

---

# QUANTIFY THE IMPACT OF BUILDING LOAD FORECAST ON ENERGY MANAGEMENT OF MICROGRID \*

---

**Lunlong Li**  
HKUST  
llidlx@connect.ust.hk

**Yi Ju**  
UC Berkeley  
juy16thu@berkeley.edu

**Zhe Wang**  
HKUST  
cezhewang@ust.hk

## ABSTRACT

Optimizing the operation of microgrids is important for achieving economic efficiency, environmental sustainability, and grid stability. Model Predictive Control (MPC) emerges as a promising strategy to accomplish these objectives. The cost structure of microgrid utility bills typically comprises both electricity usage charges and demand charges, with the latter being linked to peak power consumption during billing cycles. As a predictive control strategy, to implement MPC requires accurate load forecasting during the control horizon. In this research endeavor, we employ representative machine learning algorithms and heuristic models for load forecasting, with a specific focus on understanding how forecast uncertainties would impact control performance, particularly in the presence of demand charges. Our findings demonstrate that MPC, when operating without taking into account demand charges, exhibits robustness against forecast uncertainties, yielding nearly optimal performance even when forecast errors reach 10% Mean Absolute Percentage Error (MAPE). However, when demand charges are introduced, even minor prediction errors as low as 1% MAPE lead to significantly worse control outcomes. We further uncover that this disparity arises from the asymmetric impact of forecasting errors, where underestimations of load have a more detrimental effect on MPC performance compared to overestimations. Moreover, we identify the limitations associated with employing singular aggregate metrics such as MAPE as indicators for MPC forecasts. We emphasize the importance of considering residual patterns and temporal correlations among consecutive steps to obtain a comprehensive evaluation of forecasting accuracy. Consequently, this study advocates for a thorough examination of forecasting features to establish a clear objective for upstream forecasting tasks, in order to enhance the performance and effectiveness of MPC in microgrid energy management.

**Keywords** Prediction uncertainty · Model predictive control · Building load forecast · Smart grid operation

## 1 Introduction

### 1.1 Background

Smart microgrids that incorporate distributed energy resources (DERs) play a crucial role in the pursuit of achieving carbon neutrality by integrating renewable energy generation [1]. However, this integration introduces significant challenges to the design, planning, and operation of the energy system. One prominent challenge is the growing temporal mismatch between renewable energy generation and peak electricity demand, known as the "Duck Curve" [2]. The Duck Curve exhibits two distinct features. Firstly, there is a substantial decline in net demand (total electricity demand minus renewable generation) during midday when solar energy generation is high [3], represented by the curve's "belly." This excess electricity production often leads to renewable curtailment. Secondly, there is a rapid increase in net demand, resembling the curve's "head." Meeting this sudden demand requires grid operators to rely on

---

\*Corresponding author: Z. Wang (cezhewang@ust.hk). L. Li and Y. Ju contributed equally to this work. This study was supported by the National Science Fund for Excellent Young Scholars (Grant number 52322813) and Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone (HZQB-KCZYB-2020083).

large-scale fossil fuel generators as backup, which is costly and environmentally detrimental. This temporal imbalance strains the grid and poses challenges to the operation of smart grids [4].

Addressing the challenges posed by the duck curve involves implementing energy storage solutions such as electrical energy storage (e.g., batteries) or thermal energy storage (e.g., chilled water tanks or ice storage). However, integrating renewable generation and energy storage increases system complexity, necessitating dedicated algorithms for system operation. Advanced control algorithms like Model Predictive Control (MPC) [5] and Reinforcement Learning (RL) [6] have been employed to optimize the coordination between building HVAC systems and energy storage, demonstrating superior performance compared to traditional rule-based control (RBC) strategies.

To optimize the design and operation of multiple heterogeneous but interconnected energy subsystems in an effective and reliable way is challenging [7], as this optimization is information-intensive. The predicted information about the building load, renewable generation is needed in this predictive control paradigm. How to obtain and process the extensive information to make reasonable predictions in an online manner is crucial for making informed decisions in microgrids [8].

## 1.2 Existing Works

### 1.2.1 Smart grid operation

Smart grid operation involves the coordination of multiple energy systems, such as energy storage charging/discharging, electric vehicle charging, and space pre-cooling/pre-heating, in an efficient and reliable way [9]. Within this domain, three main control paradigms are commonly employed: rule-based control (RBC), model-predictive control (MPC), and reinforcement learning (RL). RBC relies on a predefined set of rules to govern the operation of the smart grid. Examples of RBC strategies include maximizing self-consumption (MSC) and Time-of-Use (TOU) approaches discussed in [10]. Currently, RBC strategies dominate the industry due to their simplicity and interpretability. However, a drawback of RBC is that the predetermined rules often fail to lead to optimal control actions.

MPC and RL are two emerging advanced control strategies in the field of smart grid management research and applications. MPC formulates an optimization control problem based on a physical model of the dynamic system and repeatedly re-optimizes as new system states are observed [11]. In contrast, RL aims to learn a control strategy directly from reward maximization using a Markov decision process (MDP), without explicit knowledge of the system dynamics [12]. Although both MPC and RL are optimization-based control strategies, they differ in their approach to decision-making. MPC relies on a model of the system dynamics to make informed decisions, while RL takes a model-free approach by relying on data. Compared to RL, MPC is considered more data-efficient and interpretable for energy system management [13]. As a result, it has been extensively explored in the literature for various applications such as electric vehicle charging coordination [14], thermal storage operations [15], and hydrogen energy systems [16]. MPC has also been demonstrated in field studies, as exemplified by [17].

Within the microgrid context, economic Model Predictive Control (MPC) aims to optimize various control objectives, which may include utility costs, carbon emissions, system stability, and occupant satisfaction [18]. Utility costs typically comprise two primary components: energy charges and demand charges. Energy charges are determined based on the amount of electricity consumed, and the pricing scheme may consider factors such as time of day, season, and overall consumption levels [19]. On the other hand, demand charges are calculated based on the maximum power demand observed by the consumer during a billing cycle, and for certain commercial electricity consumers, they can account for more than half of the total electric bill. Houben et al. [20] highlighted that the inability to accurately predict peak demand may lead to insufficient energy stored in the battery energy storage system (BESS) to avoid a new peak demand, resulting in additional costs.

### 1.2.2 Building load prediction

The initial step towards predictive control involves making predictions regarding building load and renewable generation. Building load prediction is an area of active research due to its broad applications in optimizing HVAC control [21], operating thermal energy storage systems [22], planning energy distribution systems [23], and managing smart grids [24].

Approaches for forecasting building thermal load can be broadly classified into three categories: white-box physics-based models, gray-box reduced-order models, and black-box data-driven models [25]. White-box models utilize detailed heat and mass transfer equations to predict building loads. Commercial software tools like EnergyPlus, Dest, and TRNSYS are available for setting up white-box models [26]. Black-box models, on the other hand, rely solely on historical data to predict building thermal load. Gray-box models simplify the building's thermal dynamics by using reduced-order Resistance and Capacity (RC) models, and then estimate parameter values using historical data. For smart grid operations, black-box data-driven models are widely preferred for two main reasons. Firstly, advancements

in machine learning algorithms and the increased availability of historical data have made black-box models more accurate [27]. Secondly, developing physics-based models requires extensive input data, including thermal properties of the building envelope, occupancy and equipment schedules, which can be time-consuming, if not totally impossible, to gather. Some widely used black-box load prediction models will be introduced in greater details in Section. 2.2.

However, due to limited access to high-quality training data, potential changes in energy usage patterns, and system behaviors, uncertainties may arise in both demand and generation predictions [28]. Additionally, limited research directly addresses load prediction accuracy in specific downstream tasks. Studies have shown that under Model Predictive Control (MPC) or its surrogate neural networks (NN), the requirement for prediction accuracy appears to be moderate [29, 30]. However, these evaluations do not consider demand charges. Gust et al. suggest that such conclusions may not hold when demand charges are taken into account [31]. Stochastic programming is a general solution for addressing these challenges, but it often leads to a substantial increase in computational complexity.

### 1.3 Scope and objectives

Our literature review revealed a significant gap in the analysis of load forecast information's value in microgrid energy management, particularly in relation to the presence or absence of demand charges. This gap represents a notable disparity between two prominent clusters of research in the field. It is important to note that this concern has garnered attention from researchers within the broader context of *integrated learning & optimization* (ILO). For instance, Silwal et al. [32] propose a loss function, along with its convex surrogate, SPO+, which explicitly incorporates the optimization problem structure into the prediction loss. These pioneering works offer valuable intellectual insights; however, their applicability is unfortunately limited to a specific class of problems.

In this work, we aim at investigating the impact of building load forecasts on MPC of the microgrid, and examining the influences of demand charge in microgrid management. Our specific objectives are as follows:

1. Perform representative machine learning and heuristic models for load forecasting.
2. Compare the performance of MPC with different forecast models, when considering demand charge and not considering it.
3. Examine how do forecasting errors impact the control performance of MPC.

The rest of the manuscript is organized as follows: Sec. 2 details the workflow of this study, including forecasting approaches, control model for microgrid energy management and configurations for the simulations. Then in Sec. 3, we present the results of the simulations. In Sec. 4, we discuss and extend the results, before we conclude in Sec. 5.

## 2 Method

In this section, we present the work stream involved in MPC of microgrid energy management. The application of MPC consists of two main stages in this study: forecasting and optimization. In order to be coherent, we first introduce the optimization task and related concepts in Sec. 2.1, then detail the load forecasting in Sec. 2.2. In the initial stage, predictions for the loads are made based on historical data and other relevant information. Machine learning approaches and heuristic methods are used for forecasting in this study. Several lump metrics are included to evaluate the forecasting accuracy. Then an optimization problem is formulated for MPC, which takes various control variables into account and is subjected to specific constraints. Its objective is to find the control actions that minimize the performance criterion (i.e., utility cost in this study). In order to compare the control performance of MPC with different forecasts, we consider rule base control (RBC) as a benchmark. RBC is a controlling technique that does not rely on forecasts, based on which the concept value of information is proposed as a metric. At the end, two experimenting scenarios considering different billing schemes and corresponding configurations are introduced in Sec. 2.3.

### 2.1 Microgrid energy management

#### 2.1.1 System Configuration

In this paper, we denote the set of real numbers by  $\mathbb{R}$  and the set of integers by  $\mathbb{Z}$ . We use functions  $\mathbf{1}\{\cdot\}$ ,  $[\cdot]^+$ ,  $[\cdot]^-$ ,  $\lfloor \cdot \rfloor$  in their conventional meanings, i.e.,  $\mathbf{1}\{x\} = 1$  if  $x$  is true otherwise  $\mathbf{1}\{x\} = 0$ ,  $[x]^+ := \max\{0, x\}$ ,  $[x]^- := \min\{0, x\}$ .<sup>2</sup>

We consider a typical external-grid-connected microgrid, which is distributed with solar panel (PV) generation and battery energy storage system (BESS). The loads of the microgrid is building energy consumption. Specifically:

<sup>2</sup>For consistency, we adopt the convention that using lowercase letters (e.g.,  $x$ ) for (decision) variables, uppercase letters (e.g.,  $X$ ) and Greek letters (e.g.,  $\alpha$ ) for parameters, and calligraphic uppercase letters (e.g.,  $\mathcal{X}$ ) for sets.

- At time  $t$ , building load  $P_t^D$  and PV generation  $P_t^{PV}$  are assumed to be uncontrollable.
- The microgrid can import electricity power from, or export to the external grid, which is  $p_t^G$  at time  $t$  (negative for export), at time-of-use (TOU) tariff  $\gamma_t^+$  (purchase) or  $\gamma_t^-$  (sell).
- A battery with capacity  $B$  and maximum discharging (charging) power  $\bar{P}^B$  ( $\underline{P}^B$ ). At time  $t$ , its discharging power is  $p_t^B$ , for which negative value indicates charging.
- A *demand charge* term is included as part of the utility bill, which has a constant marginal rate  $\bar{\gamma}$  (per kW) applied to the maximum consumed power  $p^{DC}$  within the billing cycle.

### 2.1.2 Optimal Energy Management

The goal of the optimization is to minimize the electricity bills over horizon  $\mathcal{T} = \{0, 1, \dots, T\}$ , i.e.,:

$$\sum_{t=0}^{T-1} \left\{ \gamma_t^+ [p_t^G]^+ - \gamma_t^- [p_t^G]^- \right\} \Delta_t + \bar{\gamma} p^{DC} \quad (1a)$$

which is the consumption-based cost plus the demand charge.  $p^{DC} = \max_t \{p_t^G\}$ , which is the maximum grid import power within a billing cycle.<sup>3</sup> We can reformulate the maximum equality as:

$$p^{DC} \geq p_t^G, \quad \forall t \quad (1b)$$

Supply-demand balance should be observed at every time step:

$$p_t^G + p_t^B + P_t^{PV} = P_t^D, \quad \forall t \quad (1c)$$

Power imported from or exported to external grid is limited:

$$\underline{P}^G \leq p_t^G \leq \bar{P}^G, \quad \forall t \quad (1d)$$

If  $\underline{P}^G = \bar{P}^G = 0$ , the microgrid is isolated. In general, such limits are determined by the transformer capacity, per certain microgrid codes.

For the battery, the following constraints (dynamics) need to be satisfied<sup>4</sup>:

$$b_{t+1} = b_t + \left( \eta^- [p_t^B]^- - \frac{1}{\eta^+} [p_t^B]^+ \right) \Delta_t, \quad \forall t \quad (1e)$$

$$\underline{P}^B \leq p_t^B \leq \bar{P}^B, \quad \forall t \quad (1f)$$

$$\underline{B} \leq b_t \leq \bar{B}, \quad \forall t \quad (1g)$$

$$b_0 = B_0, \quad b_T \geq B_T \quad (1h)$$

where  $\eta^+$  ( $\eta^-$ ) is the discharging (charging) efficiency.

To conclude, a general formulation of optimal energy management is:

$$\min \quad \text{obj. (1a)} \quad (2a)$$

$$\text{s.t.} \quad \text{constr. (1b)-(1h)} \quad (2b)$$

With some trivial reformulations, the program (2) is a linear programming (LP).<sup>5</sup>

### 2.1.3 Model predictive control (MPC)

Compactly, we write the constrained finite-time optimal control (CFTOC) problem as:

$$\tilde{u}_{t_{[K]}} = \arg \min J(u_{t_{[K]}} \mid x_t, w_{t_{[K]}}, \theta_{t_{[K]}}) \quad (3)$$

where  $z_{t_{[K]}} = \{z_t, z_{t+1}, \dots, z_{t+K}\}$ , and  $x, u, w, \theta$ , by convention of control literature, are states, actions, disturbances, and parameters respectively. Disturbances  $w$  are unknown at the moment of solving CFTOC, so we substitute with the forecasts  $\hat{w}$  in the problem formulation. Under MPC, CFTOC is solved, but we only implement the first  $K^{\text{exe}} \leq K$  steps of its solutions before re-optimize. The procedure is shown in Alg. 1.

<sup>3</sup>For simplicity, we consider  $\mathcal{T}$  contains one billing cycle, which is typically one month.

<sup>4</sup>We consider an ideal battery model here for simplicity. We will plug in more complicated models, e.g., considering battery degradation, in our future work for sensitivity analysis.

<sup>5</sup>Some assumptions, e.g.,  $\gamma_t^+ \geq \gamma_t^-$  are required, which are also trivial.

**Algorithm 1:** model predictive control

---

```

 $t \leftarrow T_0, x_t \leftarrow X_0$ 
while  $t \leq T_{\text{end}}$  do
  forecast  $\hat{w}_{t[K]}$ 
  solve  $\tilde{u}_{t[K]} = \arg \min J(u_{t[K]} \mid x_t, \hat{w}_{t[K]}, \theta_{t[K]})$ 
  for  $k = 0, \dots, K^{\text{exe}} - 1$  do
    execute  $\tilde{u}_{t+k}$ , update  $x_{t+k}$ 
   $t \leftarrow t + K^{\text{exe}}$ 

```

---

In our specific context, states  $x$  are battery charge  $b_t$  and the maximum imported power in the billing cycle up to  $t$ , denoted as  $P_t^{\text{max}}$ . Actions  $u$  include schedules for battery charging  $p_{t[K]}^{\text{B}}$  and grid power exchange  $p_{t[K]}^{\text{G}}$ . Disturbances  $w$  refer to uncontrollable PV generations and building loads, whose estimates are denoted as  $\hat{P}_{t[K]|t}^{\text{PV}}$  and  $\hat{P}_{t[K]|t}^{\text{D}}$ .

In real execution, the battery will follow what are solved from the CFTOC.<sup>6</sup> When building load and PV generation are realized, the microgrid imports/exports the amount of energy from the grid that synchronously balances the supply and demand (instead of executing the one solved in CFTOC with predicted load and generation). Formally:

$$\hat{p}_t^{\text{G}} = \hat{P}_t^{\text{D}} - \hat{P}_t^{\text{PV}} - \hat{p}_t^{\text{B}} \quad (4)$$

### 2.1.4 Rule-based control (RBC)

Rule-based control (RBC) is a classical and widely used control paradigm. Different from optimization-based control (e.g. model predictive control), RBC relies on predefined rules to make decisions for controlling, and does not integrate the forecasts into decision making. One particularly notable RBC strategy, known as 'Maximize Self-Consumption' (MSC), has found widespread application in the realm of microgrid management, particularly in systems featuring renewable energy sources. The core tenet of MSC revolves around elevating the priority of locally generated power, specifically from PV generations and BESS, with the ultimate aim of curtailing dependence on the external grid. To be specific, PV generations will first satisfy the building load. When there is surplus power from PV, it will charge the battery, and lastly export to the external grid. When PV generations can not satisfy the energy requirement of building, the battery will discharged to provide the needed energy, and then importing from the external grid will be the last option if the power is still not enough. The control performance of RBC served as part of the benchmark when evaluating the performance of MPC with different forecasts.

### 2.1.5 Model predictive control performance

*Value of information* (VoI) is a useful concept for evaluating the impact of forecast accuracy on control performance. Let  $u^*(\hat{w})$  be the control sequence implemented under MPC when the estimates for disturbances  $w$  are  $\hat{w}$ , and  $J(u^*(\hat{w}); w)$  be the realized outcome of this control sequence. We define the value of forecast  $\hat{w}$  as the difference of it control performance and that when ground truth information  $w$  is available, i.e.,

$$V(\hat{w}; w) = J(u^*(w); w) - J(u^*(\hat{w}); w) \quad (5)$$

From the definition,  $V(\hat{w}; w) \leq 0$ , and the higher value  $V$  is, the more valuable  $\hat{w}$  is for the control task. In order to make meaningful comparisons across different scenarios, we define a *normalized value of information* (VoI\*) as:

$$\text{VoI}^* = \frac{J(u^*(\hat{w}); w) - J(u^{\text{RBC}}; w)}{J(u^*(w); w) - J(u^{\text{RBC}}; w)} \quad (6)$$

where  $u^{\text{RBC}}$  is the control sequence made maximize self-consumption strategy as introduced in Sec. 2.1.4. The forecast  $\hat{w}$  achieves 100% VoI\* if it has the same control performance as MPC with ground truth forecasts (MPC-GT) and it cannot do better. Forecast  $\hat{w}$  is useless if MPC with it cannot outperform a RBC strategy that does not require any forecast, in which case its VoI\* is 0%. VoI\* may even be negative, which means  $\hat{w}$  provides misleading information which leads MPC to make far suboptimal actions so that it is outperformed by RBC.

---

<sup>6</sup>We assume that such operations are always feasible and can be precisely implemented, since we focus on the uncertainty of building load.

## 2.2 Building load forecasts<sup>7</sup>

To diversify the forecasting algorithms and their forecast residual distributions, we selected two distinct categories of forecasting methods: *machine learning approaches* and *heuristic methods*. Four distinct and illustrative techniques for time series forecasting were utilized in the construction of the load prediction model: linear regression (LR), extreme gradient boosting (XGBoost), probabilistic forecasting incorporating autoregressive recurrent networks (DeepAR), and the temporal fusion transformer (TFT). Model training of LR, DeepAR and TFT is carried out based on the open source software *Darts* from *Unit8* [33], and the model working for XGBoost is performed on open source software from [34]. In addition, we denote MPC with machine learning forecasts by **MPC-ML**.

### 2.2.1 Load data

We utilized an open source dataset obtained from the microgrid at the University of California, San Diego (UCSD) [35]. The dataset consists of building load and PV generation data spanning from 2017 to 2019, sampled every 15 minutes. After conducting the required data quality control, we selected 8 buildings that were representative, and calculated their average load as a virtual profile. The average load, denoted by  $\mu_{\text{bld}}$ , was 73.6 kW. Evaluations on control performances were carried out on the data from the year 2019, while forecast models were trained (and validated) using data from 2017 and 2018.

### 2.2.2 Covariates

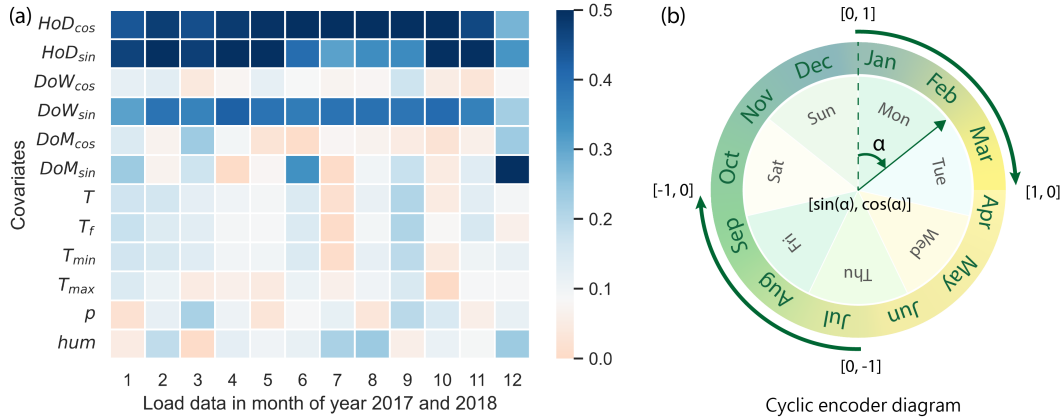


Figure 1: *Pearson correlation and cyclic encoder diagram (a)*. Absolute Pearson correlation values between covariates and load data in a month. Month is given in  $x$ -axis and type of covariates labeled in  $y$ -axis. Deeper blue indicates higher value of correlation, and orange indicates value below 0.1. *(b)*. The diagram shows the encoding rule of cyclic encoding. For simplicity, only month of year (MoY) and day of week (DoW) are shown in the diagram.

In the context of the forecasting tasks of this study, covariates pertain to external data that can be utilized as inputs for model enhancement seeking to improve forecasts. These external data encompass meteorological data ( $CO^M$ ) and time feature data ( $CO^T$ ). The target to predict is the series of building loads, and the covariates themselves are not predicted.

*Meteorological data* was retrieved from *OpenWeather* with an hourly granularity. The measured data includes: ambient air temperature ( $T$ ,  $^{\circ}C$ ), feels-like temperature ( $T_f$ ,  $^{\circ}C$ ), atmospheric pressure ( $p$ , mmHg) and relative humidity ( $hum$ , %). Having perceived that the daily load patterns of the profile are associated with the activity intensity during workday or holiday, we encode *time feature* covariates including hour of day ( $HoD$ ), day of week ( $DoW$ ), month of year ( $MoY$ ) and day of month ( $DoM$ ). These inputs use a cyclic encoder (see Fig. 1-b), which can help the model describe the periodic temporal features more objectively compared with the integer and one-hot approaches [36]. Furthermore, a boolean variable ( $D_{work}$ ) is included to indicate whether the day is a workday.

<sup>7</sup>For simplicity and consistency, we only consider building load forecast in this work, meanwhile assuming PV generations demand known. The error in PV generation forecasts can be easily combined with that of building load (in other words, forecasting the *net* load).



As shown in Fig. 1-a, most covariates we selected have correlation value higher than 0.1. Among all covariates, time features, particularly **HoD** and **DoW**, show strong relations with monthly load data. Although *meteorological data* has limited correlation with load data, it still provide information which can be useful for forecasting in several months.

### 2.2.3 Linear regression (LR)

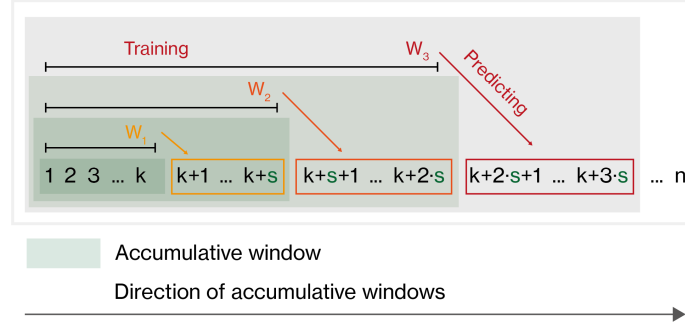


Figure 2: *Accumulative training diagram*.  $W$  represents the windows for each training,  $s$  is the prediction steps. In this study, the initial  $W_1$  is 61080 steps (data of year 2017 and 2018) and  $s$  is 96 steps (data of one day).

Linear regression is a statistical modeling technique used to establish a linear relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the variables and aims to find the best-fit line that minimizes the difference between the predicted values and the actual values. Linear regression is widely used for forecasting and prediction tasks because of its capability to achieve reliable predictive performance and cross-compatibility [37], especially when the relationship between the variables is believed to be linear. Compared with other complicated machine learning techniques, training a LR model is fast, even when dealing with large data sets. Thus, online forecasting is feasible, which is also known as accumulative training. The process of accumulative training is shown in Fig. 2. This method gradually expands the size of training data without overflowing historical data. Increasing training data can reduce the prediction errors [38]. We denote MPC with forecasts generated by this method as **MPC-LR**. Furthermore, we execute an additional forecasting technique utilizing LR by incorporating covariates namely time characteristics and meteorological data. The designation for MPC utilizing this forecast is **MPC-LRCo**.

### 2.2.4 Probabilistic forecasting with autoregressive recurrent networks (DeepAR)

Probabilistic forecasting with autoregressive recurrent networks (DeepAR) is a method that combines autoregressive models and recurrent neural networks (RNNs) to make probabilistic predictions. In some cases, long short-term memory (LSTM) can be used to replace the RNN. AR models capture dependencies between past and current observations, while RNNs are designed to model sequential data. By combining these two techniques, probabilistic forecasting with AR networks can generate predictions along with a measure of uncertainty, providing a probabilistic distribution of future values. DeepAR is good at learning seasonal behaviors and dependencies on given covariates across time series. At the same time, little manual intervention in providing covariates is needed in order to capture complex behaviors [39]. In this method, only future covariates (i.e. time features) are encoded due to the limitation of RNN model. MPC with this prediction is denoted as **MPC-DeepAR**.

### 2.2.5 Temporal Fusion Transformer (TFT)

The Temporal Fusion Transformer (TFT) is a forecasting model that combines elements of recurrent neural networks and transformers. It is designed to capture complex temporal patterns in time series data. TFT uses a hierarchical architecture that includes encoder-decoder layers, attention mechanisms, and gating mechanisms. It can handle multiple time series inputs and can model long-range dependencies effectively. TFT has been shown to deliver accurate and interpretable predictions for various forecasting tasks [40]. Major constituents of TFT related to our forecasting task include gating mechanisms, variable selection networks and temporal processing. Gating mechanisms and variable selection networks help select relevant input variables at each time step. In the temporal processing mechanism, a sequence-to-sequence layer is employed for local processing of short-term dependencies, while long-term temporal relations are captured using a novel interpretable multi-head attention block. MPC with this prediction is denoted as **MPC-TFT**.

Model parameters	TFT	DeepAR	XGBoost
used covariates	$CO^T$	$CO^T$	$CO^T, CO^M$
epochs	[1, 100]	[1, 50]	[1, 250]
optimization trials	100	100	30
learning rate	$[10^{-4}, 1]$	$[10^{-4}, 1]$	$[10^{-3}, 1]$
batch size	96	[8, 672]	-
dropout rate	[0.01, 1]	[0.01, 1]	-
RNN layers	-	[1, 8]	-
LSTM layers	[1, 8]	-	-
hidden dimensions	[1, 8]	-	-
attention heads	[1, 8]	-	-
minimum child weight	-	-	[5, 10]
maximum depth	-	-	[1, 5]
gamma	-	-	[0.25, 1]
subsample	-	-	[0.01, 0.2]
colsample bytree	-	-	[0.01, 0.2]
regularization alpha	-	-	$[10^{-8}, 1]$
regularization lambda	-	-	$[10^{-8}, 1]$

Table 1: Hyperparameters for *Optuna* searching

### 2.2.6 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is a popular machine learning algorithm known for its efficiency and predictive power. XGBoost is an ensemble method that combines multiple weak prediction models, typically decision trees, into a strong predictive model. It uses gradient boosting, which sequentially trains models by focusing on the errors made by previous models [34]. XGBoost incorporates regularization techniques to prevent overfitting and provides high accuracy for various prediction tasks, including regression and classification. XGBoost is widely used for building-scale [25] and city-scale [41] load forecasting, and is also a frequent winner in Kaggle forecasting competitions [42]. Its performance mainly attributes to these innovations: the tree-learning algorithm boosting sparse data process, a theoretically justified weighted quantile sketch procedure enabling approximate tree learning, and parallelisation together with distributed computing shrinking the training time. Benefited by the tree-learning algorithm, both meteorological and time feature covariates are utilized in XGBoost training. MPC with this prediction is denoted as **MPC-XGBoost**.

### 2.2.7 Hyperparameter tuning

For machine learning tasks, hyperparameter searching is a cumbersome and time consuming process. Due to model complexity and large size of training data, conventional grid-search approach is not applicable in this study. Thus we chose *Optuna*, an efficient way of hyperparameter tuning adopting efficient searching and pruning algorithms [43], as our hyperparameter tuning method. Corresponding configurations for DeepAR, TFT and XGBoost are listed in Table. 1.

### 2.2.8 Adaptive moving average (AMA)

In contrast with *machine learning* approaches, *heuristic* methods are simple and intuitive. We adopt a quite simple heuristic model for building load forecast. Specifically, the estimate for time step  $t + k$  forecast at time  $t$ :

$$\hat{P}_{t+k|t}^D = \alpha^k P_{t-1}^D + (1 - \alpha^k) \bar{P}_{\mathcal{T}(t+k)}^D \quad (7)$$

where  $\bar{P}_{\mathcal{A}}^D = \frac{1}{|\mathcal{A}|} \sum_{t \in \mathcal{A}} P_t^D$  is the average of some selected historical data, and for our specific choice,  $\mathcal{T}(t)$  is the set of the most recent four steps whose day-of-week (DoW) and time-of-day (ToD) are the same as  $t$ . We pick  $\alpha = 0.1$ . Since it combines last observed data with exponentially decaying historical data as the prediction window moving forward, we call this implementation as adaptive moving average (AMA). MPC with this prediction is denoted as **MPC-AMA**.

### 2.2.9 Forecast accuracy

The forecasting performances are primarily evaluated by using four (lump) metrics, as defined in Eqs. (8)-(11), where  $y_t$  and  $\hat{y}_t$  are actual and forecasting values. They are the mean absolute error (MAE), the root mean squared error



(RMSE), the mean absolute percentage error (MAPE) and the coefficient of variation of the root mean squared error (CV-RMSE). MAE and RMSE are scale-dependent which can provide a straight-forward quantification to the readers, while MAPE and CV-RMSE are normalized metrics so it is more comparable across specific buildings. In addition, existing studies and guidelines provided some benchmarks for evaluating the model by using CV-RMSE. Specifically, resulting CV-RMSE below 30% indicates the model is calibrated and sufficiently close to physical reality for engineering purposes when using hourly data [44].

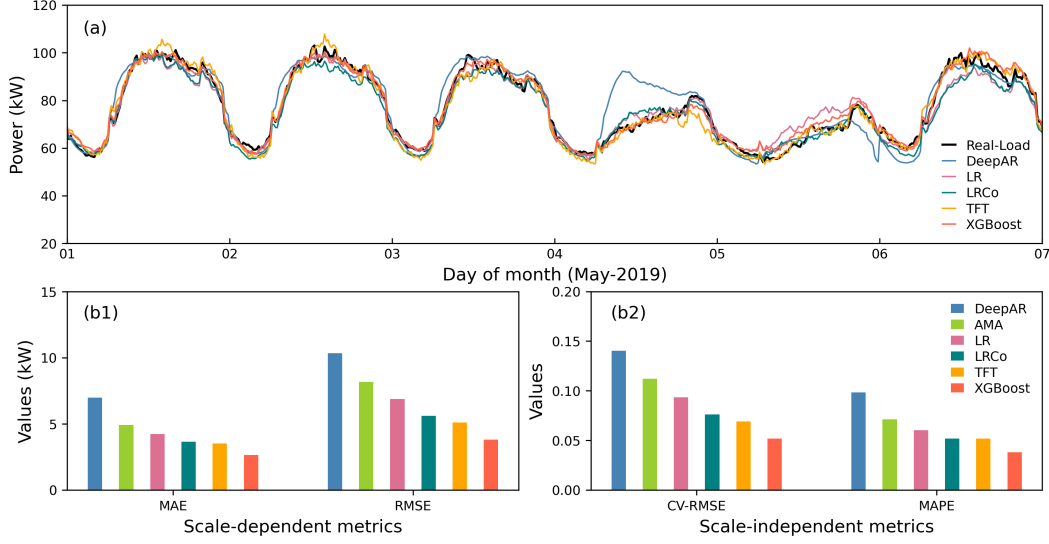


Figure 3: *Sample forecasts and error metrics (a)*. Real load values are represented by black line, and forecasts of different machine learning models are colored lines as tagged in legend. The AMA generates adaptive forecasts, which are not plotted in this figure. *(b1)-(b2)* Visualizing the scale-dependent and scale-independent metrics of different forecasts.

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (8)$$

$$RMSE = \frac{1}{n} \sqrt{\sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (9)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (10)$$

$$CV - RMSE = \frac{\sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}}}{\frac{\sum_{t=1}^n y_t}{n}} \quad (11)$$

Fig. 3. provides sample forecasts of a week, and summarizes the resulting MAE, RMSE, CV-RMSE and MAPE. In general, all forecasting models outperform the threshold of 30% in terms of CV-RMSE, showcasing a range from 5.17% to 14.04%. Compared to performances provided by other related research [45], the resulting CV-RMSE is lower. This enhanced performance can be attributed to the nature of the building load data employed in this study, which represents the average demand of eight individual buildings. This aggregation serves to mitigate the inherent uncertainty in load patterns, resulting in more robust forecasting outcomes. Among all accessed methods, XGBoost emerges as the frontrunner, exhibiting superior performance with a MAPE of 3.80%. TFT has the resulting MAPE of 5.19% inferior to XGBoost. Though the structure of LR is rather simple, it exerts good performance of 6.01% (MAPE), as a result of constantly retraining. By encoding covariates, the MAPE of LRCo is enhanced to 5.18%. Of particular note is the heuristic method AMA, which defies expectations with its robust performance, despite its simplicity. Leveraging observations from the preceding time step, AMA attains a MAPE of 7.13%, surpassing the performance of the more intricate DeepAR model, which registers a MAPE of 9.83%.

### 2.3 Simulation configurations

Based on streamed PV and building load data and corresponding forecasts in the year of 2019, the simulations are performed under two scenarios ( $\text{DC}^0$  and  $\text{DC}^+$ ) to examine the impact of pricing schemes. Under  $\text{DC}^0$  (without demand charge), the utility bills consists of only TOU (time of use) costs, where the demand charge rate  $\bar{\gamma}$  was set as 0 \$/kW. Under  $\text{DC}^+$  (with demand charge), the demand charge rate  $\bar{\gamma}$  was set as 18 \$/kW in a billing cycle of one month, representing a typical scenario in California. Apart from the demand charge rate  $\bar{\gamma}$ , other configurations for both scenarios are set as follows:

In order to make our analysis more generalizable, we intentionally rescaled the annual PV generation and the battery capacity, then they are proportional to the annual building load. The factors, namely  $\phi_{\text{pv}}$  and  $k_{\text{bat}}$  were set as 50% and 6h in this study. Battery (dis)charging efficiencies were set as 0.98. We used the day-ahead price (DA) of San Diego Gas & Electric (SDGE) as the grid TOU  $\gamma^+$ <sup>8</sup>. Selling prices  $\gamma^-$  were set as fixed proportion to buying prices  $\gamma^+$ , and the factor was 0.6 in our base case.

## 3 Results

### 3.1 The impact of pricing schemes

The first finding of this study reveals that the performance of MPC varies under different pricing schemes, namely  $\text{DC}^0$  and  $\text{DC}^+$ , despite having the same forecast. Under  $\text{DC}^0$ , enhancing forecast accuracy does not yield equitable progress for MPC in terms of VoI\*. And MPC with inferior forecasts attains near optimal control performance. In contrast, under  $\text{DC}^+$ , MPC with highly precise forecasts exhibits VoI\* that is inferior compared to RBC.

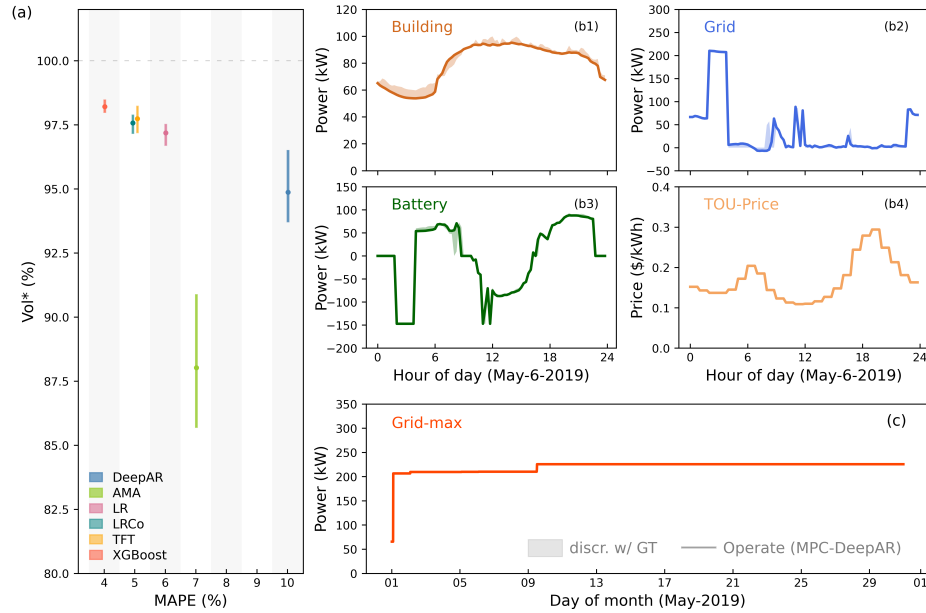


Figure 4: *Control performance and sample control sequences under  $\text{DC}^0$ .* (a) Performance of MPC with all available predictions as stated in previous sections.  $x$ -axis shows the forecasting error (MAPE), and  $y$ -axis is the corresponding normalized value of information (VoI\*). Each bar marks the 25% and 75% percentiles of trials for 12 months of 2019. (b1)-(b4) Sample control sequences of MPC-DeepAR. (c) Peak imported power tracking till the moment. (In (b1)-(b4) and (c), solid line represents actions that are actually executed, and shaded area shows discrepancy between executed actions and optimal actions by MPC-GT. For building, shaded area means prediction error.)

### Under $\text{DC}^0$

Fig. 4 presents the quantified control performance and sample control sequences under  $\text{DC}^0$ . Fig. 4-a gives the performance of MPC with all forecasts, whose prediction errors (MAPE) vary significantly, ranging from 3.8%

<sup>8</sup>Considering the price difference between wholesale markets and business plans, we rescaled TOU so that its average was 0.17 \$/kWh. We assume DA is known.

(XGBoost) to 9.83% (DeepAR). The control performances for the majority of monthly trials achieved near-optimal levels of VoI\* around 90%.

Fig. 4-b1 to b4 present control sequences of **MPC-DeepAR**, which demonstrate near-identical control behaviors to the optimal ones. The TOU tariff peaks at around 7:00 and 19:00, when the external grid importation is decreased to almost zero by using the BESS as a substitute, which was charged beforehand during low-TOU-tariff periods. Throughout the process, the shaded area in Fig. 4-b1. demonstrates that the forecast errors do not result in significant sub-optimality. This suggests that the forecasts provide ample information for the CFTOC problem to effectively utilize the gap between peak and valley of TOU tariff, thereby reducing the utility cost to near-optimal levels.

Under the  $DC^0$  scenario, enhancing the forecast accuracy does not result in significant control performance enhancement. Despite reducing the MAPE from 9.83% (DeepAR) to 3.80% (XGBoost), the average VoI\* increased by only 3.34% from an already high level of 94.87% to 98.21%. Furthermore, compared to machine learning approaches, AMA is an ideal method for obtaining the near-future loads forecasts under  $DC^0$ . Since AMA requires limited historical data<sup>9</sup> and is computationally efficient. It can be easily implemented whilst exhibiting a strong performance of 87.9% with respect to VoI\*.

#### Under $DC^+$

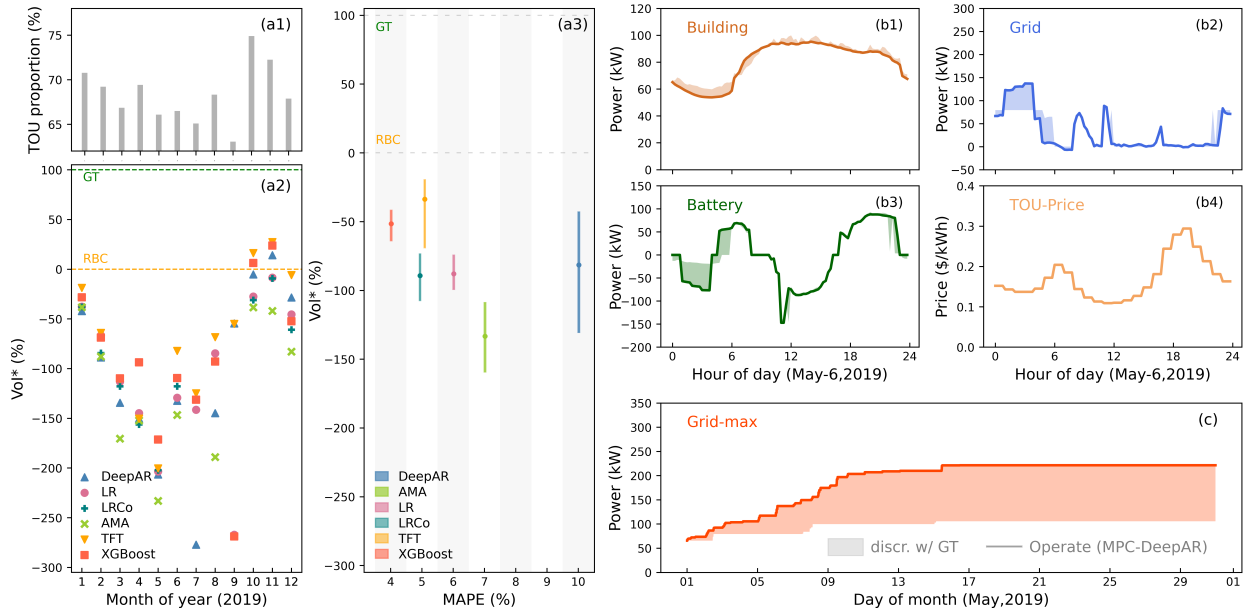


Figure 5: Control performance and sample control sequences under  $DC^+$ . (a1) Bars show the proportion of TOU cost in the utility bill of month in year 2019. The values are calculated based on the **MPC-GT**. (a2) Each marker shows the performance of a monthly trial. Different marker styles represent prediction methods.  $x$ -axis shows months in year 2019 and  $y$ -axis is the corresponding normalized value of information (VoI\*). (a3), (b1)-(b4) and (c) follow the same rules as in Fig. 4.

Under the scenario of  $DC^+$ , we applied a demand charge rate of 18\$/kW per month while maintaining other configurations unchanged. The corresponding results are presented in Fig.5. The control performance is significantly worse than under  $DC^0$ . And even the MPC with the best forecasts, XGBoost in this study, has inferior control performance in comparison to the baseline RBC.

Based on the sample control sequences shown in Fig. 5-b1 to b4, the majority of control actions show similarities to optimal ones, but fail in establishing the proper boundary of peak demand. In the initial 6 hours of the day, excessive power than needed was imported from the external grid to pre-charge the battery. These operations can yield benefits by storing energy for high-TOU-tariff hours. However, raising peak demand unnecessarily leads to a serious decline in economic performance. Throughout the billing cycle, the peak demand was pushed up successively up to 221.3 kW, which is much higher than the optimal level of 105.6 kW. Since demand charge makes up a significant proportion (around 25% to 60%) of utility bills, the OPEX rises higher than that of RBC. As the proportion of demand charge

<sup>9</sup>Typically, utilizing data from the previous month is sufficient for achieving relatively accurate forecasts.

increases (i.e. the proportion of TOU cost decreases), the control performance worsens, which is demonstrated in Fig. 5-a1 and a2.

### 3.2 Asymmetric impact of forecast errors

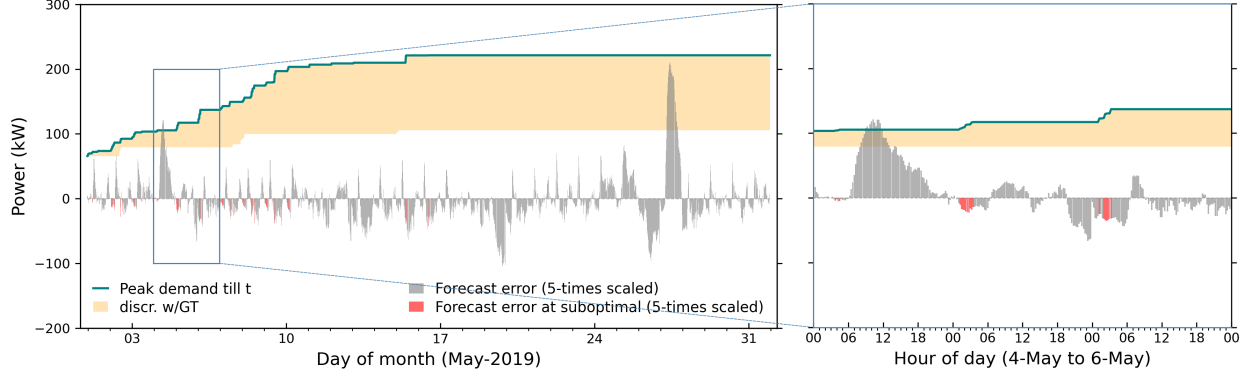


Figure 6: Track peak demand and forecast errors of **MPC-DeepAR** under  $DC^+$ . Green line is the peak demand till time  $t$  of **MPC-DeepAR** control sequence. Yellow shaded area is the discrepancy of peak demand with MPC-GT. Gray bars are the forecast error at every time step. When the discrepancy (yellow shaded area) increases, the error bars are highlighted as red. All forecast errors are **5-times scaled up** for better visibility.

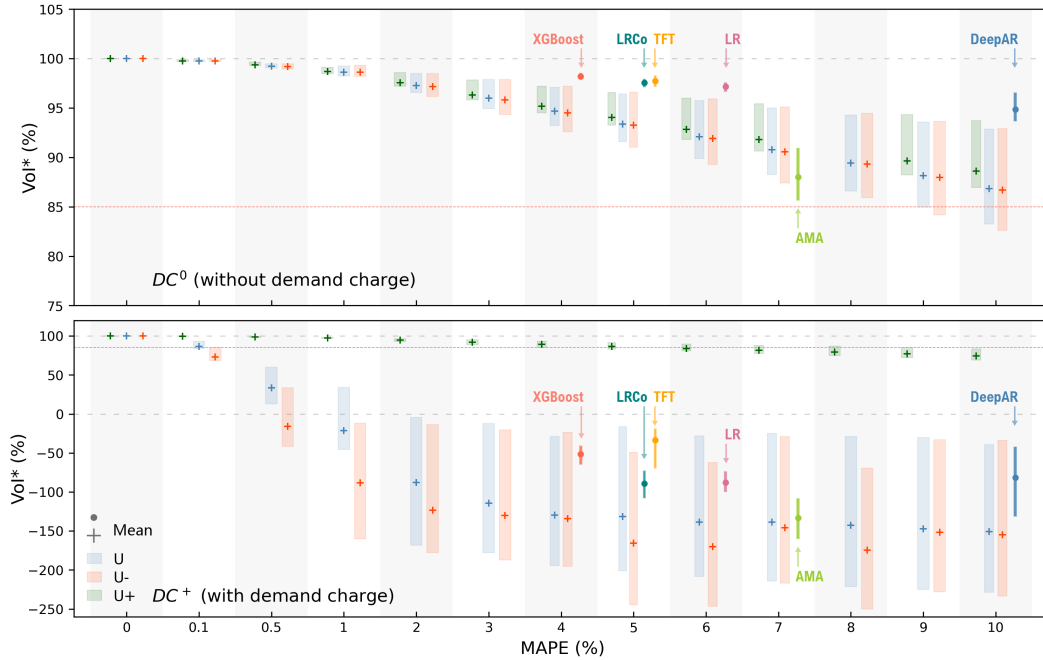


Figure 7: Control performance of **MPC- $\epsilon$**  and **MPC-ML** under  $DC^+$  and  $DC^0$ .  $x$ -axis shows the magnitude of forecasts, and  $y$ -axis is the corresponding normalized value of information ( $Vol^*$ ). Semi-transparent bars are performances of **MPC- $\epsilon$** , colors represent noise directions (blue: unbiased, orange: underestimate, green: overestimate). Narrow bars with emphasized colors are performances of **MPC-ML**, as noted in the tags. Each bar marks the 25% and 75% percentiles of trials for 12 months in 2019. “+” and “-” marks their mean values. Note the  $x$ -axis ticks are not even. Gaped gray shading areas merely help with identifying bars within one MAPE group.

When examining the trend of peak demand in conjunction with the forecast errors under  $DC^+$ , as depicted in Fig. 6, it becomes evident that the forecast error exerts a biased influence. The tracked peak demand remains consistent with the optimal curve under load overestimates (up to 40 kW, i.e., 54.3% of the yearly average load). It only deviates from and

exceeds the optimal one at time steps when the load is underestimated (marked as red bars). The explanation for this is that if the load is underestimated, more power needs to be imported from the external grid to maintain energy balance. This results in a higher executed  $p_t^G$  than expected in the CFTOC solutions, consequently pushing up the tracked peak demand  $P_t^{\max}$  irreversibly.

To further reveal the asymmetric impact of the forecast errors by decoupling it with the error magnitude, we generate synthesized forecasts by artificially adding random noises on true values. We consider adding three types of *uniformly distributed* noises: for a given MAPE  $\epsilon$ ,  $\mathbf{U} = \mathbf{U}(-2\epsilon, 2\epsilon)$ ,  $\mathbf{U}^+ = \mathbf{U}(0, 2\epsilon)$ ,  $\mathbf{U}^- = \mathbf{U}(-2\epsilon, 0)$ , and forecasts

$$\hat{P}_{t+k|t}^D = (1 + \xi_{t+k})P_{t+k}^D, \xi_{t+k} \sim \mathbf{U}(\text{or } \mathbf{U}^+, \mathbf{U}^-) \text{ i.i.d.} \quad (12)$$

According to corresponding noises sampling rules,  $\mathbf{U}^+$  always overestimates the load whereas  $\mathbf{U}^-$  always underestimates the load, and  $\mathbf{U}$  generates unbiased forecasts. Hereafter, we call MPC control with artificial noise as **MPC- $\epsilon$** . Then we vary the magnitude of forecasting MAPE from 0.1% to 10%<sup>10</sup> for  $\mathbf{U}$ ,  $\mathbf{U}^+$  and  $\mathbf{U}^-$ , and run the simulation month by month in 2019. Corresponding results are provided in Fig. 7. We also mark the control performances with practical forecast models, i.e., ML and AMA models, on the same subplots in Fig. 7.

Under  $\mathbf{DC}^0$ , the control performances do not exert notable asymmetric impact. All tested cases have near-optimal performances, and corresponding average Vol\* is higher than 85%. However, for  $\mathbf{DC}^+$ , the asymmetric impact of forecasting error is notable. **MPC- $\epsilon$**  with purely overestimating forecasts ( $\mathbf{U}^+$ ) outperforms the others significantly, and it still has near-optimal performances. While MPC with other forecasts containing (more or less) underestimating errors has performance inferior to RBC for most cases. In general, the relationship between forecasting error (MAPE) and Vol\* revealed by our simple **MPC- $\epsilon$**  model is consistent with other “real” models (i.e. **MPC-ML** and **MPC-AMA**). Though these “real” models generate nearly unbiased forecasts, their performance is slightly better than **MPC- $\epsilon$**  with  $\mathbf{U}$  which has unbiased forecasts as well. And it is noteworthy that their variances across different months are significantly smaller. Moreover, Vol\* is not monotonically increasing with the decrease of MAPE. For instance, **MPC-XGBoost** has lower MAPE than **MPC-TFT**, but its Vol\* is not higher than that of **MPC-TFT** under  $\mathbf{DC}^+$ . Such nuances hint us that more metrics besides MAPE need to be considered to characterizing the relationship between Vol\* and load forecast accuracy. And we extend this part in the discussions.

## 4 Discussion

### 4.1 Forecast residuals and control performance

The preceding observations suggest that the MAPE may not effectively serve as an adequate indicator of MPC’s performance, especially when considering demand charges. Therefore, we incorporated three additional metrics that encapsulate aspects of the *worst-case* scenario, *variance*, and *magnitude of mean values* for forecast residuals, as outlined in Eqs. (13) to (15), where  $\hat{y}_t$  and  $y_t$  represent the forecasted and actual values, respectively. The minimum value of underestimation:

$$\text{Min}^- = \min\{\hat{y}_t - y_t | 0 < t < n\} \quad (13)$$

The standard deviation  $\sigma$  indicating the fluctuation of residuals:

$$\sigma = \sqrt{\frac{1}{n} \sum_{k=1}^n (\hat{y}_t - \bar{y})^2} \quad (14)$$

The average value of forecasting errors:

$$\text{ME} = \frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t) \quad (15)$$

In order to observe more samples, we present the monthly values individually, as depicted in Figure 8. Notably, our chosen metrics exhibit a rather significant correlation with the Vol\* metric under  $\mathbf{DC}^0$ , with  $\sigma$  and MAPE revealing a discernible relationship. However, under  $\mathbf{DC}^+$ , increased uncertainties obscure such correlations, with only the ME metric indicating that higher ME values tend to correspond to higher Vol\* outcomes.

Moreover, analyzing the distribution of forecast residuals may provide more insightful information for interpreting the influence of forecast errors, as shown in Fig. 9. It is evident that the forecast residuals distribute normally and

<sup>10</sup>This range is selected intentionally to cover the MAPE of all forecast methods used in previous sections from 3.80% (XGBoost) to 9.83% (DeepAR). The lower bound of 0.1% aims at gapping forecast and ground truth.

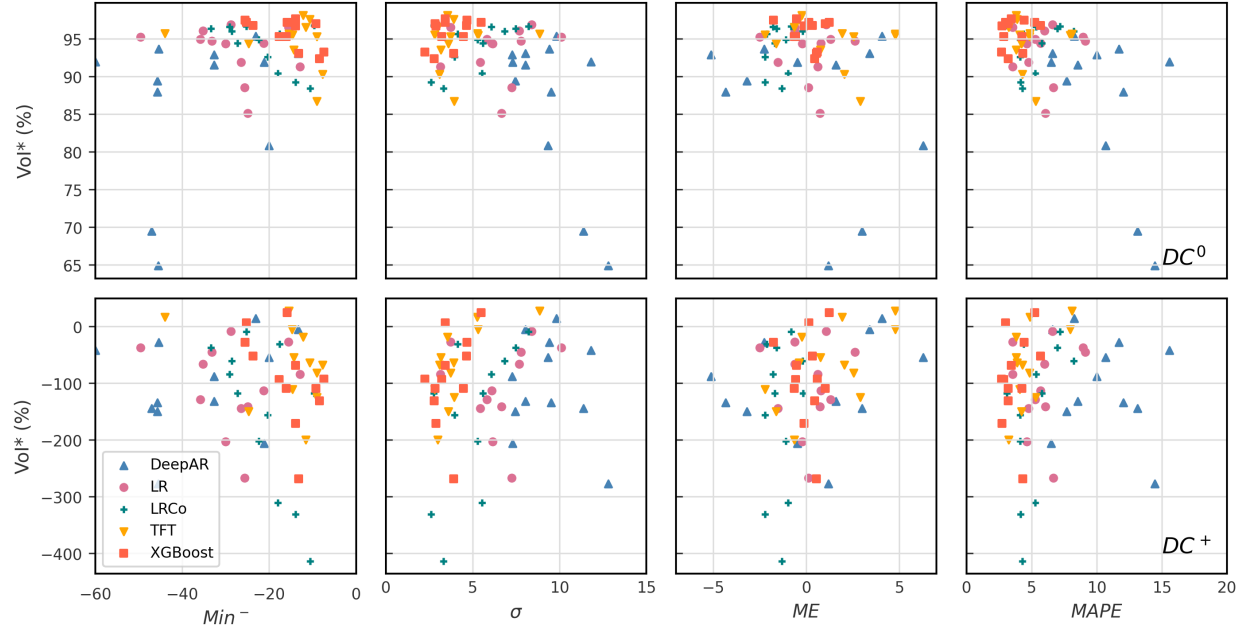


Figure 8:  $x$ -axis is values of metrics differed in each column.  $y$ -axis is the corresponding normalized value of information (Vol\*). Each marker represents one monthly trial. Different forecast models are marked by varied markers.

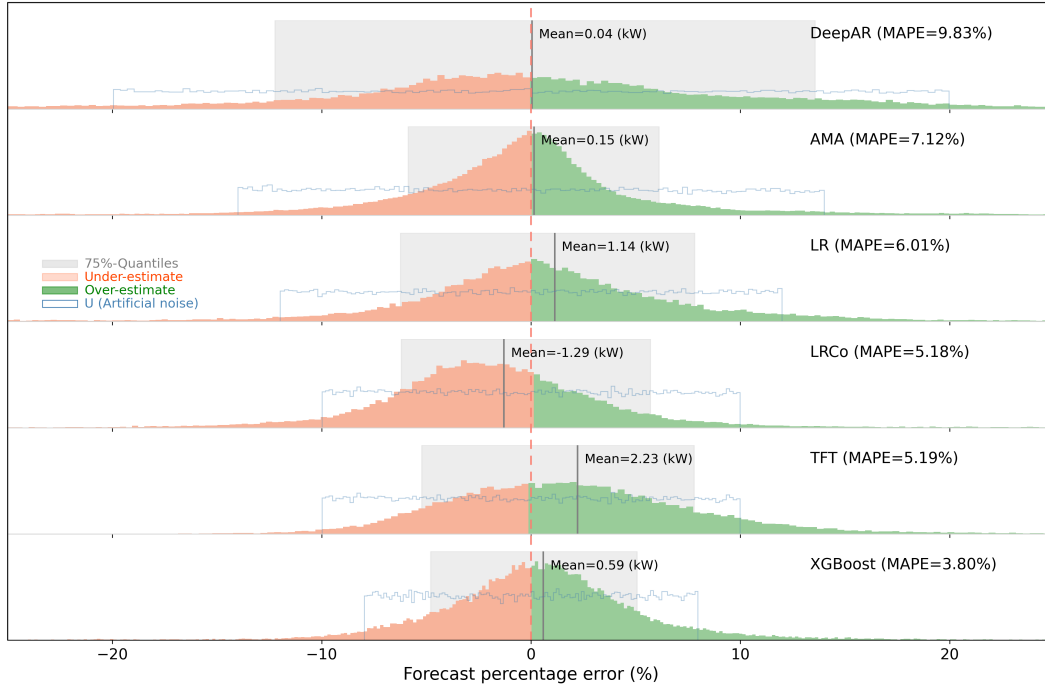


Figure 9: Forecast error distribution compared with artificial noise ( $U$ ).  $x$ -axis shows the value of forecast percentage error. Histogram in each subplot shows the error distribution of one specific prediction model tagged in the upper right. Orange bars are underestimated part, and green ones are overestimated part. Gray shading area marks the 75% percentiles of the distribution. Mean value of forecasting errors is tagged. For comparison, the error distribution of unbiased artificial noise ( $U$ ) is plot with blue lines, and corresponding magnitude, namely MAPE, is close to that of the model in the same subplot.



have a more centralized distribution compared to artificial noise. Additionally, there are variations in the distribution pattern among different forecasts. DeepAR, AMA and XGBoost exhibit relatively symmetrical distribution. TFT and LR demonstrate a greater amount of residuals in the overestimated section, clarifying why **MPC-TFT** and **MPC-LR** exceed performance expectations. This is attributed to overestimated forecasts shown to perform better under  $DC^+$ , as analysed in Sec. 3.2. LRCO, in contrast, have more underestimated residuals, thus **MPC-LRCO** exhibits performance inferior to TFT which has similar MAPE.

These observations highlight the need for the development of synthesized error metrics to more comprehensively uncover the intricate relationship between forecast residuals and MPC control performance within the domain of microgrid management. Singular and lump metrics, while informative, can capture only a single dimension of the residual characteristics. Other factors tied to time series patterns hold greater significance for MPC control challenges, as they exert a direct influence on the execution stage of MPC strategies. Our future research endeavors will be dedicated to further exploring and elucidating these relationships.

#### 4.2 Value of forecasts in the scope of time

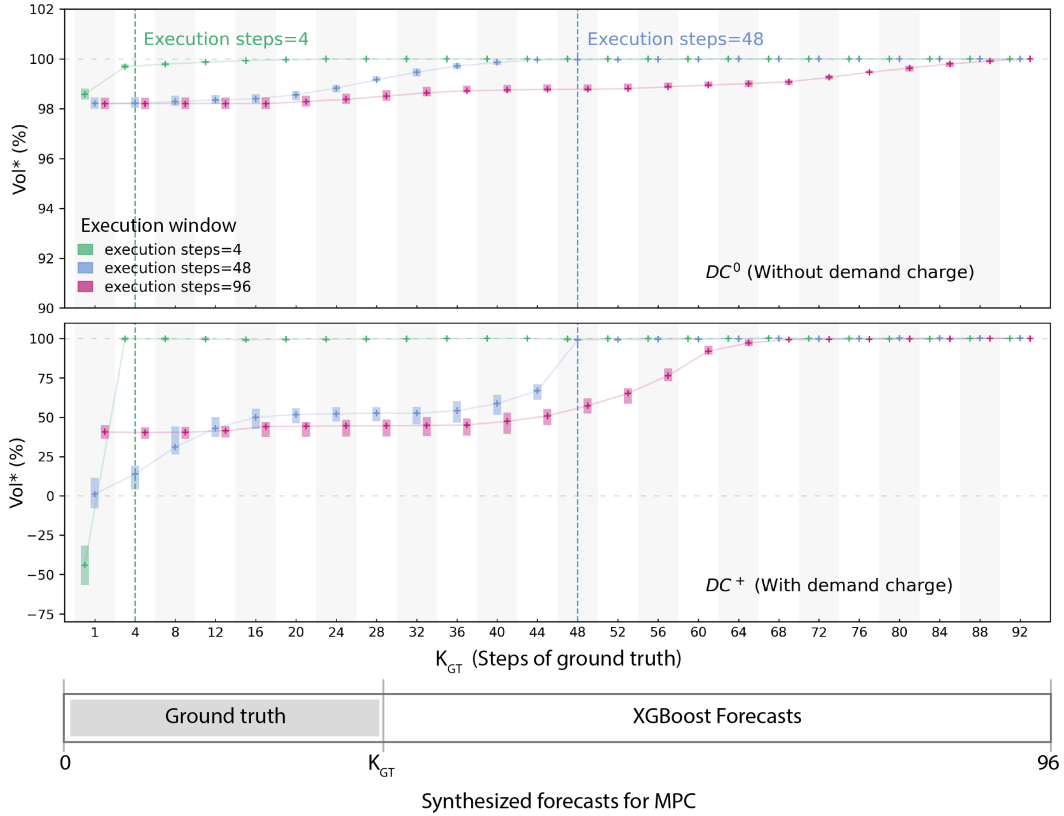


Figure 10:  $Vol^*$  with varied forecast in time-scope.  $x$ -axis is the steps of ground truth given to each CFTOC problem. How the forecasts are concatenated is shown in the diagram at the bottom.  $y$ -axis is the corresponding normalized value of information ( $Vol^*$ ). Each bar marks the 25% and 75% percentiles of trials for 12 months in 2019, and "+" marks their mean values.

In this study, the MPC optimization is a dynamic process. Within a single control window,  $K$  steps of forecasts are considered in the linear programming (LP), and only  $K_{exe}$  steps of solutions are executed<sup>11</sup>. Consequently, it is pertinent to examine the value of the prediction over time. A straightforward approach would be to create a combined prediction by merging accurate data with forecasts. As shown in Fig. 10, we concatenate ground truth  $\{P_t^D \mid 0 \leq t \leq K_{GT}\}$  and forecasts  $\{\hat{P}_t^D \mid K_{GT} < t < K\}$  as synthesized forecasts for each step, and then conduct monthly simulations in 2019 while varying  $K_{GT}$  between 0 and 96. The simulations were conducted in three groups for  $K_{exe}$  values of 4, 48, and 96, as depicted by three colors.

<sup>11</sup>In this study,  $K$  is set as 96 and  $K_{exe}$  is 4, which correspond to 24h and 1h.

The key observation is that the forecast error taking place within the execution window exerts a significant impact on control performance, compared to the negligible influence of errors occurring outside this window. This is because forecasts beyond the execution window help the LP detect and adapt to future trends, and associated errors do not have a direct effect on control performance. In contrast, forecast errors occurring within the execution window cause a direct mismatch between the solutions and executions, resulting in sub-optimality.

Furthermore, there is another phenomenon that is somewhat counter-intuitive. In the case of  $\text{DC}^0$ , frequent re-optimization yields better performance as expected albeit with limited improvement. However, under  $\text{DC}^+$ , frequent re-optimization leads to deteriorating control performance. As depicted in Fig. 10, when  $K_{GT}$  is 1 and under  $\text{DC}^+$ , MPC with execution window of 96 steps (purple bars) outperform that with 4 steps (green bars). As examined in Sec. 3.2, frequent re-optimisation leads to offering inaccurate information more frequently. Therefore, decreasing the frequency of optimization when highly-accurate forecasts are unavailable under  $\text{DC}^+$  refines performance control and diminishes the computational load.

## 5 Conclusion and future work

In this research paper, we investigate the influence of forecast accuracy on the control performance in the context of smart grid operation. This is of utmost importance because the ultimate goal is to achieve effective control performance rather than merely accurate predictions. Furthermore, the presence of different pricing schemes adds complexity to optimizing control performance. Particularly, the pricing scheme that includes demand charges in the utility bill poses a significant obstacle. The longer billing cycle associated with demand charges compared to time-of-use costs makes it challenging to effectively control demand charges within a short optimization window for Model Predictive Control (MPC). Therefore, we thoroughly analyze the impact of prediction uncertainties with and without considering demand charges, and our findings reveal the following three key points:

1. When demand charges are not taken into account, MPC achieves near-optimal performance even when the Mean Absolute Percentage Error (MAPE) of forecast reaches 10%. Improving the accuracy of the forecasts has a limited contribution to enhancing control performance. Moreover, the simple Adaptive Moving Average (AMA) model proves to be an ideal forecasting model as it is computationally efficient, requires minimal historical data, and achieves satisfactory forecasting performance. However, when considering demand charges, even 1% MAPE significantly deteriorates control performance to a level lower than Rule-Based Control (RBC).
2. Through simulations involving artificial noise, including underestimation and overestimation, we uncover the asymmetric impact of forecasts when demand charges are considered. Underestimating future load leads to a substantial increase in peak demand, resulting in poor control performance.
3. Singular lump metrics such as MAPE cannot serve as reliable indicators of MPC control performance, especially when demand charges are involved. These metrics overlook the residual distribution patterns and temporal correlations, which hold valuable information for MPC. Therefore, more intricate forecast features should be considered, aiming to provide a clear objective for upstream forecasting tasks.

Based on these findings, it is important to acknowledge certain limitations in this study. The testbed formulated in this research is relatively simplistic, as it does not account for electric vehicles, despite the growing trend of incorporating charging stations in microgrid management. Additionally, we did not investigate the battery degradation in the Battery Energy Storage System (BESS). Furthermore, we only examined a few system configurations, which may limit the generalizability of our findings.

In addition to the MPC strategy implemented in this study, there are several solutions available to address prediction uncertainties, such as robust MPC and stochastic MPC. As a next step, we plan to evaluate their performance. Both robust MPC and stochastic MPC incorporate forecast uncertainty into the optimization stage, which may help mitigate the asymmetric impact of forecasting errors revealed in this study.

## 6 Acknowledgement

This study was supported by the National Science Fund for Excellent Young Scholars (Grant number 52322813). This work was supported in part by the Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone (HZQB-KCZYB-2020083).

## 7 Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used *DeepL* in order to polish the language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## References

- [1] Y. Li, X. Zhang, F. Xiao, W. Gao, and Y. Liu, “Modeling and management performances of distributed energy resource for demand flexibility in japanese zero energy house,” in *Building Simulation*. Springer, 2023, pp. 1–16.
- [2] O. Kwon, S. Lee, and J. Park, “A numerical study to compensate duck curve of ess integrated gas turbine system with reused-battery,” *Journal of Energy Storage*, vol. 55, p. 105422, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352152X22014141>
- [3] Z. Wang, M. B. Luther, P. Horan, J. Matthews, and C. Liu, “On-site solar pv generation and use: Self-consumption and self-sufficiency,” in *Building Simulation*. Springer, 2023, pp. 1–15.
- [4] R. F. Widjaja, W. Wu, Z. Zhou, R. Sun, H. C. Fontenot, and B. Dong, “A general spatial-temporal framework for short-term building temperature forecasting at arbitrary locations with crowdsourcing weather data,” in *Building Simulation*. Springer, 2023, pp. 1–20.
- [5] D. Kim, Z. Wang, J. Brugger, D. Blum, M. Wetter, T. Hong, and M. A. Piette, “Site demonstration and performance evaluation of mpc for a large chiller plant with tes for renewable energy integration and grid decarbonization,” *Applied Energy*, vol. 321, p. 119343, 2022.
- [6] S. Touzani, A. K. Prakash, Z. Wang, S. Agarwal, M. Pritoni, M. Kiran, R. Brown, and J. Granderson, “Controlling distributed energy resources via deep reinforcement learning for load flexibility and energy efficiency,” *Applied Energy*, vol. 304, p. 117733, 2021.
- [7] L. Kotzur, L. Nolting, M. Hoffmann, T. Groß, A. Smolenko, J. Priesmann, H. Büsing, R. Beer, F. Kullmann, B. Singh, A. Praktiknjo, D. Stolten, and M. Robinius, “A modeler’s guide to handle complexity in energy systems optimization,” *Advances in Applied Energy*, vol. 4, p. 100063, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S266679242100055X>
- [8] N. I. Bazenkov, B. A. Boldyshev, S. V. Dushin, S. A. Frolov, M. V. Goubko, V. O. Korepanov, and L. A. Sereda, “Intensive data collection system for smart grid and smart building research,” in *2019 1st International Conference on Control Systems, Mathematical Modelling, Automation and Energy Efficiency (SUMMA)*, 2019, pp. 411–415.
- [9] H. Dong, C. Xu, and W. Chen, “Modeling and configuration optimization of the rooftop photovoltaic with electric-hydrogen-thermal hybrid storage system for zero-energy buildings: Consider a cumulative seasonal effect,” in *Building Simulation*. Springer, 2023, pp. 1–21.
- [10] B. Zou, J. Peng, S. Li, Y. Li, J. Yan, and H. Yang, “Comparative study of the dynamic programming-based and rule-based operation strategies for grid-connected pv-battery systems of office buildings,” *Applied energy*, vol. 305, p. 117875, 2022.
- [11] J. Hu, Y. Shan, J. M. Guerrero, A. Ioinovici, K. W. Chan, and J. Rodriguez, “Model predictive control of microgrids—an overview,” *Renewable and Sustainable Energy Reviews*, vol. 136, p. 110422, 2021.
- [12] Z. Wang and T. Hong, “Reinforcement learning for building controls: The opportunities and challenges,” *Applied Energy*, vol. 269, p. 115036, 2020.
- [13] D. Wang, W. Zheng, Z. Wang, Y. Wang, X. Pang, and W. Wang, “Comparison of reinforcement learning and model predictive control for building energy system optimization,” *Applied Thermal Engineering*, vol. 228, p. 120430, 2023.
- [14] Y. Shi, H. D. Tuan, A. V. Savkin, T. Q. Duong, and H. V. Poor, “Model predictive control for smart grids with multiple electric-vehicle charging stations,” *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2127–2136, 2018.
- [15] R. Tang and S. Wang, “Model predictive control for thermal energy storage and thermal comfort optimization of building demand response in smart grids,” *Applied Energy*, vol. 242, pp. 873–882, 2019.
- [16] C. Avilés A., S. Oliva H., and D. Watts, “Single-dwelling and community renewable microgrids: Optimal sizing and energy management for new business models,” *Applied Energy*, vol. 254, p. 113665, 2019.
- [17] M. Goldsworthy, T. Moore, M. Peristy, and M. Grimeland, “Cloud-based model-predictive-control of a battery storage system at a commercial site,” *Applied Energy*, vol. 327, p. 120038, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261922012958>

- [18] M. F. Zia, E. Elbouchikhi, and M. Benbouzid, "Microgrids energy management systems: A critical review on methods, solutions, and prospects," *Applied Energy*, vol. 222, pp. 1033–1055, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261918306676>
- [19] C. Fernández Bandera, G. Bastos Porsani, and M. Fernández-Vigil Iglesias, "A demand side management approach to increase self-consumption in buildings," in *Building Simulation*, vol. 16, no. 2. Springer, 2023, pp. 317–335.
- [20] N. Houben, A. Cosic, M. Stadler, M. Mansoor, M. Zellinger, H. Auer, A. Ajanovic, and R. Haas, "Optimal dispatch of a multi-energy system microgrid under uncertainty: A renewable energy community in austria," *Applied Energy*, vol. 337, p. 120913, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261923002775>
- [21] A. Kusiak and M. Li, "Cooling output optimization of an air handling unit," *Applied Energy*, vol. 87, no. 3, pp. 901–909, 2010.
- [22] N. Luo, T. Hong, H. Li, R. Jia, and W. Weng, "Data analytics and optimization of an ice-based energy storage system for commercial buildings," *Applied Energy*, vol. 204, pp. 459–475, 2017.
- [23] L. Pedersen, J. Stang, and R. Ulseth, "Load prediction method for heat and electricity demand in buildings for the purpose of planning for mixed energy distribution systems," *Energy and Buildings*, vol. 40, no. 7, pp. 1124–1134, 2008.
- [24] X. Xue, S. Wang, Y. Sun, and F. Xiao, "An interactive building power demand management strategy for facilitating smart grid optimization," *Applied Energy*, vol. 116, pp. 297–310, 2014.
- [25] Z. Wang, T. Hong, and M. A. Piette, "Building thermal load prediction through shallow machine learning and deep learning," *Applied Energy*, vol. 263, p. 114683, 2020.
- [26] X. Li and J. Wen, "Review of building energy modeling for control and operation," *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 517–537, 2014.
- [27] Y. Liu, H. Ming, X. Luo, L. Hu, and Y. Sun, "Timetabling optimization of classrooms and self-study rooms in university teaching buildings based on the building controls virtual test bed platform considering energy efficiency," in *Building Simulation*, vol. 16, no. 2. Springer, 2023, pp. 263–277.
- [28] C. Tian, Y. Ye, Y. Lou, W. Zuo, G. Zhang, and C. Li, "Daily power demand prediction for buildings at a large scale using a hybrid of physics-based model and generative adversarial network," in *Building Simulation*, vol. 15, no. 9. Springer, 2022, pp. 1685–1701.
- [29] C. Sun, F. Sun, and S. J. Moura, "Nonlinear predictive energy management of residential buildings with photovoltaics & batteries," *Journal of Power Sources*, vol. 325, pp. 723–731, 2016.
- [30] J. Wu, C. Lu, and C. Wu, "Learning-aided framework for storage control facing renewable energy," *IEEE Systems Journal*, vol. 17, no. 1, pp. 652–663, 2022.
- [31] G. Gust, T. Brandt, S. Mashayekh, M. Heleno, N. DeForest, M. Stadler, and D. Neumann, "Strategies for microgrid operation under real-world conditions," *European Journal of Operational Research*, vol. 292, no. 1, pp. 339–352, 2021.
- [32] A. N. Elmachtoub and P. Grigas, "Smart "predict, then optimize"," *Management Science*, vol. 68, no. 1, pp. 9–26, 2022.
- [33] J. Herzen, F. Lässig, S. G. Piazzetta, T. Neuer, L. Tafti, G. Raille, T. V. Pottelbergh, M. Pasioka, A. Skrodzki, N. Huguenin, M. Dumonal, J. Kościsz, D. Bader, F. Gusset, M. Benheddi, C. Williamson, M. Kosinski, M. Petrik, and G. Grosch, "Darts: User-friendly modern machine learning for time series," *Journal of Machine Learning Research*, vol. 23, no. 124, pp. 1–6, 2022. [Online]. Available: <http://jmlr.org/papers/v23/21-1177.html>
- [34] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 785–794. [Online]. Available: <https://doi.org/10.1145/2939672.2939785>
- [35] S. Silwal, C. Mullican, Y.-A. Chen, A. Ghosh, J. Dilliot, and J. Kleissl, "Open-source multi-year power generation, consumption, and storage data in a microgrid," *Journal of Renewable and Sustainable Energy*, vol. 13, no. 2, p. 025301, Mar. 2021, publisher: American Institute of Physics. [Online]. Available: <https://aip.scitation.org/doi/full/10.1063/5.0038650>
- [36] S. Jung, J. Moon, S. Park, and E. Hwang, "An attention-based multilayer gru model for multistep-ahead short-term load forecasting," *Sensors*, vol. 21, no. 5, 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/5/1639>

- [37] Z. Wang, T. Hong, H. Li, and M. Ann Piette, “Predicting city-scale daily electricity consumption using data-driven models,” *Advances in Applied Energy*, vol. 2, p. 100025, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666792421000184>
- [38] S. Chen, X. Zhou, G. Zhou, C. Fan, P. Ding, and Q. Chen, “An online physical-based multiple linear regression model for building’s hourly cooling load prediction,” *Energy and Buildings*, vol. 254, p. 111574, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378778821008586>
- [39] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, “Deepar: Probabilistic forecasting with autoregressive recurrent networks,” *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169207019301888>
- [40] B. L. et., “Temporal fusion transformers for interpretable multi-horizon time series forecasting,” *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169207021000637>
- [41] Z. Wang, T. Hong, H. Li, and M. A. Piette, “Predicting city-scale daily electricity consumption using data-driven models,” *Advances in Applied Energy*, vol. 2, p. 100025, 2021.
- [42] C. S. Bojer and J. P. Meldgaard, “Kaggle forecasting competitions: An overlooked learning opportunity,” *International Journal of Forecasting*, vol. 37, no. 2, pp. 587–603, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169207020301114>
- [43] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.
- [44] I. M. T. Agami Reddy and C. Panjapornpon, “Calibrating detailed building energy simulation programs with measured data—part i: General methodology (rp-1051),” *HVAC&R Research*, vol. 13, no. 2, pp. 221–241, 2007. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/10789669.2007.10390952>
- [45] C. Fan, F. Xiao, and Y. Zhao, “A short-term building cooling load prediction method using deep learning algorithms,” *Applied Energy*, vol. 195, pp. 222–233, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261917302921>