Data-Driven Building Energy Modeling with Feature Selection and Active Learning for Data Predictive Control

Liang Zhang*
National Renewable Energy Laboratory, Golden, Colorado, U.S.
* Corresponding author: Liang.Zhang@nrel.gov

Highlights

- A framework of building energy modeling for data predictive control is developed
- The framework integrates active learning strategy with feature selection technique
- The framework is demonstrated in a virtual testbed (EnergyPlus and BCVTB)
- The case study shows improvement in energy modeling accuracy and DPC performance

Abstract

Three gaps impede the development of cost-effective and accurate data-driven building energy modeling/models (DBEM) for energy forecasting and predictive control strategies. Gap 1: data bias is common in building operation data, but this topic is hardly studied in DBEM; Gap 2: high data dimensionality is common in DBEM, but a systematic and scalable methodology is lacking to solve the problem; Gap 3: the interactions between data bias and high dimensionality have not been systematically studied for DBEM and predictive control in buildings.

In this paper, to address the three gaps mentioned above, we develop a framework that integrates active learning and feature selection for DBEM used for whole building data predictive control (or DPC, which is a branch of model predictive control). The framework provides a systematic methodology and automatic workflow that starts with raw data from building automation systems to the establishment of data-driven energy models for DPC controllers. The developed strategies and framework are evaluated in a virtual testbed based on EnergyPlus and BCVTB. Improved performance and reduced computational complexity are observed from the DBEM built with the developed framework, as well as the DPC controller based on that DBEM, indicating the effectiveness of the developed framework.

Keywords

data predictive control; data-driven building energy modeling; machine learning; active learning; feature selection; joint active learning and feature selection

1. Introduction

In 2019, 32% of the primary energy in the U.S. was consumed by the building sector [1]. Operation strategies and building controls have a significant impact on occupant comfort and building energy efficiency [2-4]. Moreover, building controls are increasingly challenging and important with the rapid development of grid-interactive efficient buildings in the recent years.

Model predictive control (MPC) has received extensive attention in the field of whole building control and operation strategies [5-8]. MPC is an optimal-control based method to select control inputs by minimizing an objective function: it searches for optimal trade-offs between conflicting objectives [9]. For whole building control, the conflicting objectives are typically energy efficiency/cost and thermal comfort. MPC can consider disturbances (e.g., weather, internal gains, etc.) and give constraints (e.g., comfort and setpoint ranges) to optimize building operation based on the forecasted building energy consumption and defined control goals.

The development of the model, used in the MPC objective function, is the most essential and challenging task of MPC design [10]. To develop MPC for whole building operation and control, an accurate building energy modeling/model (BEM) for building energy forecasting is the cornerstone. Data-driven model is one of the modeling methodologies to capture the mathematical relationship between building operation data and energy consumption using machine learning and statistical methods. Data-driven building energy modeling/model, or DBEM, have been increasingly studied due to it characteristics of simplicity, high level of automation, and low development engineering effort [11, 12], compared with physics-based model and grey-box model [13]. Moreover, the rapid development of machine learning techniques and the increasing data accessibility in buildings further empower the applications of DBEM for building operation and control. MPC built with the DBEM is also termed as data predictive control (DPC) [14]. Owing to the facts above, DBEM and DPC are increasingly studied in academia and applied in the industry.

In spite of the promising prospect, the DBEM for DPC has limitations: it learns from data, and it heavily depends on data quality [15], while building operation data face two major issues: data bias and high data dimensionality. Data bias, or sample selection bias, means that the sample is collected in such a way that some members of the intended population are less likely to be included than others, causing that the sample obtained is not representative of the population intended to be analyzed [16]. For DBEM, data bias refers to the fact that most of the building operation data are generated under limited operational modes, weather conditions, and setpoints (often one or two fixed values, such as a constant zone temperature setpoint) [17]. For nonlinear systems, a data-driven model generated from biased data has poor ability to forecast unseen situations such as uncommon building operation conditions and weather in the future. The ability to forecast unseen situation is often termed as generalization or extendibility.

High data dimensionality, in the context of DBEM, refers to the hundreds of candidate features (or inputs) from the equivalent number of BAS measurements (sensors and meters). DBEM built with unselected and high dimensional features have severe overfitting problem and poor extendibility or generalization. Moreover, DBEM with high dimensional inputs have low computation speed and cost-effectiveness in model training and prediction, which hampers the real-time control in future DPC applications.

As can be observed from the advantages and limitations of DBEM introduced earlier, it is essential to effectively collect unbiased building operation data and deal with high dimensional data to generate cost-effective and accurate DBEM for DPC.

Active learning is a concept in machine learning. It is similar to experiment design in statistics and excitation in system identification: it effectively generate unbiased and informative training data [18]. Specifically, active learning is a pre-modeling procedure to iteratively query the user (or the information source) to obtain the desired labels (or outputs) at new data points [19]. The active learning aims at: (1) generating informative and unbiased training data, and (2) improving model generalization and extendibility to defy data bias problem among training data. To explain it in the context of DBEM, active learning iteratively perturbs the building operation to obtain building energy consumption at different operation scenarios as new building operation data for later training process.

Feature selection (or attribute selection, variable selection) can effectively select and identify features for a data-driven model and improve model generalization and extendibility. Feature selection aims at: (1) improving model generalization and extendibility by defying high dimensionality of training data, and (2) improving the cost-effectiveness and computation speed of the model [20]. Research of feature selection is the key to defy high data dimensionality problem in building operation data.

Extendible energy forecasting model and unbiased training data for the model are critical to the accuracy of DBEM and the DPC built with that DBEM. This paper focuses on a systematic methodology or framework to integrate active learning and feature selection for DBEM in the context of DPC applications.

2. Literature Review

In this section, the literature of the latest development of active learning, feature selection, and the integration of them in DBEM and DPC is reviewed.

Active learning is widely studied in machine learning: uncertainty sampling from Lewis and Gale [21], Query-by-committee (QBC) from Seung et al.[22], expected model change from Settles et al.[23], and expected error reduction from Roy and McCallum [24]. Though well developed in machine learning field, sophisticated active learning algorithms are hardly applied to DBEM study.

Only a few papers are found to apply active learning to the building field. Zhang and Wen [18] explored the application of active learning in DBEM. The developed active learning strategy in this paper considered two key factors of active learning using two modules. The first module efficiently generates informative training data in a short time, and the second module passively considers weather disturbances at the same time. The developed active learning strategy was applied and evaluated in both virtual and real building testbeds against traditional DBEM. In the comparison study, data bias problem in building operation data was identified, improved performance brought by active learning strategy was observed, the effectiveness of block design module was validated, and the extendibility of DBEM developed with the strategy was demonstrated [18].

In terms of feature selection, a review paper [25] summarized feature selection applications in DBEM including filter method (statistics-based scoring and ranking method: [26-35]), wrapper method (exhaustive feature set selection method: [36-40]), embedded method (feature selection module embedded in machine learning algorithms: [41-45]), and hybrid method (combination of two or more methods mentioned above: [46, 47]). The topic of feature selection has been widely applied and studied in this field.

Active learning and feature selection are often intertwined with each other: the result of feature selection is different with different training data (with and without active learning); the result of active learning is also different with different features selected (with and without feature selection). Noisy and high dimensional features (without feature selection) brings adverse effect on active learning, while informative or representative samples (with active learning) are beneficial to feature selection [48]. The intertwining between active learning and feature selection often happens when both are model-based. If active learning is conducted first, a model for active learning is needed, while at this time, feature selection is not conducted and the features for the model are unknown.

How to effectively combine active learning and feature selection has been studied by several researchers in the field of machine learning and information science but not in the field of DBEM. They used the terminology of “joint active learning with feature selection,” “active sampling for feature selection”, and “active feature selection.” Liu et al. [49, 50] introduced the concept of active feature selection and investigated a selective sampling approach to active feature selection in a filter model setting. Further, they showed how it realized active feature selection and reduced the required number of training instances to achieve time savings without performance deterioration. Veeramachaneni and Avesani [51] proposed a technique to actively sample the feature values with the ultimate goal of choosing between alternative candidate features with minimum sampling cost. Their algorithm was based on extracting candidate features in a region of the instance space where the feature value was likely to alter domain knowledge the most. An experimental evaluation of a standard database showed that it was possible to outperform a random subsampling policy in terms of the accuracy in feature selection. Li et al. [48] presented an unsupervised learning approach for simultaneous sample and feature selection. Their proposed approach was in contrast to existing works which mainly tackle these two problems separately. They proposed a framework to jointly conduct active learning and feature selection based on the CUR matrix decomposition [52]. From the data reconstruction perspective, both the selected samples and features can best approximate the original dataset respectively, such that the selected samples characterized by the features were highly representative. To sum up, it will be of great interest to extend the integration of active learning and feature selection, which currently only exists in the literature on machine learning and information science, to the DBEM field.

Table 1 summarizes the literature review on active learning, feature selection and their integration. Concluded from the introduction and literature review, the fact that active learning and feature selection have not been systematically integrated for DBEM and DPC application impedes the development of accurate DBEM and DPC. To address the gap, this paper aims to develop a framework that integrates active learning and feature selection to systematically defy data bias and high data dimensionality of building data in the process of DBEM used for whole building DPC application.

Table 1. Summary of the literature of active learning and feature selection reviewed in theoretical studies, DBEM, and DPC for building control

Field	**Active learning**	**Feature Selection**	**Integration of active learning and feature selection**
Theoretical studies	Well studied	Well studied	Studied
DBEM	Studied	Well studied	Rarely studied
DPC for building control	Studied	Well studied	Rarely studied

In this paper, Section 3 develops a framework integrating active learning and feature selection to provide a systematic methodology and automatic workflow for DBEM and DPC. In Section 4, the case study of a virtual testbed applying the developed framework is introduced. In Section 5, the DBEM and DPC controller developed by the proposed framework is evaluated by being compared with other modeling methodologies. In Section 6, conclusions and future work are summarized.

3. A Framework that Integrates Active Learning and Feature Selection for DPC

As discussed in Section 2, though the topic of joint active learning and feature selection has been studied in the field of machine learning, this has not been studied in the field of DBEM. The main objective of this paper is to study the integration of active learning with feature selection for DBEM and whole building DPC. The detailed research questions of incorporating active learning with feature selection in the context of DPC include:

- 1. What is the order to apply active learning and feature selection? If the model-based active learning is applied first, how to decide the features of the model used in active learning without applying feature selection? If feature selection is applied first, will the selected feature be the same as the features selected after active learning? Active learning and feature selection are tangled with each other, and the same issue is stated in [48]. A mechanism is in great needs to deal with this problem.
- 2. How to design the integration of active learning and feature selection systematically in the context of DPC? DBEMs are eventually used for DPC. A systematic methodology or automated workflow is in need to guide the design of the integration of active learning and feature selection to specific DPC applications. For example, the framework should refer to the forecasting horizon of DPC (e.g., forecasting the next hour building energy consumption) to decide the output of the model used in active learning and feature selection. In another example, controllable inputs to be manipulated in active learning should be the same as those in future DPC; modelers should not remove controllable inputs of future DPC during feature selection. A workflow is in great need to coordinate the design of active learning and feature selection with future DPC applications.

To fill the research gap, we develop a framework to systematically incorporate active learning with feature selection. The framework considers the compatibility among active learning, feature selection, DBEM, and DPC. The goal of the framework is to improve DBEM and finally improve the performance of DPC.

As can be seen in Figure 1, the framework starts with the data collected from BAS. Sensor data are categorized into controllable inputs, disturbances/state variables, and outputs. Active learning introduced in [18] is conducted first to perturb controllable inputs (also considers disturbances using block design) to generate informative training data. Then, a systematic feature selection procedure as described in [46] is conducted to select the feature set that works the best with the controllable inputs, from disturbances/state variables. The final outcomes of this integrated framework are DBEMs (used for DPC) that are trained with the information-rich data generated from the active learning strategy and model features (inputs) that are selected by the feature selection procedure based on this data.

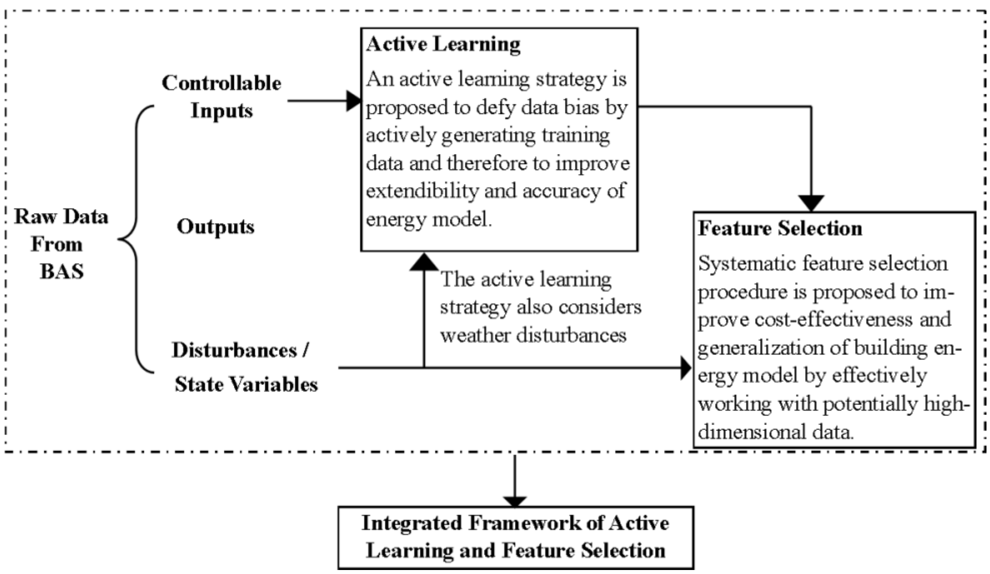

Figure 1. Brief diagram of the framework that integrates active learning and feature selection

Figure 2 provides a more detailed workflow showing the procedures from collecting data from BAS to generating the final DBEMs used for whole building DPC. Sensor data from BAS are firstly categorized into disturbances/state variables, controllable inputs, and outputs. The controllable inputs here should be the same as the controllable inputs in the future DPC application. The outputs here should correspond with the outputs of DBEM in the future DPC application. Among all disturbances, weather-related ones (such as outdoor air dry-bulb temperature) along with the outputs, are used to build a model for disturbance categorization in active learning [18]. Weather disturbances, along with controllable inputs, features selected based on domain knowledge (optional) and outputs, are used to build the preliminary DBEM used for active learning. Feature selection is not conducted yet, and features do not need to be perfectly selected because this step only focuses on manipulating controllable inputs, instead of selecting disturbances using feature selection.

Then, the active learning strategy introduced in [18] is conducted. Three components of active learning: a DBEM for active learning, a setpoint pool, and an active learning algorithm in machine learning (such as expected error reduction, uncertainty sampling, QBC, and expected model change) are decided, and active learning strategy will generate informative training data within the budget for active learning.

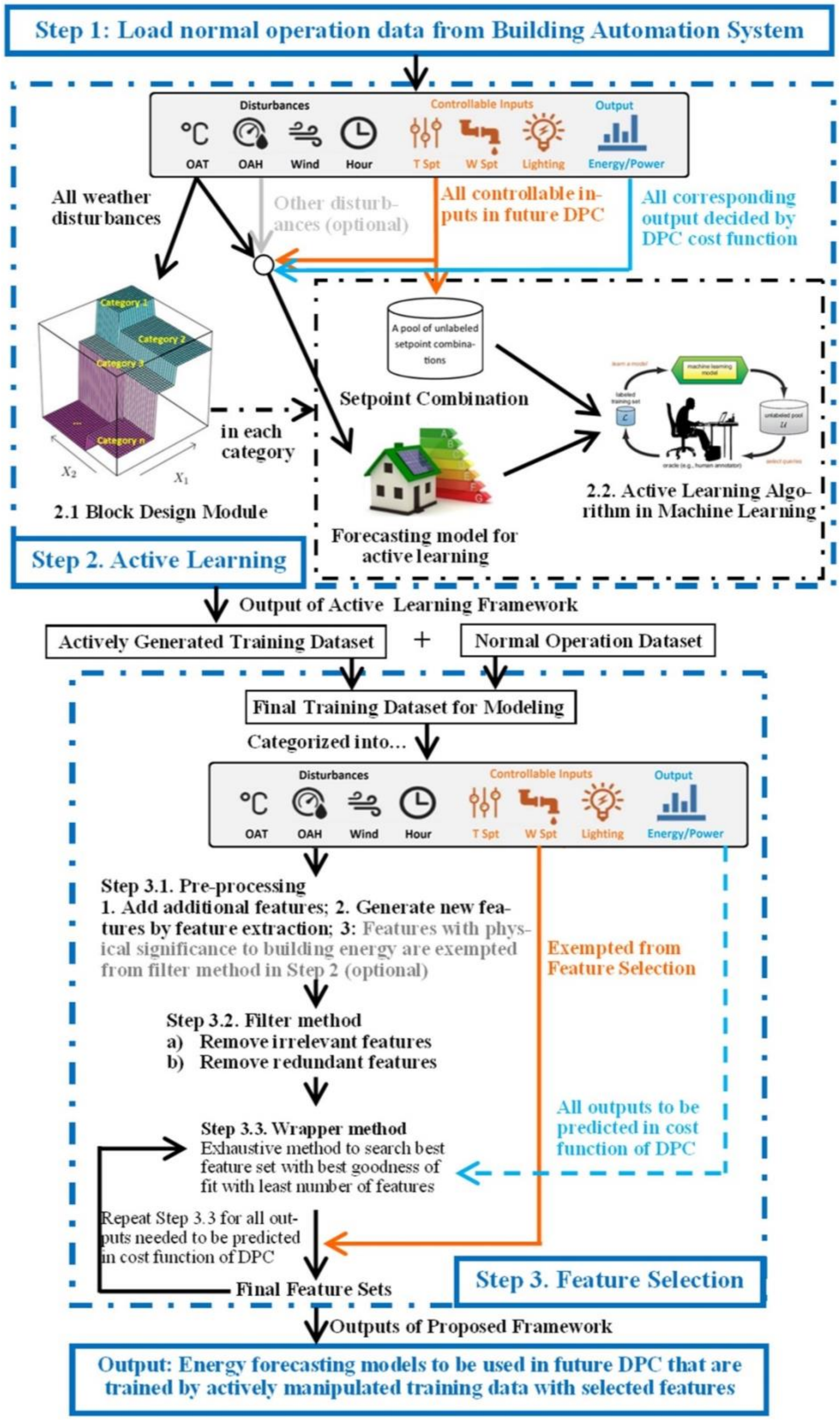

Figure 2. A detailed diagram of the framework that integrates active learning and feature selection

After active learning, the generated data are merged with existing normal operation data. In the cases where normal operation data greatly outnumber active learning data, weight is added on training data to increase their impact on the whole dataset. Then, a systematic feature selection procedure decides the final feature set from the merged data. With the same categorization of features in active learning, all features in the merged dataset are also categorized into disturbances, controllable inputs, and outputs. Controllable inputs are exempted from selection and directly selected into the final feature set. For the disturbances, a three-step feature selection procedure as described in [46] is conducted. The developed feature selection process is conducted individually for all DBEMs that will be used in the future DPC application; the feature sets selected for each forecasting models are different. For example, if the forecasting horizon in the future DPC application is three hours and requires three DBEMs to forecast the building energy consumptions in the future first, second and third hour, the systematic feature selection procedure will be conducted for three times to decide three different feature sets for three DBEMs. The final outcomes of this integrated framework are DBEMs (used for DPC) that are trained with the information-rich data generated from the active learning strategy and model features (inputs) that are selected by the feature selection procedure based on this data.

It is worth mentioning that active learning, feature selection, and DPC are model-based. The data-driven modeling algorithm of the models (like ANN, SVM, decision tree, etc.) in active learning, feature selection, and future DPC should be the same.

We summarize the key characteristics of the developed framework: (1) active learning is conducted before feature selection; (2) model features for active learning consist of weather disturbances, controllable inputs and some disturbances based on domain knowledge (if there is any). Systematic feature selection techniques are not required during active learning; (3) the outputs, controllable inputs, forecasting horizon, and data-driven modeling algorithms in active learning and feature selection should be the same as those in future DPC; (4) the developed feature selection process is conducted independently for all DBEMs that will be used in the future DPC application; the feature sets selected for each forecasting model are different.

In the next section, the developed framework is applied to a virtual testbed (small reference office building) to demonstrate the effectiveness of the integrated framework.

4. Case Study

4.1 Virtual Testbed: Small Reference Office Building and BCVTB

The object building in the case study is the small reference office building developed by National Renewable Energy Laboratory [53] under the environment of EnergyPlus [54]. It is a one-story 511-square-meter office building; the packaged air-conditioning unit is used for cooling and furnace is used for heating; the air distribution system is single-zone constant air volume; floor-to-ceiling height is 3.05 meter; the window-to-wall ratio is 0.21 meter; lighting follows the code of ASHRAE 90.1-2004. Figure 3 shows the outlook and zone configuration of the building. More details of the small office building can be found in [53]. The building model is simulated under the TMY (typical meteorological year) weather file "USA_IL_Chicago-OHare.Intl. TMY3" in the modeling stage and is simulated under AMY (actual typical meteorological year) weather file "USA_IL_Chicago-OHare.Intl. AMY: 2007" in the control and testing stage.

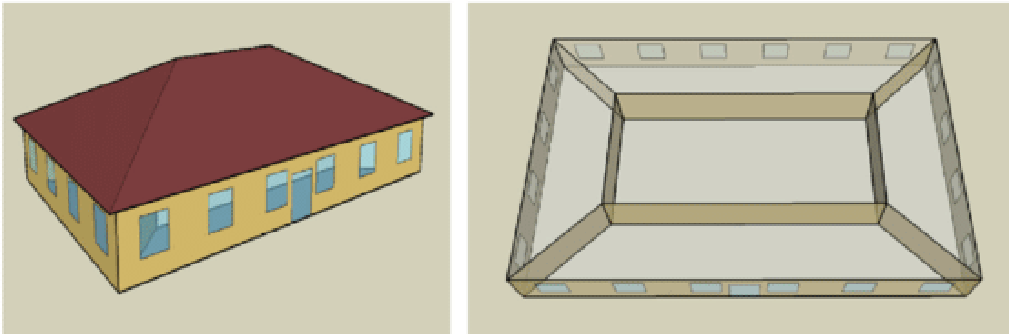

Figure 3. The appearance and zone configuration of the small office building

We use BCVTB to connect the DPC controller (modeled with MATLAB SIMULINK) and virtual testbeds (modeled with EnergyPlus). BCVTB is a software environment that allows users to couple different simulation programs for co-simulation [55]. Figure 4 shows the modularized user interface in BCVTB, where the virtual building modeled in EnergyPlus is connected with active learning and DPC controller modeled in Simulink. The simulation time interval and duration in BCVTB, Simulink, and EnergyPlus are set to the same values.

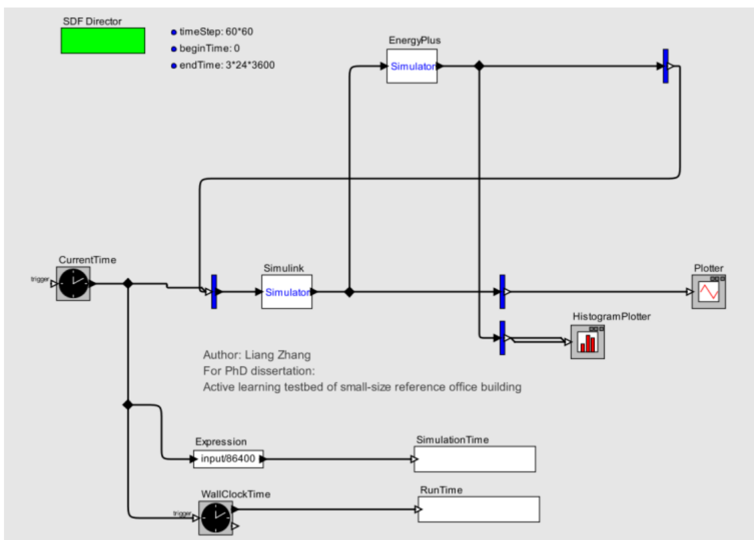

Figure 4. BCVTB interface connecting virtual building (EnergyPlus) and active learning/DPC controller (MATBLAB Simulink) as virtual testbed for small reference office building

4.2 Framework Applied in DBEM

The proposed framework is applied to the DBEM of the small office building. Table 2 lists the virtual sensors and their categories. Features are generated from those sensors. The output of the DBEM is whole building electricity consumption in the future first, second, and third hour, which will be in accordance with the DPC application where the forecasting horizon is three hours. The details of DPC settings are introduced in Section 4.3.

Table 2. Virtual sensors in the small office building

Sensor Name	**Details**	**Category**
ZTSP	Zone temperature setpoint	Controllable inputs
OADT	Site outdoor air dry-bulb temperature	Weather disturbances

OAWT	Site outdoor air wet-bulb temperature	Weather disturbances
SR_DIF	Site diffuse solar radiation rate per area	Weather disturbances
SR_DIR	Site direct solar radiation rate per area	Weather disturbances
ZSF1_ZT*	ZSF1 zone air temperature	State variables
ZNF1_ZT	ZNF1 zone air temperature	State variables
ZSF2_ZT	ZSF2 zone air temperature	State variables
ZNF2_ZT	ZNF2 zone air temperature	State variables
OCC	Occupancy schedule	Other disturbances
LIT	Lighting schedule	Other disturbances
EQP	Equipment schedule	Other disturbances
ZSF1_INLET_T	ZSF1 zone supply air temperature	State variables
ZNF1_INLET_T	ZNF1 zone supply air temperature	State variables
ZSF2_INLET_T	ZSF2 zone supply air temperature	State variables
ZNF2_INLET_T	ZNF2 zone supply air temperature	State variables
Other node temperature	EnergyPlus system node temperature	State variables
Other node flow rate	EnergyPlus system node flow rate	State variables
ELECTRICITY_WHOLE	Whole building electricity consumption from the current time step	Output from the current time step

* ZSF1, ZNF1, ZSF2, and ZNF2 are zone names in the reference building

4.2.1 Application of Active Learning

The active learning in this study applies the same strategy presented in [18] and is shown in Figure 2. The more detailed diagram is shown in Figure 5. The developed strategy in [18] considers two key factors of active learning with two modules. The first module, setpoint selection module, efficiently generates informative training data in a short time using expected error reduction algorithm; the second module, called weather disturbance module, passively considers weather disturbances at the same time using block design [18]. More details can be found in [18].

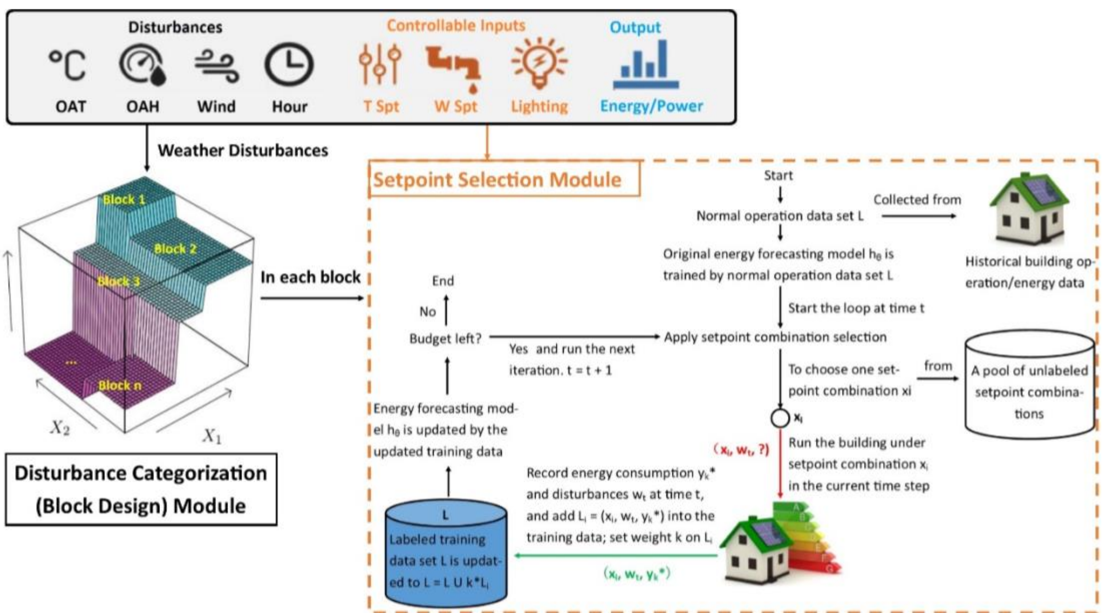

Figure 5. The diagram of active learning strategy for DBEM (source: [18])

Offline weather disturbance categorization (or block design) is the first step of the developed active learning strategy. A classifier is built with weather features based on the data collected during normal operation. The purpose is to create classes of weather conditions for block design to passively consider weather conditions in the experiment design or active learning. In the case study, weather features include site outdoor air dry-bulb temperature (OADT, ℃), site outdoor air wet-bulb temperature (OAWT, ℃), site diffuse solar radiation rate per area (SR_DIF, W/m^2), site direct solar radiation rate per area (SR_DIR, W/m^2), as shown in Table 2. Decision tree is the algorithm for the classifier. The inputs of the decision tree model are the four weather features mentioned earlier; the output is the next-hour whole building electricity consumption. In the decision tree classification process, the weather space, i.e., "USA_IL_Chicago-OHare.Intl. TMY3" weather data, is divided into several sub-spaces (blocks) for the virtual testbed. With the classifier, weather disturbances are categorized into three blocks based on four weather features mentioned earlier. The weather condition of the first block is SR_DIF < 81.5 W/m^2; the weather condition of the second block is SR_DIF ≥ 81.5 W/m^2 and OADT < 23.05℃; the weather condition of the third block is SR_DIF ≥ 81.5 W/m^2 and OADT ≥ 23.05℃. More details about the block design can be found in [18].

The setpoint selection module is then conducted. Expected error reduction is the active learning algorithm in the setpoint selection module. According to Figure 5, the first step of expected error reduction is to build an initial DBEM using collected normal operation data. To build the model, model features (inputs) are decided based on domain knowledge. In the case study, the features of the DBEM for expected error reduction are: controllable input zone temperature setpoint (ZTSP), selected weather disturbances: outdoor air dry-bulb temperature (OADT), diffuse solar radiation (SR_DIF), and other features by domain knowledge: zone temperature of zone ZSF1 (ZSF1_ZT), occupancy schedule (OCC), zone temperature setpoint in the last time step (ZTSPt-1), and current whole building electricity consumption (ELECTRICITY_WHOLE). Multivariate adaptive regression spline (MARS) is the data-driven modeling algorithm of the model for active learning. The original DBEM built by the above features and the data-driven modeling algorithm is not the final DBEM for DPC: it is only an initial or preliminary DBEM trained by normal operation data as the first step of estimated error reduction algorithm.

In normal operation data, the only controllable input, zone temperature setpoint, has extremely limited values (24℃ and 26.7 ℃) in normal operation period. This schedule is the same as the default zone temperature setpoint schedule in the small reference office building developed by U.S. Department of Energy. The initial or preliminary DBEM is built with normal operation data. The DBEM gradually includes more actively selected training data to get updated in every step of active learning. Expected error reduction is a pool-based algorithm: it selects setpoint from a limited set of setpoints. The pool consists of a full factorial design of zone temperature setpoint. With zone temperature setpoint ranges from 20℃ to 30℃ with an equal interval of 0.5℃, the pool size is 21. The summarized settings of the active learning strategy applied in small office building is summarized in Table 4.

Within the limited time of active learning (36 budgets, or the timestep limit for active learning process), setpoints are selected as shown in Figure 6. Using the data generated from the active learning strategy, the developed MARS model has experienced not only almost every possible value in the setpoint pool but also experienced quite different dynamic patterns (fast changing and slow changing) which is beneficial for the model to recognize different building operations and dynamics.

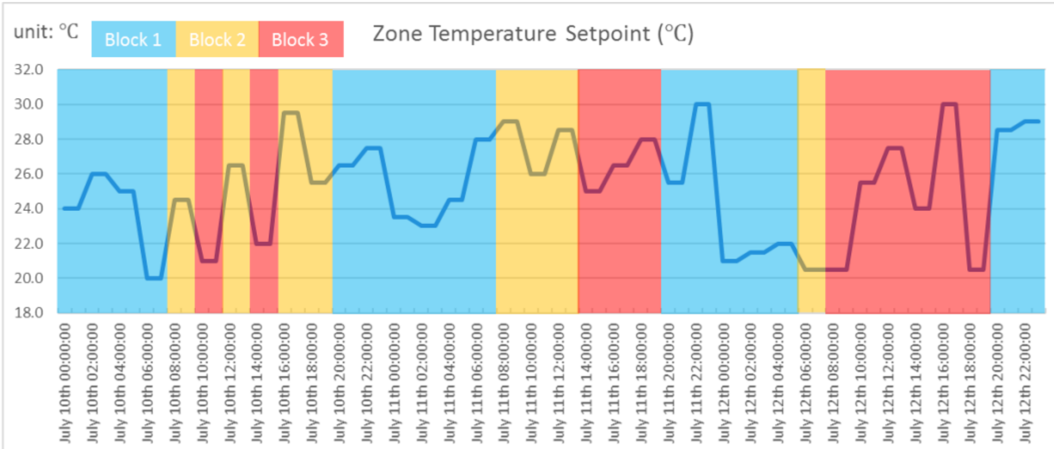

Figure 6. Zone temperature setpoint decided by the developed strategy with block number during the active learning period in the small office building

4.2.2 Application of Feature Selection

Training data have been informatively expanded by adding active learning data to normal operation data, and the next step is the application of feature selection on the training data generated from normal operation and active learning. The feature selection in this study is in accordance with the feature selection framework for DBEM presented in [46]. In the first step, new features: hour of day and weekday indicator, are generated from time stamp (e.g., 12/2/2018 12:03:00). Then, 1-hour lag features are extracted from all raw disturbances and state variables. The number of features after Step 1 is 35. We use Pearson correlation coefficient for filter method in Step 2. The top five features with highest absolute Pearson correlation coefficient are: zone temperature setpoint for control (0.897), diffuse solar radiation (0.884), Direct solar radiation (0.878), zone temperature of zone ZSF2 (0.790), and supply air temperature of zone ZSF1 (0.777). We use a forward wrapper method in Step 3. MARS is used as the data-driven modeling algorithm for the wrapper method. Root mean squared error (RMSE) of 4-fold cross validation is used as evaluation metrics in the wrapper method. To sum up, Table 4 lists the details of the integrated framework for the reference office building. The final feature set used for next-hour DBEM is listed in Table 3.

Following the joint active learning and features selection framework introduced in Section 3 and with the perturbed values using active learning and the selected features using systematic feature selection method, DBEMs can be trained using MARS algorithm to forecast building energy for each forecasting horizon (future 1st, 2nd, and 3rd hour, according to DPC settings which will be introduced in Section 4.3 in detail).

Table 3. The final feature set in small office building using the integrated framework of active learning and feature selection for next-hour DBEM

No	Features	Description
1	$ZTSP_{t+1}$	Zone temperature setpoint for control
2	SR_DIF	Diffuse solar radiation
3	SR_DIR	Direct solar radiation
4	ZSF2_ZT	Zone temperature of zone ZSF2*
5	ZSF1_INLET_T	Supply air temperature of zone ZSF1
6	ZSF2_INLET_T	Supply air temperature of zone ZSF2
7	ELECTRICITY_WHOLE	Whole building electricity consumption in the current hour
8	ZNF1_INLET_T_{t-1}	Supply air temperature of zone ZNF1 in the last time step
9	$OADT_{t-1}$	Outdoor air dry-bulb temperature in the last time step
10	ZNF1_ZT	Zone temperature of zone ZNF1
11	Hour of day	Hour of day
12	Weekday/end Indicator	Weekday and weekend indicator

13	Hour of day$_{t-1}$	Hour of day in the last time step
14	Weekday/end Indicator$_{t-1}$	Weekday and weekend indicator in the last time step
15	SR_DIF$_{t-1}$	Diffuse solar radiation in the last time step
16	SR_DIR$_{t-1}$	Direct solar radiation in the last time step
17	ZNF2_ZT$_{t-1}$	Zone temperature of zone ZNF2 in the last time step
18	ZNF2_ZT	Zone temperature of zone ZNF2
19	OCC$_{t-1}$	Occupancy schedule in the last time step
20	ZNF2_ZT	Zone temperature of zone ZNF2
21	ZSF2_ZT	Zone temperature of zone ZSF2
22	ZSF1_ZT	Zone temperature of zone ZSF1
23	OAWT$_{t-1}$	Outdoor air wet-bulb temperature in the last time step
24	OCC	Occupancy schedule

* ZSF1, ZNF1, ZSF2, and ZNF2 are zone names in the reference building

Table 4. Details of the integrated framework of active learning and feature selection in the case study

Active Learning	
Block design variables	OADT, SR_DIF, OAWT, SR_DIR
Block classification algorithm	Regression tree
Number of blocks in block design	3
Normal operation dates	July 1 to Sept 30
Weather	USA_IL_Chicago-OHare.Intl. TMY3
Features for DBEM	ZTSP (controllable), ZTSP$_{t-1}$, OADT, ZSF1_ZT, SR_DIF, OCC, ELECTRICITY_WHOLE$_{t-1}$
Machine learning algorithm for DBEM	MARS
Setpoint selection algorithm	Expected error reduction
Setpoint selection interval	Two hours
Training data weights	551 to make training data account for 60% of total training
Setpoint pool size in each block	Zone temperature setpoint: 20 to 30℃ with 0.5℃ interval; pool size: 21
Active learning dates	July 10 to July 12
Weather file	USA_IL_Chicago-OHare.Intl. AMY: 2007
Feature Selection	
Additional features by domain knowledge	Time of day, weekday indicator, time-lag features for all disturbances
Filter method settings	Skipped due to enough computation power for direct wrapper method
Wrapper method searching method	The forward feature selection method
Wrapper method algorithm	MARS
Wrapper method evaluation metrics	RMSE of 4-fold cross-validation
Integrated Framework	
Weights of training data from the active learning	10 (equivalent 30 days of active learning data for training)

4.3 Framework Applied in DPC

The developed DBEM introduced in Section 4.2 can be used as the forecasting model to build DPC controller. In this section, a DPC controller is designed as a supervisory control strategy for a demand-response (DR) scenario, i.e., the DPC only modifies setpoints, rather than controlled devices. The objective is to minimize the whole building electricity cost during a forecasting horizon while ensuring thermal comfort by maintaining zone temperature within a comfortable range. Four major settings, including cost function, optimization method, and DPC performance metrics, are introduced in this section.

4.3.1 Cost Function

The cost function in this study considers trade-offs between two conflicting objectives: building electricity consumption and thermal comfort. The cost function of DPC in this study is defined as:

$$minimize: J = \sum_{k=1}^{3}[R(k) * f_k(u(t+k-1), u(t+k), w_k(t)) + \varphi(u_1)]$$
$$subject\ to:\ u_{lb} \le u \le u_{ub}. \quad \text{(Eq.1)}$$

R(k) (unit: \$/kWh) is the electricity tariff (or cost rate) within the kth hour in the future, which is available from the electricity tariff shown in Figure 9. u is zone temperature setpoint. u_{lb} (20 °C) and u_{ub} (30 °C) are lower and upper boundaries of controllable inputs. "k=1:3" indicates that the forecasting and control horizon is 3 hours, and the control sampling time is 1 hour. The selection of the forecasting horizon and control sampling time is introduced in Section 4.3.1.1. $\varphi(u)$ is the penalty function of zone temperature setpoint that is introduced in Section 4.3.1.2.

The key component in the cost function is f_k. f_k (unit: kWh) is the function to forecast whole building electricity consumption within the kth future hour. This function is defined by the DBEM developed in Section 4.2. The independent variables of the function are present controllable inputs $u(t+k)$, previous controllable inputs $u(t+k-1)$ and previous disturbances $w(t)$. As shown in Eq.1 and Figure 7, f_1 is the function to forecast the energy consumption in the 1st future hour, whose features (input variables) are controllable inputs at time t+1 and t, and disturbances at time t; f_2 is the function to forecast the energy consumption in the 2nd future hour, whose features are controllable inputs at time t+2 and t+1 and disturbances at time t; f_3 is the function to forecast the energy consumption in the 3rd future hour, whose features are controllable inputs at time t+3 and t+2 and disturbances at time t. In this study, weather is not forecasted during the optimization process.

For each forecasting model with different forecasting horizon, the disturbances are individually decided by the systematic feature selection procedure. Thus, for the three forecasting models (f_1, f_2 *and* f_3), disturbances and state variables selected for each model are different. k in w_k indicates different selected feature sets for models with different forecasting horizons.

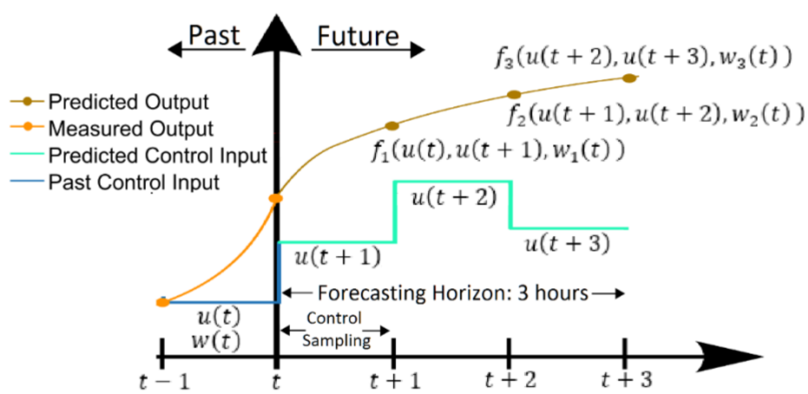

Figure 7. Diagram of DPC with 3-hour control/forecasting horizon and 1-hour control sampling time

4.3.1.1 Forecasting Horizon, Control Horizon and Control Sampling Time

The forecasting horizon refers to the length of time for which system output is computed by the MPC or DPC, whereas the control horizon denotes the length of time for which the control signal is computed; the control sampling time is the time during which the control signal remains unchanged [56]. For typical whole building DPC, the forecasting horizon is 3-48 hours, the control horizon is 3-5 hours, and the control sampling time is 1-3 hours [56]. In this study, the forecasting horizon is 3 hours; the control horizon is 3 hours; the control sampling time is 1 hour.

4.3.1.2 Penalty Function of Thermal Comfort

Penalty function of thermal comfort is widely used in the existing whole building DPC research to introduce an energy-comfort trade-off to the cost function [57, 58]. In this study, thermal comfort is considered by adding a penalty function of zone temperature setpoint to the cost function. The penalty function quantifies deviation of zone temperature setpoint from the center of the comfort range (25 ℃ in this study), adds a penalty to the cost function to avoid large deviation from the comfort temperature, and introduce an energy-comfort trade-off to the cost function. In this study, the penalty function is in the cubic form [59], which is shown in Eq.2 and Figure 8.

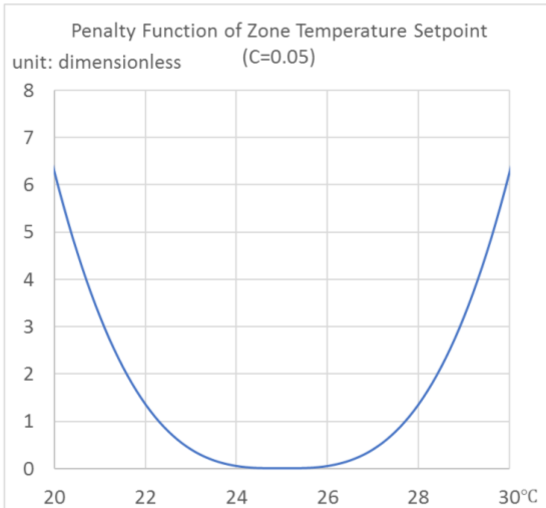

Figure 8. Penalty function of zone temperature setpoint in the cost function of DPC (C=0.05)

$$\varphi(u) = C * |u - 25|^3 \qquad \text{(Eq.2)}$$

u is zone temperature setpoint. The center of zone temperature comfort range is 25℃. When C = 0.05, the penalty of 1℃ deviation from 25℃ is 0.05 (dimensionless); the penalty of 2℃ deviation is 0.4; the penalty of 5℃ deviation is 6.25. The value of parameter C is customized to reflect a user-selected trade-off between discomfort penalty and building electricity consumption.

4.3.1.3 Electricity Tariff

To introduce demand-response events in this study, we use a dynamic electricity tariff shown in Figure 9. The tariff is also used in a building MPC study [60]. The tariff utilizes a step-wise electricity pricing mechanism to reflect the electricity supply-demand market.

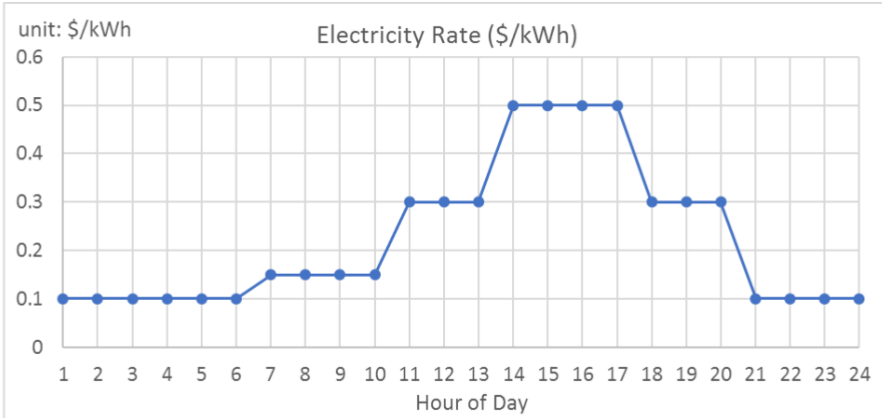

Figure 9. Stepwise dynamic electricity tariff in the cost function of DPC

4.3.2 Optimization Method

In each time step of DPC runtime, the control sequence (setpoint) is calculated by minimizing the cost function. The exhaustive searching method is used as the optimization method in this study. The exhaustive searching method first discretizes the setpoints by 0.5℃ throughout the control horizon. Then it calculates the cost function of every combination of setpoints throughout the control horizon. Finally, the exhaustive searching method uses the combination of setpoints with the minimum cost function as the control signal in the current time step lasting for one hour in the future.

4.3.3 DPC Performance Metrics

Among the common performance metrics of DPC controller, total electricity cost, peak load shifting capability, thermal comfort penalty, and computation time are considered in this study.

Total electricity cost is the most straightforward index to reflect the energy-saving ability of a DPC controller. The equation of this metrics is shown in Eq.3. n is the total hour of DPC runtime; R(k) is the electricity tariff at time k; E(k) is the whole building electricity consumption at time k.

$$M_1 = \sum_{k=1}^{n}[R(k) * E(k)] \qquad \text{(Eq.3)}$$

Peak load shifting capability is a more specific reflection of system response to high electricity tariff. It equals the highest building electricity consumption when the electricity tariff reaches its highest value. The mathematical expression of peak load shifting is defined as follows. R(i) is the electricity tariff at time i; E(i) is the whole building electricity consumption at time i.

$$M_2 = \underset{E(i)}{argmax}\, R(i) \qquad \text{(Eq.4)}$$

The third metrics is the thermal comfort penalty $\varphi(T_k)$, which is the same penalty function used in the cost function. It quantifies the accumulated penalty of deviation of zone temperature setpoint from the center of comfort range. n is the total hour of DPC runtime; $\varphi(T_k)$ is the penalty of zone temperature setpoint at time k.

$$M_3 = \sum_{k=1}^{n}[\varphi(T_k)] \qquad \text{(Eq.5)}$$

Computation time is also important in DPC because the control signal calculation time should be less than the control sampling time. As a result, in this study, average CPU time used to generate future setpoints for the next control signal is one of the evaluation metrics. The average CPU time is calculated under the following computer configuration: Intel(R) Xeon(R) CPU E3-1505M v5 @ 2.8GHz with 16.0 GB installed memory.

4.3.4 Developed Framework Applied to build DPC Controller

To sum up, the developed integrated framework introduced in Section 3 is applied to the virtual testbed to generate DBEMs. Details of the building are summarized in Section 4.1. The application of the integrated framework to the DBEM of the small office building is introduced in Section 4.2 in detail. The control testbed for DPC built in BCVTB is introduced in Section 4.1. DPC runtime for the small office building is from July 13th to July 17th (5 days), which is right after active learning conducted on July 10th to July 12th. Table 5 summarizes detailed DPC settings in the virtual testbed.

Table 5. Summary of the detailed DPC settings in virtual testbed

Contents	Settings
Cost function	$J = \sum_{k=1}^{3}[R(k) * f_k(u(t+k-1), u(t+k), w_k(t)) + \varphi(u_1)]$
Constraints	None
Control horizon	3 hours
Control sampling time	1 hour
Forecasting horizon	3 hours
Zone temperature setpoint penalty function	$\varphi(u_1) = C * \lvert u - 25 \rvert^3, C = 5E4$
Electricity tariff	Shown in Figure 9
Optimization method	Exhaustive searching method
Performance metrics	Total electricity cost, peak load shifting capability, thermal comfort penalty, average CPU time per step
DPC runtime	July 13th to July 17th (5 weekdays)
Weather files	USA_IL_Chicago-OHare.Intl. AMY: 2007
Testbed	BCVTB connecting EnergyPlus and Simulink (Figure 4)

The control signals that are calculated by the DPC controller is shown in Figure 10. It can be seen from the figure that the zone temperature setpoint varies among 26, 27 and 28°C to respond to the dynamic electricity tariff. Calculation and comparison of DPC performance among different DPC controllers are introduced in Section 5.

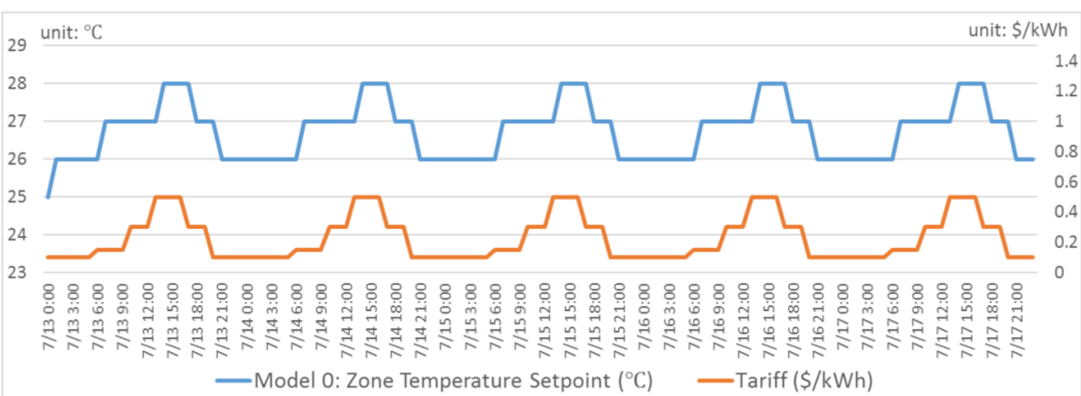

Figure 10. Control signals (setpoints) and electricity tariff of DPC in the small office building

5. Comparison Study: Results and Discussion

To demonstrate the effectiveness of the developed framework, two comparison studies are conducted. First, the DBEM developed by the framework is compared with other DBEMs (Section 5.1). Then, the DPC controller developed by the DBEM by the proposed framework is compared with other DPC controller (Section 5.2).

5.1 DBEM Evaluation

To quantify the impact of the developed joint active learning and feature selection framework on DBEM accuracy and performance, two metrics are adopted: extendibility and computation cost. The metric of extendibility is calculated by averaged DBEM testing accuracy when the building is operated under different and even unseen operation strategies. In the first testing scenario, the virtual building is operated under normal operation strategy as the default schedules in the reference building [53] shown in **Error! Reference source not found.**Figure 11.

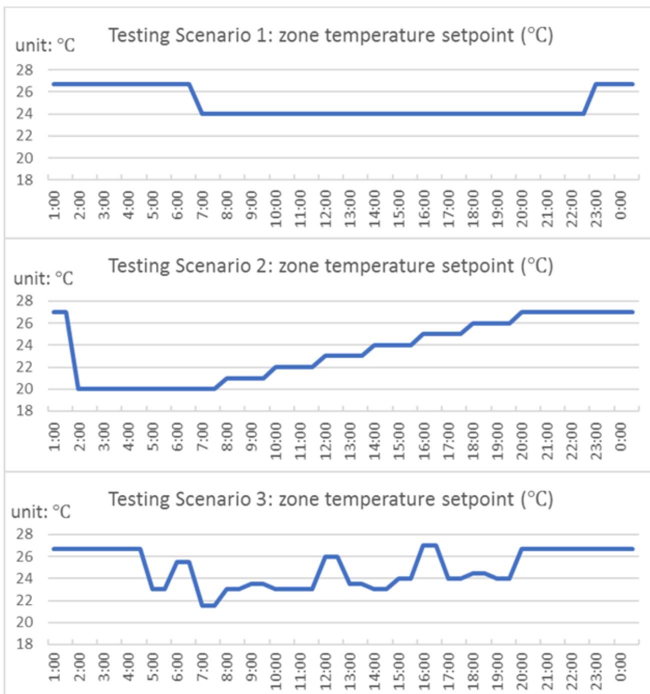

Figure 11. Daily zone temperature setpoint schedules in Testing Scenario 1, 2 and 3

In the second testing scenario, the building is operated under a precooling scenario, where the zone temperature setpoint is set low before building operation hours and it rises slowly, similar to those described in [61] (Figure 11). In the third testing scenario, the building is operated under a fast-changing setpoint schedule to mimic demand response, which is similar to those described in [60] (Figure 11). Each testing scenario lasts seven days simulated with the weather of “USA_IL_Chicago-OHare.Intl. TMY3” under the virtual testbed introduced in Section 4. The metric of computation cost of the active learning is evaluated by CPU time to generate setpoints in each time step, and the computation cost of feature selection is calculated by the time to search for the optimal features for DBEM modeling. The CPU time is calculated under the computer configuration of Intel(R) Xeon(R) CPU E3-1505M v5 @ 2.8GHz with 16.0 GB installed memory

To show the effectiveness of the integrated framework of active learning and feature selection, we compare the performance of the model built by the developed framework with the models built by other literaturereported modeling processes. As shown in Table 6, the model built by the developed integrated framework of active learning and feature selection is marked as Model 0. A model that represents typical field practices without any active learning nor feature selection is denoted as Model 1. Model 1 is trained by normal operation data. To demonstrate the effectiveness of active learning, Model 2 uses the same feature selection procedure as that for Model 0 without any active learning process. To demonstrate the effectiveness of feature selection, Model 3 uses the same active learning process as that for Model 0 without any feature selection process. To examine how scalable the developed active feature framework is, when it is used with different modeling structure, Model 4 is developed which applies the active feature framework with a decision tree modeling structure. To examine the significance of sequence, i.e., whether active learning should be used before feature selection process, Model 5 is developed with the same active learning and feature selection methods, but with an opposite sequence, that is, feature selection is applied first using normal operation data, active learning process is then applied with those selected features.

Table 6. Models and metrics for the comparison study

Models for Comparison	**Metrics**
(1) Model 0: trained by data generated by the developed framework with MARS algorithm **(2) Model 1:** trained by normal operation data **(3) Model 2:** trained by data with active learning without feature selection **(4) Model 3:** trained by data with feature selection without active learning **(5) Model 4:** trained by data generated by the developed framework with regression tree **(6) Model 5:** trained by data generated by feature selection first and active learning second	**(1) Extendibility:** model forecasting accuracy in three testing scenarios **(2) Computation Cost** (a) Active learning (b) Feature selection

Table 8 and Figure 12 summarized the comparison results. Model 0 has a similar error under Test Scenario 1 (normal operation) but has a much lower error under Test Scenario 2 and 3 (pre-cooling and demand-response setpoint schedules) than Model 1. The result indicates that Model 0 has better model extendibility than Model 1 without much compromise on the forecasting accuracy for normal operation (Test Scenario 1). In other words, training data enriched by active learning strategy can greatly improve model extendibility (forecasting building energy under various and unseen operation modes). Model 1 has poor performance in Test Scenario 2 at nighttime when zone temperature setpoint is 24℃, because Model 1 is trained by normal operation schedule where zone temperature setpoint is kept only at 26.7℃ at nighttime. Hence, it causes inaccuracy to use data-driven models trained by biased data (such as the normal operation schedule in the comparison study) to forecast building energy pattern in operation modes such as precooling, preheating, and demand response.

Model 0 with integrated active learning and feature selection outperforms Model 2 with only active learning techniques. However, only a little increment of model performance is observed because MARS algorithm is used as the machine learning algorithm in the case study, and it has intrinsic feature selection module with cross-validation test. In this way, Model 2 has experienced feature selection inside MARS but not with the systematic feature selection procedure. The large increment of performance can be observed from the algorithm without intrinsic feature selection (like decision tree) when adding feature selection to active learning.

Similar to the comparison between Model 0 and Model 1, Model 0 with integrated active learning and feature selection outperforms Model 3 with only feature selection techniques in Test Scenario 2 and Test Scenario 3, indicating the importance of active learning in the integrated framework in improving the model extendibility.

Model 0 (modeled by MARS) has worse performance than Model 4 (modeled by regression tree) only in Test Scenario1, but Model 0 has much better performance in Test Scenario 2 and Test Scenario 3, because decision tree has a higher tendency of overfitting than MARS. High performance in Test Scenario 1 is a sign of overfitting. Overfitting also leads to high variance problem, which makes the model lose accuracy in unseen cases of Test Scenario 2 and 3.

Model 0 (active learning first and feature selection second) outperforms Model 5 (feature selection first and active learning second) especially in Test Scenario 2 and Test Scenario 3. Table 7 shows that fewer features are selected in Model 5. Besides, Model 5 has lower accuracy in three testing scenarios. These two pieces of evidence indicate that Model 5 underestimates the model complexity, leading to an underfitting problem.

The high model performance of both Model 0 and Model 4 indicates the scalability of the framework to be compatible to various data-driven modeling algorithms.

Table 7. Comparison of the feature selected in Model 0 (active learning first and feature selection second) and Model 5 (feature selection first and active learning second)

No	Features in Model 0	Features in Model 5
1	Zone temperature setpoint for control	Zone temperature setpoint for control
2	Diffuse solar radiation	Diffuse solar radiation
3	Direct solar radiation	Direct solar radiation
4	Zone temperature of zone ZSF2	Supply air temperature of ZNF1 in the last time step
5	Supply air temperature of zone ZSF1	Equipment schedule
6	Supply air temperature of zone ZSF2	Hour of day
7	Whole building electricity consumption in the past hour	Supply air temperature of ZSF1 in the last time step
8	Supply air temperature of zone ZNF1 in the last time step	Zone temperature of ZSF1 in the last time step
9	Outdoor air dry-bulb temperature in the last time step	Diffuse solar radiation in the last time step
10	Zone temperature of zone ZNF1	Zone temperature of ZSF2 in the last time step
11	Hour of day	
12	Weekday and weekend indicator	
13	Hour of day in the last time step	
14	Weekday and weekend indicator in the last time step	
15	Diffuse solar radiation in the last time step	
16	Direct solar radiation in the last time step	
17	Zone temperature of zone ZNF2 in the last time step	
18	Zone temperature of zone ZNF2	
19	Occupancy schedule in the last time step	
20	Zone temperature of zone ZNF2	
21	Zone temperature of zone ZSF2	
22	Zone temperature of zone ZSF1	
23	Outdoor air wet-bulb temperature in the last time step	
24	Occupancy schedule	

For the CPU time of active learning to generate setpoints for the next time step, since the development of Model 1 and Model 3 do not include an active learning process, Model 1 and Model 3 do not cost any CPU time in active learning. For Model 0 and Model 2 with same active learning process, they have higher computational burden than Model 3 due to the higher complexity of MARS calculation, but both are safely within the computation time limit of 1-hour to be less than real-time control interval. For the CPU time of offline feature selection, since the Model 1 and Model 2 do not conduct feature selection, they do not cost any CPU time in feature selection. Since Model 0 conduct feature selection after active learning, the data size for feature selection is larger than Model 3 without active learning, and naturally, it costs more time than Model 3 under the same feature selection framework. CPU time of feature selection in Model 4 is lower than Model 0 because the decision tree has much lower computation cost than MARS.

Table 8. Results of comparison studies for the integrated framework of active learning and feature selection in the reference office building

Small Building	Test Scenario 1**	Test Scenario 2	Test Scenario 3		

	RMSE 10^8	R^2	RMSE 10^6	R^2	RMSE 10^6	R^2	**Active Learning CPU Time (s/step)**	**Feature Selection Time (s)**
Model 0*	8.0	0.94	9.9	0.89	11.5	0.88	57.3	7445
Model 1	9.6	0.92	16.2	0.70	16.1	0.77	0	0
Model 2	9.2	0.93	10.7	0.87	11.8	0.87	57.3	0
Model 3	10.5	0.90	13.5	0.82	14.3	0.80	0	6837
Model 4	7.2	0.96	11.3	0.86	12.2	0.86	47.0	4992
Model 5	8.0	0.94	10.6	0.87	12.5	0.86	59.9	7070

*Model 0: developed framework; Model 1: normal operation; Model 2: only active learning; Model 3: only feature selection; Model 4: decision tree; Model 5: first feature selection then active learning.
** RMSE: root mean squared error; R^2: r-squared.

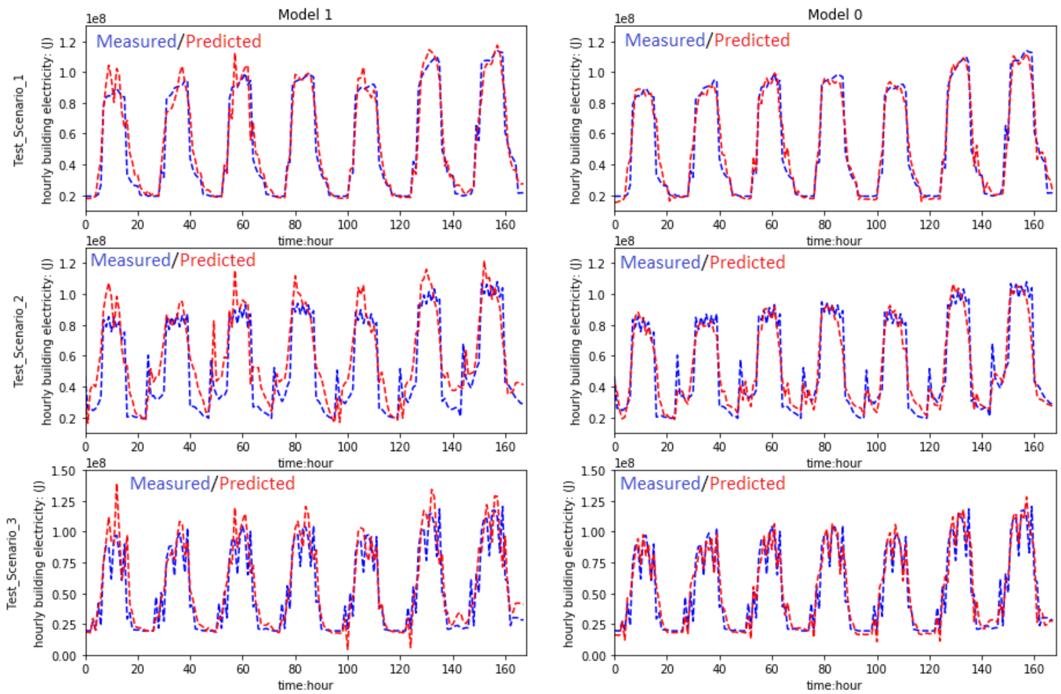

Figure 12. Results of comparison studies (Model 0 and Model 1) in the small reference office building (blue: measured; red: forecasted)

5.2 DPC Controller Evaluation

To further demonstrate the impact of the developed integrated framework on improving the performance of DPC controllers, a comparison study is conducted to compare DPC performance between two controllers. Controller 0 is the DPC controller whose DBEM is built by Model 0 (the model using developed integrated framework of active learning and feature selection). Controller 1 is the DPC controller whose DBEM is built by Model 1 (the model using normal operation data without active learning and feature selection techniques).

Four evaluation metrics (total electricity cost, peak load shifting capability, thermal comfort penalty, and computation time) for the comparison study are introduced in Section 4.3.3.

As shown in Table 9, Controller 0 outperforms Controller 1 in all metrics except zone temperature penalty. Controller 0 is more energy efficient than Controller 1 because of the lower total electricity cost and better peak load shifting capability. Average CPU time is also shorter in Controller 0 because the feature selection process in the developed framework has greatly reduced the model complexity. Though thermal comfort penalty is worse in Controller 0, the zone temperature setpoint has been constrained under 28℃ in the reference office building, which is within the comfort range of 22 - 28℃.

Table 9. DPC performance metrics of Controller 0 and Controller 1 in the virtual testbeds

	Total Electricity Cost ($)	**Peak Load Shifting Capability (kWh)**	**Thermal Comfort Penalty (C = 5E4)**	**Average CPU Time per Step (s/step)**
Controller 0	511.1	29.04	4.76E7	55.0
Controller 1	520.5	29.69	2.89E7	58.0

Next, Controller 0 and Controller 1 are compared in detail. As shown in Figure 13, the key disagreement of the control sequence (zone temperature setpoint) happens when tariff reaches 0.5. The reason that Controller 1 still generate a 27℃ zone temperature setpoint during that period is that Model 1 overestimates the energy consumption (4.2% higher than the energy forecasted by Model 0). Hence Controller 1 does not choose to sacrifice thermal comfort to further raise the temperature setpoint to 28℃ like what Controller 0 does. The key reason for the overestimation of building energy consumption in Model 1 is that it has not been trained on setpoints other than 24 and 26.7℃. Model 1 extrapolates to estimate the energy consumption at 28℃, which is a source of model inaccuracy. Model 0 with active learning has already been trained on different setpoints and have a more accurate estimation of building energy consumption, leading to better performance in DPC.

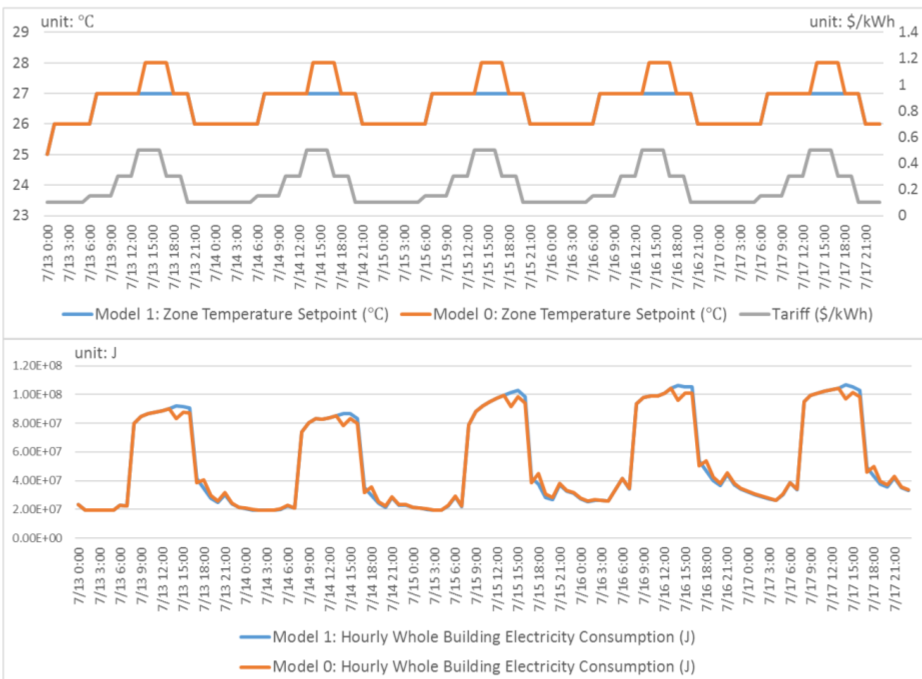

Figure 13. Setpoints and whole building electricity consumption of reference office building operated under Controller 0 (using Model 0) and Controller 1 (using Model 1)

6. Conclusions and Future Work

In this paper, we develop an integrated active learning and feature selection framework to systematically defy data bias and high data dimensionality of building data in the process of DBEM in the context of whole building DPC application.

We first demonstrate that the developed integrated framework of active learning and feature selection can improve the accuracy and extendibility of DBEM under various testing scenarios including pre-cooling and demand response. We also demonstrate that the developed integrated framework can improve the performance of a DPC controller. Comparison studies using the virtual testbed based on reference office building has shown improved performance of DPC controller built by the developed framework in terms of total electricity cost, peak load shifting capability, and average CPU time.

The developed framework can be used as a guidance for the development and modeling of building DPC applications. It can also guide the streamline and workflow of a standardized and automated building DPC modeling process, so that DPC can be easily and automatically implemented in the real world, thus casting a larger impact on the industry.

Although the developed framework provides an automatic workflow to improve DBEM, a small part of the framework relies on modelers' domain knowledge, which could be an obstacle for full automation of the whole modeling process. The major domain knowledge needed in the framework exists in the feature extraction process to decide how much time-lag is needed for time-lag variables, which is currently decided by on-site stabilization tests. Moreover, the integrated framework and whole building DPC are not applied to a real building testbed. In the future, it is desirable to apply the developed framework and DPC to a real building, where many practical problems need to be addressed in the real-building applications.

Acknowledgements

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Building Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

References

1. EIA, Monthly energy review – June 2020. US Energy Information Administration, 2020.
2. Oldewurtel, F., et al., Use of model predictive control and weather forecasts for energy efficient building climate control. Energy and Buildings, 2012. 45: p. 15-27.
3. Braun, J.E., Reducing energy costs and peak electrical demand through optimal control of building thermal storage. ASHRAE transactions, 1990. 96(2): p. 876-888.
4. Dounis, A.I. and C. Caraiscos, Advanced control systems engineering for energy and comfort management in a building environment—A review. Renewable and Sustainable Energy Reviews, 2009. 13(6): p. 1246-1261.
5. Oldewurtel, F., et al. Increasing energy efficiency in building climate control using weather forecasts and model predictive control. in Clima-RHEVA World Congress. 2010.
6. Henze, G.P., et al., Impact of adaptive comfort criteria and heat waves on optimal building thermal mass control. Energy and Buildings, 2007. 39(2): p. 221-235.
7. Donaisky, E., et al. PMV-based predictive algorithms for controlling thermal comfort in building plants. in Control Applications, 2007. CCA 2007. IEEE International Conference on. 2007. IEEE.
8. Cigler, J., et al., Optimization of predicted mean vote index within model predictive control framework: Computationally tractable solution. Energy and Buildings, 2012. 52: p. 39-49.
9. Rawlings, J.B. and D.Q. Mayne, Model predictive control: Theory and design. 2009.
10. Serale, G., et al., Model predictive control (MPC) for enhancing building and HVAC system energy efficiency: Problem formulation, applications and opportunities. Energies, 2018. 11(3): p. 631.

11. Li, X. and J. Wen, Review of building energy modeling for control and operation. Renewable and Sustainable Energy Reviews, 2014. 37: p. 517-537.
12. Zhang, L., M. Alahmad, and J. Wen, Comparison of time-frequency-analysis techniques applied in building energy data noise cancellation for building load forecasting: A real-building case study. Energy and Buildings, 2021. 231: p. 110592.
13. Li, Y., et al., Grey-box modeling and application for building energy simulations-A critical review. Renewable and Sustainable Energy Reviews, 2021. 146: p. 111174.
14. Jain, A., M. Behl, and R. Mangharam. Data Predictive Control for building energy management. in American Control Conference (ACC), 2017. 2017. IEEE.
15. Zhang, L., A Pattern-Recognition-Based Ensemble Data Imputation Framework for Sensors from Building Energy Systems. Sensors, 2020. 20(20): p. 5947.
16. Institute, N.C., Dictionary of cancer terms. 2007, National Institutes of Health Bethesda^ eMD MD.
17. Zhang, L. and J. Wen, Apply Active Learning in Short-term Data-driven Building Energy Modeling. 2018.
18. Zhang, L. and J. Wen, Active learning strategy for high fidelity short-term data-driven building energy forecasting. Energy and Buildings, 2021. 244(C).
19. Settles, B., Active learning literature survey. 2010. Computer Sciences Technical Report, 2014. 1648.
20. Guyon, I. and A. Elisseeff, An introduction to variable and feature selection. Journal of machine learning research, 2003. 3(Mar): p. 1157-1182.
21. Lewis, D.D. and W.A. Gale. A sequential algorithm for training text classifiers. in Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval. 1994. Springer-Verlag New York, Inc.
22. Seung, H.S., M. Opper, and H. Sompolinsky. Query by committee. in Proceedings of the fifth annual workshop on Computational learning theory. 1992. ACM.
23. Settles, B., M. Craven, and S. Ray. Multiple-instance active learning. in Advances in neural information processing systems. 2008.
24. Roy, N. and A. McCallum, Toward optimal active learning through monte carlo estimation of error reduction. ICML, Williamstown, 2001: p. 441-448.
25. Zhang, L., et al., A review of machine learning in building load prediction. Applied Energy, 2021. 285: p. 116452.
26. Dodier, R.H. and G.P. Henze, Statistical analysis of neural networks as applied to building energy prediction. Journal of solar energy engineering, 2004. 126(1): p. 592-600.
27. Zhao, H.-x. and F. Magoulès, Feature selection for predicting building energy consumption based on statistical learning method. Journal of Algorithms and Computational Technology, 2012. 6(1): p. 59-77.
28. Chae, Y.T., et al., Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. Energy and Buildings, 2016. 111: p. 184-194.
29. Tsanas, A. and A. Xifara, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy and Buildings, 2012. 49: p. 560-567.
30. Liu, Y., W. Wang, and N. Ghadimi, Electricity load forecasting by an improved forecast engine for building level consumers. Energy, 2017. 139: p. 18-30.
31. Chitsaz, H., et al., Short-term electricity load forecasting of buildings in microgrids. Energy and Buildings, 2015. 99: p. 50-60.
32. Sholahudin, S. and H. Han, Simplified dynamic neural network model to predict heating load of a building using Taguchi method. Energy, 2016. 115: p. 1672-1678.
33. Sala-Cardoso, E., et al., Activity-aware HVAC power demand forecasting. Energy and Buildings, 2018. 170: p. 15-24.
34. Ghofrani, M., et al., A hybrid short-term load forecasting with a new input selection framework. Energy, 2015. 81: p. 777-786.

35. Ceperic, E., V. Ceperic, and A. Baric, A strategy for short-term load forecasting by support vector regression machines. IEEE Transactions on Power Systems, 2013. 28(4): p. 4356-4364.
36. Fan, C., F. Xiao, and S. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Applied Energy, 2014. 127: p. 1-10.
37. Kolter, J.Z. and J. Ferreira Jr, A large-scale study on predicting and contextualizing building energy usage. 2011.
38. Jovanović, R.Ž., A.A. Sretenović, and B.D. Živković, Ensemble of various neural networks for prediction of heating energy consumption. Energy and Buildings, 2015. 94: p. 189-199.
39. Massana, J., et al., Short-term load forecasting in a non-residential building contrasting models and attributes. Energy and Buildings, 2015. 92: p. 322-330.
40. Son, H. and C. Kim, Forecasting short-term electricity demand in residential sector based on support vector regression and fuzzy-rough feature selection with particle swarm optimization. Procedia engineering, 2015. 118: p. 1162-1168.
41. Kusiak, A., M. Li, and Z. Zhang, A data-driven approach for steam load prediction in buildings. Applied Energy, 2010. 87(3): p. 925-933.
42. Yu, Z., et al., A decision tree method for building energy demand modeling. Energy and Buildings, 2010. 42(10): p. 1637-1646.
43. Ahmad, M.W., M. Mourshed, and Y. Rezgui, Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings, 2017. 147: p. 77-89.
44. Hsiao, Y.-H., Household electricity demand forecast based on context information and user daily schedule analysis from meter data. IEEE Transactions on Industrial Informatics, 2014. 11(1): p. 33-43.
45. Jurado, S., et al., Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques. Energy, 2015. 86: p. 276-291.
46. Zhang, L. and J. Wen, A systematic feature selection procedure for short-term data-driven building energy forecasting model development. Energy and Buildings, 2019. 183: p. 428-442.
47. Mohammadi, M., et al., Small-scale building load forecast based on hybrid forecast engine. Neural Processing Letters, 2018. 48(1): p. 329-351.
48. Li, C., et al., Joint Active Learning with Feature Selection via CUR Matrix Decomposition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.
49. Liu, H., H. Motoda, and L. Yu, A selective sampling approach to active feature selection. Artificial Intelligence, 2004. 159(1-2): p. 49-74.
50. Liu, H., H. Motoda, and L. Yu. Feature selection with selective sampling. in ICML. 2002.
51. Veeramachaneni, S. and P. Avesani. Active sampling for feature selection. in null. 2003. IEEE.
52. Mahoney, M.W. and P.J.P.o.t.N.A.o.S. Drineas, CUR matrix decompositions for improved data analysis. 2009. 106(3): p. 697-702.
53. Deru, M., et al., US Department of Energy commercial reference building models of the national building stock. 2011.
54. Crawley, D.B., et al., EnergyPlus: creating a new-generation building energy simulation program. Energy and buildings, 2001. 33(4): p. 319-331.
55. Wetter, M., Co-simulation of building energy and control systems with the Building Controls Virtual Test Bed. Journal of Building Performance Simulation, 2011. 4(3): p. 185-203.
56. Afram, A. and F. Janabi-Sharifi, Theory and applications of HVAC control systems–A review of model predictive control (MPC). Building and Environment, 2014. 72: p. 343-355.
57. Martinčević, A., M. Vašak, and V. Lešić. Model predictive control for energy-saving and comfortable temperature control in buildings. in Control and Automation (MED), 2016 24th Mediterranean Conference on. 2016. IEEE.
58. Basalaev, A.A., T.A. Barbasova, and D.A. Shnayder, Simulation study on supply temperature optimization of university campus heating system. Procedia engineering, 2015. 129: p. 587-594.

59. Bryan, K. and Y. Shibberu, Penalty functions and constrained optimization. Dept. of Mathematics, Rose-Hulman Institute of Technology. http:// www. rosehulman. edu/~ bryan/lottamath/penalty. pdf, 2005.
60. Lee, Y.M., R. Horesh, and L. Liberti, Optimal HVAC control as demand response with on-site energy storage and generation system. Energy Procedia, 2015. 78: p. 2106-2111.
61. Braun, J.E., Load control using building thermal mass. Journal of solar energy engineering, 2003. 125(3): p. 292-301.