Advances in Applied Energy 9 (2023) 100124

Contents lists available at ScienceDirect

Advances in Applied Energy

journal homepage: www.elsevier.com/locate/adapen

Multi-zone building control with thermal comfort constraints under disjunctive uncertainty using data-driven robust model predictive control

Guoqing Hu[a], Fengqi You[a,b,*]

[a] Systems Engineering, Cornell University, Ithaca, NY 14853, USA
[b] Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Keywords:
Robust model predictive control
Machine learning
Multi-zone building
Relative humidity control
Uncertainty

ABSTRACT

This paper proposes a novel data-driven robust model predictive control (MPC) framework for a multi-zone building considering thermal comfort and uncertain weather forecast errors. The control objective is to maintain each zone's temperature and relative humidity within the specified ranges by minimizing the energy usage of the underlying heating system. A state-space model is developed to use a hybrid physics-based and data-driven method for the multi-zone building's temperature and relative humidity. The temperature and humidity RMSEs between the state-space model and the EnergyPlus-based model are less than 0.25 ˚C and 5.9%, respectively. The uncertainty space is based on historical weather forecast error data, which are clustered by using a k-means clustering algorithm. Machine learning approaches, including principal component analysis and kernel density estimation, are used to construct each basic uncertainty set and reduce the conservatism of resulting robust control action under disturbances. A robust MPC framework is built upon the proposed state-space model and data-driven disjunctive uncertainty set. An affine disturbance feedback rule is employed to obtain a tractable approximation of the robust MPC problem. Besides, the feasibility and stability of the proposed MPC are discussed in detail. A case study of controlling temperature and relative humidity of a multi-zone building in Ithaca, New York, USA, is presented. The results demonstrate that the proposed framework can reduce up to 8.8% of total energy consumption compared to conventional robust MPC approaches. Moreover, the proposed framework can essentially satisfy the thermal constraints that certainty equivalent MPC and robust MPC largely violate.

1. Introduction

Around 40% of the global energy is used for building temperature control [1]. In particular, the energy used for controlling the temperature and relative humidity is responsible for 30% of average U.S. household power consumption [2]. This level of energy consumption makes measures targeted on energy saving of these building systems very engaging. Model predictive control (MPC) is a powerful method for building energy control. MPC is a model-based control strategy that determines the optimal control sequence by solving a sequence of numerical optimization problems with constraints over a specific horizon based on the prediction model. The first input in the optimal sequence is sent into the system, and the entire computation is repetitively performed at subsequent control intervals following a receding horizon approach [3]. In building control applications, MPC has been reported to save a tremendous amount of energy usage compared to the rule-based control strategies [4,5].

Despite the success of MPC application in building energy systems and existing literature in this area [6], there are some limitations in the previous studies. In [6], only single-zone control was considered. When MPC is applied to the multi-zone building control, the resulting dynamic optimization problem could be large-scale due to its multi-input, multi-output (MIMO) characteristic [7]. For the large-scale system, it is challenging to derive an accurate system model and descriptions of its operating conditions [8]. In MPC approaches, uncertainties are often modeled as additive disturbances acting on the system. In the building control cases, the uncertainty of measured disturbances, such as inaccurate ambient weather prediction, can also affect the control performance of MPC [9], as this uncertainty was not considered in [4,10]. A possible approach to mitigate the effect of uncertainties on MPC is the robust techniques, which considers uncertainties of disturbances during solving the optimization problem [11] and ensures constraints satisfaction under stochastic conditions in a compact set [12]. Studies use robust MPC (RMPC) to make building controllers robust by modeling uncer-

* Corresponding author at: Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA
E-mail address: fengqi.you@cornell.edu (F. You).

https://doi.org/10.1016/j.adapen.2023.100124
Received 12 September 2022; Received in revised form 29 January 2023; Accepted 31 January 2023
Available online 1 February 2023

Nomenclature

A	system matrix of system states in compact form
B_u	system matrix of control inputs in compact form
B_v	system matrix of deterministic disturbances in compact form
B_w	system matrix of uncertain disturbances in compact form
c_p	specific heat of air
F_u	coefficient matrix for control input constraints in compact form
F_x	coefficient matrix for state variable constraints in compact form
f_u	coefficient vector for control input constraints in compact form
f_x	coefficient vector for state variable constraints in compact form
$m_{air,t}$	mass of supply air during the period from time t to time $t+1$
$m_{dehum,t}$	mass of water vapor taken by dehumidifier during the period from time t to time $t+1$
$m_{hum,t}$	mass of water vapor provided by humidifier during the period from time t to time $t+1$
$m_{heat,in,t}$	mass of air flow provided by air circulation during the period from time t to time $t+1$
$m_{heat,out,t}$	mass of air flow taken by air circulation during the period from time t to time $t+1$
Q	heat energy
RH_t	relative humidity in the zone at time t
$RH_{out,t}$	ambient relative humidity at time t
$T_{air,t}$	ambient temperature at time t
$T_{zone,t}$	zone temperature at time t
V_{zone}	zone volume
ΔT	temperature difference between heated supply air and ambient
δ_T	temperature difference between heated supply air and zone
$\rho_{heat,sat,t}$	saturated water vapor density at ambient temperature at time t
$\rho_{abs,t}$	absolute water vapor density at time t
ρ_{air}	air density
$\rho_{water,sat,t}$	saturated water vapor density at zone temperature at time t
$\rho_{sat,t}$	saturated water vapor density at time t

tainties using conventional box uncertainty sets [9,11]. However, a box uncertainty set can potentially lead to over-conservative solutions to the RMPC problems [13].

A promising alternative to RMPC is data-driven RMPC, which is capable of capturing the high-density region of uncertainties in decision-making for each time step. This method has established itself as an effective approach to monitor, control, and optimize industrial processes [14]. One of the promising methods is from [15], which constructed the data-driven robust optimization framework using principal component analysis (PCA) in conjunction with kernel density estimation (KDE). PCA can capture the correlations between uncertain parameters and identify the uncertainty sources. Meanwhile, KDE can help in extracting distributional information of latent uncertainties which are projected onto each principal component [15]. On the other hand, the weather forecast errors can have a complex disjoint-set data structure [16]. Hence, the introduction of data-driven disjunctive uncertainty sets can alleviate the over-conservatism issue confronted by the most commonly used single uncertainty set-based RMPC frameworks. However, to the best of our knowledge, no existing literature focused on introducing disjunctive uncertainty sets to data-driven RMPC for multi-zone building temperature and humidity control. Therefore, there is a knowledge gap present in systematically and effectively handling uncertain weather forecast errors with disjunctive data structures in multi-zone building control problems by introducing machine learning algorithms with the RMPC.

In order to fill this knowledge gap, we propose a novel data-driven RMPC framework that minimizes energy consumption for controlling the temperature and relative humidity of multi-zone buildings under uncertain disturbances. To develop the state-space model (SSM) required for the proposed MPC, we start by studying the dynamics of temperature and relative humidity within each thermal zone based on the building geometry and airflow dynamics. Afterward, we use data-driven techniques and system identification methods to construct the SSM and then incorporate the SSM into the proposed data-driven RMPC framework. Historical weather forecast data and historical weather measurement data about ambient temperature and relative humidity are gathered. Historical data on weather forecast errors that represent uncertain disturbances is obtained by calculating the difference between forecast values and measurement values. The RMPC framework uses the k-means clustering algorithm to formulate the disjunctive uncertainty set to tackle the complex data structure of weather forecast errors [16]. Each basic uncertainty set is then constructed by PCA coupled with KDE. Based on the proposed linear SSM and the disjunctive uncertainty set, we propose a novel data-driven RMPC framework, which is capable of controlling the multi-zone building's temperature and relative humidity. The proposed framework not only is computationally efficient but also mitigates the conservatism issue of conventional robust optimization based MPC approaches. To avoid infeasibility cases, a soft constraint is introduced to ensure the feasibility for the entire sequence of the control optimization problems. The linear constraints for the control optimization problems are formulated and presented in Section A.3 of Appendix. Then we establish system stability in a mean-square sense under mild assumptions on the properties of disturbances. The feasibility and stability issues of the proposed control scheme are addressed properly in Section A.1 of Appendix. Finally, we carry out simulations for a multi-zone building located in Ithaca, New York, USA, and compare the proposed framework with other MPC approaches, including certainty equivalent MPC (CEMPC), RMPC with conventional box uncertainty set, data-driven RMPC with disjunctive conventional box uncertainty sets clustered by k-means (KM-RMPC), and RMPC with PCA and KDE based data-driven non-clustered uncertainty set (PKRMPC).

The main contributions of this work are summarized as follows:

- A new data-driven RMPC framework for multi-zone building control capable of capturing disjunctive uncertainty data from uncertain weather forecast errors for both temperature and humidity control;
- A formal feasibility and stability guarantee of the data-driven RMPC framework for multi-zone building control;
- A simulation study utilizing historical weather data to control a multi-zone building located in Ithaca, New York, USA.

The rest of this paper is organized as follows. Section 2 presents the proposed SSM. Section 3 proposes the data-driven RMPC framework with disjunctive data-driven uncertainty sets. The comparison results with other MPC approaches are presented in Section 4. The conclusion of this paper is summarized in Section 5.

2. State-space model formulation

The aims of the following text are (i) to develop the nonlinear physics-based model for the multi-zone building control, and (ii) to reduce it to a linear SSM. The physics-based model, which predicts the zone temperature and relative humidity in the building, is constructed. Those two factors are considered since they contribute the most to the predicted mean value (PMV), which is widely used for determining thermal comfort [17]. However, the computation effort of PMV index requires is substantial by considering its complex and highly nonlinear

correlation between various building parameters [18]. The temperature and relative humidity are selected to be controlled in this work, with regarding to their importance in computing PMV [19]. Nevertheless, the nonlinear physics-based model requires large computational efforts. Thus, we reduce the order of the model via the means of system identification, which can sustain the computation efficiency and retain an accurate description of the physical correlation between the parameters.

2.1. Physics-based model construction

Measuring the temperature values within each zone is crucial to develop the MPC for the multi-zone building. Energy forecasting models for building energy systems are important for developing better control strategies for energy management [20]. Thus, we derive a Resistance-Capacitance (R.C.) model from describing the temperature within the multi-zone building by using the Building Resistance-Capacitance Modeling (BRCM) MATLAB toolbox is constructed based on the EnergyPlus[21]. The R.C. modeling methodology of BRCM uses an electric analogy to equalize the building energy flow. It uses resistances and capacitances to represent the thermal resistances and thermal capacitances of the building elements, and voltages and currents represent the temperatures of the building elements and heat transfers between those. The air handling unit includes a return air heat recovery system and a heater. The heat from heater is generated using a gas boiler. The assumptions of the air handling unit include (i) there is no air mixing taking place in the heat exchanger [22] and (ii) air handling unit receiving the constant volume of air [23]. The models simulated by EnergyPlus and TRNSYS, and their accompanying solutions, have high nonlinearity which could be complicate and not be capable of being applied to real-time control and optimization [23]. According to work [24], both EnergyPlus and TRNSYS calculate the building performance on the basis of the solution of thousands of equations for every control interval even for a medium-sized building [25]. EnergyPlus can have the computation time around 600 s for the similar multi-zone building control [26]. Such a long computation time is unsuitable for the real-time MPC control because the optimization framework may iterate the simulation model multiple times, usually more than 15 times, until the optimal solution is found [27]. On the other hand, EnergyPlus has limitation in incorporating the advanced control algorithms in the built-in software [28]; hence the tremendous efforts are needed to add robust MPC based control strategies into the simulation. Therefore, reduced order model is preferred in the real-time control framework [29]. In this work, BRCM toolbox can return an accurate description of the self-designed building's temperature values with a linearized SSM for MPC with a relatively high computational efficiency. Such a long computation time is not suitable for the real-time control which has 15 min interval and needs recursive simulation this toolbox can return an accurate description of the self-designed building's temperature values with a linearized SSM for MPC with a relatively high computational efficiency [30]. This toolbox widely is adopted in building control works, including the perspective of developing MPC approach [31], data-driven robust MPC approach [32], building energy simulation [33].The form of linear SSM returned by BRCM [21] is shown as follow:

$$x_{t+1} = Ax_t + B_u u_t + B_v v_t + B_w w_t \tag{1}$$

where x_t, u_t, v_t, w_t are system states, control inputs, ambient disturbances, and forecast uncertainties at time t, respectively. x stands for the state variables which includes floor temperature, ceil temperature, wall temperature, room temperature. The power consumption rates of the heater, radiator, and humidifier are included in u. v stands for environmental disturbances, which include forecast ambient temperature, underground temperature, and forecast ambient humidity. w stands for the forecast uncertainties, which are the differences between forecast and measurement of the temperature and humidity. In this control process, the constraints are set throughout the prediction horizon H as following:

$$x_{\min,t} \le x_t \le x_{\max,t}, \forall t \in \{1, 2, \ldots, H-1\} \tag{2}$$

$$u_{\min,t} \le u_t \le u_{\max,t}, \forall t \in \{1, 2, \ldots, H-1\} \tag{3}$$

$x_{min,t}$ and $x_{max,t}$ are the lower and upper limits of the zone temperature and relative humidity within individual thermal zone. Such state constraints are introduced to maintain the zones' thermal comfort. $u_{min,t}$ and $u_{max,t}$ are the lower and upper bound for the control actuators. The upper bounds of temperature are introduced in the control framework to avoid excessive heating. The upper bound is not applied to humidifier to ensure the feasibility of the solution. In addition, polytopic input constraints are also considered to describe the nonnegativity of the energy supply. The comfort bounds are enforced only during the occupied time (9:00 AM – 5:00 PM), but the input constraints are set throughout the controlling process [34,35].

Considering that relative humidity is crucial to the human health condition [36,37], we extend this physics-based model to consider the relative humidity within each zone, which is not included in the BRCM. We build the control volume thermodynamic model to study the airflow and relative humidity values inspired by [23]. Fig. 1 illustrates the thermodynamic model of each thermal zone.

In [38], several assumptions are made in the derivation of the physics-based model: (i) ideal gas behavior, (ii) perfect mixing, (iii) constant pressure process, (iv) negligible infiltration and exfiltration effects. Based on assumption (iv), the water vapor can be assumed to be exclusively adjusted by the amount of air supply and humidifiers. In this work, we set the constant air flow rate V_{air} to ensure that air in each zone is constantly circulating [23,39].

The total mass of water vapor is used because the discrete-time model is adopted in this work. The discrete-time model usually considers the integration of the state change rate and is widely used in building control [31,40]. The mass of water vapor brought by and taken away from the air circulation can be further calculated from (4)–(7):

$$\rho_{zone,sat,t-1} = f\left(T_{zone,t-1}\right) \tag{4}$$

$$\rho_{out,sat,t-1} = f\left(T_{out,t-1}\right) \tag{5}$$

$$m_{air,in,t-1} = \rho_{out,sat,t-1} \cdot RH_{out,t-1} \cdot V_{air} \tag{6}$$

$$m_{air,out,t-1} = \rho_{zone,sat,t-1} \cdot RH_{t-1} \cdot V_{air} \tag{7}$$

where $\rho_{zone,sat,t-1}$ is the saturated water vapor density in the zone at t-1. $\rho_{out,sat,t-1}$ is the saturated water vapor density from the ambient environment at *t-1*. f is a linear function over the saturated water vapor density at different zone temperatures. The saturated vapor density values and their corresponding temperature values are collected in [41], and [23] proves that the linear approximation is in good agreement with the results from the nonlinear equations between saturated vapor density and temperature. The linear regression result is also shown in Section 1 of Appendix. $T_{zone,t-1}$ is the zone temperature at *t-1*, and $T_{out,t-1}$ is the ambient temperature at *t-1*.

Besides, (8) can also help in finding the $m_{water,t-1}$, mass of the water vapor within the thermal zone during the period from time *t-1* to time *t*,

$$m_{water,t-1} = \rho_{water,sat,t-1} \cdot RH_{t-1} \cdot V_{room} \tag{8}$$

Finally, the mass conservation equation can be used to determine $m_{water,t}$, the mass of the water vapor in the zone at time *t* based on (9).

$$m_{water,t} = m_{water,t-1} + m_{hum,t-1} + m_{air,in,t-1} - m_{air,out,t-1} \tag{9}$$

Therefore, relative humidity at time *t*, RH_t, can be determined by using (10).

$$\rho_{abs,t} = \frac{m_{water,t}}{V_{zone}}, \ \rho_{sat,t} = f\left(T_{room,t}\right), \ RH_t = \frac{\rho_{abs,t}}{\rho_{sat,t}} \tag{10}$$

Fig. 1. Individual thermal zone's climate is adjusted by artificial control actuators (including a heater, a radiator, and humidifiers) and climate disturbances (including humidity, air temperature, and ground temperature).

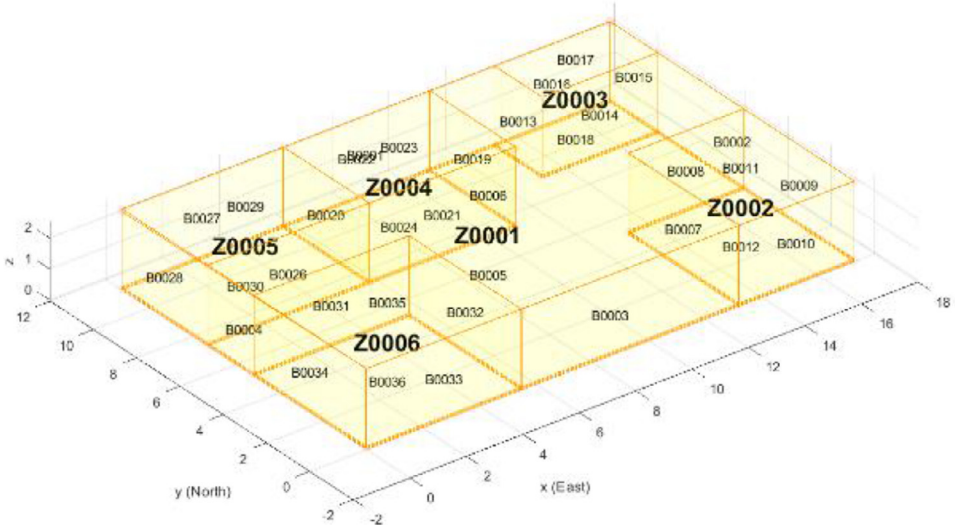

Fig. 2. 3-D modeling of a multi-zone building in BRCM. A single-story building is chosen because this type of building is focused and studied by many studies [47,48].

In this humidity model, following [35,42,43], we do not consider the effects from occupancy, lighting, etc., but can be added in our model. The trend of those effects can be found in [44] and inserted into (9).

This nonlinear physics-based model has a high computational cost, which can be mitigated by linearization [45]. Therefore, to maintain the prediction accuracy and improve the computational efficiency, the system identification method is used to reduce the order of the physics-based model.

2.2. System identification

A multi-zone building is self-designed and shown in Fig. 2, constructed based on the BRCM [21] toolbox found in MATLAB. The major goal is to minimize energy usage while maintaining thermally comfortable environments for all zones. Each zone has a humidifier to control the relative humidity and is assumed to receive the same fraction of the supplied air. Therefore, the amount of airflow received by each zone is constant. This model assumes the constant volume airflow intake because it is computationally efficient and can be adopted in real-time control. This assumption can also be found and justified by [46]. Besides, the ambient temperature and relative humidity have impact on each thermal zone. Temperature and relative humidity adjustments are necessary to maintain thermal comfort. Excessive heating should be prevented for the purpose of saving energy and preserving a thermally comfortable environment.

The temperature values in each zone are determined based on the BRCM toolbox. The discrete-time SSM is commonly used for building modeling, and a shorter time interval is preferred in the simulation process [23,49]. Therefore, the control interval is set to 15 min, which is the shortest interval available in BRCM. This time interval is found to be a good compromise between a sufficiently fast reacting control and the complexity of the result optimization problem [31]. Moreover, with shorter-term control intervals, the framework requires shorter-term weather forecast data, which is justified to be more accurate [50]. We feed 1 h ahead forecast values into the control system to develop the optimal control strategy over each time interval. Consequently, the relative humidity of each zone is incorporated into the integrated SSM by the nonlinear physics-based model proposed above. Then the integrated linear SSM is constructed on the basis of both the BRCM SSM and linearization to the relative humidity model by using 'greyest' in MATLAB system identification toolbox. The linear SSM is computed because of its computational efficiency for real-time control [51,52]. The SSM has the form shown in (1), where the matrices have the following specifications: A is a 30 × 30 matrix (6 individual thermal zones with zone temperature, wall temperature, ceil temperature, floor temperature, and relative humidity values within each); B_u is a 30 × 8 matrix which has heater and radiator combined temperature-control system and six sets of humidifiers installed to control the relative humidity. B_v is updated to the dimension of 30 × 3, combining the impact from ambient temperature, underground temperature, and ambient relative humidity. B_w,

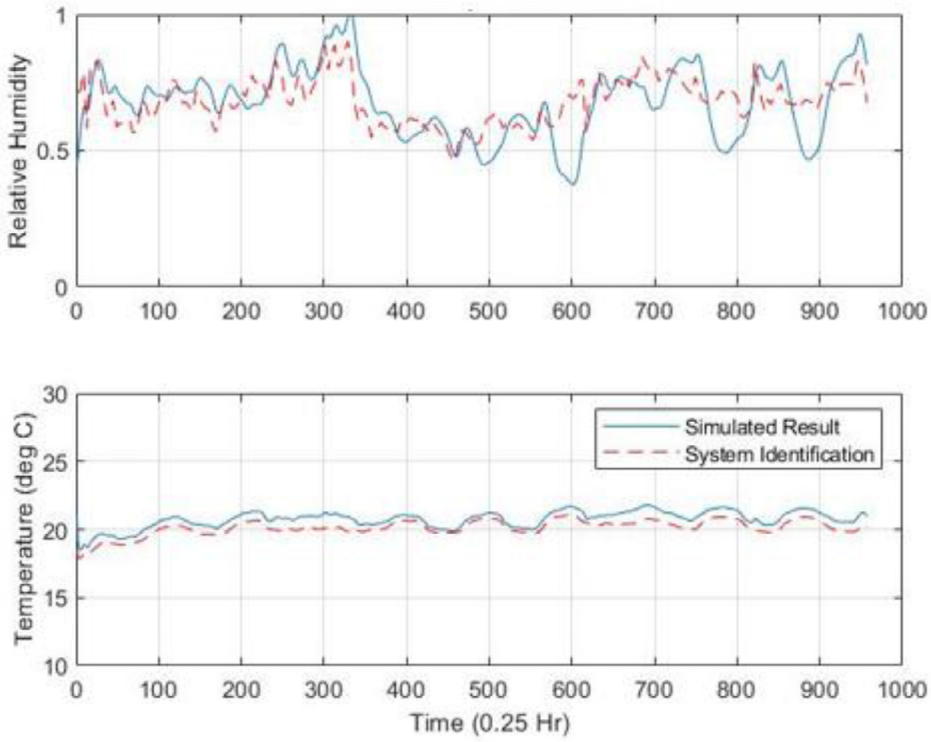

Fig. 3. The comparison of system identification results and EnergyPlus-based model for the worst-case scenario. It can be observed that the red line (predicted by linear SSM) essentially follows the blue line (simulated by the EnergyPlus-based model). This plot is for the case study building's room 5, which is selected because it has the worst linearization results compared to other rooms.

on the other hand, considers two sources of forecasting error: from ambient temperature and from ambient relative humidity, thereby having the dimension of 30×2.

System identification is used to obtain a linearized model for MPC [53]. Here we assume that the indoor relative humidity has a negligible impact on the wall, floor, and ceil temperature, as in [38]. We apply the system identification to the nonlinear physics-based model to retrieve the linearized SSM that can be used for multi-zone building control. The five-zone building model described in Section 4 is adopted to identify the linear SSM. The Fig. 3 demonstrates the comparison of the prediction of this linear SSM with the results from the EnergyPlus-based model. On the other hand, in Section A.2 of the Appendix, The root mean square error (RMSE) between the prediction from linear SSM and simulated results from EnergyPlus-based model with training data and testing data are demonstrated.Fig. 4.

When linear SSM is used to predict the zone temperature and relative humidity with the data which are beyond the range of the sample data, the RMSEs of indoor zone temperature are between 0.20 °C and 0.25 °C, and for relative humidity is between 4.89% and 5.88%. The error values are relatively close to those in [19], indicating a good agreement of the linearized SSM to the EnergyPlus-based model. This level of RMSEs has minimal impact on MPC performance in achieving a thermal comfort environment in the building, as those are tested in [9,54]. Therefore, this SSM can be used for formulating the following control framework.

3. Data-driven robust model predictive control

After SSM is formed, the disjunctive data-driven uncertainty sets should be constructed for the MPC framework. From the control perspective, the major challenge lies in the inherent uncertainty of weather forecasting due to the stochastic nature of atmospheric processes and the imperfection of the weather model [55]. In previous studies, the uncertainties from the weather forecast errors are modeled as the differences between the forecast values and measured ones [6,56–58]. To tackle the issue arising from the forecast uncertainties' complex disjoint-set data structure [16], we use the k-means algorithm to cluster the uncertainty data into multiple groups. Within each basic uncertainty set, the polytope is identified to compactly enclose the high-density region of disturbance via the means of PCA and KDE.

Inspired from [59], we construct the disjunctive uncertainty sets to better learn the trend of the uncertainty data. Clustering is necessary because of the complex disjoint-data structure of the forecast uncertainties [16]. Specifically, k-means clustering has been found to be an effective tool in data classification [60,61] and is adopted in this work. First, normalization to the uncertainty data is recommended, shown in (11), to facilitate the convergence of the k-means algorithm [62].

$$w_0 = w - 1\mu_0^T \tag{11}$$

μ_0^{T} is the mean values of the uncertainty sets, w is the original uncertainty set and w_0 is the normalized uncertainty set. Afterward, the groups are identified by minimizing the sum of intra-cluster variances, i.e., squared Euclidean distance [63].

$$D^* = \arg\min \left(\sum_{i=1}^{k} \sum_{w \in D_i} \|w - \mu_i\|^2 \right) \tag{12}$$

Despite the number of uncertainty set groups, the conventional box uncertainty set can lead to over-conservative solutions [64]. Therefore, PCA and KDE are adopted to cope with polyhedral shapes' data . Thanks to the previous normalization process, PCA [65] can then maximize the variance of the uncertainty under the same scale. The covariance matrix can be approximated as,

$$S_i = \frac{1}{N-1} w_i^T w_i, \ \forall i \in k \tag{13}$$

where i is the index for clusters of uncertainty data, and k represents the set of resulting clusters. PCA is a powerful approach on modeling high dimensional data [66] by projecting the correlated raw datasets in to their corresponding uncorrelated principal components [65]. As the covariance matrix S_i can be further decomposed as $S_i = Q_i \Lambda_i Q_i^T$, where Q_i's column contains all the eigenvectors corresponding to the eigenvalues stored in the diagonal matrix Λ_i [67]. The individual eigenvalue represents the variance of this axis if data is projected on this eigenvector.

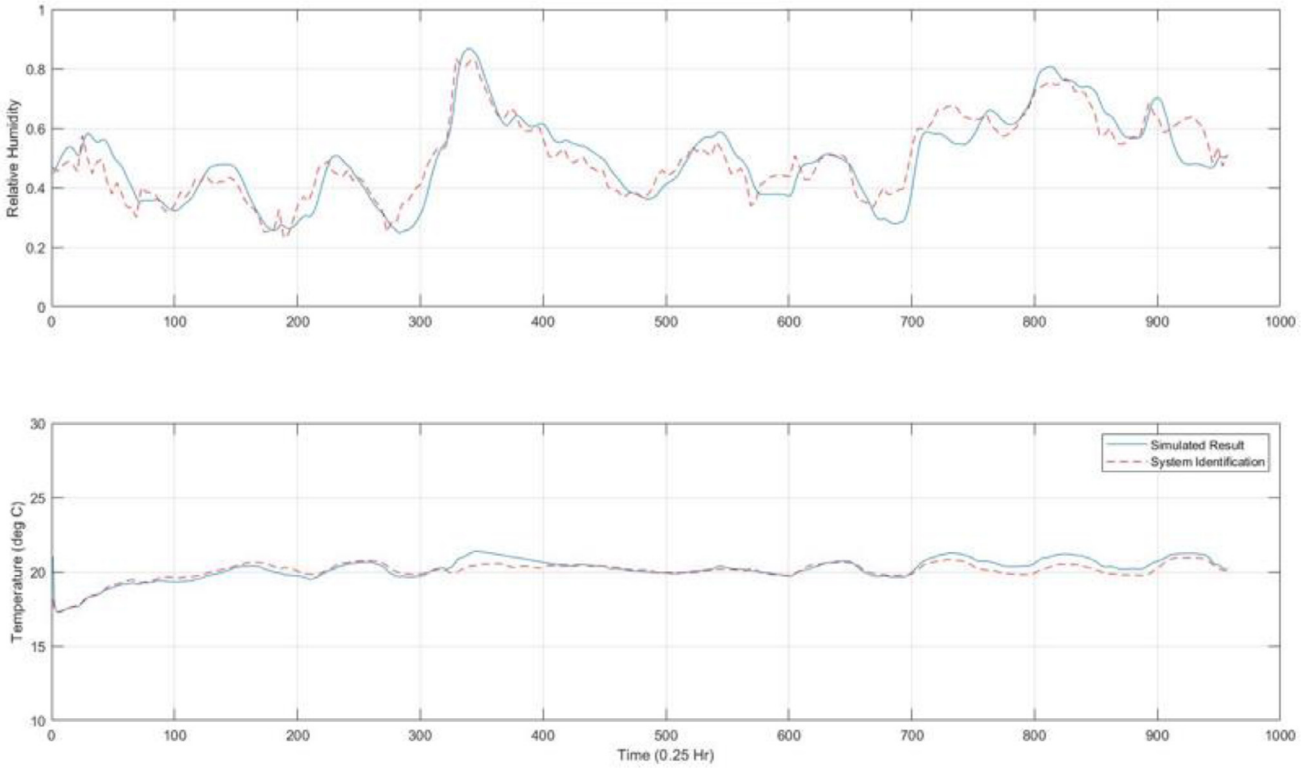

Fig. 4. The comparison of system identification results and EnergyPlus-based model for the worst-case scenario. The fitting curve (shown in a red dashed line) is observed to accurately predict the values simulated from the EnergyPlus model, indicating the feasibility of applying this linear model for control framework development.

Finally, it can be studied the distributional information of the uncertainty dataset of each component j within the ith clustering group via the KDE approach [68]:

$$f_{j,i} = \frac{1}{N}\sum_{n=1}^{N} K(\xi_{j,i}, p_{j,i}^{(n)}), \forall i \in k \tag{14}$$

where $\xi_{j,i}$ is the inferred latent variable, and $f_{j,i}$ is the acquired probability density function of $\xi_{j,i}$. K stands for the standard Gaussian density function. $p_{j,i}$ is the projection of uncertainty data point w_i onto the principal components, which are found from the eigenvectors of matrix S_i in (13).

With knowing the probability density function, the cumulative density function can be written as follows:

$$F_{j,i}^{-1}(\alpha) = \min\{\xi_{j,i} | F_{j,i}(\xi_{j,i} \geq \alpha)\}, \forall i \in k \tag{15}$$

whereα is the pre-specified small quantile parameter, ranging from 0 to 0.5. $F_{i,j}$ is the cumulative probability function of $f_{i,j}$ found in equation (18) [68]. The uncertainty set $\mathbb{W}^k$within a cluster k can be formulated under the introduction of forward and backward deviation variables z^+ and z^- [15]:

$$\mathbb{W}^k = \left\{ \mathbf{w}^k \in R^H \middle| \begin{array}{c} \mathbf{w}^k = \hat{\boldsymbol{\mu}}_k + Q_k \boldsymbol{\xi}_k, \boldsymbol{\xi}_k = \underline{\boldsymbol{\xi}}_k z^- + \bar{\boldsymbol{\xi}}_k z^+ \\ \mathbf{0} \leq z^+, z^- \leq \mathbf{1}, z^+ + z^- \leq \mathbf{1}, \mathbf{1}^T\left(z^+ + z^-\right) \leq \Gamma \\ \underline{\boldsymbol{\xi}} = \left[\hat{F}_{1,k}^{-1}(\alpha), ..., \hat{F}_{1,k}^{-1}(\alpha)\right]^T \\ \bar{\boldsymbol{\xi}} = \left[\hat{F}_{1,k}^{-1}(1-\alpha), ..., \hat{F}_{1,k}^{-1}(1-\alpha)\right]^T \end{array} \right\} \tag{16}$$

The illustration of the proposed disjunctive data-driven uncertainty sets can be observed in Fig. 5. These data-driven disjunctive uncertainty sets are formed based on the weather forecast error between Feb 1st and Feb 21st, with the sample size of 2000.

After the data-driven disjunctive uncertainty sets are established, the optimization problem of the proposed data-driven RMPC can be cast as follows:

$$\begin{aligned} &\min \sum_{i \in B_u} c_i u_i + \lambda^T L \lambda \\ &\text{s.t. } F_u[Mw + h] \leq f_u, \ \forall w \in W^k \\ &\quad F_x[Ax_0 + B_u h + B_v v + (B_w + B_u M)w] \leq f_x + \lambda, \ \forall w \in W^k \\ &\quad u_i = h_i + \sum_{j=0}^{i-1} M_{i,j} w_j, \ \forall w \in W^k \end{aligned} \tag{17}$$

On the one hand, M and h are the decision variables because affine disturbance feedback [69] is used for formulating the linear robust optimization counterpart. λ is the soft constraint proposed in [70]. Both parameters are created to ensure the feasibility of the model. L, on the other hand, is a positive weighted diagonal matrix that heavily penalizes the violation of constraints. c_i is the cost coefficient. The decision variables in this optimization framework are inputs (u) and the amount of violation to the hard constraints (λ). The linear constraints for the optimization problem are formulated in Section 3 of Appendix, and feasibility and stability guarantee is presented in Section 4 of Appendix.

4. Simulation results and discussion

In this section, the physics-based model of the multi-zone building is designed based on building geometry and the study of air dynamics. Afterward, the nonlinear physics-based model is incorporated into SSM by the system identification toolbox in MATLAB for temperature and humidity control. Finally, it is reported on how the different controllers are assessed. The multi-zone building MPC performance of the controllers for one week in Ithaca, New York, is considered.

4.1. Problem statement

The forecasted weather data in Ithaca, New York, is obtained from [71], and the historical measured data is acquired from [72]. The uncertainty set can be constructed from weather forecast error and regarded

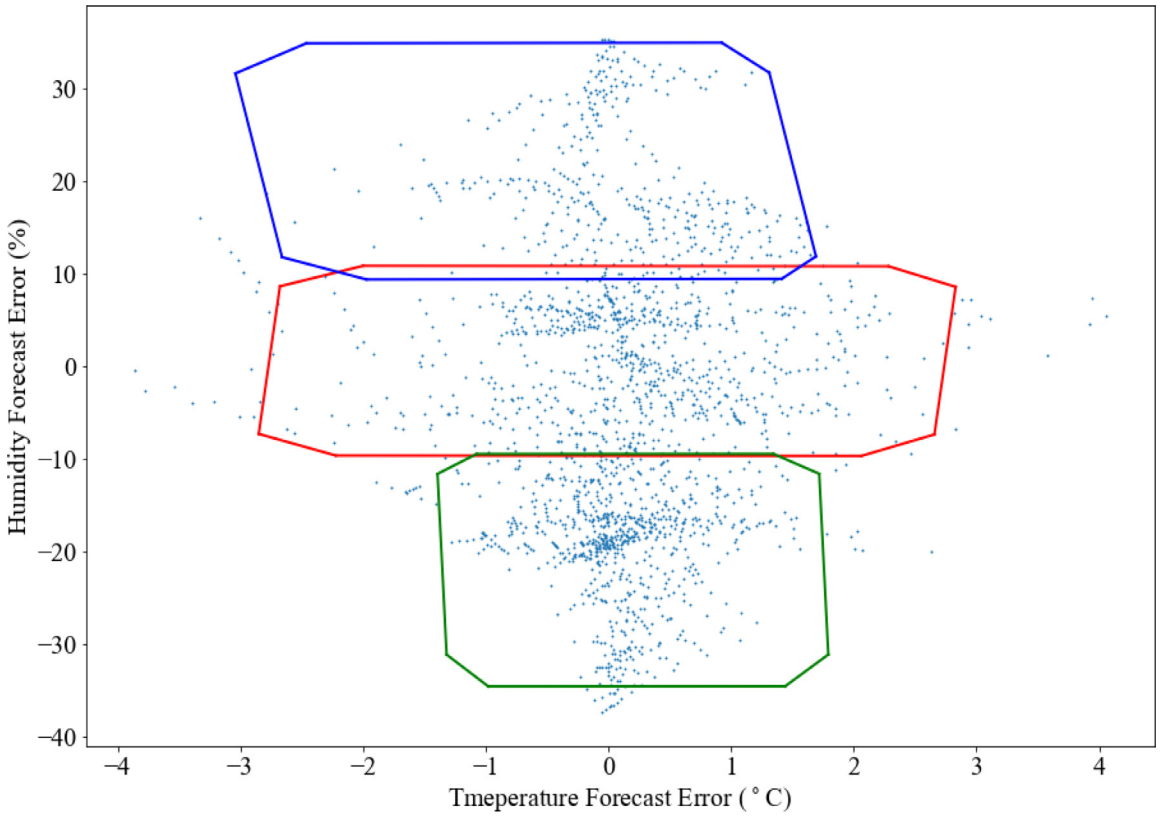

Fig. 5. Uncertainty sets construction based on the temperature forecast error and humidity forecast error at Ithaca, New York USA, starting from Feb 1st, 2016, with a sample size of 2000. The uncertainty set is clustered into three groups (enclosed by green, blue, red lines), within each basic uncertainty set, $\alpha = 0.008$ and $\Gamma = 1.7$.

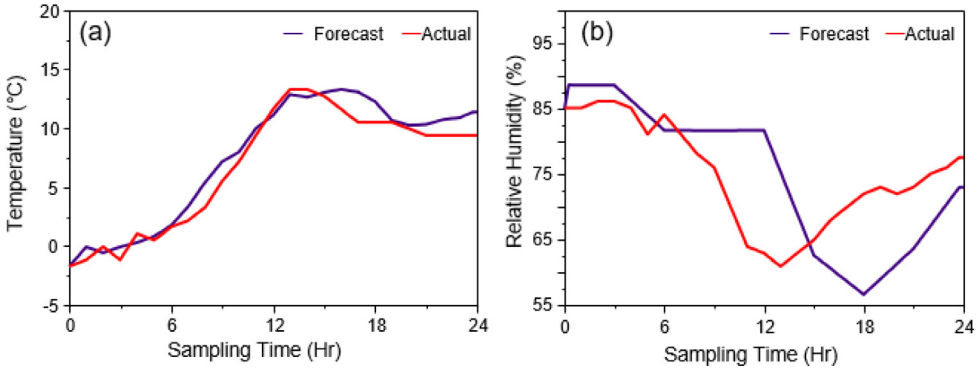

Fig. 6. (a) The temperature forecast on Nov 1st, 2016, at Ithaca, New York, USA. (b) The humidity forecast on Nov 1st, 2016, at Ithaca, New York, USA. The actual weather measurements are shown in red, and the forecast weather data is shown in blue.

as identical and independently distributed, following to [35]. Fig. 6 exemplifies the profile of real measurements and 1 h ahead predictions on Nov. 1st, 2016 of Ithaca, NY, USA, and prediction errors can be observed, especially for relative humidity which could be up to 15%. Such large forecast uncertainties could significantly affect the MPC's performance.

To tackle the issue arising from the complex disjoint weather forecast uncertainty data structure, we group the uncertainty data with k-means clustering algorithm and enclose the high-density region in each basic uncertainty set using PCA coupled with KDE. We apply the proposed framework to this multi-zone building, which considers both zone temperature and relative humidity.

Based on [35,44], the constraints in the MPC are:

- The zone temperature should be within 21 °C to 24 °C,
- The relative humidity should be between 0% and 65%.
- The maximum power consumption rate is capped at 45 W/m^2.

The same parameters and initial conditions are assigned to make a fair comparison. The initial condition for the simulation is that all temperature parameters should be 21 °C, and relative humidity values should be set to 40%. The GUROBI is used to solve the MPC optimization problems, and all computations are performed on XPS 17 9710 equipped with Intel Core i7–11,800 CPU @ 2.30 GHz and 16 G.B. of RAM.

4.2. Control results

The summary of the thermal comfort performances of five controllers for the entire first week of November is shown in Table 1. The overall heating power consumption considers the sum of fan power and heating power installed in AHU as specified in BRCM [21].

We report control performances of simulations from Nov 1st, 2016, to Nov 8th, 2016, under five different control strategies, namely CEMPC, RMPC, KM-RMPC, PKRMPC, and our proposed framework (KM-PKRMPC). As introduced before, CEMPC is certainty equivalent MPC; RMPC is RMPC with box uncertainty set; KM-RMPC is k-means clus-

Table 1
Performance of five MPC-based controllers on the multi-zone building control in the first week of November 2016 in Ithaca, NY. CEMPC is certainty equivalent MPC; RMPC is RMPC with box uncertainty set; KM-RMPC is k-means clustered data-driven RMPC with disjunctive box uncertainty sets; PKRMPC is data-driven RMPC using PCA coupled with KDE; KM-PKRMPC is data-driven RMPC using k-means clustering and PCA coupled with KDE. The mean total power consumption rate is the mean of the power consumption of all equipment, including humidifiers, heater, and radiator. The mean humidity control power consumption rate is calculated from the mean of humidifier and dehumidifier power consumption. The mean heating power consumption rate is calculated from the mean of the heater and radiator power consumption.

	CEMPC	RMPC	KM-RMPC	PKRMPC	KM-PKRMPC
Mean temperature violation	29.36%	6.32%	7.90%	6.00%	4.25%
Mean relative humidity violation	0%	0%	0%	4.45%	0%
Total power cost[1]	$710.85	$784.58	$729.77	$751.71	$720.13
Total humidity control cost	$33.14	$95.76	$61.42	$81.40	$38.94
Total heating control cost	$677.71	$688.82	$668.35	$670.31	$681.19
Mean computation time	0.10 s	0.47 s	0.38 s	0.97 s	1.63 s

[1] Cost: According to nyserda.ny.gov, an average of 17.7 ¢/kWh is used in this work. The data source can be retrieved from: www.nyserda.ny.gov/Researchers-and-Policymakers/Energy-Prices/Electricity/Monthly-Avg-Electricity-Residential.

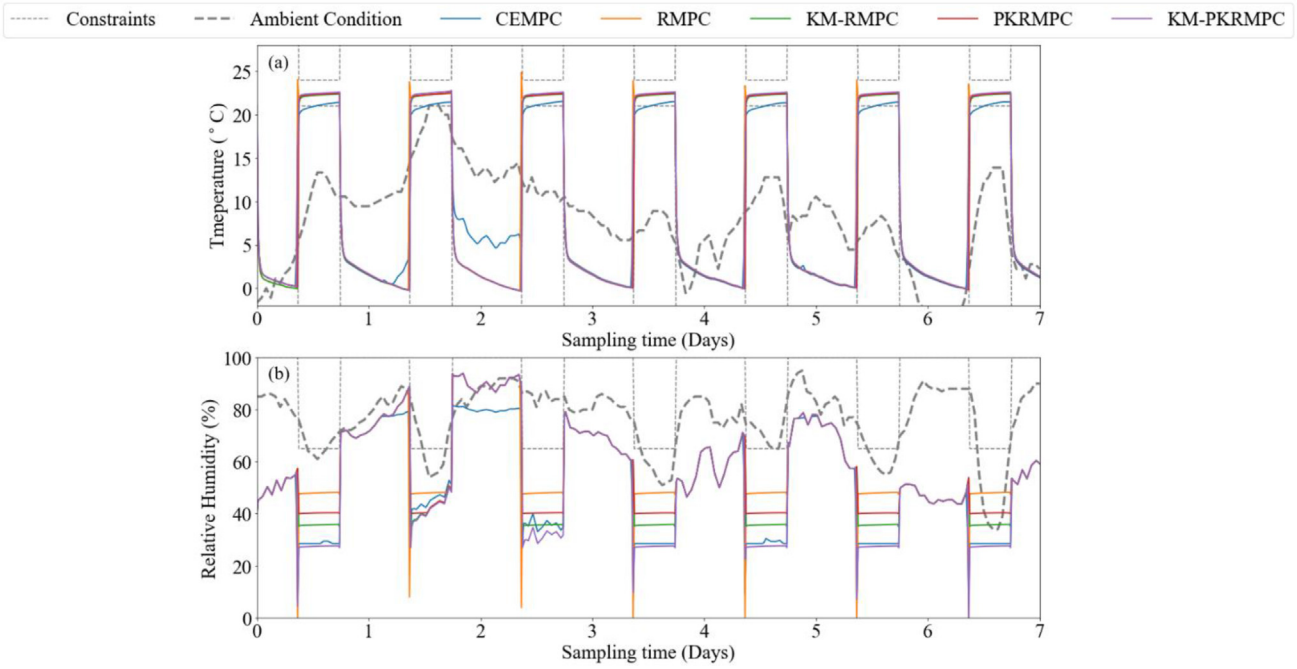

Fig. 7. One of room's temperature profile (shown in (a)) and relative humidity profile (shown in (b)) in the first week of Nov. 2016 in Ithaca, NY, USA. The initial condition for zone temperature is set to 21 °C and relative humidity to 40%. The ambient weather conditions are demonstrated in gray lines. CEMPC is certainty equivalent MPC; RMPC is RMPC with box uncertainty set; KM-RMPC is k-means clustered data-driven RMPC with disjunctive box uncertainty sets; PKRMPC is data-driven RMPC using PCA coupled with KDE; KM-PKRMPC is data-driven RMPC using k-means clustering and PCA coupled with KDE. The comforts bounds for temperature and humidity are enforced from 9:00 AM to 5:00 PM.

tered data-driven RMPC with disjunctive box uncertainty sets; PKRMPC is data-driven RMPC using PCA coupled with KDE; PCA coupled with KDE based data-driven RMPC; KM-PKRMPC is the proposed approach, namely data-driven RMPC using k-means clustering and PCA coupled with KDE. The winter season in 2016 is selected based on the consideration of the severe weather condition in New York State during that period [73]. The scarcity of water and other natural resources led to a great concern for the usage of natural resources. Therefore, it is strongly desired to save power consumption without compromising on thermal comfort, which can effectively contribute to saving limited resources.

Based on the results shown in Figs. 7 and 8 CEMPC violates the constraints more severely, as shown in Fig. 9. CEMPC, which only considers the deterministic conditions, fails to satisfy constraints on temperature and relative humidity. Another observation is the control benefits from disjunctive uncertainty sets. For temperature control, KM-PKRMPC manages to lower the violation frequency and reduce energy usage compared to PKRMPC. Similarly, for relative humidity control, KM-PKRMPC uses less energy to maintain relative humidity within a comfortable range than PKRMPC. KM-RMPC also shows an advantage over RMPC by lowering 7.51% of the power consumption, and KM-PKRMPC can lower 4.38% of the power consumption compared to PKRMPC. On the other hand, the humidity control profiles are different among the control frameworks, as shown in Fig. 7(b). Both KM-RMPC, PKRMPC, and RMPC have steady control curves when the constraints are enforced. Therefore, they have a higher humidity energy consumption compared to KM-PKRMPC and CEMPC for stabilizing the humidity states within each thermal zone.

Therefore, it is safe to acknowledge that disjunctive uncertainty sets can help improve the controller's ability to hedge against uncertainty disturbances.

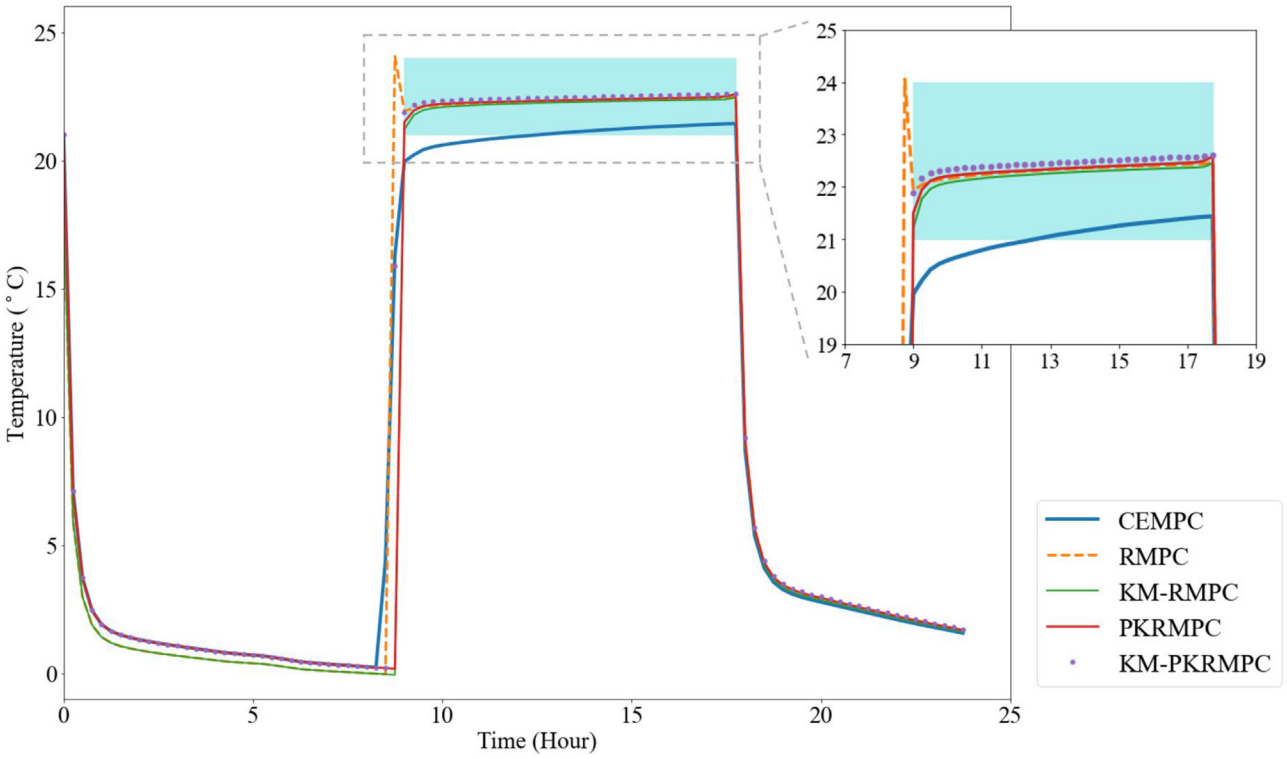

Fig. 8. The temperature profile on Nov. 1st,2016. The temperature constraints are marked in light blue. CEMPC is certainty equivalent MPC; RMPC is RMPC with box uncertainty set; KM-RMPC is k-means clustered data-driven RMPC with disjunctive box uncertainty sets; PKRMPC is data-driven RMPC using PCA coupled with KDE; KM-PKRMPC is data-driven RMPC using k-means clustering and PCA coupled with KDE. It can be observed that CEMPC have a large violation in the morning.

The performance of our proposed control framework was additionally tested with the forecast errors, which are outside of the disjunctive data-driven uncertainty sets. 153 control intervals are found to contain the forecast errors which are outside the disjunctive data-driven uncertainty sets. Our proposed control framework tested these time intervals. No violation cases are found for both temperature or relative humidity control. The control profiles and statistic summary of the KM-PKRMPC are summarized in Appendix A.5. These results suggest our framework's strong performance in maintaining environmental conditions even in the worst scenario. The control profiles are shown in Fig. A2 and Fig. A3 of the Appendix.

On the other hand, the computational times of solving the control optimization problem in various MPC methods are presented in Table 1. Though compared with other MPC approaches, the mean computation time for our proposed framework is 1.63 CPUs per optimization problem, it suffices to complete the optimization procedure within the sampling interval of 15 min.

From these comparisons, the advantages of the proposed framework can be summarized as follows. By clustering the uncertainty data, the k-means algorithm can tackle the problem arising from the complexity of the disjoint-set data structure. Moreover, by extracting useful information from each set of the disjunctive uncertainty sets, a desirable balance is created between utilizing the weather for adjusting the indoor environment and hedging against the uncertainty in the forecast error. In Table 1, although CEMPC achieves lower energy consumption, our proposed framework eventually pays off because a satisfactory performance is achieved in substantially reducing the 25.11% violation to the zone temperature and relative humidity level constraints by only using 1.31% more energy. In Fig 7, CEMPC is observed to violate when constraints are enforced. In comparison with KM-RMPC and PKRMPC control strategies, improved control performances can also be obtained by the proposed framework. Specifically, the proposed framework desirably lowered 4.38% to 1.33% of control cost while avoiding the constraint violation.

In this sense, by leveraging the power of uncertainty data, our proposed framework can achieve a desirable tradeoff between using forecast information for saving energy and hedging against uncertainty in forecast error. Finally, though being slower than other MPC approaches, the proposed framework's computation efficiency still suffices to be applied in practice.

5. Conclusion

In this paper, we developed a novel data-driven RMPC framework for effectively controlling multi-zone buildings. The nonlinear physics-based model is formed with the help of BRCM and the study of air dynamics. We made use of system identification to linearize the physics-based model to retrieve the SSM for the multi-zone building to predict and control the indoor thermal comfort parameters, including temperature and relative humidity. This SSM was incorporated into the data-driven RMPC framework, which optimizes the trajectory of temperature and relative humidity in each zone with minimal energy usage. The disjunctive uncertainty set was formed based on the k-means clustering algorithm to tackle the complex disjoint-set structure of uncertainty data from weather forecast errors. We used PCA and KDE in each set to capture the support of disjunctive uncertainty distribution more effectively. Optimal control decisions balancing between temperature and relative humidity constraints could then be determined by solving the data-driven RMPC optimization problem at each time step. The feasibility and stability issues were also formally addressed. The simulation results of a case study of a multi-zone building in Ithaca, New York, USA showed that the proposed data-driven RMPC framework could reliably maintain the zone temperature and relative humidity within the thermally comfortable level, which CEMPC and RMPC failed to follow. Additionally, from the power saving perspective, the proposed framework could reduce the energy consumption by 1.33% compared to the KM-RMPC approach and by 4.38% than the PKRMPC approach.

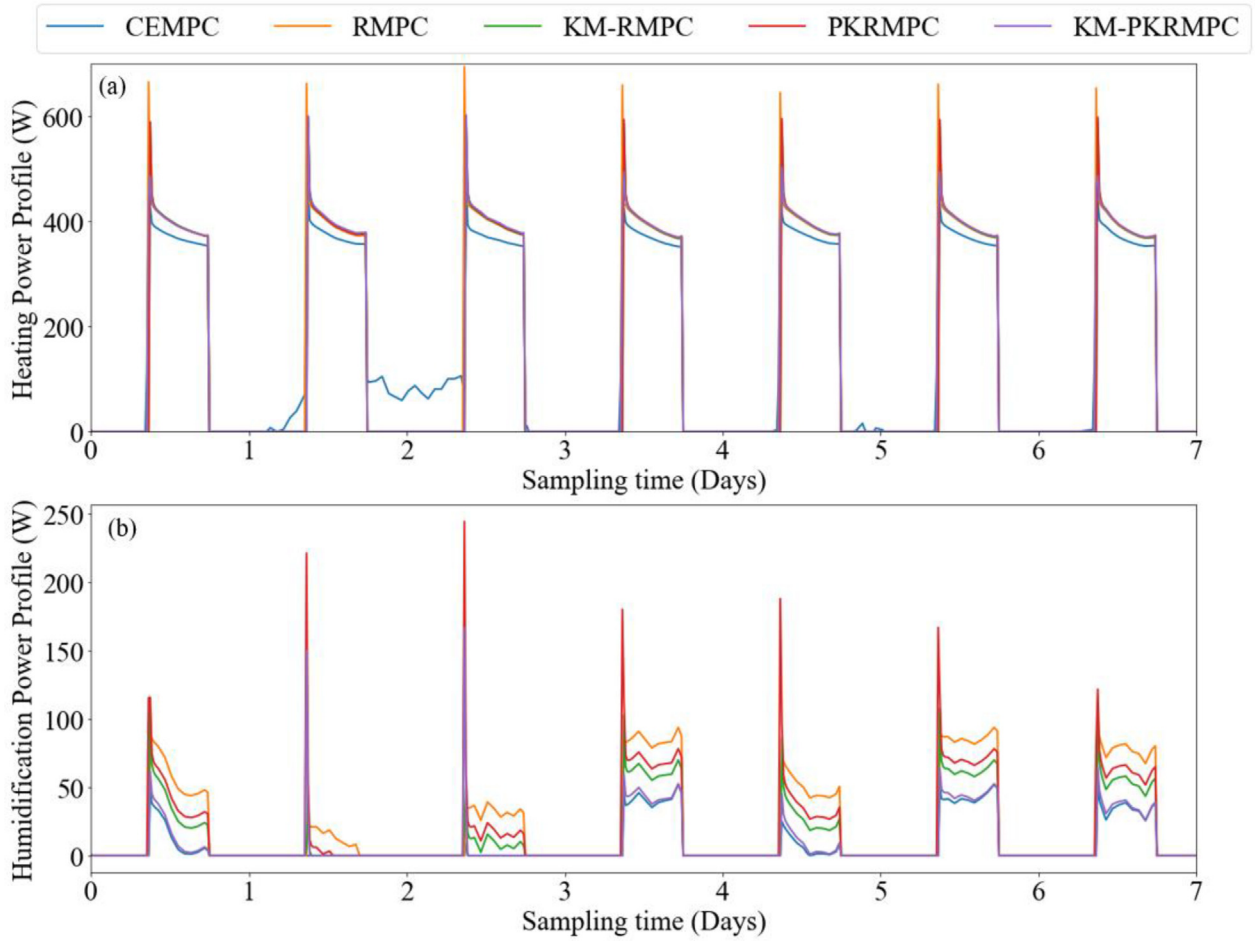

Fig. 9. The input trajectory of each controller's control efforts. Heating control is the sum of the power consumption from the heater and radiator (shown in (a)), and humidity control is the mean of the humidifier consumption (shown in (b)). CEMPC is certainty equivalent MPC; RMPC is RMPC with box uncertainty set; KM-RMPC is k-means clustered data-driven RMPC with disjunctive box uncertainty sets; PKRMPC is data-driven RMPC using PCA coupled with KDE; KM-PKRMPC is data-driven RMPC using k-means clustering and PCA coupled with KDE. The mean heating power consumption rate is calculated from the mean of the heater and radiator power consumption. The mean humidity control power consumption rate is calculated from the mean of humidifier and dehumidifier power consumption.

In future work, more efforts could be made to develop a more comprehensive building model, which can consider additional sources of internal heat and moisture gains, such as occupancy, lighting, etc. Another direction can be to investigate cooling capability in the multi-zone building and add this into the control problem. The key issue arises from the complicated mechanistic models of the building control, where the control problem is computationally expensive and hence poses challenges in real-time optimizations.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix

This Appendix includes four sections. Section A.1 presents the linear function of relative humidity and temperature. Section A.2 demonstrates the statistic results of system identification. Section A.3 shows the formulation of linear constraints within the optimization problems of the control framework. The details of stability and feasibility guarantees are provided in Section A.4.

A.1. Linear function over saturated water vapor and temperature

Fig. A1 shows the linear regression and its result. The blue dots are the results of the nonlinear equation between the saturated water vapor density and room temperature, and the dashed line is the linear equation. R^2 is 0.9835, indicating that 98.35% of the variation of the original nonlinear equation results has been explained. Therefore, this linear regression shows a good agreement with the original nonlinear equations and can be adopted in this work

A.2. Statistic results comparison to validate the performance of system identification to the EnergyPlus-based model simulation

Table A1 RMSE of system identification of training sample data. Each "room" represents a "zone". The RMSEs of linearized SSM for predicting relative humidity are between 3.36% and 5.59%, and those for temperature are between 0.62% and 0.76%.

As shown in Table S2, when linear SSM is used to predict the zone temperature and relative humidity with the data which are beyond the range of the sample data, the RMSEs of indoor zone temperature are between 0.20 °C and 0.25 °C, and for relative humidity is between 4.89%

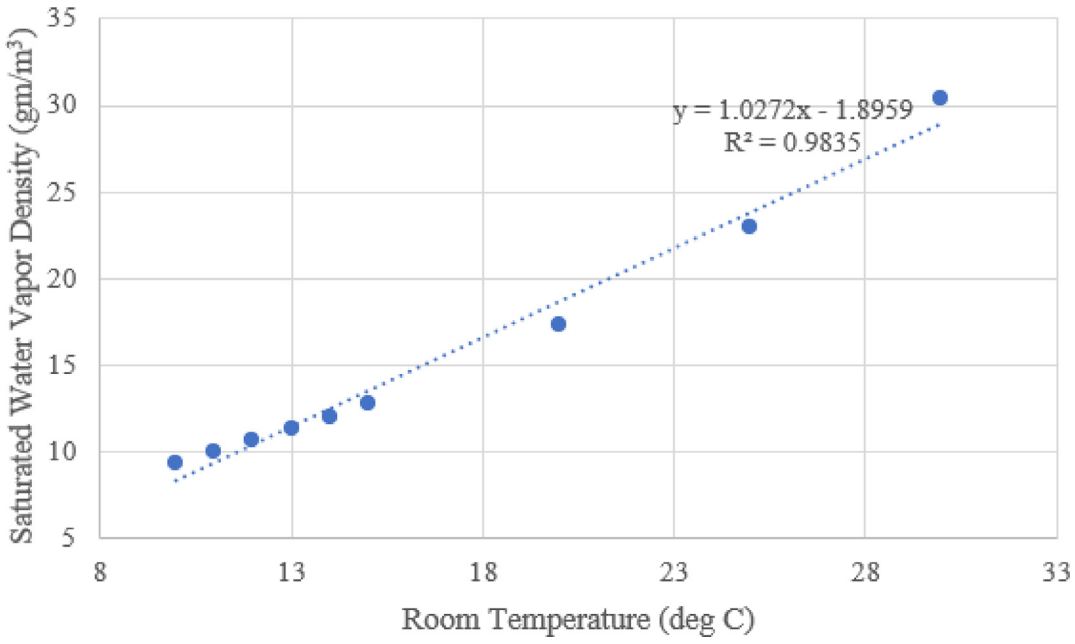

Fig. A1. Linear regression of saturated water vapor density equation with zone temperature. The blue dots are the nonlinear equation retrieved from [41]. The R2 value of the linear regression model is around 0.9835, meaning that this linear equation explains 98.35% of the variation of the original nonlinear data.

Table A1
RMSE of system identification of training sample data. Each "room" represents a "zone". The RMSEs 23 of linearized SSM for predicting relative humidity are between 3.36% and 5.59%, and those for temperature are between 0.62% and 0.76%.

	Relative Humidity	Temperature
Room 1	5.40%	0.21 °C
Room 2	4.79%	0.19 °C
Room 3	5.42%	0.21 °C
Room 4	5.30%	0.21 °C
Room 5	5.05%	0.22 °C

Table A2
RMSE of system identification of testing data. The strong performance of the SSM in accurately predicting the physical process proves its reliability for building control in practice.

	Relative Humidity	Temperature
Room 1	5.33%	0.23 °C
Room 2	5.88%	0.20 °C
Room 3	5.38%	0.23 °C
Room 4	5.17%	0.24 °C
Room 5	4.89%	0.25 °C

Table A1 and Table A2 demonstrate the root mean squared error (RMSE) between the prediction from linear SSM and simulated results from EnergyPlus model with training data and testing data. For system identification, the EnergyPlus-based model simulation"s power profile of heater, radiator and humidifier are extracted. Those energy profiles are and used in simulating the room temperature and humidity dynamics and system identification.

and 5.88%. The error values are relatively close to those in [19], indicating a good agreement of the linearized SSM to the EnergyPlus model.

A.3. Linear constraints formulation

After adopting the ADF [74], we use $u = \pi(w) = h + Mw$ to avoid redundancy in expression. The main benefit of using ADF is that the resulting optimization problem can be formulated as a convex problem, which can be solved by the off-the-shelf convex optimization solvers efficiently [75]. leading to the following optimization problem:

$$\begin{aligned} &\min_{\pi(w),\lambda} \left\{ \max_{w\in W^k} \left\{ c^T \pi(w) \right\} + \lambda^T L\lambda \right\} \\ &s.t. F_u \pi(w) \le f_u, \forall w \in W^k \\ &Fx\left(Ax_0 + B_u\pi(w) + B_v v + B_w W\right) \le f_x + \lambda, \forall w \in W^k \end{aligned} \tag{S18}$$

by introducing a variable t to the objective function. We can further convert (S18) into the following problem to form our robust constraints:

$$\begin{aligned} &\min_{t,\boldsymbol{h},\boldsymbol{M},\lambda} t + \lambda^T L\lambda \\ &\text{s.t.} \max_{z\in\mathbb{W}^k} \boldsymbol{a_w w} \le \boldsymbol{b_w} \end{aligned} \tag{S19}$$

where

$$\begin{aligned} a_z &= \begin{bmatrix} c^T M \\ F_u M \\ F_x(B_u M + B_w) \end{bmatrix}, \\ b_z &= \begin{bmatrix} t - c^T h \\ f_u - F_u h \\ f_x - F_x\left(Ax_0 + B_u h + B_v v\right) + \lambda \end{bmatrix} \end{aligned} \tag{S20}$$

To incorporate our uncertainty set, we use $z = [z^+, z^-]^T$ to obtain the following optimization problem,

$$\begin{aligned} &\min_{t,\boldsymbol{h},\boldsymbol{M}} t + \lambda^T L\lambda \\ &\text{s.t.} \max_{z\in\mathbb{W}^k} \boldsymbol{a}\cdot\boldsymbol{z} \le \boldsymbol{b} \end{aligned} \tag{S21}$$

with

$$\boldsymbol{a} = \begin{bmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_k \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_1 \\ \vdots \\ \boldsymbol{b}_k \end{bmatrix} \tag{S22}$$

where,

$$a_i = \begin{bmatrix} c^T M \\ F_u \cdot M \\ F_x \cdot \left(B_u M + B_w\right) \end{bmatrix} \cdot \begin{bmatrix} Q\left(1 \cdot \bar{\xi}_i^T\right) & Q\left(1 \cdot \underline{\xi}_i^T\right) \end{bmatrix},$$
$$b_i = \begin{bmatrix} t - c^T\left(h + M\hat{\mu}\right) \\ f_u - F_u\left(h + M\hat{\mu}\right) \\ f_x - F_x\left(Ax_0 + B_u h + \left(B_u M + B_w\right)\cdot\hat{\mu} + B_v\right) + \lambda \end{bmatrix} \tag{S23}$$

Thanks to the uncertainty set's polyhedral structure, the maximization problem on the left-hand side of the robust constraints in (S21) can be equivalently dualized in the following compact form:

$$\begin{aligned} &\min_{t,h,M,Z,\lambda} t + \lambda^T L \lambda \\ &\text{s.t.} \boldsymbol{Z}^T \boldsymbol{r} \le \boldsymbol{b} \\ &\boldsymbol{Z}^T \boldsymbol{R} = \boldsymbol{a} \\ &\boldsymbol{Z} \ge \boldsymbol{0} \end{aligned} \tag{S24}$$

where $\boldsymbol{Z}$ stands for the matrix consisting of all dual variables. We note that the optimization problem (S19) with constraints (S20), (S22), and (S23) is always feasible, as the polyhedral uncertainty set $\mathbb{W}^k$ is trivially non-empty by our construction. Hence, the strong duality theorem ensures there is no duality gap, namely $\max_{w\in\mathbb{W}^k} \boldsymbol{a}\cdot w = \min_{\boldsymbol{Z}^T\boldsymbol{R}=\boldsymbol{a},\boldsymbol{Z}\ge\boldsymbol{0}} \boldsymbol{Z}^T\boldsymbol{r}$. At this point, this MPC problem is computationally efficient since the constraints are both linear and convex. The data-driven RMPC framework is completed. The feasibility and stability of the framework can also be ensured, and the details of which are discussed in the next subsection.

A.4. Feasibility and stability

Feasibility and stability are crucial to the MPC under the stochastic condition. In case of unbounded disturbance from the forecast error, system states may be driven toward the unstable states, resulting in infeasible optimization problems. To tackle this problem, soft constraints [70] are introduced in (S24) to ensure the feasibility of the control optimization problem. With the introduction of soft constraints and slack variables in the convex constraints, the optimization problem can always remain feasible [76,77].

Besides the feasibility issue, the stability of the stochastic system is determined mainly from the property of matrix A. If A is Schur-stable, then all eigenvalues are within the unit circle; it can be concluded that the mean square boundedness $\sup_{t\in\mathbb{H}_0}\mathbb{E}\{\|x_t\|^2\} < \infty$can be supported under the condition in which both inputs and covariance $\mathbb{E}\{w_t w_t^T\}$of disturbances are bounded [78]. The second condition is that if A is not Schur-stable, equivalently at least one of the eigenvalues out of the unit circle, the mean square boundedness cannot be reached. The only case that needs to be tackled is when A is Lyapunov stable. In this case, it suffices to consider when A is orthogonal [79]. In order to better deal with this case, we define $X_{.F.}$ as a set of initially feasible states of the MPC problem. Some standing assumptions are then made as follows.

Assumption 1. The stochastic process $\{w_t\}_{t\in\mathbb{H}_0}$ is bounded, that

$$\sup_{t\in\mathbb{H}_0}\mathbb{E}\left\{\|w_t\|^4\right\} = C < \infty \tag{S25}$$

Assumption 2. For the optimal control strategy $\hat{u}$, there is a constant$c > \|B_w\|_2 \cdot \sup_{t\in\mathbb{H}_0}\mathbb{E}\{\|w_t\|^4\}$ which is

$$\left\|Ax_t + B_u\hat{u}\right\| - \|x_t\| \le -c, \ \forall x_t \notin X_F \tag{S26}$$

where X_F is the set of feasible solutions.

Assumption 1 requires the fourth moment of the disturbance to be a finite value, which is not restrictive. Assumption 2 asks the controller to control the model to the origin when no feasible solutions are found. This assumption can be understood as overcoming the effect from additive disturbances through exceeding the largest expectation of $\|B_w\|_2 \cdot \sup_{t\in\mathbb{H}_0}\mathbb{E}\{\|w_t\|^4\}$. Assumption 2 is not required to be met throughout the control sequence, and it must be satisfied when $\|x_t\|$ is too large and the problem without soft constraint is infeasible.

Lemma 1. Let X_H be a sequence of nonnegative random variables and suppose that there exist constants $b > 0$and $J, M < \infty$, such that $X_0 \le J$, and for all H,

$$\mathbb{E}\left(X_H - X_{H-1}|F_H\right) \le b \text{ on the event that} \{X_H > J\} \tag{S27}$$

and

$$\mathbb{E}\left(\left|X_{H+1} - X_H\right|^4 | X_0, \ldots, X_H\right) \le M \tag{S28}$$

Then there exists a $\zeta = \zeta(b, V, M) > 0$such that $\sup_{t\in H}\mathbb{E}\{X_H\} < \zeta$.

Details of the proof can be found in [80] and omitted here, and now the stability guarantee can be readily established.

Theorem 1. (Mean-Square Boundedness) Suppose that admissible control input is bounded $\sup_{t\in F_u}||u|| < b_u < \infty$and A is an orthogonal matrix. If Assumptions 1 and 2 are met, then for all initial states, there is always a constant $c > 0$ such that the closed-loop system admits mean-square boundedness:

$$\sup_{t\in\mathbb{H}_0}\mathbb{E}\left\{\|x_t\|^2\right\} = \zeta < \infty \tag{S29}$$

Proof. The condition (S27) is to be verified first. This nonnegative stochastic process is defined as $\{X_t = x_t\}_{t\in H}$. It can be set that $V = \sup_{x\in X_F}||x|| < \infty$ since X_F is bounded. In the event$X_t \ge V$, we have:

$$\begin{aligned} &E\left(X_{t+1}X_{t\backslash Ft}\right) \\ &= E\left(Ax_t + B_u\hat{u} + B_w w_t - x_t\backslash F_t\right) \\ &\le E\left(Ax_t + B_u\hat{u} + B_w w_t - x_t\backslash F_t\right) \\ &\le -c + E\left(B_{w2}\cdot\sup_{wt} w_t\right) \end{aligned} \tag{S30}$$

Meanwhile,

$$\begin{aligned} y_{t+1} &= (A^T)^{t+1}x_{t+1} \\ &= (A^T)^{t+1}x_t + (A^T)^{t+1}[B_u\hat{u} + B_w w_t] \\ &= y_t + (A^T)^{t+1}[B_u\hat{u} + B_w w_t] \end{aligned} \tag{S31}$$

from $\|y_t\|^2 = y_t^T y_t = x_t^T A^T A x_t = x_t^T x_t = \|x_t\|^2$, we can conclude$\|y_t\| = \|x_t\|$. Also, according to triangle inequality $-\|y_{t+1} - y_t\| \le \|y_{t+1}\| - \|y_t\| \le \|y_{t+1} - y_t\|$, it holds that $|\|y_{t+1}\| - \|y_t\||^4 \le \|y_{t+1} - y_t\|^4$.

Therefore, we have,

$$\begin{aligned} &\left|X_{t+1} - X_t\right|^4 \\ &= \left|\|y_{t+1}\| - \|y_t\|\right|^4 \\ &\le \left|\|y_{t+1} - y_t\|\right|^4 \\ &\le \left\|(A^T)^{t+1}[B_u\hat{u} + B_w w_t]\right\|^4 \\ &\le \left\|B_u\hat{u} + B_w w_t\right\|^4 \\ &\le \left(\|B_u\|_2 \cdot \|\hat{u}\| + \|B_w\|_2 \cdot \|w_t\|\right)^4 \end{aligned} \tag{S32}$$

Thus, we can obtain the following inequality: $\mathbb{E}(|X_{t+1} - X_t|^4 | X_0, \ldots, X_t) \le \mathbb{E}\{(\|B_u\|_2 \cdot \|\hat{u}\| + \|B_w\|_2 \cdot \|w_t\|)^4\}$.Because the fourth moment $\|w_t\|^4$ is bounded, it can be deduced that there is a constant $M = M(\|B_u\|_2, \hat{u}, \|B_w\|_2, C) > 0$ such that

$$\mathbb{E}\left\{\left(\|B_u\|_2 \cdot \|\hat{u}\| + \|B_w\|_2 \cdot \|w_t\|\right)^4\right\} \le M \tag{S33}$$

which yields the condition (S28).

At this moment, all constants *c, V, M* are defined from the equations above. Therefore, from the Lemma 1, there is a constant $\zeta = \zeta(b, V, M) > 0$ which is

$$\sup_{t\in\mathbb{H}_0}\mathbb{E}\left\{\|x_t\|^2\right\} = \sup_{t\in\mathbb{H}_0}\mathbb{E}\left\{\|X_t\|^2\right\} = \zeta < \infty \tag{S34}$$

This completes the proof.

A.5. Test for control performance out of disjunctive data-driven uncertainty sets

The controller's performance is tested with the forecast errors which are outside of the disjunctive data-driven uncertainty sets. The results are demonstrated in the table and figure below.

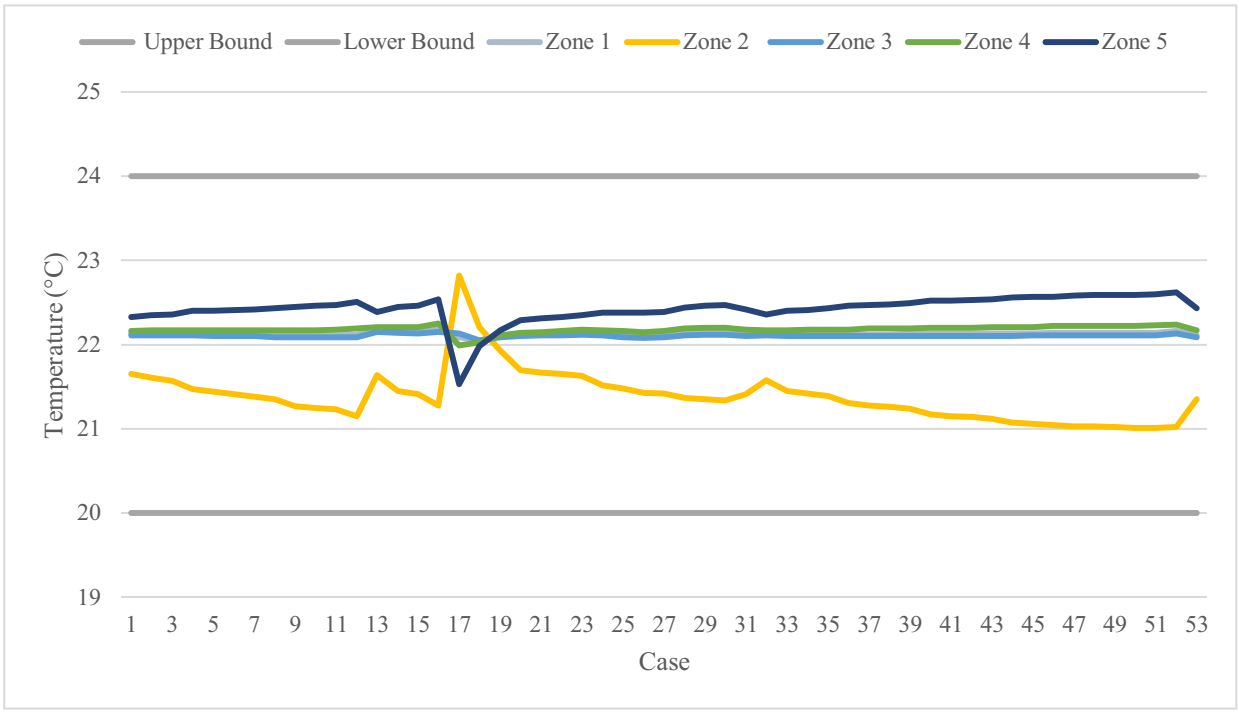

Fig. A2. The temperature control profile of KM-PKRMPC under the forecast errors which are out of the disjunctive data-driven uncertainty sets.

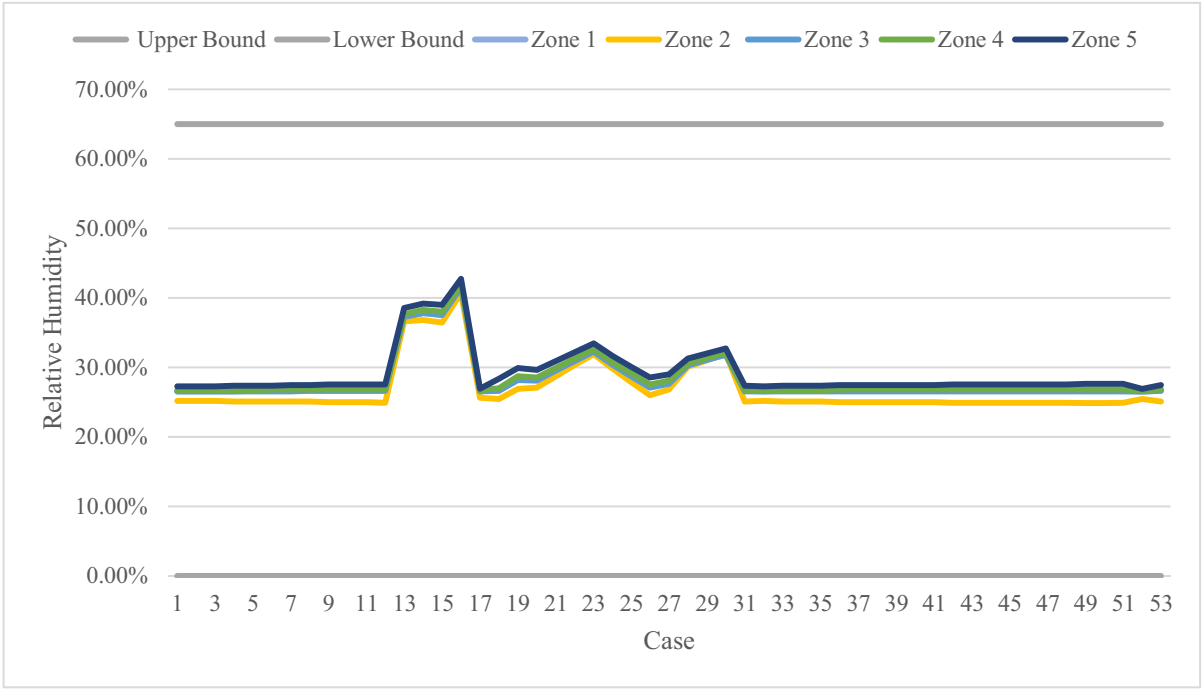

Fig. A3. The temperature control profile of KM-PKRMPC under the forecast errors which are out of the disjunctive data-driven uncertainty sets.

References

[1] Shaikh PH, Nor NBM, Nallagownden P, Elamvazuthi I, Ibrahim T. A review on optimized control systems for building energy and comfort management of smart sustainable buildings. Renew Sustain Energy Rev 2014;34:409–29.
[2] E.I.A. Use of energy explained. Energy use in homes: U.S. energy information administration; 2019.
[3] Qin SJ, Badgwell TA. A survey of industrial model predictive control technology. Control Eng Pract 2003;11:733–64.
[4] Široký J, Oldewurtel F, Cigler J, Prívara S. Experimental analysis of model predictive control for an energy efficient building heating system. Appl Energy 2011;88:3079–87.
[5] Yang S, Oliver Gao H, You F. Model predictive control for Demand- and Market-Responsive building energy management by leveraging active latent heat storage. Appl Energy 2022;327:120054.
[6] Oldewurtel F, Parisio A, Jones CN, Gyalistras D, Gwerder M, Stauch V, et al. Use of model predictive control and weather forecasts for energy efficient building climate control. Energy Build 2012;45:15–27.
[7] Chen WH, You F. Sustainable building climate control with renewable energy sources using nonlinear model predictive control. Renew Sust Energy Rev 2022;168:112830.
[8] Muntwiler S, Wabersich KP, Hewing L, Zeilinger MN. Data-driven distributed stochastic model predictive control with closed-loop chance constraint satisfaction. CoRR 2020.

[9] Yang S, Wan MP, Chen W, Ng BF, Zhai D. An adaptive robust model predictive control for indoor climate optimization and uncertainties handling in buildings. Build Environ 2019;163:106326.
[10] Prívara S, Široký J, Ferkl L, Cigler J. Model predictive control of a building heating system: the first experience. Energy Build 2011;43:564–72.
[11] Maasoumy M, Razmara M, Shahbakhti M, Vincentelli AS. Handling model uncertainty in model predictive control for energy efficient buildings. Energy Build 2014;77:377–92.
[12] Bemporad A, Morari M. Robust model predictive control: a survey. robustness in identification and control. Springer; 1999. p. 207–26.
[13] Shang C, You F. A data-driven robust optimization approach to scenario-based stochastic model predictive control. J Process Control 2019;75:24–39.
[14] Shang C, You F. Data analytics and machine learning for smart process manufacturing: recent advances and perspectives in the big data era. Engineering. 2019;5:1010–16.
[15] Ning C, You F. Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods. Comput Chem Eng 2018;112:190–210.
[16] Fay D, Ringwood JV. On the influence of weather forecast errors in short-term load forecasting models. IEEE Trans Power Syst 2010;25:1751–8.
[17] Fanger P.O. Thermal comfort. Analysis and applications in environmental engineering. Thermal comfort Analysis and applications in environmental engineering. 1970.
[18] Klaučo M, Kvasnica M. Explicit MPC approach to PMV-based thermal comfort control. In: Proceedings of the 53rd IEEE conference on decision and control; 2014. p. 4856–61.
[19] Martínez-Mariño S, Eguía-Oller P, Granada-Álvarez E, Erkoreka-González A. Simulation and validation of indoor temperatures and relative humidity in multi-zone buildings under occupancy conditions using multi-objective calibration. Build Environ 2021;200:107973.
[20] Li X, Wen J. Review of building energy modeling for control and operation. Renew Sustain Energy Rev 2014;37:517–37.
[21] Sturzenegger D, Gyalistras D, Semeraro V, Morari M, Smith RS. BRCM Matlab Toolbox: model generation for model predictive building control. In: Proceedings of the American control conference; 2014. p. 1063–9. 2014.
[22] Sturzenegger D. Model predictive building climate control. ETH-Zürich: Steps towards practice; 2014.
[23] Yang S, Wan MP, Ng BF, Zhang T, Babu S, Zhang Z, et al. A state-space thermal model incorporating humidity and thermal comfort for model predictive control in buildings. Energy Build 2018;170:25–39.
[24] Kim D, Braun JE. A general approach for generating reduced-order models for large multi-zone buildings. J Build Perform Simul 2015;8:435–48.
[25] Chen Q, Li N. Fast simulation and high-fidelity reduced-order model of the multi-zone radiant floor system for efficient application to model predictive control. Energy Build 2021;248:111210.
[26] Lee D, Ooka R, Ikeda S, Choi W, Kwak Y. Model predictive control of building energy systems with thermal energy storage in response to occupancy variations and time-variant electricity prices. Energy Build 2020;225:110291.
[27] Kämpf JH, Wetter M, Robinson D. A comparison of global optimization algorithms with standard benchmark functions and real-world applications using EnergyPlus. J Build Perform Simul 2010;3:103–20.
[28] Yang S, Oliver Gao H, You F. Model predictive control in phase-change-material-wallboard-enhanced building energy management considering electricity price dynamics. Appl Energy 2022;326:120023.
[29] Fang X, Gong G, Li G, Chun L, Peng P, Li W, et al. Deep reinforcement learning optimal control strategy for temperature setpoint real-time reset in multi-zone building HVAC system. Appl Therm Eng 2022;212:118552.
[30] Yao Y, Yang K, Huang M, Wang L. A state-space model for dynamic response of indoor air temperature and humidity. Build Environ 2013;64:26–37.
[31] Sturzenegger D, Gyalistras D, Morari M, Smith RS. Model predictive climate control of a swiss office building: implementation, results, and cost–benefit analysis. IEEE Trans Control Syst Technol 2016;24:1–12.
[32] Darivianakis G, Georghiou A, Smith RS, Lygeros J. The power of diversity: data–driven robust predictive control for energy-efficient buildings and districts. IEEE Trans Control Syst Technol 2019;27:132–45.
[33] Finck C, Li R, Kramer R, Zeiler W. Quantifying demand flexibility of power-to-heat and thermal energy storage in the control of building heating systems. Appl Energy 2018;209:409–25.
[34] Mantovani G, Ferrarini L. Temperature control of a commercial building with model predictive control techniques. IEEE Trans Ind Electron 2015;62:2651–60.
[35] Oldewurtel F, Jones CN, Parisio A, Morari M. Stochastic model predictive control for building climate control. IEEE Trans Control Syst Technol 2014;22:1198–1205.
[36] Arundel AV, Sterling EM, Biggin JH, Sterling TD. Indirect health effects of relative humidity in indoor environments. Environ Health Perspect 1986;65:351–61.
[37] Baughman AV, Arens EA. Indoor humidity and human health - Part I: literature review of health effects of humidity-influenced indoor pollutants. ASHRAE Trans 1996;102:193–211.
[38] Rentel-Gomez C, Velez-Reyes M. Decoupled control of temperature and relative humidity using a variable-air-volume HVAC system and non-interacting control. In: Proceedings of the 2001 IEEE international conference on control applications; 2001. p. 1147–51. (CCA'01) (Cat No01CH37204).
[39] Vereecken E, Roels S, Janssen H. *In situ* determination of the moisture buffer potential of room enclosures. J Build Phys 2010;34:223–46.
[40] Ma Y, Matuško J, Borrelli F. Stochastic model predictive control for building HVAC systems: complexity and conservatism. IEEE Trans Control Syst Technol 2015;23:101–16.
[41] Nave C.R. Hyperphysics, Georgia State University. Canada. 2012.
[42] Kim B, Ahn H. Consensus-based coordination and control for building automation systems. IEEE Trans Control Syst Technol 2015;23:364–71.
[43] Liao Z, Dexter AL. An inferential model-based predictive control scheme for optimizing the operation of boilers in building space-heating systems. IEEE Trans Control Syst Technol 2010;18:1092–102.
[44] Standard 62-2001, Ventilation for acceptable indoor air quality. American Society of Heating, Refrigerating and Air Conditioning Engineers. 2001.
[45] He L, Lei B, Bi H, Yu T. Simplified building thermal model used for optimal control of radiant cooling system. Math Probl Eng 2016;2016:2976731.
[46] Goyal S, Barooah P. A method for model-reduction of non-linear thermal dynamics of multi-zone buildings. Energy Build 2012;47:332–40.
[47] Zabalza Bribián I, Aranda Usón A, Scarpellini S. Life cycle assessment in buildings: state-of-the-art and simplified LCA methodology as a complement for building certification. Build Environ 2009;44:2510–20.
[48] Ortiz O, Bonnet C, Bruno JC, Castells F. Sustainability based on LCM of residential dwellings: a case study in Catalonia, Spain. Build Environ 2009;44:584–594.
[49] Hu J, Karava P. A state-space modeling approach and multi-level optimization algorithm for predictive control of multi-zone buildings with mixed-mode cooling. Build Environ 2014;80:259–73.
[50] Lorite IJ, Ramírez-Cuesta JM, Cruz-Blanco M, Santos C. Using weather forecast data for irrigation scheduling under semi-arid conditions. Irrig Sci 2015;33:411–427.
[51] Dehkordi V, Candanedo J. State-space modeling of thermal spaces in a multi-zone building. In: Proceedings of the 4th high performance buildings conference; 2016.
[52] Cole WJ, Powell KM, Hale ET, Edgar TF. Reduced-order residential home modeling for model predictive control. Energy Build 2014;74:69–77.
[53] Ljung L. System Identification. Wiley Encyclopedia of Electrical and Electronics Engineering. 2017:1–19.
[54] Ma X, Fang C, Ji J. Prediction of outdoor air temperature and humidity using Xgboost. IOP Conf Ser Earth Environ Sci 2020;427:012013.
[55] Chen W-H, You F. Semiclosed greenhouse climate control under uncertainty via machine learning and data-driven robust model predictive control. IEEE Trans Control Syst Technol 2022;30:1186–97.
[56] Chen W-H, Shang C, Zhu S, Haldeman K, Santiago M, Stroock AD, et al. Data-driven robust model predictive control framework for stem water potential regulation and irrigation in water management. Control Eng Pract 2021;113:104841.
[57] Chen WH, You F. Smart greenhouse control under harsh climate conditions based on data-driven robust model predictive control with principal component analysis and kernel density estimation. J Process Control 2021;107:103–13.
[58] Shang C, Chen W-H, Stroock AD, You F. Robust model predictive control of irrigation systems with active uncertainty learning and data analytics. IEEE Trans Control Syst Technol 2020;28:1493–504.
[59] Zhao N, You F. Sustainable power systems operations under renewable energy induced disjunctive uncertainties via machine learning-based robust optimization. Renew Sustain Energy Rev 2022;161:112428.
[60] Mason JC, Price RK, Tem'Me A. A neural network model of rainfall-runoff using radial basis functions. J Hydraul Res 1996;34:537–48.
[61] Lin GF, Chen LH. A non-linear rainfall-runoff model using radial basis function network. J Hydrol 2004;289:1–8.
[62] Bottou L., Bengio Y. Convergence properties of the k-means algorithms. Advances in neural information processing systems 1995. p. 585–92.
[63] Hartigan JA, Wong MA. A K-means clustering algorithm. J R Stat Soc Ser C 1979;28:100–8 (Applied Statistics).
[64] Ning C, You F. Data-driven adaptive nested robust optimization: general modeling framework and efficient computational algorithm for decision making under uncertainty. AlChE J 2017;63:3790–817.
[65] Wold S, Esbensen K, Geladi P. Principal component analysis. Chemom Intell Lab Syst 1987;2:37–52.
[66] Murphy KP. Machine learning: a probabilistic perspective. MIT press; 2012.
[67] Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans R Soc A Math Phys Eng Sci 2016;374:20150202.
[68] Zhang X, King ML, Hyndman RJ. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. Comput Stat Data Anal 2006;50:3009–3031.
[69] Oldewurtel F, Jones CN, Morari M. A tractable approximation of chance constrained stochastic MPC based on affine disturbance feedback. In: Proceedings of the 47th IEEE conference on decision and control; 2008. p. 4731–6.
[70] Meseguer P, Rossi F, Schiex T, Rossi F, van Beek P, Walsh T. Chapter 9 - Soft Constraints. In: Foundations of artificial intelligence. Elsevier; 2006. p. 281–328.
[71] Herzmann D., Arritt R., Todey D. Iowa environmental mesonet. Available at mesonet agron iastate edu/request/coop/fe phtml (verified 27 sept 2005) Iowa State Univ, Dep of Agron, Ames, IA. 2004.
[72] Diamond HJ, Karl TR, Palecki MA, Baker CB, Bell JE, Leeper RD, et al. U.S. climate reference network after one decade of operations: status and assessment. Bull Am Meteorol Soc 2013;94:485–98.
[73] Sweet SK, Wolfe DW, DeGaetano A, Benner R. Anatomy of the 2016 drought in the Northeastern United States: implications for agriculture and water resources in humid climates. Agric For Meteorol 2017;247:571–81.
[74] Mayne DQ, Seron MM, Raković SV. Robust model predictive control of constrained linear systems with bounded disturbances. Automatica 2005;41:219–24.
[75] Goulart PJ, Kerrigan EC, Maciejowski JM. Optimization over state feedback policies for robust control with constraints. Automatica 2006;42:523–33.
[76] Calafiore GC, Fagiano L. Robust model predictive control via scenario optimization. IEEE Trans Automat Contr 2013;58:219–24.

[77] Lu S, Lee JH. You F. Soft-constrained model predictive control based on data-driven distributionally robust optimization. AlChE J 2020;66:e16546.
[78] Paulson JA, Buehler EA, Braatz RD, Mesbah A. Stochastic model predictive control with joint chance constraints. Int J Control 2020;93:126–139.
[79] Chatterjee D, Hokayem P, Lygeros J. Stochastic receding horizon control with bounded control inputs: a vector space approach. IEEE Trans Automat Control 2011;56:2704–10.
[80] Pemantle R, Rosenthal JS. Moment conditions for a sequence with negative drift to be uniformly bounded in Lr. Stoch Process Appl 1999;82:143–55.