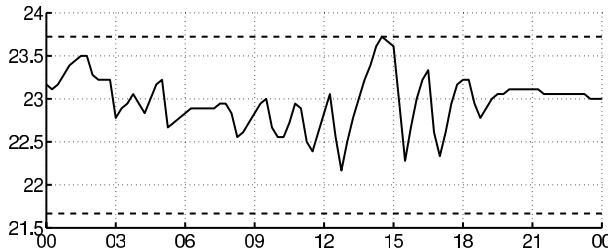


Fig. 20. Lab temperature.

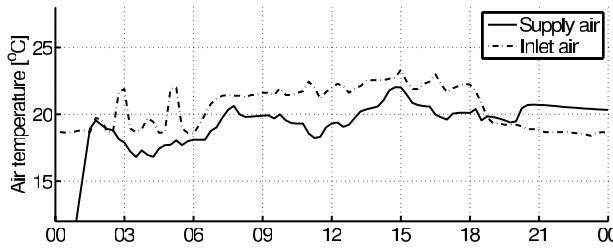


Fig. 21. Lab supply air temperature.

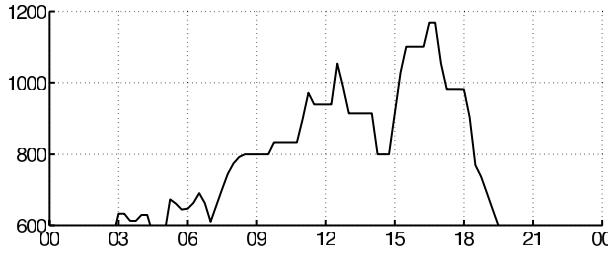


Fig. 22. Lab supply air mass flow rate.

observed that the air flow is decreased when energy price is higher from 13:00 to 14:00 so that the total energy bill is minimized.

V. CONCLUSION

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

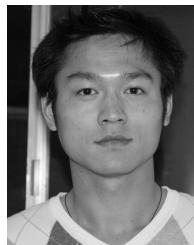
ACKNOWLEDGMENT

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors would like to thank the anonymous reviewers for their helpful comments on the original version of the manuscript.

REFERENCES

- [1] E. Arens, H. Zhang, and C. Huizenga, "Partial- and whole-body thermal sensation and comfort—Part I: Uniform environmental conditions," *J. Thermal Biol.*, vol. 31, nos. 1–2, pp. 53–59, 2006.
- [2] E. Arens, H. Zhang, and C. Huizenga, "Partial- and whole-body thermal sensation and comfort—Part II: Non-uniform environmental conditions," *J. Thermal Biol.*, vol. 31, nos. 1–2, pp. 60–66, 2006.
- [3] A. Aswani, N. Master, J. Taneja, D. Culler, and C. Tomlin, "Reducing transient and steady state electricity consumption in HVAC using learning-based model-predictive control," *Proc. IEEE*, vol. 100, no. 1, pp. 240–253, Jan. 2012.
- [4] A. Aswani *et al.*, "Identifying models of HVAC systems using semi-parametric regression," in *Proc. Amer. Control Conf.*, Jun. 2012, pp. 3675–3680.
- [5] A. Bemporad, F. Borrelli, and M. Morari, "Min-max control of constrained uncertain discrete-time linear systems," *IEEE Trans. Autom. Control*, vol. 48, no. 9, pp. 1600–1606, Sep. 2003.
- [6] J. E. Braun, "Reducing energy costs and peak electrical demand through optimal control of building thermal storage," *ASHRAE Trans.*, vol. 96, no. 2, pp. 976–887, 1990.
- [7] G. C. Calafiore, "Random convex programs," *SIAM J. Optim.*, vol. 20, no. 6, pp. 3427–3464, 2010.
- [8] G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 742–753, May 2006.
- [9] M. C. Campi and S. Garatti, "A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality," *J. Optim. Theory Appl.*, vol. 148, no. 2, pp. 257–280, 2011.
- [10] M. Cannon, P. Couchman, and B. Kouvaritakis, "MPC for stochastic systems," in *Assessment and Future Directions of Nonlinear Model Predictive Control* (Lecture Notes in Control and Information Sciences), vol. 358, R. Findeisen, F. Allgöwer, and L. Biegler, Eds. Berlin, Germany: Springer-Verlag, 2007, pp. 255–268.
- [11] M. Cannon, B. Kouvaritakis, and D. Ng, "Probabilistic tubes in linear stochastic model predictive control," *Syst. Control Lett.*, vol. 58, nos. 10–11, pp. 747–753, 2009.
- [12] M. Cannon, B. Kouvaritakis, S. V. Rakovic, and Q. Cheng, "Stochastic tubes in model predictive control with probabilistic constraints," *IEEE Trans. Autom. Control*, vol. 56, no. 1, pp. 194–200, Jan. 2011.
- [13] M. Cannon, D. Ng, and B. Kouvaritakis, "Successive linearization NMPC for a class of stochastic nonlinear systems," in *Nonlinear Model Predictive Control* (Lecture Notes in Control and Information Sciences), vol. 384, L. Magni, D. Raimondo, and F. Allgöwer, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 249–262.
- [14] A. Charnes and W. W. Cooper, "Chance-constrained programming," *Manag. Sci.*, vol. 6, no. 1, pp. 73–79, 1959.
- [15] E. Cinquemani, M. Agarwal, D. Chatterjee, and J. Lygeros. (2011). On convex problems in chance-constrained stochastic model predictive control. Swiss Federal Inst. Technol. Zürich, Zürich, Switzerland [Online]. Available: <http://control.ee.ethz.ch/index.cgi?page=publications;action=details;id=3317>
- [16] E. M. Constantinescu, V. M. Zavala, M. Rocklin, S. Lee, and M. Anitescu, "A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 431–441, Feb. 2011.
- [17] S. Goyal and P. Barooah, "A method for model-reduction of nonlinear building thermal dynamics," *Energy Buildings*, vol. 47, pp. 332–340, Apr. 2012.
- [18] S. Goyal, H. Ingle, and P. Barooah, "Effect of various uncertainties on the performance of occupancy-based optimal control of HVAC zones," in *Proc. IEEE 51st Annu. Conf. Decision Control*, Dec. 2012, pp. 7565–7570.
- [19] S. Goyal, H. Ingle, and P. Barooah, "Zone-level control algorithms based on occupancy information for energy efficient buildings," in *Proc. ACC*, Jun. 2012, pp. 3063–3068.

- [20] P. Haves *et al.*, "Model predictive control of HVAC systems: Implementation and testing at the University of California, Merced," Lawrence Berkeley National Laboratory, Univ. California at Berkeley, Berkeley, CA, USA, Tech. Rep. LBNL-3906E, 2010.
- [21] G. P. Henze, C. Felsmann, and G. Knabe, "Evaluation of optimal control for active and passive building thermal storage," *Int. J. Thermal Sci.*, vol. 43, no. 2, pp. 173–183, 2004.
- [22] G. P. Henze, D. E. Kalz, S. Liu, and C. Felsmann, "Experimental analysis of model-based predictive optimal control for active and passive building thermal storage inventory," *HVAC&R Res.*, vol. 11, no. 2, pp. 189–214, 2005.
- [23] P. Hokayem, E. Cinquemani, D. Chatterjee, F. Ramponi, and J. Lygeros, "Stochastic receding horizon control with output feedback and bounded control inputs," in *Proc. IEEE Conf. Decision Control*, Dec. 2010, pp. 6095–6100.
- [24] J. Jang, "System design and house dynamic signature identification for intelligent energy management in residential buildings," Ph.D. dissertation, Dept. Mech. Eng., Univ. California at Berkeley, Berkeley, CA, USA, 2008.
- [25] A. Kelman and F. Borrelli, "Bilinear model predictive control of a HVAC system using sequential quadratic programming," in *Proc. 18th IFAC World Congr.*, vol. 18, 2011.
- [26] A. Kelman, Y. Ma, and F. Borrelli, "Analysis of local optima in predictive control for energy efficient buildings," in *Proc. 50th IEEE CDC-ECC*, Dec. 2011, pp. 5125–5130.
- [27] A. Kelman, Y. Ma, and F. Borrelli, "Analysis of local optima in predictive control for energy efficient buildings," *J. Building Perform. Simulat.*, vol. 6, no. 3, 1–20, 2012.
- [28] S. Koehler and F. Borrelli, "Building temperature distributed control via explicit MPC and 'trim and respond' methods," in *Proc. ECC*, Jul. 2013, pp. 4334–4339.
- [29] B. Kouvaritakis, M. Cannon, S.V. Raković, and Q. Cheng, "Explicit use of probabilistic distributions in linear predictive control," *IEEE Trans. Autom. Control*, vol. 46, no. 10, pp. 1719–1724, Oct. 2010.
- [30] B. Kouvaritakis, M. Cannon, and R. Yadlin, "On the centres and scalings of robust and stochastic tubes," in *Proc. Int. Conf. Control UKACC*, 2010, pp. 565–570.
- [31] J. H. Lee and B. L. Cooley, "Optimal feedback control strategies for state-space systems with stochastic parameters," *IEEE Trans. Autom. Control*, vol. 43, no. 10, pp. 1469–1475, Oct. 1998.
- [32] Y. Ma, F. Borrelli, B. Hencey, B. Coffey, S. Bengea, and P. Haves, "Model predictive control for the operation of building cooling systems," in *Proc. Amer. Control Conf.*, Jun. 2010, pp. 5106–5111.
- [33] Y. Ma, F. Borrelli, B. Hencey, B. Coffey, S. Bengea, and P. Haves, "Model predictive control for the operation of building cooling systems," *IEEE Trans. Control Syst. Technol.*, vol. 20, no. 3, pp. 796–803, May 2012.
- [34] Y. Ma, F. Borrelli, B. Hencey, A. Packard, and S. Bortoff, "Model predictive control of thermal energy storage in building cooling systems," in *Proc. 48th IEEE Conf. CDC/CCC*, Dec. 2009, pp. 392–397.
- [35] Y. Ma, A. Kelman, A. Daly, and F. Borrelli, "Predictive control for energy efficient buildings with thermal storage: Modeling, simulation, and experiments," *IEEE Control Syst. Mag.*, vol. 32, no. 1, pp. 44–64, Feb. 2012.
- [36] Y. Ma, J. Matuško, and F. Borrelli. (2013). Stochastic model predictive control for building HVAC systems: Complexity and conservatism. Dept. Mech. Eng., Univ. California at Berkeley, Berkeley, CA, USA [Online]. Available: http://www.mpc.berkeley.edu/people/sergey-vichik/file/SystemID_DOE_Library.pdf
- [37] P. May-Ostendorp, G. Henze, C. Corbin, B. Rajagopalan, and C. Felsmann, "Model predictive control of mixed-mode buildings with rule extraction," *Building Environ.*, vol. 46, no. 2, pp. 428–437, 2011.
- [38] T. Nagai, "Optimization method for minimizing annual energy, peak energy demand, and annual energy cost through use of building thermal storage," *ASHRAE Trans.*, vol. 108, no. 1, pp. 976–887, 2002.
- [39] F. Oldewurtel *et al.*, "Energy efficient building climate control using stochastic model predictive control and weather predictions," in *Proc. ACC*, Jul. 2010, pp. 5100–5105.
- [40] Z. O'Neill, S. Narayanan, and R. Brahme, "Model-based thermal load estimation in buildings," in *Proc. 4th Nat. Conf. IBPSA-USA, SimBuild*, Aug. 2010.
- [41] J. A. Primbs, "A soft constraint approach to stochastic receding horizon control," in *Proc. 46th IEEE Conf. Decision Control*, Dec. 2007, pp. 4797–4802.
- [42] G. Schildbach, L. Fagiano, and M. Morari, "Randomized solutions to convex programs with multiple chance constraints," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2479–2501, 2012.
- [43] B. W. Silverman, "Density estimation for statistics and data analysis," in *Monographs on Statistics and Applied Probability*, London, U.K.: Chapman & Hall, 1986.
- [44] J. Skaf and S. P. Boyd, "Design of affine controllers via convex optimization," *IEEE Trans. Autom. Control*, vol. 55, no. 11, pp. 2476–2487, Nov. 2010.
- [45] 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 *Proc. 41st IEEE Conf. Decision Control*, vol. 4, Dec. 2002, pp. 4643–4648.
- [46] 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 *Proc. 40th IEEE Conf. Decision Control*, vol. 5, Dec. 2001, pp. 4228–4233.
- [47] S. Vichik and F. Borrelli. (2012). Identification of thermal model of DOE library. Dept. Mech. Eng., Univ. California at Berkeley, Berkeley, CA, USA [Online]. Available: http://www.mpc.berkeley.edu/people/yudong-ma/files/TCST_report.pdf
- [48] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Math. Program.*, vol. 106, no. 1, pp. 25–57, 2006.
- [49] V. M. Zavala, E. M. Constantinescu, T. Krause, and M. Anitescu, "On-line economic optimization of energy systems using weather forecast information," *J. Process Control*, vol. 19, no. 10, pp. 1725–1736, 2009.
- [50] V. M. Zavala, "Real-time optimization strategies for building systems," *Ind. Eng. Chem. Res.*, vol. 52, no. 9, pp. 3137–3150, 2013.



Yudong Ma received the B.S. degree in mechanical engineering with a minor in mathematics from Tsinghua University, Beijing, China, and the Ph.D. degree in mechanical engineering from the University of California at Berkeley, Berkeley, CA, USA, in 2007 and 2012, respectively.

He is currently a Mechanical Engineering Director at Brightboxtech, Inc., Berkeley, CA, USA. His current research interests include constrained optimal control, MPC, and its application to energy efficient building operation.



Jadranko Matuško received the B.S., M.S., and Ph.D. degrees from the University of Zagreb, Zagreb, Croatia, in 1999, 2003, and 2008, respectively.

He is currently an Associate Professor with the Faculty of Electrical Engineering and Computing, University of Zagreb. His current research interests include system identification and estimation, model predictive control, intelligent control, and mechatronic systems.



Francesco Borrelli received the Laurea degree in computer science engineering from the University of Naples Federico II, Naples, Italy, and the Ph.D. degree from the Automatic Control Laboratory, ETH-Zurich, Zurich, Switzerland, in 1998 and 2002, respectively.

He is currently an Associate Professor with the Department of Mechanical Engineering, University of California at Berkeley, Berkeley, CA, USA. He has authored more than 100 publications in the field of predictive control, and the book *Constrained Optimal Control of Linear and Hybrid Systems* (Springer Verlag). His current research interests include constrained optimal control, model predictive control, and its application to advanced automotive control and energy efficient building operation.

Dr. Borrelli was the Chair of the IEEE Technical Committee on Automotive Control, in 2008. He was a recipient of the 2009 NSF CAREER Award and the 2012 IEEE Control System Technology Award.