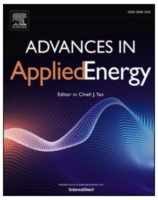
Advances in Applied Energy 20 (2025) 100246

Contents lists available at ScienceDirect

Advances in Applied Energy

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



Prediction-based control of energy storage systems using dynamic accuracy weighting

Xiao Wang [a], Xue Liu [b], Xuyuan Kang [a,*], Fu Xiao [c], Da Yan [a,*]

[a] Building Energy Research Center, School of Architecture, Tsinghua University, Beijing 100084, China
[b] School of Architecture, Southwest Jiaotong University, Chengdu 611756, China
[c] Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Keywords:
Energy storage system
Prediction-based control
Prediction accuracy metric
Artificial intelligence
Domain knowledge

ABSTRACT

Integrating domain knowledge into artificial intelligence models is increasingly recognized as essential for improving energy storage system control based on load predictions. Commonly used accuracy metrics for load prediction models, such as mean absolute percentage error, coefficient of variation of mean absolute error, and coefficient of variation of root mean squared error, are not monotonically correlated with final control performance; in other words, the model with the highest prediction accuracy does not necessarily yield optimal control outcomes. This study introduces a dynamically weighted error metric, which incorporates the attributes of energy storage systems and the temporal dynamics of prediction-based control by leveraging domain knowledge from heating, ventilation, and air conditioning systems. The proposed dynamically weighted error metric enhanced the selection of load prediction models, and these models reduced the operating cost of six energy storage systems by up to 6.5 % compared to those using traditional prediction accuracy metrics. The scalability of dynamically weighted error metric was further validated across 10 energy storage capacities and 18 Time-of-Use tariffs in the six building cases, achieving 93.9 %–97.2 % of the ideal cost reductions and outperforming traditional metrics (86.4 %–95.4 %). The applicability of dynamically weighted error metric to common energy storage systems is discussed and confirmed. Additionally, a web-based tool was developed to facilitate dynamically weighted error calculation in practical applications. This study demonstrates that incorporating domain knowledge through dynamic accuracy weighting evidently enhances the whole-process performance of artificial intelligence in energy storage system control.

1. Introduction

Energy storage systems are important for buildings to save energy [1], reduce operation cost [2], and mitigate carbon emission [3]. For example, ice-based thermal energy storage (ITES) systems are effective in cooling load shifting in buildings [8]. Under the incentives of the Time-of-Use (TOU) tariff and carbon emission factors provided by the grid, buildings can use energy storage systems to shift loads [9], reduce operation costs [10], and support the decarbonization of the grid [11]. Load prediction-based control has been widely investigated [4] to support operation controls of electrical [5], heating [6], and cooling [7] energy storage systems.

Accurate day-ahead prediction of the building energy load is important in the operation control of energy storage systems [12]. TOU tariffs [13] usually present a regular variation trend during the day [14]. Therefore, accurate day-ahead hourly [15] or sub-hourly [16] load prediction can help better match energy demand with TOU tariff. The operation cost of the energy storage system can be further reduce with accurate load predictions [17]. Existing studies on load prediction-based control for energy storage systems [18] have generally followed a similar flow. According to this flow, the building energy load was predicted using input features such as weather forecasts [19], occupancy predictions [20], and historical load patterns [21]. The predicted load was then used for the control optimization of energy storage systems and determination of control actions such as the operation modes of energy storage, energy release, and direct energy supply [22]. Thus, the accuracy of the load prediction significantly influenced the control performance of an energy storage system [23]. Numerous studies [24] have investigated the load predictions in buildings [25]. In comparison to statistical methods, such as exponential smoothing [26] and

---

* Corresponding authors at: Room 201, Building Energy Research Center, School of Architecture, Tsinghua University, Beijing 100084, China.
E-mail addresses: kangxuyuan@tsinghua.edu.cn (X. Kang), yanda@tsinghua.edu.cn (D. Yan).

https://doi.org/10.1016/j.adapen.2025.100246
Received 22 September 2025; Received in revised form 28 September 2025; Accepted 28 September 2025

autoregressive methods [27], artificial intelligence (AI)-based methods have become mainstream for load prediction owing to their accuracy [28]. The frequently investigated AI-based load prediction methods include nonlinear regression [29], neural networks [30], decision trees [31], and ensemble learning [32]. These methods demonstrated the ability of AI to improve the accuracy of load prediction.

However, recent studies have reached a bottleneck in improving the accuracy of AI-based load predictions. Zheng et al. [33] used a sparrow search algorithm to enhance a combined convolutional neural network and support vector machine model; the mean absolute percentage error (MAPE) was only reduced by 1.06 %. Li et al. [34] investigated a transformer neural network for cooling load prediction, which only improved the R squared (R^2) by 0.02 compared with an ensemble learning model of decision trees and neural networks. Enhancing AI with domain knowledge is an emerging approach in heating, ventilation, and air conditioning (HVAC) engineering field [35]. Physics-informed learning [36] is a popular approach for combining AI with HVAC domain knowledge [37]. The applications of these models include room temperature prediction [38], chiller energy performance prediction [39], and cooling load prediction [40]. Physics-informed learning improves the robustness and scalability of AI models by adding physics constrains to AI models [41]. However, it brings limited improvements in prediction accuracy [42]. Chen et al. [43] integrated a resistance-capacitance model into a neural network; in comparison to conventional neural networks, the improvement in prediction accuracy was only between 1.2 % and 2.0 % in the coefficient of variation of root mean squared error (CV-RMSE). After decades of research, the AI-based load prediction models have reached a high level of accuracy. Further improving AI model accuracy as measured by conventional metrics does not necessarily lead to significant enhancements in control performance.

An alternative way to further enhance AI-based control performance is to leverage domain knowledge in selecting appropriate AI models for downstream applications. In the context of load prediction-based energy storage system control, the most accurate load prediction model was selected for the downstream control optimization of the energy storage system. Therefore, a metric evaluating the accuracy of load prediction is critical to the application performance of AI-based load prediction models for energy storage system control. However, a discrepancy exists between prediction accuracy and control performance, as Fig. 1 shows. Our previous study [44] found that the MAPE, coefficient of variation of mean absolute error (CV-MAE), and CV-RMSE, might not properly select load prediction models, causing a gap of 3.3 % in control performance by cost reduction. This means an “accurate” prediction model under the conventional accuracy metrics does not necessarily result in a “good” control performance of energy storage systems. The gap between the commonly used prediction accuracy metrics and control performance metrics weakens the whole-process performance of load prediction-based control.

To bridge the gap, application-oriented prediction accuracy metrics should be designed. Our previous study [44] reported that a good prediction accuracy metric exhibited good monotonicity and certainty. In other words, a higher prediction accuracy may result in better control performance. Under a certain level of prediction accuracy, the control performance is certain. The Spearman correlation coefficient can be used to quantitatively evaluate the goodness of a prediction accuracy metric. The evaluation approach for the goodness of prediction accuracy

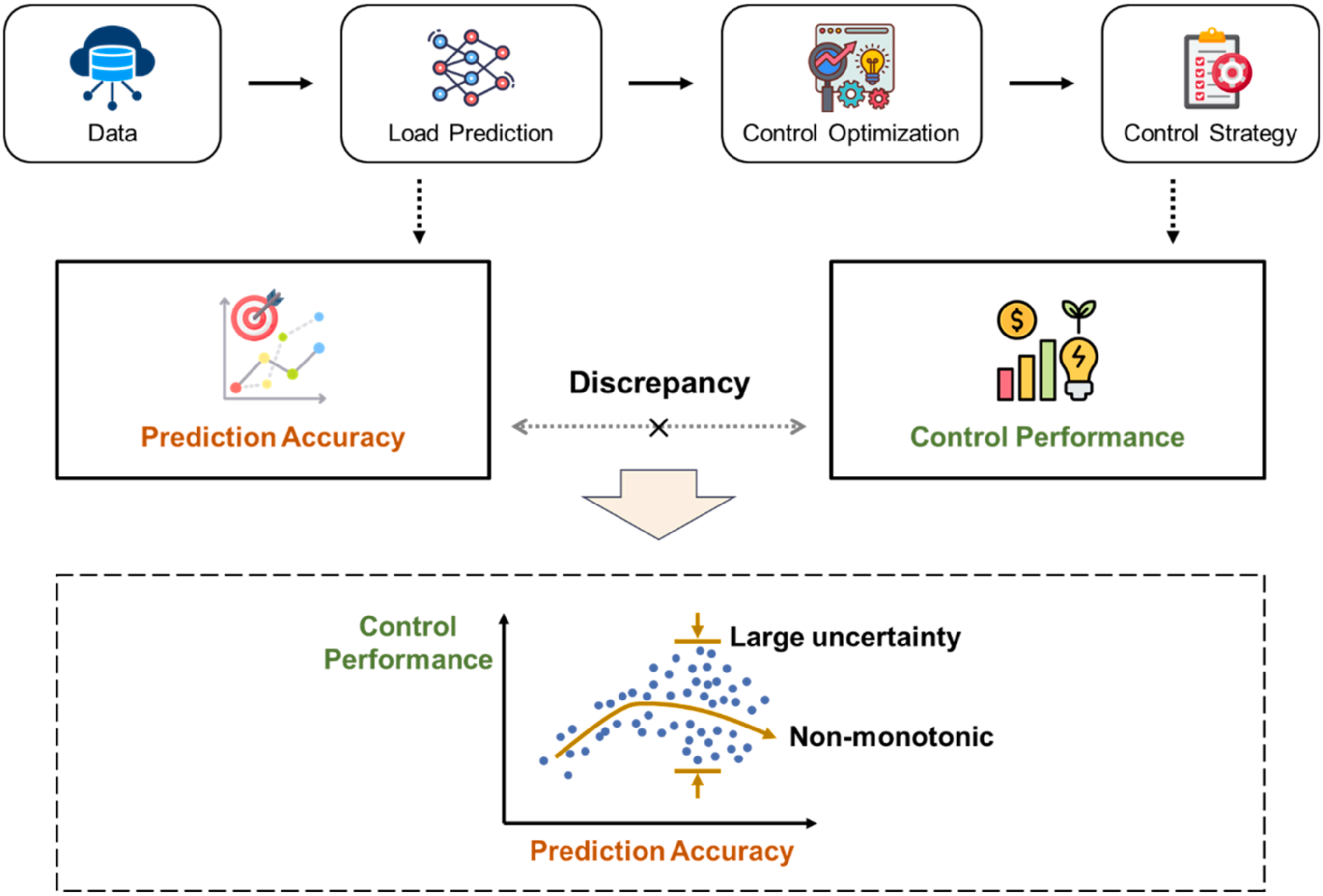


Fig. 1. The discrepancy between prediction accuracy and control performance in load prediction-based control of energy storage systems. This discrepancy is reflected in two aspects: (1) the control performance is not monotonically increased with the increase of prediction accuracy; and (2) the control performance under a certain prediction accuracy is uncertain. This discrepancy greatly limits the whole-process performance of load prediction-based control of energy storage systems.

metric supported the proposal of application-oriented prediction accuracy metrics. This incorporation of HVAC domain knowledge into prediction accuracy metrics can improve prediction model selection and enhance the overall performance of prediction-based control.

Prediction accuracy metrics are commonly used to evaluate AI-based load prediction models during training [45], testing [46], and validation [47]. Each prediction accuracy metric has different characteristics. The mean absolute error (MAE) offers a balanced measure by indicating the average deviation of predictions [48] and is widely used to evaluate the prediction models [49]. The mean bias error (MBE) evaluates the systematic deviation of predictions [50] and is typically used in training evaluations [51]. The mean squared error (MSE) emphasizes larger deviations of predictions [52] and is usually used for model training [53]. The root mean squared error (RMSE) provides a measure of deviation with a clear physical significance [54], which is used as widely as MAE [55]. The MAPE is sensitive to small deviations in the dimensionless evaluation metrics of the prediction accuracy [56]. By integrating the coefficient of variation (CV) into the MAE and RMSE, CV-MAE [57] and CV-RMSE [58] can be used to assess the relative error. MAPE, CV-MAE, and CV-RMSE are widely used metrics for the cross-case evaluation of prediction accuracy [59]. R^2 is frequently used in model fitting [60] to quantify the overall impact of both systematic and random errors [61]. The existing prediction accuracy metrics are summarized in Table A3.

These prediction accuracy metrics are scalable for different cases of load prediction [62] but inflexible to the variation of control conditions. This is because they are statistical metrics with uniform weights of prediction error at different time. In the load prediction-based control of energy storage system, TOU tariffs may affect the importance of accurate load predictions at different times. Owing to the limited energy storage capacity, accurate load prediction when the cumulative load is close to the energy storage capacity may be more important. The load prediction accuracy at these key times may significantly influence the downstream control performance; however, it is not reflected in the existing prediction accuracy metrics, which causes a gap between the prediction accuracy and control performance metrics in prediction-based energy storage system control [44]. Therefore, proposing an application-oriented prediction accuracy metric using a dynamic accuracy weighting method is possible to bridge the gap between control performance and prediction accuracy metrics in terms of application-oriented prediction accuracy.

Few studies have proposed application-oriented prediction accuracy metrics for load prediction in prediction-based energy storage system control, whereas several studies have discussed the gap between the prediction accuracy and control performance metrics. Kazmi et al. [63] found that more accurate prediction models did not always yield higher values for downstream tasks. Hong et al. [64] investigated electricity price predictions and found that the use of predictions affected the valuation of accurate predictions. Koponen et al. [65] analyzed different prediction accuracy metrics for electricity load prediction for battery control in renewable energy storage systems. Statistical prediction accuracy metrics are likely to underestimate the increase in costs because of the decrease in prediction accuracy. Haben et al. [66] proposed a new prediction accuracy metric for predicting the electrical energy consumption to support battery control in residential buildings. Because the accurate prediction of the peak load and time was critical in this control application, the proposed metric reduced the penalty in off-peak hours. The results showed that the proposed prediction accuracy metric helped to train better electricity load prediction models with more similar patterns to the measured loads; however, the results were not evaluated for the electricity cost of the system. Our previous study [44] proposed a prediction accuracy metric modified from the CV-RMSE, which was weighted by the TOU tariff; however, the improvement in the cost reduction of the ITES system was not significant compared with the CV-RMSE. An application-oriented prediction accuracy metric should be proposed with HVAC domain knowledge of load prediction-based energy storage system control. The research gaps are summarized as follows.

1. A discrepancy between conventional prediction accuracy metrics and control performance metrics exists, which greatly limits the control performance of energy storage systems.
2. The existing studies have little focus on the whole-process performance from load prediction to energy storage system control.
3. Few studies investigated the dynamic adaptation between load prediction and control optimization in energy storage system controls using HVAC domain knowledge.

To bridge these gaps, this study proposes a dynamic prediction accuracy metric for load prediction-based energy storage system control. HVAC domain knowledge is used to design this metric, which is dynamically adjusted with the TOU tariff, energy storage capacity, and predicted load. The main contributions are as follows.

1. An adaptive method is proposed to dynamically generate the application-oriented prediction accuracy metric to select load prediction models for energy storage system control.
2. The improvement of the proposed metric on the control performance is analyzed in six building cases with ITES systems and compared with commonly used metrics, including MAPE, CV-MAE, and CV-RMSE.
3. The scalability of the proposed metric is analyzed for 10 energy storage capacities and 18 TOU tariffs.

This study aims to enhance the application of AI with HVAC domain knowledge in the whole process of load prediction-based control for energy storage systems. The designed methodology aims to achieve better monotonicity and certainty between load prediction accuracy metrics and control performance metrics. The details of methodology are shown as follows.

2. Methodology

This study proposes a dynamic accuracy weighting method for the load prediction-based control of energy storage systems. The design of the proposed metric is introduced in Section 2.1. The technical approach of evaluating the proposed metric is illustrated in Fig. 2. The proposed dynamically weighted error (DWE) metric is used for the selection of load prediction models for the control optimization of energy storage systems. The proposed DWE was evaluated by comparing with the benchmark statically weighted error metrics, including MAPE, CV-MAE, and CV-RMSE. The evaluation was conducted in simulation of six real buildings and 10 energy storage systems with different capacities under 18 time-of-use tariffs using 180 load prediction models. The control performance was evaluated based on the cost reduction performance of the prediction-based control. A comparative evaluation of the benchmark and proposed prediction accuracy metrics was conducted by analyzing the monotonicity and certainty under the Spearman correlation between the prediction accuracy and control performance metrics. The technical approach of evaluating the proposed DWE aims to demonstrate the scalability of the proposed dynamic accuracy weighting method for enhanced prediction-based energy storage system control.

2.1. Design of an adaptive method for generating the dynamically weighted error (DWE)

The main contribution of this study is an adaptive method to generate the application-oriented metric of load prediction accuracy for the prediction-based energy storage system control for cost-saving purposes under the TOU tariff. The proposed prediction accuracy metric DWE is defined in Eq. (1).

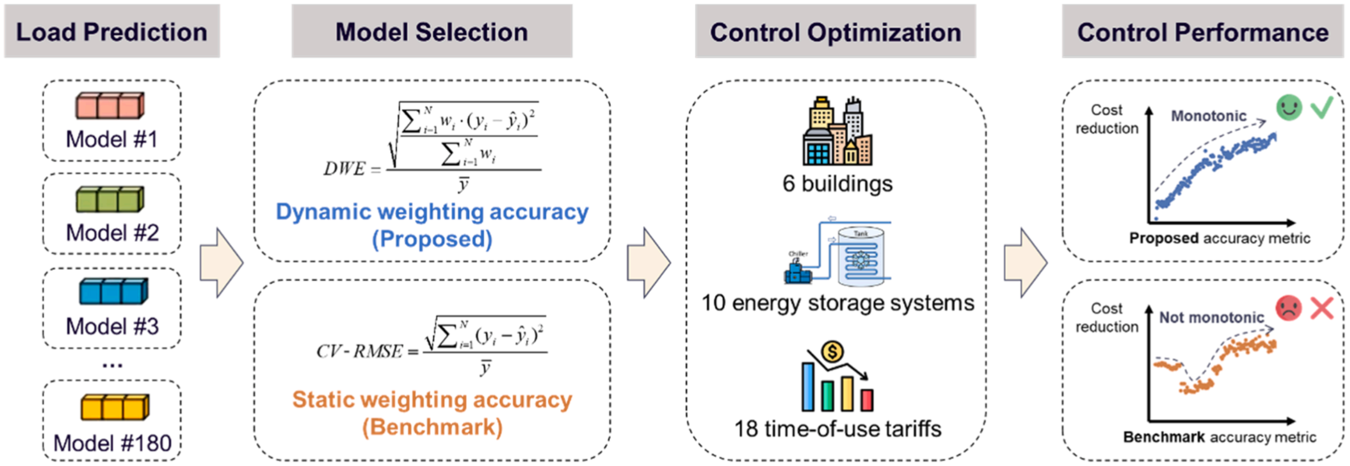


Fig. 2. Technical approach of applying and evaluating the proposed dynamic accuracy weighting method. The proposed dynamic accuracy weighting method is used to generate the dynamically weighted error (DWE) to select load prediction models for enhanced control performance of energy storage systems.

$$DWE = \frac{\sqrt{\frac{\sum_{i=1}^{N} w_i \cdot \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N} w_i}}}{\bar{y}} \tag{1}$$

where y_i and $\hat{y}_i$ are the measured and predicted loads of the i^{th} hour, respectively; $\bar{y}$ is the hourly average of the measured loads of one day, w_i is the weight for the squared error of the i^{th} hour, and N is the number of load samples in one day, which is 24 in this study, representing data in hourly time resolution.

The proposed DWE was modified from the CV-RMSE by adding weights to the squared errors for each hour. The weights of DWE were dynamically "predicted" each day, which was quite different from the existing statically weighted error metrics for prediction accuracy, such as MAPE, CV-MAE, and CV-RMSE. The essence of this design was to consider the varying conditions of the prediction-based energy storage system control for prediction model selection, which were reflected in the dynamically predicted weights of the DWE.

The day-ahead weight prediction is the key to DWE. Because the prediction-based control strategy depends on the energy storage capacity, TOU tariff, and day-ahead predictions of the loads, the weight prediction of the DWE considers the inputs of the energy storage capacity (C_s), electricity prices (P) in TOU tariff, and predicted 24-hour loads (L_p). Because the prediction-based control strategy depends on the controlled energy storage system conditions and temporal process of control optimization, the DWE weight (w) is the Hadamard product (also known as element-wise product) of the system-relevant (w_s) and temporal-relevant (w_t) factors, as shown in Eq. (2).

$$w = w_s \circ w_t \tag{2}$$

The flow chart of generating the accuracy weights of DWE is illustrated in Fig. 3. The dimensions of the variables are also marked in the

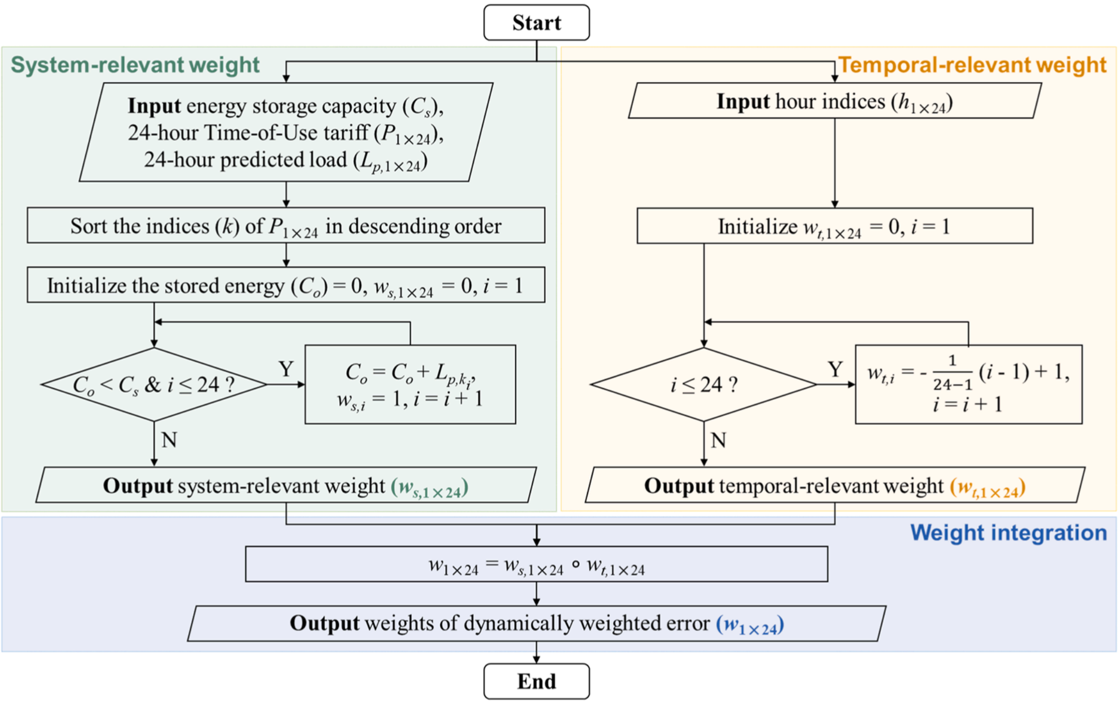


Fig. 3. The flow chart of generating the accuracy weights of dynamically weighted error (DWE). The generation of DWE has three main processes, including dynamic predictions of system-relevant weight, temporal-relevant weight, and weight integration.

figure. In comparison with the static accuracy weighting method, the proposed dynamic accuracy weighting method adjusts the prediction error weights in different hours based on system-relevant and temporal-relevant conditions for a control application-oriented selection of load prediction models. In energy storage system control, prediction accuracy during certain hours may not affect control performance [44]. Thus, the system-relevant factor excludes these hours by considering the predicted load, TOU tariff, and energy storage capacity. After the relevant hours are selected, the temporal-relevant factor gives an order of the importance of these hours. Because the detailed mechanism of the importance decay of accurate prediction over time remains unclear, a linear function is used to reflect the decay intuitively. This approach decouples the two processes of selecting the key hours and ordering the importance of the selected key hours, which formulizes a structure of dynamic accuracy weighting method for DWE. The detailed methods of determining w_s and w_t are described in the following sections.

The system-relevant factor (w_s) reflects the attributes of a controlled energy storage system. As mentioned in [44], the inconsistency between prediction accuracy and control performance metrics in prediction-based energy storage system control is caused by the variations in TOU tariffs. Because the energy storage system stores energy during off-peak hours and releases energy during peak hours to reduce operating costs, the importance of accurate load prediction is influenced by the TOU tariff. The energy storage capacity also influences the importance of accurate load prediction. When the energy storage capacity is too high or too low, the prediction-based control shows less improvement than the baseline control without load prediction [44]. Therefore, the system-relevant weight prediction involves energy storage capacity, TOU tariff, and predicted 24-hour loads. The hourly indices from 1 to 24 are sorted in descending order of a 24-hour TOU tariff. Because there is little room for control optimization during the hours when the load cannot be covered by the energy storage capacity, this helps determine the key hours for the control optimization of energy release. The stored energy is used to satisfy these hours with higher electricity prices until the maximum stored energy, which is equal to the energy storage capacity, is allocated. Therefore, w_s during these hours were set to 1, whereas those during other hours were set to 0. The predicted w_s dynamically indicates the key hours for accurate load prediction for each day.

In addition to the system-relevant factor (w_s), the temporal-relevant factor (w_t) is also considered because the allocation of energy release is also relevant to time. An inaccurate prediction of the load may affect energy allocation. For example, the underprediction of the load may lead to the early use of the stored energy, and the subsequent control strategies are unexpected. Therefore, accurate load prediction in the early hours is more important than that in the later hours of the day [44]. To reflect this attribute, w_t is designed and calculated using Eq. (3).

$$w_{t,i} = -\frac{1}{24-1}(i-1)+1 \quad i \in [1,24] \tag{3}$$

where i denotes the i^{th} hour of the day. This equation defines a linear descending function of w_t from the first to the last hour of the day. The w_t at 0:00–1:00 is the highest at 1 because it is the first hour of a day, and the inaccurate prediction of the load may influence every hour of the day. Instead, w_t at 23:00–24:00 is 0 because the inaccurate load prediction of the last hour of the day will not influence any other hours of the day. The w_t reflects the varying importance of prediction-based control strategies and accurate load predictions of temporal orders.

An intuitive example of calculating the weights of DWE is shown in Appendix B. In summary, the designed DWE can dynamically reflect the importance difference of accurate load predictions for different hours according to the conditions of the energy storage system, which is oriented toward the prediction model selection for prediction-based energy storage system control for purposes such as cost-saving under the TOU tariff.

2.2. Definition of index for comparing control performance

The proposed DWE metric was compared with MAPE, CV-MAE, and CV-RMSE. These four prediction accuracy metrics were used to select load prediction models for the load prediction-based control of energy storage systems. This study used 180 AI-based load prediction models, as described in Appendix C. The energy storage systems and energy loads from six buildings are shown in Appendix D. Each energy storage system was controlled using two strategies: a fixed schedule and a prediction-based control strategy. The control strategies are described in Appendix E. The improvement in control performance using AI-based load prediction was evaluated using the cost reduction rate (*CRR*) of the prediction-based control strategy from a fixed schedule, as shown in Eq. (4).

$$CRR = \frac{C_F - C_P}{C_F} \times 100\% \tag{4}$$

where C_P and C_F are the electricity costs of the energy storage system under the prediction-based control strategy and the fixed schedule, respectively.

In prediction-based control, a more accurate prediction model is expected to achieve a better control performance. For an effective control-oriented prediction accuracy metric, the selected accurate prediction model should have a good performance. That is, for an effective control-oriented prediction accuracy metric, the consistence of monotonicity between a prediction accuracy metric and a control performance metric is good. Because the Spearman correlation coefficient is a suitable measure for assessing the consistence of monotonicity between two variables, it is selected to evaluate the effectiveness of control-oriented prediction accuracy metrics. The Spearman correlation coefficient ranges from -1 to +1, where a value of +1 indicates a perfect positive monotonic relationship, 0 indicates a nonmonotonic relationship, and -1 indicates a perfect negative monotonic relationship. In this study, the load prediction accuracy metrics were evaluated using the Spearman correlation coefficient between the load prediction accuracy and cost reduction rate of prediction-based control in energy storage systems. In addition, under a certain level of prediction accuracy, the uncertainty of the cost reduction rate was expected to be low. Therefore, under a better prediction accuracy metric, the distribution of cost reduction rate was more compact. This study evaluated the certainty of the prediction accuracy metric using the distribution of the top ten cost reduction rates of the load prediction models selected by the prediction accuracy metric. The Spearman correlation coefficient and distribution of the top ten cost reduction rates were used to comprehensively evaluate the prediction accuracy metrics of monotonicity and certainty.

2.3. Sensitivity analysis to examine the scalability of DWE

The scalability of the proposed DWE metric is critical for its applicability. In the load prediction-based control of energy storage systems, energy storage capacity and TOU tariff are the main factors limiting control performance. To validate the scalability of the proposed DWE metric, a sensitivity analysis was conducted across cases of 10 energy storage capacities and 18 TOU tariffs. In the sensitivity analysis, an ideal scenario was used to evaluate the effectiveness of prediction accuracy metric. The ideal scenario consisted of the load prediction models with actually the highest cost reduction rates, representing the limit of the cost reduction rate using a perfect prediction accuracy metric to select load prediction models for downstream controls. The approach cost reduction (*ACR*) to ideal scenario is defined in Eq. (5).

$$ACR = \frac{C_F - C_{P,m}}{C_F - C_{P,ideal}} \times 100\% = \frac{CRR_m}{CRR_{ideal}} \tag{5}$$

where subscript m represents a certain scenario using prediction accuracy metric such as MAPE, CV-MAE, CV-RMSE, and DWE; subscript *ideal* represents the ideal scenario.

In each of the six buildings, the energy storage capacity varied from 10 % to 100 % in increments of 10 % of the maximal daily total energy load. The 18 TOU tariffs are presented in Appendix F, including nine with fixed structures and nine with hourly dynamic structures. Statistics were conducted on the improvements in the cost reduction rate using DWE compared with those using MAPE, CV-MAE, and CV-RMSE.

3. Results

The proposed DWE is analyzed in cases of cooling load prediction to support the prediction-based control of ITES systems. The performance of the proposed DWE metric was compared with those of the MAPE, CV-MAE, and CV-RMSE for six buildings with ITES systems. The results are presented in the following sections.

3.1. Analysis across multiple buildings

The results of the MAPE, CV-MAE, CV-RMSE, and proposed DWE are compared for the six building cases in Fig. A1. The scatter in this figure represents the cooling load prediction model. Each cooling load prediction model has a prediction accuracy under a certain prediction accuracy metric. The prediction accuracy ranking is in descending order of the prediction error of the 180 tested prediction models, where rank one represents the most accurate cooling load prediction model under a certain prediction accuracy metric. Each cooling load prediction model generates predicted cooling loads for the prediction-based ITES control to achieve a simulated cost reduction rate. In cooling load prediction-based ITES control, a more accurate cooling load prediction may have a higher cost reduction rate. A good prediction accuracy metric may reflect this result well, which means that the scatter of the prediction accuracy rank and cost reduction rate are compact to a monotonic trend line. The results show that in the six cases, DWE has consistently better monotonicity than MAPE, CV-MAE, and CV-RMSE. The most accurate cooling load prediction models under DWE can result in the highest cost reduction rate. Additionally, the certainty of the DWE was better than that of the MAPE, CV-MAE, and CV-RMSE in the six buildings. Under a certain prediction accuracy, the scatter distribution of cost reduction rates is compact, which means that DWE has less uncertainty than MAPE, CV-MAE, and CV-RMSE in evaluating the application performance of the cooling load prediction models. For example, in building #1, under the proposed DWE metric, the most accurate prediction model, which ranked the first in prediction accuracy, had a significantly higher cost reduction rate than those of the other prediction models. Under MAPE, CV-MAE, and CV-RMSE, the most accurate prediction models had moderate cost reduction rates among all the 180 prediction models. Consistently in other buildings, under DWE, the most accurate prediction model has obviously higher cost reduction rates than other prediction models. The proposed DWE can more properly select cooling load prediction models for control performance than MAPE, CV-MAE, and CV-RMSE. Evidently, the proposed DWE can evaluate the cooling load prediction models on the prediction-based control performance of ITES systems better than the MAPE, CV-MAE, and CV-RMSE. The proposed DWE was scalable in six buildings and 180 well-tuned cooling load prediction models.

Quantitatively, the performances of MAPE, CV-MAE, CV-RMSE, and DWE were evaluated using the Spearman correlation coefficients between the prediction accuracy and cost reduction rate, as shown in Table 1. In the cooling load prediction-based control of ITES systems, a lower cooling load prediction error may lead to a higher cost reduction rate. Thus, an applicable cooling load prediction accuracy metric should have a negative Spearman correlation coefficient with the cost reduction rate. A higher absolute value of the negative Spearman correlation coefficient indicates better prediction accuracy. In Table 1, the four prediction accuracy metrics in the six cases have negative Spearman correlation coefficients, which means that they are all applicable for evaluating the cooling load prediction model for prediction-based ITES control. In all six cases, MAPE had the lowest absolute values of negative Spearman correlation coefficients compared to the other three metrics, making it the least applicable prediction accuracy metric. In most cases, the proposed DWE had the highest absolute values of negative Spearman correlation coefficients compared with MAPE, CV-MAE, and CV-RMSE. This result indicates that the proposed DWE has a better monotonicity of prediction accuracy and cost reduction rate than MAPE, CV-MAE, and CV-RMSE. The proposed DWE outperformed the MAPE, CV-MAE, and CV-RMSE in evaluating the performance of the cooling load prediction model for prediction-based ITES control.

Furthermore, the effectiveness of the DWE was evaluated by selecting the top ten accurate cooling load prediction models for prediction-based ITES control. Fig. 4 shows the cost reduction rate distributions of the selected top ten accurate cooling load prediction models under different prediction accuracy metrics for the six building cases. The ideal scenario means the distributions of the actual top ten cost reduction rates among the 180 cooling load prediction models. A perfect prediction accuracy metric may have an identical distribution of cost reduction rate with the ideal scenario. In the six buildings, the means and medians of the cost reduction rates under DWE are all higher than those under MAPE, CV-MAE, and CV-RMSE and closest to the ideal scenario. The ranges between the quartiles of cost reduction rates under DWE are within 0.6 %, which is reliable for cost reduction applications of cooling load prediction-based ITES control. The cost reduction rates under traditional prediction accuracy metrics such as MAPE, CV-MAE, and CV-RMSE were in the range of 5.6 %–14.2 % on average. In comparison with the MAPE, CV-MAE, and CV-RMSE, the proposed DWE can further improve the percentile of the cost reduction rate by up to 2.2 % on average, with a maximum improvement of 6.5 %. The results suggest that the ITES control performance can be significantly improved using the AI-based cooling prediction models selected by the proposed prediction accuracy metric designed with HVAC domain knowledge.

3.2. Analysis across multiple energy storage capacities

In the cooling load prediction-based control of ITES systems, the ice storage capacity influences the improvement in the cost reduction rate. To validate the scalability of the proposed DWE metric, a sensitivity analysis was conducted for 10 ice storage capacities, as shown in Fig. A2. The situations shown in Fig. 4 are specific to those shown in Fig. A2 when the ice storage capacity is 70 % in each building case. The mean values in Fig. 4 were used to calculate the approach cost reductions to ideal scenario in Fig. A2 when the ice storage capacity was 70 %. The ideal scenario consisted of the cooling load prediction models with actually the highest cost reduction rates, representing the limit of the cost reduction rate using a perfect prediction accuracy metric. Under most ice storage capacities in each building case, the proposed DWE is

Table 1
Spearman correlation coefficients between prediction accuracy metrics and cost reduction rate.

	Building #1	Building #2	Building #3	Building #4	Building #5	Building #6
MAPE	-0.269	-0.712	-0.732	-0.740	-0.535	-0.569
CV-MAE	-0.758	-0.799	-0.742	-0.746	-0.555	-0.920
CV-RMSE	-0.714	-0.880	**-0.917**	-0.889	-0.715	-0.767
DWE (Proposed)	**-0.933**	**-0.973**	-0.867	**-0.913**	**-0.978**	**-0.981**

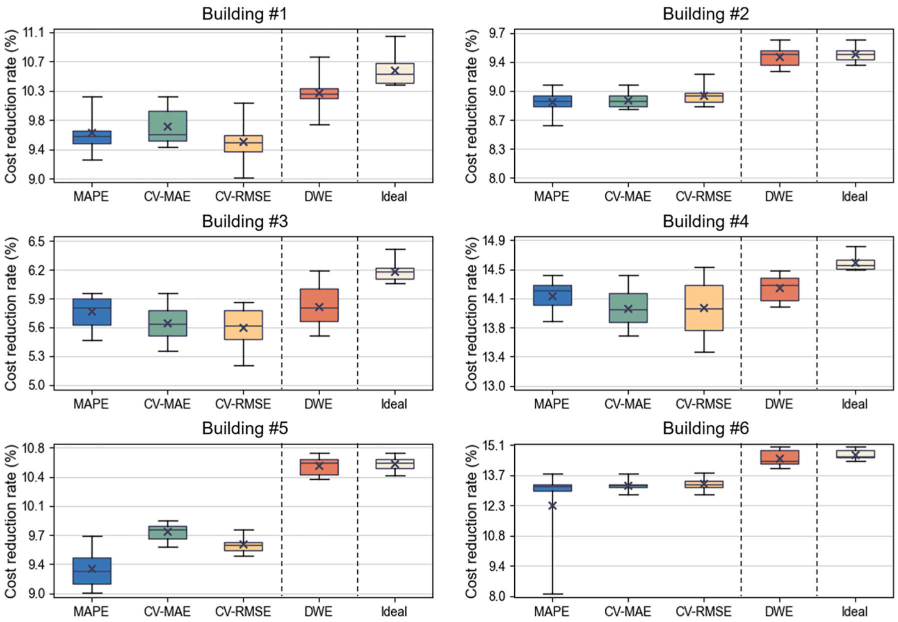


Fig. 4. Cost reduction rates (CRRs) of load prediction-based control of energy storage systems in 6 buildings. The boxes consist of the top ten accurate load prediction models under different prediction accuracy metrics including the MAPE, CV-MAE, CV-RMSE, and the proposed DWE. The ideal scenario is the distribution of the actual top ten cost reduction rates.

closer to the ideal line than the MAPE, CV-MAE, and CV-RMSE. In six building cases, the approach cost reductions to ideal scenario by DWE were over 89.2 %−98.5 %, whereas those by MAPE, CV-MAE, and CV-RMSE were over 62.0 %−84.6 %. These results show that the ITES control performance using the cooling load prediction models selected by DWE is generally better than those selected by conventional prediction accuracy metrics under various ice storage capacities. Thus, the proposed DWE is indicated to have a good scalability to multiple energy storage capacities.

3.3. Analysis across multiple time-of-use tariffs

The TOU tariff is another key factor influencing the performance improvement of the load prediction-based control of energy storage systems. This study conducted a sensitivity analysis under 18 TOU tariffs. The distributions of approach cost reductions to ideal scenario under 18 TOU tariffs are shown in Fig. A3. The average approach cost reductions in Fig. A2 are the specific situations shown in Fig. A3 for TOU #1. The ideal scenario is marked with dashed lines with an approach cost reduction of 100 %, which represents a perfect prediction accuracy metric that the cooling load prediction models with actually the highest cost reduction rates can be selected. The approach cost reductions of DWE are closer to the ideal scenario than those of MAPE, CV-MAE, and CV-RMSE in almost the 18 TOU tariffs. The average approach cost reductions by DWE were 93.9 %−97.2 %, whereas those by MAPE, CV-MAE, and CV-RMSE were 86.4 %−95.4 %. The results show that the ITES control performance using the cooling load prediction models selected by DWE is generally better than those selected by conventional prediction accuracy metrics under various TOU tariffs. The proposed DWE is demonstrated to have a good scalability to multiple TOU tariffs. In summary, the proposed DWE is scalable under 18 TOU tariffs and 10 ice storage capacities for six buildings with ITES systems. The application scalability of the proposed DWE metric is validated compared with the conventional prediction accuracy metrics of MAPE, CV-MAE, and CV-RMSE for ITES controls under TOU tariffs.

4. Discussions

4.1. Application perspectives of the proposed dynamic accuracy weighting method

This study proposed a dynamically weighted error (DWE) to select load prediction models for prediction-based control of energy storage systems. The proposed DWE was validated for six building cases with energy storage systems, 10 energy storage capacities, and 18 TOU tariffs. In comparison to traditional prediction accuracy metrics, including MAPE, CV-MAE, and CV-RMSE, the proposed DWE is scalable and can improve the application performance of AI-based load prediction models. The proposed method of generating DWE was published on the website, https://dwe.dest.net.cn/. A detailed introduction of DWE web tool is shown in Appendix G. This tool is feasible for the researchers and engineers to implement DWE in real-world applications.

The main contribution of this study is proposing a dynamic accuracy weighting method to enhance the prediction-based control of energy storage systems. Fig. 5 depicts a perspective of applications of the proposed dynamic accuracy weighting method and the DWE metric. In the conventional operation control of energy storage systems, operators determine the control objective based on incentives, and conduct prediction-based control to get predictions of control strategies of energy storage systems. As validated in this study, the conventional approach causes a discrepancy between prediction accuracy and control performance, which greatly limits the operation performance of energy storage systems. This is because the selection of load prediction models is not adapted well to the control optimization models under prediction uncertainty. This study solved this issue by proposing the DWE metric to select load prediction models. The DWE is generated dynamically to make its prediction accuracy metric reflecting the control performance, which connects the load prediction and control optimization as a whole process.

The DWE may also be adaptive to various energy storage systems including battery, phase change materials, and water energy storage systems. This is because the required data to determine the DWE is similar for different energy storage systems. To determine the DWE, load predictions, control incentive predictions, and energy storage capacities are required regardless of the type of energy storage system. The

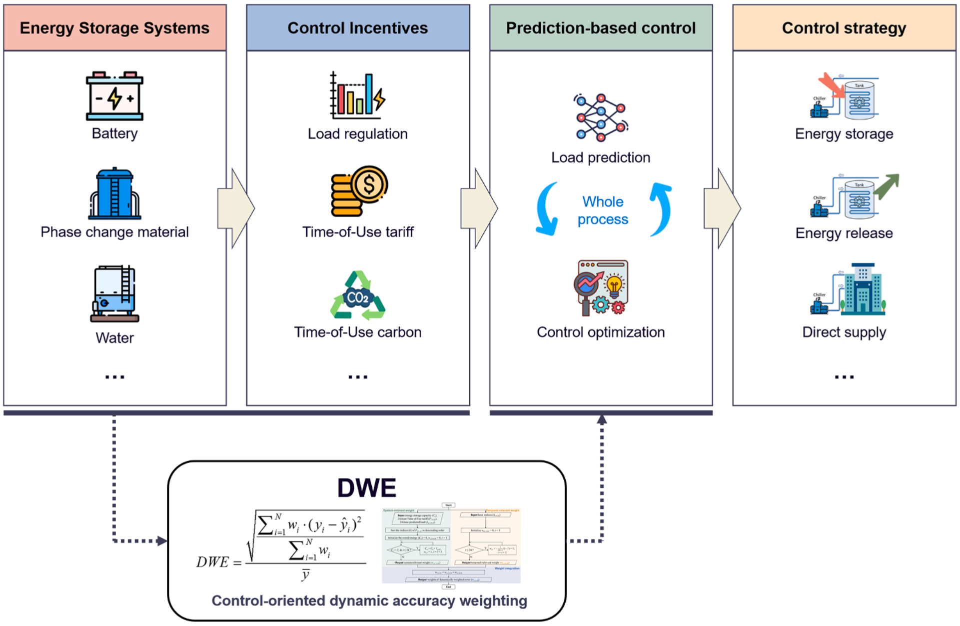


Fig. 5. Application and perspectives of the proposed dynamic accuracy weighting method. Based on the proposed method, dynamically weighted errors (DWEs) can be generated for various energy storage systems under various control incentives. By linking the whole process of load prediction and control optimization, the DWEs can further enhance the control performance of energy storage systems.

physical attributes of different types of energy storage systems are different, whereas they can be enriched by the domain knowledge under the proposed framework of DWE.

The DWE can also be applied for different control incentives, such as load regulation, TOU tariff, and TOU carbon. This is because day-ahead prediction of control incentive is an input of DWE weight calculation, regardless of the type of control incentive. The type of control incentive determines the objective for control-oriented selection of load prediction models, which informs the DWE to track the control performance index.

The proposed DWE is efficient for real-time control applications. For the computation efficiency, a template case is provided on the website of DWE calculation. It takes less than one second to finish the hourly DWE calculation one week ahead. The effectiveness of DWE was validated by the control performances of multiple cases of ITES control under TOU tariff in this paper. Based on the types of energy storage systems and control incentives, the DWE can be determined to enhance the whole process of prediction-based control, and making control strategies effectively and efficiently.

The DWE is proposed for dynamic selections of load prediction models for the whole process of prediction-based control. The selected load prediction model using DWE is validated to be more adaptive to control optimization models to generate more suitable control strategies for enhanced control performance. With the developed web tool, the performance of the proposed dynamic accuracy weighting method is promising to be validated in real-world applications of energy storage system controls of various system types and control incentives.

4.2. Limitations and future works

In this study, the proposed prediction accuracy metric DWE was validated as scalable for the selection of cooling load prediction models for prediction-based ITES control. In comparison to the MAPE, CV-MAE, and CV-RMSE, the DWE can further improve the application performance of the AI-based cooling load prediction model. This study was a novel attempt to enhance the whole-process performance of AI with HVAC domain knowledge for load prediction-based control of energy storage systems. Still, a few limitations exist that are discussed as follows.

This study focuses on a dynamic accuracy weighting method for prediction accuracy metrics to select load prediction models for better control performance of energy storage systems. This contribution facilitates the engineering application of prediction-based control. More fundamentally, new machine learning algorithm or modeling architecture can be proposed based on the proposed DWE. Future studies may investigate the performance of machine learning models using the DWE as a loss function for, or develop new structures of neural networks based on the DWE to enhance the performance of load prediction models fundamentally.

The scalability of the proposed DWE was tested as shown in Figs. A2 and A3. In certain situations, improvements in the proposed DWE require further validation. The proposed DWE will be deployed in building #1 to test its real-world performance. An on-site comparison between DWE and CV-RMSE will be conducted on the ITES system for the cost reduction. The robustness of DWE under input noise and forecast error may be analyzed in this on-site implementation and simulations of other ITES systems. For specific cases, deeper analyses of the structure of the TOU tariff and the mechanism of the prediction-based control process can be conducted to fine-tune the proposed DWE. The proposed metric should also be validated and applied in more cases that differ in building type, climate location, ITES system, and TOU tariff.

Furthermore, the proposed prediction accuracy metric is validated for the cooling load prediction models in prediction-based ITES control. In other energy storage systems with similar energy storage and release rates of ITES systems, such as thermal energy storage with water tanks [67] and electric energy storage with batteries [68], the proposed DWE is also possibly applicable but not validated in this study. For borehole, pit, and aquifer energy storage systems with low energy storage and release rates, the DWE may be not applicable. Future studies may validate the performance of the proposed DWE in wider types of energy storage systems. Also, for hourly or sub-hourly decision-makings, such as the optimal control of operating chiller combinations, the proposed DWE is not applicable either. In hourly or sub-hourly decision-makings, the prediction uncertainty is the main obstacle to the control performance. Future studies may investigate the influence of prediction uncertainty on the control performance of energy storage systems requiring hourly or sub-hourly decision-makings. Based on the understanding of prediction uncertainty, novel application-oriented prediction accuracy metrics may be proposed based on the confidence interval of prediction uncertainty to further enhance the prediction-based control performances of more energy storage systems.

Additionally, under different incentives such as the peak charge of electricity and carbon emission factors, application-oriented prediction accuracy metrics may differ. This study proposed a concept of dynamic accuracy weighting for prediction accuracy metrics under a financial control performance metric of electricity cost. For other control performance metrics such as carbon emissions [69], occupant comfort [70], equipment lifespan [71], and the combinations of different control performance metrics, the dynamic accuracy weighting method of prediction accuracy metrics need further investigations. An automatic optimization approach for the weights of prediction accuracy metrics can be proposed for various control performance metrics. Proposing an approach for the selection of AI-based prediction models for downstream control applications may be also possible. A possible method is to formulate the model selection problem into an optimization problem under uncertainties. Optimization solvers can be used to obtain a proper prediction accuracy metric dynamically under certain boundary conditions. The advanced machine learning approaches, such as ensemble learning and reinforcement learning, are also possible to select load prediction models for prediction-based controls, which needs further investigations. The selection approach of control-oriented load prediction model is also promising to enhance AI-based prediction models during the entire prediction-based control process.

From the perspective of physics-informed learning, the use of application-oriented prediction accuracy metrics to train AI-based prediction models is another research topic. This is a concept of physics-informed learning that adds physical constraints to the loss functions of AI models. Application-oriented prediction accuracy metrics can be used as loss functions to directly train an AI model. As a loss function for AI training, simplicity can be a key consideration for the computational efficiency and convergence of AI training; i.e., this type of prediction accuracy metric should be simple and efficient, and requires a deeper understanding of the prediction-based control process with domain knowledge. Based on the progress on load prediction models using transformer [34] and ensemble learning [40] technologies, the modification of loss function may further improve the application-oriented model performances. Future investigations can be conducted on the reasons for the gap between the prediction accuracy and control performance metrics. With HAVC domain knowledge, proposing application-oriented loss functions for the training of AI-based prediction models is promising for further enhancing the whole-process performance of AI models.

5. Conclusions

This study highlights a critical issue with the existing prediction accuracy metrics, such as mean absolute percentage error, coefficient of variation of mean absolute error, and coefficient of variation of root mean squared error, used to evaluate the performance of artificial intelligence-based load prediction models for energy storage system control. It was found that models with higher load prediction accuracy did not consistently lead to greater cost savings in the operation of these systems. This discrepancy between prediction accuracy and control performance arises from the varying impact of predictions on control performance at different times due to the dynamic inhomogeneous of energy storage capacity and Time-of-Use pricing on the time scale. However, the existing accuracy metrics used to evaluate prediction models are based on equally weighted average values over the training period, which may not adequately address the dynamic inhomogeneity. This study addresses the discrepancy by introducing a dynamic accuracy weighting method and a novel prediction accuracy metric, dynamically weighted error. This new metric evaluates model performance by incorporating control performance considerations in practical applications. In comparison to traditional prediction accuracy metrics such as mean absolute percentage error, coefficient of variation of mean absolute error, and coefficient of variation of root mean squared error, the proposed dynamically weighted error improved cost reduction rates of up to 6.5 % in six building cases with energy storage systems. A web tool was developed for researcher and engineers to easily calculate the dynamically weighted error. This study provided a novel paradigm of using domain knowledge-informed artificial intelligence models to enhance the whole-process control performance of energy storage system.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors declare that they did not use generative AI and AI-assisted technologies in the writing process of this paper.

CRediT authorship contribution statement

Xiao Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Xue Liu:** Writing – review & editing, Supervision, Conceptualization. **Xuyuan Kang:** Writing – review & editing, Supervision, Methodology. **Fu Xiao:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Da Yan:** Supervision, Project administration, Methodology, Conceptualization.

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.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grant Number: 52225801, 52408116), the Postdoctoral Fellowship Program of CPSF under Grant Number GZB20230369, 2024M761691 and the Shuimu Tsinghua Scholar Program of Tsinghua University.

Appendix A. Scalability analysis of the proposed dynamically weighted error

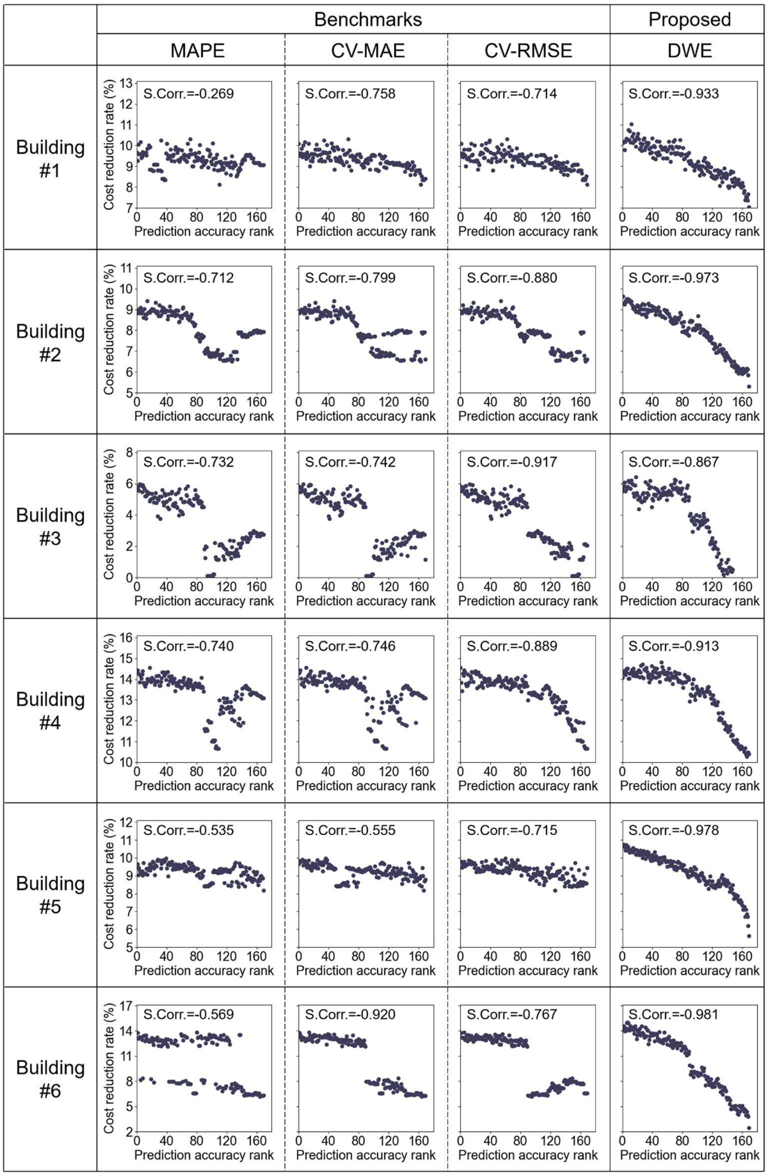


Fig. A1. Correlation between prediction accuracy and cost reduction rates under benchmark and proposed prediction accuracy metrics. The validation is conducted in six building cases with energy storage systems using 180 load prediction models. S. Corr. denotes the Spearman correlation coefficient between prediction accuracy and cost reduction rate. Higher absolute value of a negative Spearman correlation coefficient indicates a better prediction accuracy metric.

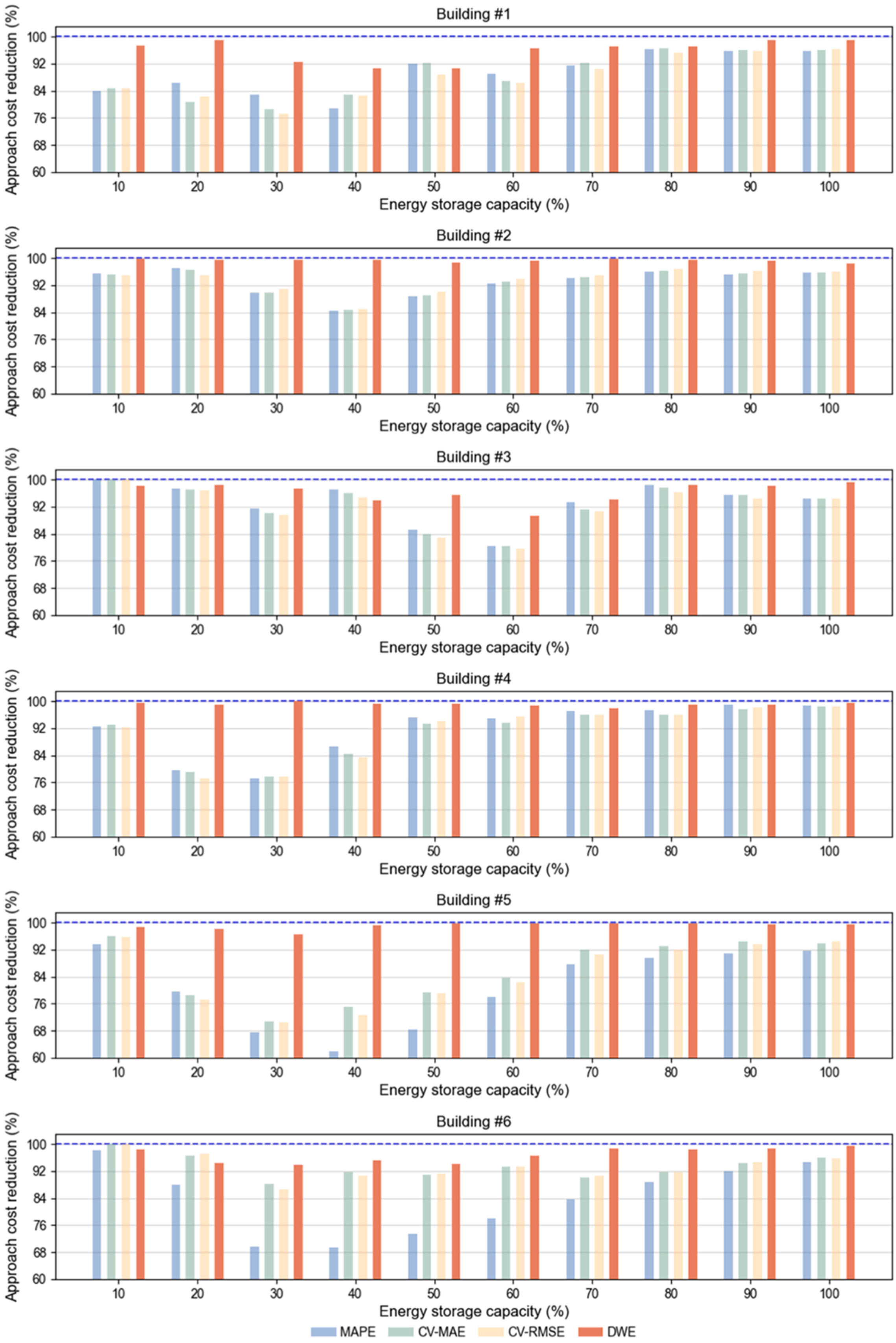


Fig. A2. Approach cost reductions (ACRs) to ideal scenario under 10 energy storage capacities. The ACR of 100 % denotes the ideal scenario where the maximal cost reduction rate is realized when load prediction is completely deterministic. The energy storage capacity is a relative value to the maximal daily total energy load of a building energy storage system.

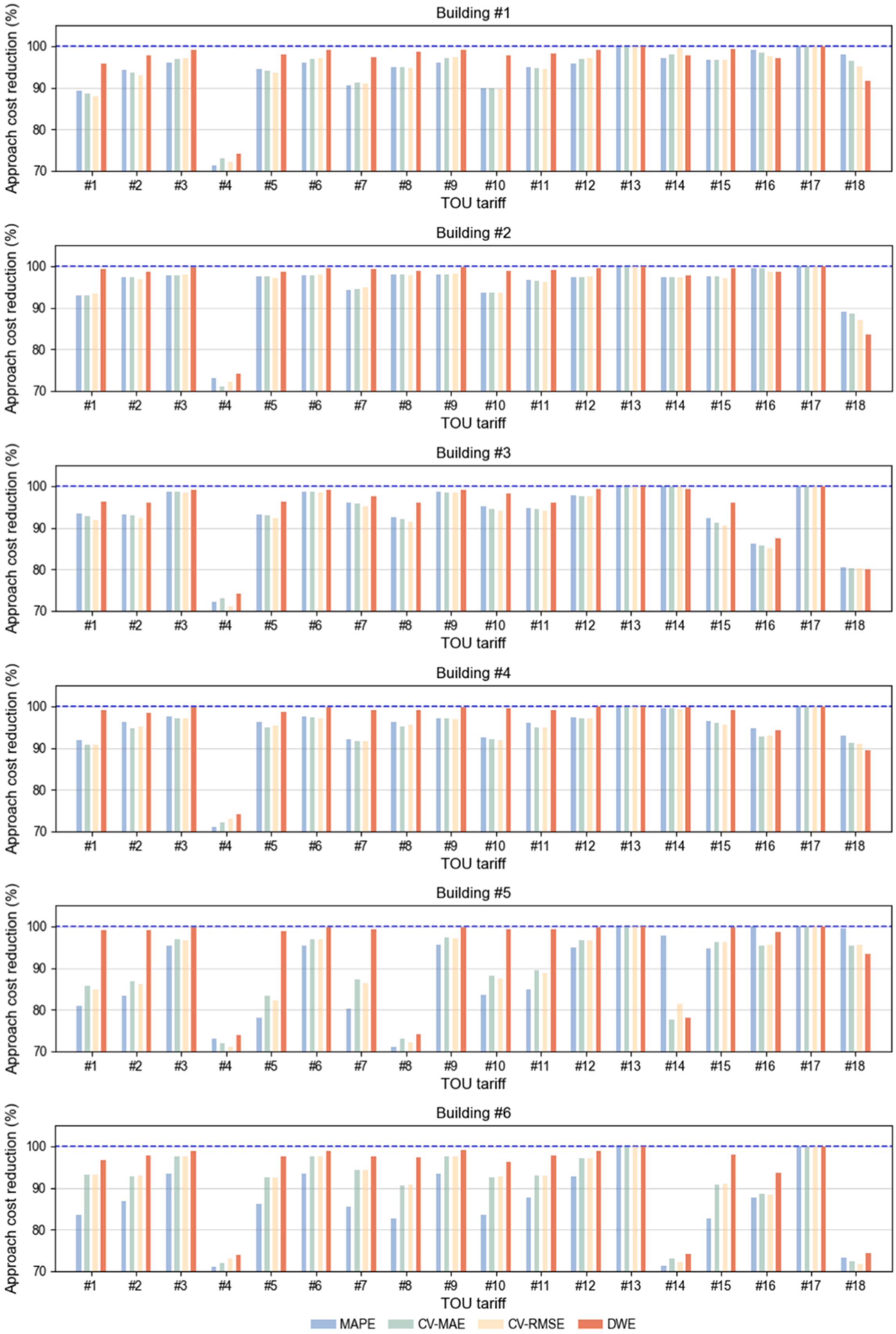


Fig. A3. Approach cost reductions (ACRs) to ideal scenario under 18 Time-of-Use (TOU) tariffs. The ACR of 100 % denotes the ideal scenario where the maximal cost reduction rate is realized when load prediction is completely deterministic.

Appendix B. An example of calculating the weights of dynamically weighted error

Fig. A4 shows an example of calculating the weights of DWE. w_s is a Boolean variable with 24 dimensions that represent the key hours of a day for load prediction-based energy storage system control. It is determined according to the combined influence of the TOU tariff, energy storage capacity, and load prediction of the day. For example, the hours with the lowest electricity price have less importance of accurate cooling load prediction, because the control strategy during this time is usually energy storage regardless of load prediction. Thus, the w_s is obtained by sorting the electricity price and accumulating the energy loads at high electricity prices until the maximal energy storage capacity is reached. w_t is a variable that decreases from the start to the end of the day and describes the temporal characteristics of the load prediction-based energy storage system control. In other words, the hours after are less important than the hours before on a given day. w is the Hadamard product of w_s and w_t, reflecting their combined effects, which is the DWE weight.

Fig. A4. An intuitive example of calculating the weights of the proposed dynamically weighted error (DWE). Step 1: Receive the hourly electricity prices ahead of a day. Step 2: Sort the electricity prices in a descending order to get a sorted series of hour indices. Step 3: Accumulate the hourly energy loads in the order of the sorted hour indices until the accumulated load is just higher than the energy storage capacity. Step 4: Set value 1 at the hours of accumulated energy loads, and set value 0 at other hours of a day, to obtain the system-relevant weight (ws). Step 5: Set values at hours based on Eq. (3) to obtain the temporal-relevant weight (wt). Step 6: Calculate the Hadamard product between ws and wt to obtain the integrated weight (w) of the proposed DWE.

Appendix C. Load prediction models

The load prediction models were identical to those in [44], with 180 models. The models predicted the building cooling load for the next 24 hours using input features, including weather forecasts of dry bulb temperature, relative humidity, solar radiation, weekdays, holidays, hours, and the historical cooling load at the previous 24th hour. The cooling load prediction models were trained using historical data from the past 30 days using the sliding window method from [72]. The algorithms of the cooling load prediction models include the elastic net (EN), support vector machine (SVM), multilayer perceptron (MLP), random forest (RF), extreme gradient boosting (XGB), and light gradient boosting (LGB). For each algorithm, 30 hyperparameter combinations were selected. The selected models were validated in [44] to cover a reliable range to mitigate the influence of the differences among prediction algorithms on the effectiveness evaluation of prediction accuracy metrics. The algorithms and hyperparameters are listed in Table A1.

Table A1
Hyperparameters of load prediction models in this study.

Algorithm	Hyperparameters
EN	alpha=[0, 0.2, 0.5, 1, 2, 5] l1_ratio=[0, 0.2, 0.4, 0.6, 0.8, 1]
SVM	C=[100, 200, 500, 1000] epsilon=[0.1, 0.5, 1, 5, 10] tol=[0.001, 0.0001]
MLP	num_of_layers=[1, 2, 3, 4, 5, 6] neurons_per_layer=[10, 20, 50, 100, 150, 200]
RF	max_depth=[6, 10, 15] n_estimators=[50, 100, 200] min_samples_split=[5, 20, 50] min_samples_leaf=[2, 5, 10]
XGB	max_depth=[3, 5, 7] min_child_weight=[0.01, 0.1, 1, 10] reg_alpha=[0, 0.01, 1, 10] reg_lambda=[0, 0.01, 1, 10] learning_rate=[0.01, 0.1, 0.2] objective=['tweedie', 'squarederror']
LGB	max_depth=[-1, 3, 6] n_estimators=[50, 100, 500] min_child_weight=[0.001, 0.1, 10] reg_alpha=[0, 1, 10] reg_lambda=[0, 1, 10] learning_rate=[0.01, 0.1, 0.2]

Appendix D. Case buildings with energy storage systems

This study used six real building cases from different types and locations to validate the scalability of the proposed prediction accuracy metrics, as listed in Table A2. The buildings were office buildings, commercial buildings, hotels, or combinations of them. The building cases were from different locations with different climates from across China. The cooling loads of the six buildings were measured, as shown in Fig. A5. The data were measured hourly, with lengths from three months to a year. The cooling load patterns of the six buildings were significantly different, providing various situations for validating the scalability of the proposed prediction accuracy metric. The corresponding weather [73] and holiday data [74] for the cooling load prediction were collected from web databases, according to the locations of the cases [75].

The energy storage systems were derived from the ITES system described in [76], which was a real ITES system for building #1. It consisted of a normal chiller, three duplex chillers, an ice storage tank, four chilled water pumps, four cooling water pumps, three ethylene glycol pumps, and five cooling towers. The ITES system model was calibrated and validate using measured data. The simulation error of daily total electricity consumption of the ITES system was 4.7 % in CV-MAE. The ice storage tank capacities for the six cases were set to 70 % of the maximum daily total cooling load of the buildings. The capacities of the other devices were set according to an automatic sizing approach [77]. The TOU tariff for the six cases is TOU #1 in Fig. A6 in Appendix F, which is the actual TOU tariff for building #1. The prediction accuracy and control performance of the cooling load prediction-based ITES control were analyzed in six cases via simulation.

Table A2
Building information.

	Type	Location	Cooling area (m^2)
Building #1	A complex (a commercial building and an office building)	Beijing, China	141,000
Building #2	A complex (a commercial building, two office buildings, and a hotel)	Guangzhou, China	358,000
Building #3	An office building	Hong Kong, China	143,000
Building #4	Two office buildings	Hong Kong, China	36,000
Building #5	An office building	Zhuhai, China	4,400
Building #6	An office building	Shenzhen, China	210,000

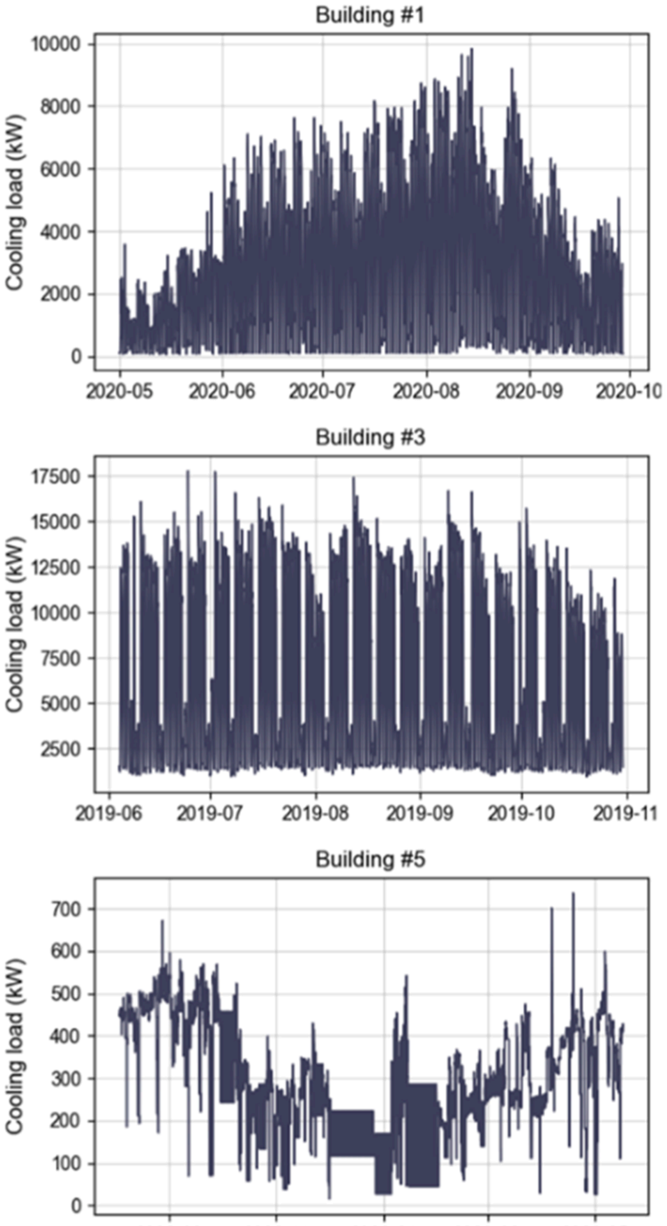


Fig. A5. Cooling loads of case buildings in this study. Six case buildings in different locations from three types were investigated for the scalability of the results.

Appendix E. Control strategies

In this study, the control strategies of the ITES system were used to determine the control actions for ice storage, ice release, and cooling by the chiller over the next 24 hours. The control strategies of the fixed schedule and prediction-based control were consistent with those in [44].

The fixed schedule used a fixed control strategy for each day, which was a benchmark control strategy for calculating the cost reduction rate described as follows.

1. Ice storage during 23:00–7:00.
2. Ice release during 10:00–15:00, 16:00–17:00, and 18:00–22:00.
3. Use of chillers for cooling during the other hours of the day.

The prediction-based control used the predicted cooling load to determine the control action, following a rule with four steps described as follows.

1. Ice storage occurred when the electricity price for an hour was less than 1/3 of the maximal electricity price. The total amount of stored ice was calculated.
2. Calculation of the electricity cost per cooling load of the other hours.
3. Arrangement of the hours of ice release according to the descending order of electricity cost per cooling load per hour until all stored ice was arranged.
4. Use of chillers for cooling during the other un-arranged hours.

Appendix F. Time-of-Use tariffs

In this study, 18 TOU tariffs were used to validate the scalability of the proposed prediction accuracy metric, as shown in Fig. A6. TOU #1 is the real TOU tariff for building #1. Other TOU tariffs are derived from TOU #1 by adjusting the peak ratio, changing the temporal structures, or adding Gaussian noise [78]. For TOU #1–#9, the TOU tariffs in each row have the same peak ratio (ratio of maximal to minimal prices). The TOU tariffs in each column have the same temporal structure (i.e., hours of electricity price level). From the first to the third row, the peak ratios are 12, 8, and 4, respectively. With the super-peak price fixed, the off-peak, partial-peak, and peak prices are scaled using a linear approach to guarantee the peak ratios. From the left to right columns, the off-peak/partial-peak/peak/super-peak hours are 8/7/6/3, 10/9/4/1, and 6/7/6/5, respectively. TOU #10–#18 were generated by adding Gaussian noise hourly to TOU #1–#3, respectively, using the corresponding columns. From the fourth row to the bottom row, the squared standard deviations of the Gaussian noise are 0.01, 0.02, and 0.05, respectively. The 18 TOU tariffs were used to test the scalability of the proposed prediction accuracy metric under different fixed and dynamic structures of TOU tariffs.

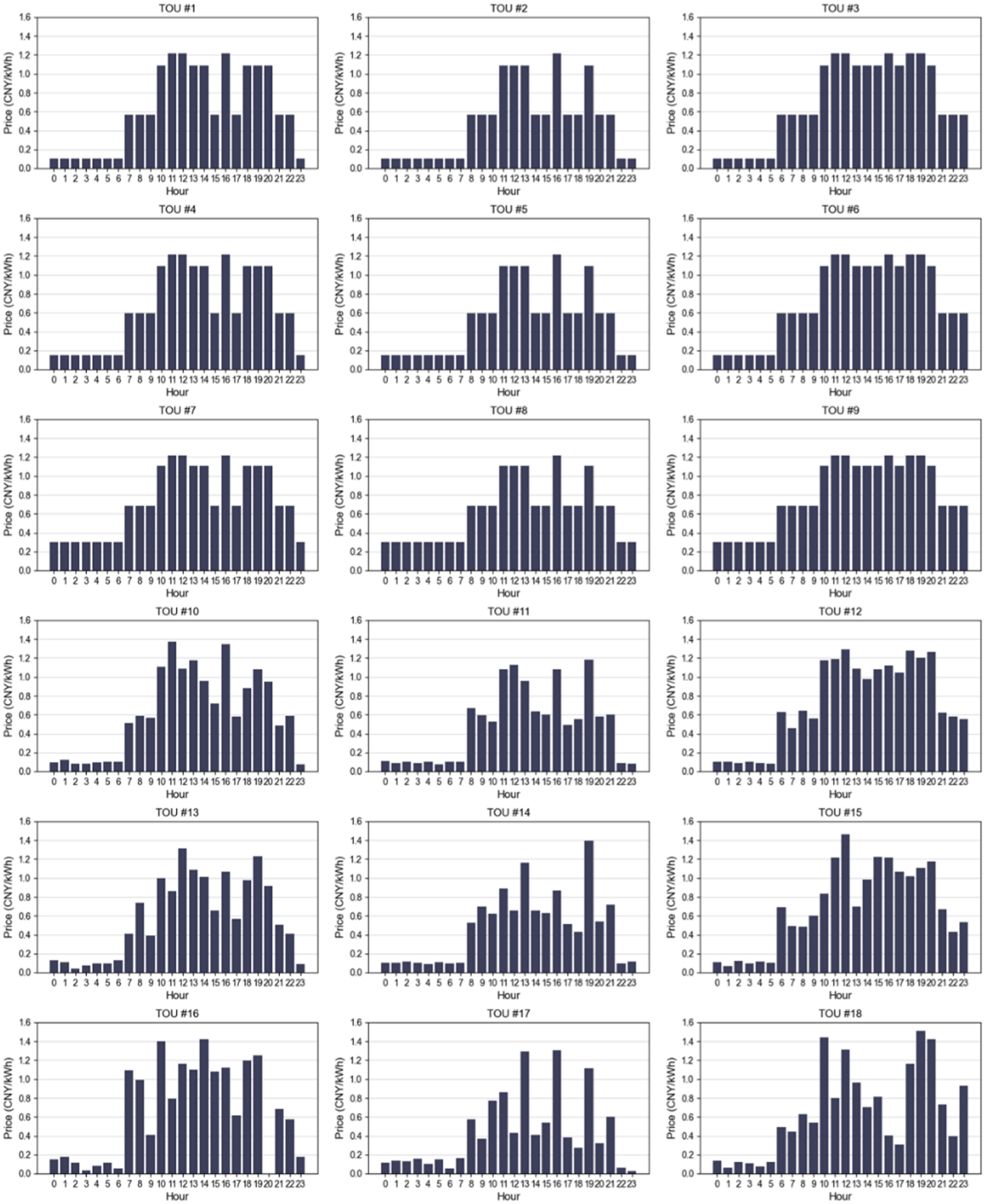
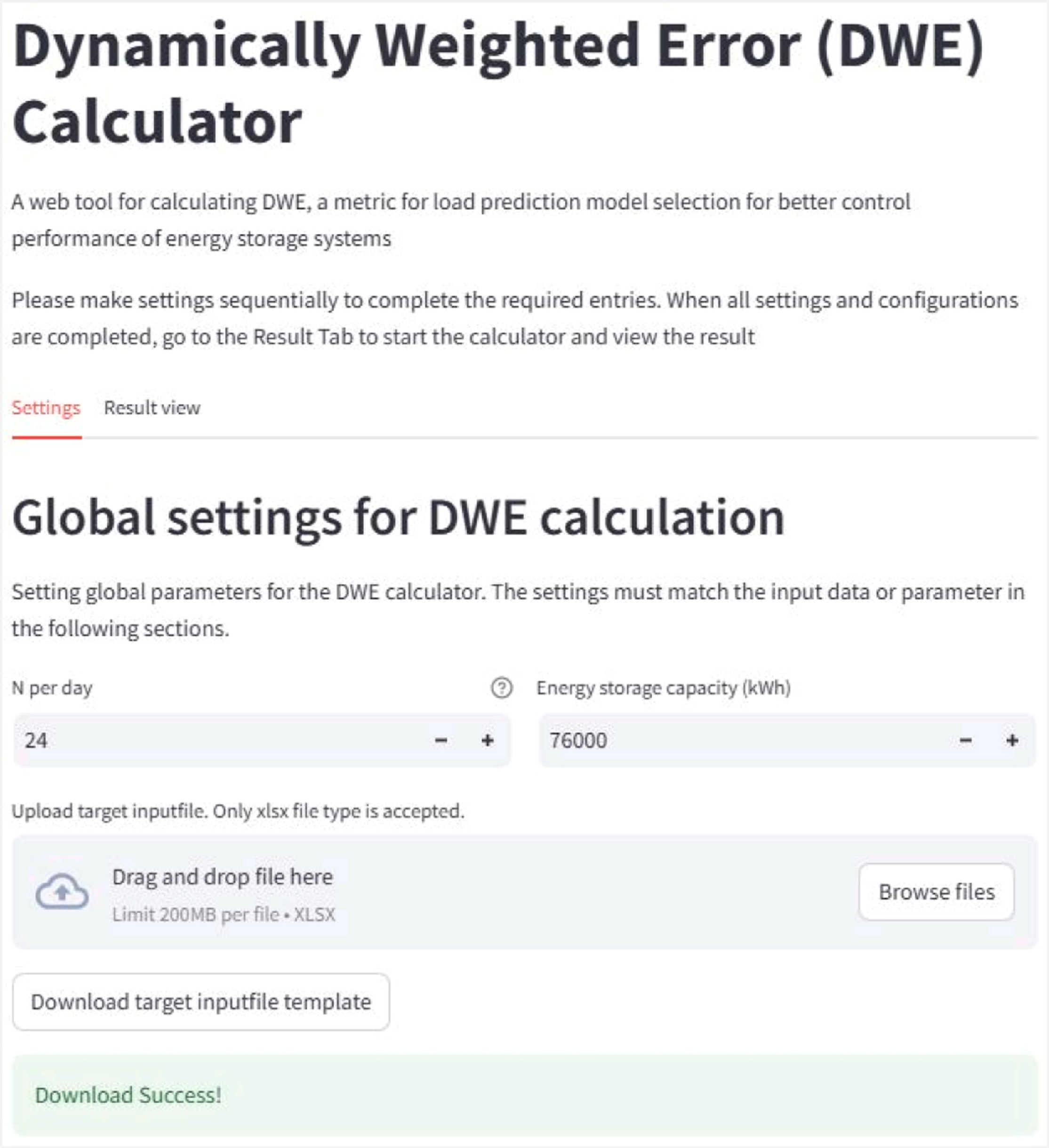


Fig. A6. Time-of-Use tariffs investigated in this study. The 18 types of Time-of-Use tariffs covered wide ranges of peak ratios and temporal structures, which supported comprehensive validations of the scalability of the proposed DWE metric.

Appendix G. A web tool for dynamically weighted error calculation

The proposed method in this study of generating the DWE was published on a website, https://dwe.dest.net.cn/, for feasible applications in research and engineering projects. It is an expansion to the DeST tool box (https://app.dest.net.cn/) [79]. The interface of DWE web tool is shown in Fig. A7. The users can get a suitable DWE for specific cases of load prediction-based energy storage system control via this tool. The inputs include number of samples per day (e.g. 24 hours), energy storage capacity (e.g. 76,000 kWh), and a time series file consisting of TOU tariff and predicted loads. The explanations of the inputs can be found on the website. The users can also download the input file template on the website. After uploading the required information in “Settings”, the users can move to “Result view” to start calculation and view the results of DWE. If the user is expected to directly get a selected load prediction model, the measured loads should be uploaded. After calculation, the measured and predicted loads of the selected model will be shown on the “Result view”. The computation time of the web tool is tested to be within a second for time series with 24 hours of a day. For further integrations of the DWE calculator with specific engineering projects, the users can contact the authors for an access to an application program interface of DWE via the Internet. The DWE calculator provides a user-friendly tool for researcher and engineers to easily improve the prediction-based control performance of energy storage systems by selecting application-oriented load prediction models.


Dynamically Weighted Error (DWE) Calculator

A web tool for calculating DWE, a metric for load prediction model selection for better control performance of energy storage systems

Please make settings sequentially to complete the required entries. When all settings and configurations are completed, go to the Result Tab to start the calculator and view the result

Global settings for DWE calculation

Setting global parameters for the DWE calculator. The settings must match the input data or parameter in the following sections.

N per day: 24

Energy storage capacity (kWh): 76000

Upload target inputfile. Only xlsx file type is accepted.

Download Success!

Fig. A7. An open access web tool for the calculation of dynamically weighted error (DWE). The calculator provides detailed instructions and templates for users to implement.

Appendix H. Summary of the existing prediction accuracy metrics

Table A3
Existing prediction accuracy metrics.

Metrics	Abbreviation	Definition
Mean absolute error	MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert$
Mean bias error	MBE	$MBE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)$
Mean squared error	MSE	$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$
Root mean squared error	RMSE	$RMSE = \sqrt{MSE}$
Mean absolute percentage error	MAPE	$MAPE = \frac{1}{n}\sum_{i=1}^{n} \lvert \frac{y_i - \hat{y}_i}{y_i} \rvert$
Coefficient of variation of MAE	CV-MAE	$CV - MAE = \frac{MAE}{\bar{y}}$
Coefficient of variation of RMSE	CV-RMSE	$CV - RMSE = \frac{RMSE}{\bar{y}}$
R squared	R^2	$R^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2}$

*y_i and $\hat{y}_i$ are measured and predicted loads; $\bar{y}$ is the average of measured loads; n is the number of load samples.

Data availability

Data will be made available on request.

References

[1] Pan Y, Zhu M, Lv Y, Yang Y, Liang Y, Yin R, et al. Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies. Adv Appl Energy 2023:10.

[2] Yang S, Gao HO, You F. Demand flexibility and cost-saving potentials via smart building energy management: opportunities in residential space heating across the US. Adv Appl Energy 2024:14.

[3] Wald D, Johnson K, King J, Comden J, Bay CJ, Chintala R, et al. Shifting demand: reduction in necessary storage capacity through tracking of renewable energy generation. Adv Appl Energy 2023:10.

[4] Heidarykiany R, Ababei C. HVAC energy cost minimization in smart grids: A cloud-based demand side management approach with game theory optimization and deep learning q. Energy Ai 2024:16.

[5] Kemp JM, Millstein D, Kim JH, Wiser R. Interactions between hybrid power plant development and local transmission in congested regions. Adv Appl Energy 2023: 10.

[6] Huang W-C, Zhang Q, You F. Impacts of battery energy storage technologies and renewable integration on the transition in the New York State. Adv Appl Energy 2023:9.

[7] Peterssen F, Schlemminger M, Lohr C, Niepelt R, Hanke-Rauschenbach R, Brendel R. Impact of forecasting on energy system optimization. Adv Appl Energy 2024:15.

[8] Huang R, Mahvi A, James N, Kozubal E, Woods J. Evaluation of phase change thermal storage in a cascade heat pump. Appl Energy 2024:359.

[9] Coraci D, Brandi S, Hong TZ, Capozzoli A. Online transfer learning strategy for enhancing the scalability and deployment of deep reinforcement learning control in smart buildings. Appl Energy 2023;333:24.

[10] Niu H, Liu X, Wang B, Shi W. Development, research and policy status of logistics cold storage in the context of carbon neutrality: an overview. Energy Build 2024; 320:114606.

[11] Vivian J, Heer P, Fiorentini M. Optimal sizing and operation of seasonal ice thermal storage systems. Energy Build 2023;300:113633.

[12] Odufuwa OY, Kusakana K, Numbi BP, Tartibu LK. Optimal energy management of grid-connected PV for HVAC cooling with ice thermal storage system. J Energy Storage 2024:77.

[13] Yang S, Gao HO, You F. Building electrification and carbon emissions: integrated energy management considering the dynamics of the electricity mix and pricing. Adv Appl Energy 2023;10:100141.

[14] Li Y, Wang Z, Xu W, Gao W, Xu Y, Xiao F. Modeling and energy dynamic control for a ZEH via hybrid model-based deep reinforcement learning. Energy 2023;277: 127627.

[15] Guo H, Huang B, Wang J. Probabilistic load forecasting for integrated energy systems using attentive quantile regression temporal convolutional network. Adv Appl Energy 2024:14.

[16] Guo J, Yun S, Meng Y, He N, Ye D, Zhao Z, et al. Prediction of heating and cooling loads based on light gradient boosting machine algorithms. Build Environ 2023; 236:110252.

[17] Irshad K, Algarni S. Novel optimized hybrid neuro-fuzzy approach for analysis of cold thermal storage system-assisted air conditioning system performance. J Build Eng 2023;65:105729.

[18] Odufuwa OY, Tartibu LK, Kusakana K, Hohne PA, Numbi BP. Application of artificial neural networks in predicting the performance of ice thermal energy storage systems. J Energy Storage 2024;95:112547.

[19] Lian H, Ji Y, Niu M, Gu J, Xie J, Liu J. A hybrid load prediction method of office buildings based on physical simulation database and LightGBM algorithm. Appl Energy 2025:377.

[20] Chen S, Huang Y, Zhang X, Kuznik F, He X, Ma Y, et al. Occupant-centric dynamic heating and cooling loads simplified prediction model for urban community at energy planning stage, 90. Sustainable Cities and Society; 2023, 104406.

[21] Lu C, Gu J, Lu W. An improved attention-based deep learning approach for robust cooling load prediction: Public building cases under diverse occupancy schedules, 96. Sustainable Cities and Society; 2023, 104679.

[22] Yu H, Zhong F, Du Y, Xie Xe, Wang Y, Zhang X, et al. Short-term cooling and heating loads forecasting of building district energy system based on data-driven models. Energy Build 2023;298:113513.

[23] Tang H, Yu J, Geng Y, Liu X, Lin BR. Optimization of operational strategy for ice thermal energy storage in a district cooling system based on model predictive control. J Energy Storage 2023;62:18.

[24] Zhang C, Lu J, Huang J, Zhao Y. End-to-end data-driven modeling framework for automated and trustworthy short-term building energy load forecasting. Build Simul 2024;17(8):1419–37.

[25] Xiao Z, Zhang J, Xiao F, Chen Z, Xu K, So PM, et al. An AI-enabled optimal control strategy utilizing dual-horizon load predictions for large building cooling systems and its cloud-based implementation. Energy Build 2025:330.

[26] He X, Zhao W, Gao Z, Zhang Q, Wang W. A hybrid prediction interval model for short-term electric load forecast using Holt-Winters and Gate Recurrent Unit. Sustain Energy, Grids Networks 2024;38:101343.

[27] Maleki N, Lundström O, Musaddiq A, Jeansson J, Olsson T, Ahlgren F. Future energy insights: time-series and deep learning models for city load forecasting. Appl Energy 2024;374:124067.

[28] Balali Y, Chong A, Busch A, O'Keefe S. Energy modelling and control of building heating and cooling systems with data-driven and hybrid models—A review. Renew Sustain Energy Rev 2023;183:113496.

[29] Zhou M, Yu J, Wang M, Quan W, Bian C. Research on the combined forecasting model of cooling load based on IVMD-WOA-LSSVM. Energy Build 2024;317: 114339.

[30] Xu K, Chen Z, Xiao F, Zhang J, Zhang H, Ma T. Semantic model-based large-scale deployment of AI-driven building management applications. Automation in Construction; 2024. p. 165.

[31] Chen Z, Zhang J, Xiao F, Xu K, Chen Y. Development of a probabilistic cooling load prediction-based robust chiller sequencing strategy and its real-world implementation. Appl Energy 2025:382.

[32] Cakiroglu C, Aydın Y, Bekdaş G, Isikdag U, Sadeghifam AN, Abualigah L. Cooling load prediction of a double-story terrace house using ensemble learning techniques and genetic programming with SHAP approach. Energy Build 2024;313:114254.

[33] Zheng R, Lei L. A hybrid model for real-time cooling load prediction and terminal control optimization in multi-zone buildings. J Build Eng 2025:112120.

[34] Li L, Su X, Bi X, Lu Y, Sun X. A novel Transformer-based network forecasting method for building cooling loads. Energy Build 2023;296:113409.

[35] Wang X, Xu H, Fan C, Tan S, Kang X, Shi Y, et al. Leveraging reinforcement learning for optimal control of chiller plant with complex hydraulic and thermodynamic characteristics. Energy Build 2025;344:116037.
[36] Wang X, Dong B. Long-term experimental evaluation and comparison of advanced controls for HVAC systems. Appl Energy 2024;371:123706.
[37] Jiang Z, Dong B. Modularized neural network incorporating physical priors for future building energy modeling. Patterns 2024;5(8).
[38] Wang X, Wang X, Kang X, Dong B, Yan D. Physics-consistent input convex neural network-driven reinforcement learning control for multi-zone radiant ceiling heating and cooling systems: an experimental study. Energy Build 2025;327:115105.
[39] Guo F, Li A, Yue B, Xiao Z, Xiao F, Yan R, et al. Improving the out-of-sample generalization ability of data-driven chiller performance models using physics-guided neural network. Appl Energy 2024;354:122190.
[40] Ma Z, Jiang G, Chen J. Physics-informed ensemble learning with residual modeling for enhanced building energy prediction. Energy Build 2024;323:114853.
[41] Yang Z, Gaidhane AD, Drgoňa J, Chandan V, Halappanavar MM, Liu F, et al. Physics-constrained graph modeling for building thermal dynamics. Energy and AI 2024;16:100346.
[42] Wang XZ, Dong B. Physics-informed hierarchical data-driven predictive control for building HVAC systems to achieve energy and health nexus. Energy Build 2023;291:18.
[43] Chen Y, Yang Q, Chen Z, Yan C, Zeng S, Dai M. Physics-informed neural networks for building thermal modeling and demand response control. Build Environ 2023;234:110149.
[44] Wang X, Liu X, Wang Y, Kang X, Geng R, Li A, et al. Investigating the deviation between prediction accuracy metrics and control performance metrics in the context of an ice-based thermal energy storage system. J Energy Storage 2024;91:112126.
[45] Song C, Yang H, Meng X-B, Yang P, Cai J, Bao H, et al. A novel deep-learning framework for short-term prediction of cooling load in public buildings. J Clean Prod 2024;434:139796.
[46] Yan X, Ji X, Meng Q, Sun H, Lei Y. A hybrid prediction model of improved bidirectional long short-term memory network for cooling load based on PCANet and attention mechanism. Energy 2024;292:130388.
[47] Hu Y, Qin L, Li S, Li X, Zhou R, Li Y, et al. Adaptive corrected parameters algorithm applied in cooling load prediction based on black-box model: A case study for subway station. Energy Build 2023;297:113429.
[48] Wang P, Guo M, Cao Y, Hao S, Zhou X, Zhao L. Pedestrian wind flow prediction using spatial-frequency generative adversarial network. Build Simul 2024;17(2):319–34.
[49] Zhang J, Huang Y, Cheng H, Chen H, Xing L, He Y. Ensemble learning-based approach for residential building heating energy prediction and optimization. J Build Eng 2023;67:106051.
[50] Ra SJ, Kim J-H, Park CS. Real-time model predictive cooling control for an HVAC system in a factory building. Energy Build 2023;285:112860.
[51] Afzal S, Ziapour BM, Shokri A, Shakibi H, Sobhani B. Building energy consumption prediction using multilayer perceptron neural network-assisted models; comparison of different optimization algorithms. Energy 2023;282:128446.
[52] Shen Y, Hu Y, Cheng K, Yan H, Cai K, Hua J, et al. Utilizing interpretable stacking ensemble learning and NSGA-III for the prediction and optimisation of building photo-thermal environment and energy consumption. Build Simul 2024;17(5):819–38.
[53] Fu C, Quintana M, Nagy Z, Miller C. Filling time-series gaps using image techniques: multidimensional context autoencoder approach for building energy data imputation. Appl Therm Eng 2024;236:121545.
[54] Fouladfar MH, Soppelsa A, Nagpal H, Fedrizzi R, Franchini G. Adaptive thermal load prediction in residential buildings using artificial neural networks. J Build Eng 2023;77:107464.
[55] Lu L, Huang X, Zhou X, Guo J, Yang X, Yan J. High-performance formaldehyde prediction for indoor air quality assessment using time series deep learning. Build Simul 2024;17(3):415–29.
[56] Liu H, Liu Y, Huang H, Wu H, Huang Y. Energy consumption dynamic prediction for HVAC systems based on feature clustering deconstruction and model training adaptation. Build Simul 2024;17(9):1439–60.
[57] Kang X, Wang X, An J, Liu X, Yan D. A novel physical-feature-based approach for stochastic simulation of typical building electricity use profiles. Energy Build 2024;302:113742.
[58] Ni Z, Zhang C, Karlsson M, Gong S. A study of deep learning-based multi-horizon building energy forecasting. Energy Build 2024;303:113810.
[59] Li G, Wu Y, Yan C, Fang X, Li T, Gao J, et al. An improved transfer learning strategy for short-term cross-building energy prediction using data incremental. Build Simul 2024;17(1):165–83.
[60] Ali A, Jayaraman R, Mayyas A, Alaifan B, Azar E. Machine learning as a surrogate to building performance simulation: predicting energy consumption under different operational settings. Energy Build 2023;286:112940.
[61] Pan H, Wu C. Bayesian optimization plus XGBoost based life cycle carbon emission prediction for residential buildings-an example from Chengdu, China. Build Simulat 2023;16(8):1451–66.
[62] Lu C, Li S, Penaka SReddy, Olofsson T. Automated machine learning-based framework of heating and cooling load prediction for quick residential building design. Energy 2023;274:127334.
[63] Kazmi H, Fu C, Miller C. Ten questions concerning data-driven modelling and forecasting of operational energy demand at building and urban scale. Build Environ 2023;239:110407.
[64] Hong T, Pinson P, Wang Y, Weron R, Yang D, Zareipour H. Energy forecasting: a review and outlook. Ieee Open Access J Power Energy 2020;7:376–88.
[65] Koponen P, Ikaheimo J, Koskela J, Brester C, Niska H. Assessing and comparing short term load forecasting performance. Energies (Basel) 2020;13(8).
[66] Haben S, Ward J, Greetham DV, Singleton C, Grindrod P. A new error measure for forecasts of household-level, high resolution electrical energy consumption. Int J Forecast 2014;30(2):246–56.
[67] Buenning F, Heer P, Smith RS, Lygeros J. Increasing electrical reserve provision in districts by exploiting energy flexibility of buildings with robust model predictive control. Adv Appl Energy 2023:10.
[68] Tsiklios C, Schneider S, Hermesmann M, Mueller TE. Efficiency and optimal load capacity of E-fuel-based energy storage systems. Adv Appl Energy 2023:10.
[69] Esmaeel Nezhad A, Nardelli PHJ, Sahoo S, Ghanavati F. Scheduling of energy hub resources using robust chance-constrained optimization. Ieee Access 2022;10:129738–53.
[70] Nezhad AE, Rahimnejad A, Nardelli PHJ, Gadsden SA, Sahoo S, Ghanavati F. A shrinking horizon model predictive controller for daily scheduling of home energy management systems, 10. Ieee Access; 2022. p. 29716–30.
[71] Javadi MS, Nezhad AE, Gough M, Lotfi M, Anvari-Moghaddam A, Nardelli PHJ, et al. Conditional value-at-risk model for smart home energy management systems. E-prime - advances in electrical engineering. Electr Energy 2021;1:100006.
[72] Kang XY, Wang X, An JJ, Yan D. A novel approach of day-ahead cooling load prediction and optimal control for ice-based thermal energy storage (TES) system in commercial buildings. Energy Build 2022;275:16.
[73] ECMWF. ERA5 reanalysis. Accessed August 1, 2024, https://registry.opendata.aws/ecmwf-era5/; 2024.
[74] The State Council of PRC. General holidays. Accessed August 1, 2024, https://www.gov.cn/; 2024.
[75] The Hong Kong Government. General holidays. Accessed August 1, 2024, https://www.gov.hk/; 2024.
[76] Wang X, Kang X, An J, Chen H, Yan D. Reinforcement learning approach for optimal control of ice-based thermal energy storage (TES) systems in commercial buildings. Energy Build 2023;301:113696.
[77] Kang X, Wang X, Liu Z, Wu Y, Liu X, Yan D. Capacity optimization and carbon-effective assessment of building micro-grids integrating PV, EV, and energy storages, 130. Sustainable Cities and Society; 2025, 106566.
[78] Kang X, Chen H, Dou Z, Wang X, Liu Z, Zhang C, et al. A district-level building electricity use profile simulation model based on probability distribution inferences, 115. Sustainable Cities and Society; 2024, 105822.
[79] Kang X, Xu H, Wang X, Chen S, Liu Z, Wang Y, et al. Distribution and correlation analysis of typical features of electricity use profiles in non-residential buildings. J Build Eng 2024:110025.