Toward Value-oriented Renewable Energy Forecasting: An Iterative Learning Approach

Yufan Zhang, Mengshuo Jia, Honglin Wen *Member, IEEE,* Yuexin Bian, Yuanyuan Shi *Member, IEEE.*

Abstract—Energy forecasting is an essential task in power system operations. Operators usually issue forecasts and use them to schedule energy dispatch in advance. However, forecasting models are typically developed in a way that overlooks the decision value of forecasts. To bridge the gap, we design a *value-oriented* point forecasting approach for sequential energy dispatch problems with renewable energy sources. At the training phase, we align the training objective with the decision value, i.e., minimizing the overall operating cost. The forecasting model parameter estimation is formulated as a bilevel program. Under mild assumptions, we convert the upper-level objective into an equivalent form using the dual solutions obtained from the lower-level operation problems. In addition, a novel iterative solution strategy is proposed for the newly formulated bilevel program. Under such an iterative scheme, we show that the upper-level objective is locally linear with respect to the forecasting model output and can act as the loss function. Numerical experiments demonstrate that, compared to commonly used forecasts predicting expected realization, forecasts obtained by the proposed approach result in lower operating costs. Meanwhile, the proposed approach achieves performance comparable to that of two-stage stochastic programs, but is more computationally efficient.

Keywords: Renewable energy forecasting, Decision-focused learning, Energy dispatch, Sequential operation

I. INTRODUCTION

With the ongoing transition towards net zero emissions in the energy sector, the penetration of renewable energy sources (RESs) has surged over recent decades. Due to inherent uncertainty, RESs cannot be scheduled at will, which has therefore motivated research on renewable energy forecasting to communicate information to the operation of power systems in advance [1]. Typically, forecasts are communicated in the form of point forecasts [2], prediction intervals [3], and probability distributions [4], [5], which offer a snapshot or a whole picture of the predictive distribution of renewable generation. Today, the forecasting community is increasingly relying on data-driven approaches, which estimate parameters of forecasting models based on historical data and some loss functions that measure the quality of forecasts in these data; see the comprehensive review by [6].

Usually, forecasts serve as the input to the downstream operational problems and therefore influence decisions. The resulting decision value of the downstream operational problem can also be regarded as the value of forecasts, as suggested by [7]. Several previous works have shown that forecasts with

Yufan Zhang, Yuexin Bian, and Yuanyuan Shi are with the Department of Electrical and Computer Engineering, University of California San Diego. Their work are partially supported by the CA Climate Action, Resilience, and Environmental Sustainability (CARES) grant from UC San Diego. Mengshuo Jia is with the Power Systems Laboratory, ETH Zürich, 8092 Zürich, Switzerland (jia@eeh.ee.ethz.ch). Honglin Wen is with the Department of Electrical Engineering, Shanghai Jiao Tong University. Co-corresponding authors: Honglin Wen (linlin00@sjtu.edu.cn) and Yuanyuan Shi (yyshi@ucsd.edu).

higher statistical accuracy may not necessarily lead to desirable value in the operational problems [8]–[13]. Therefore, **value-oriented forecasting** has been advocated [14], [15], which aims at developing forecasting products that lead to higher value in subsequent decision-making, instead of better statistical accuracy.

However, training forecasting models to achieve higher operating value presents significant challenges. Different lines of research in the literature address this issue in various ways; see an up-to-date review [16]. The first strand derives the gradient of optimal decision solutions w.r.t. the forecast for training forecasting models via gradient descent [17], [18], which involves a large-size matrix inverse and results in a computational burden. For operational problems where such a gradient is discontinuous or even vanishes everywhere, such as linear programs where coefficients in the objective function need to be estimated, this approach is not suitable. This motivates the second thread, which designs loss functions for training. "Smart Predict, then Optimize" (SPO) loss [15] is one of the pioneering works in this area, which relies on a hand-crafted regret loss function. To automate the process of designing loss functions, several studies propose learning the loss function from the data [13], [19].

In general, value-oriented forecasting approaches can communicate information that aligns with the operational value. Most power system operational problems are organized sequentially [20], which includes two deterministic problems made at different times, namely in the day-ahead and real-time stages (both referring to the time at which decisions are made relative to the delivery of energy). At the day-ahead stage, a single RES forecast is used as input to the operational problem to settle the quantity of energy to be dispatched the next day. The real-time stage occurs close to the actual delivery of electricity, following the realization of RES. During this stage, corrective decisions are made by solving an operational problem to balance any deviations from the initial schedules. In this setting, the decision value is commonly defined as minimizing the overall operating cost or maximizing the overall welfare/profit at both stages.

One of the most successful applications of aligning the forecasting model with such decision value definition is to settle the offering strategy of wind power producers, where the objective is to maximize the expected overall operating profit (by anticipating the balancing cost at the real-time stage). Refs [1], [21], [22] demonstrated that the wind power offering aligns with the Newsvendor problem [23], and the optimal offering strategy is to predict the quantile. The link between the decision value, i.e., the overall profit, and the forecast can be explicitly formulated, and the pinball loss can be used as a surrogate loss function to develop forecasting models.

For complex operational problems involving several decision variables and constraints, a closed-form relationship between the decision value, such as total operating cost, and the forecasts may not be explicit. One possible solution is to optimize the forecasting model parameter concurrently with the decision variables. For that, a bilevel program is formulated for the parameter estimation [24], [25]. In this context, the lower-level optimization encompasses day-ahead operations with the forecast acting as a parameter, while the upper-level problem focuses on optimizing the model parameters along with real-time decisions. To ensure the tractability of such a program, forecasting models are typically simplified to linear regression models, which might restrict their effectiveness.

In this paper, we develop a value-oriented forecasting approach in the context of day-ahead and real-time energy dispatch problems for virtual power plant (VPP) operators. At the training phase, a bilevel program is formulated for the estimation of the forecasting model parameter, where the upper level optimizes the forecasting model parameter toward the decision value, i.e., minimizing the expected sum of day-ahead and real-time operating costs. To explicitly show the relationship between the decision value and the forecast, we convert the upper-level objective to its dual form, assuming that the energy dispatch problems are linear programs. This conversion allows us to obtain the "value-oriented" loss function for training the forecast models. By explicitly deriving the loss function, sophisticated regression models, such as Neural Networks (NNs), can be employed as forecasting models. This approach surpasses the limitations of earlier studies [24], [25] that were confined to linear models. Also, the explicit loss function allows us to directly obtain the gradient of the decision value w.r.t. the forecast. These explicit gradients correspond to the optimal dual solutions for day-ahead and real-time energy dispatch, which can be acquired efficiently and avoids the computational burden of calculating the gradient of optimal solutions w.r.t. the forecast [17]. Furthermore, an iterative learning strategy is proposed, which iteratively updates the forecasting model parameters by gradient descent with the loss function and updates the loss function based on the newly updated forecasting model. Ultimately, we can achieve value-oriented RES forecasts that are considered "good" in terms of the expected overall costs for sequential operations. Our main contributions are,

1) From a practitioner perspective, we develop a computationally efficient framework for training advanced forecasting models (such as Neural Networks) that align with the decision value of a VPP operator. An iterative learning approach is proposed for training, which iteratively updates the forecasting model parameter via gradient descent and the loss function coefficients based on the newly updated forecasts.

2) From a VPP operator perspective, we explain why value-oriented forecasts are more desirable than quality-oriented forecasts that minimize the mean squared error of renewable forecast and realization. The derived value-oriented loss function indicates that the key reason is the asymmetric real-time marginal cost/utility in cases of energy shortage and surplus.

The remainder is organized as follows: Section II presents the preliminaries regarding the day-ahead and real-time energy dispatch problems with RES. Section III presents the training phase and formulates the parameter estimation method for the value-oriented forecasting model. The iterative solution strategy is proposed in Section IV. Sections V and VI present case studies with results and discussion, followed by conclusions.

II. PRELIMINARIES: DAY-AHEAD AND REAL-TIME ENERGY DISPATCH

In this work, we consider the operation of a centralized VPP operator that runs a group of assets including RES, slow-start generators, and some flexible resources (e.g., flexible generators and energy storage). The operator solves an energy dispatch problem, i.e., scheduling the assets to balance total demands at the lowest cost [26].

Usually, VPPs are operated in a sequential framework, which includes the day-ahead and real-time stages. Decisions in the day-ahead stage are solved at time t on day $d - 1$, which schedules the generation of slow-start generators for each time slot, i.e., $\tau, \forall \tau = 1, ..., T$ on the next day d. At the real-time stage, the operator dispatches the flexible resources to compensate for the deficit or surplus of the power imbalance, which is calculated throughout the day d, over time intervals of 1 h. In this paper, we assume the real-time dispatch decisions are made, after the RES realization is revealed [1]. Also, the day-ahead schedule is assumed to be fixed at the real-time stage. This is reasonable for inflexible devices, such as combined heat and power (CHP) plants. CHP plants produce both electricity and useful heat simultaneously. The heat demand often dictates their operation, making their electricity generation schedule relatively inflexible [27].

We provide a review of three existing two-stage dispatch models under the uncertainties of RES, whose illustration is shown in Fig. 1.

1) Deterministic dispatch: The operator schedules slow-start generators at the day-ahead stage, considering a single RES forecast. The real-time dispatch addresses the power imbalance caused by forecasting errors, via flexible resources. This is a simplification of how most centralized operators work.

2) Stochastic dispatch: The operator solves a two-stage stochastic program at the day-ahead stage, which schedules slow-start generators and RES, and anticipates its impact on the real-time operation. In this way, the first-stage decisions of stochastic dispatch involve the decisions regarding slow-start generators and RES schedules. The second-stage decisions, also known as recourse decisions, involve the dispatch of flexible resources under RES scenarios (which is a kind of probabilistic forecast to approximate the conditional distribution). Only the first-stage solutions will be kept at the day-ahead stage. Real-time dispatch is still needed to address the power imbalance caused by the difference between the RES realization and its schedule.

3) Improved deterministic dispatch: As conventional deterministic dispatch is widely known to lead to high overall operating costs, [28] proposed to leverage a strategic RES schedule within the deterministic operation framework, to

---

[1]In fact, the realization is approximately informed close to the time τ, as available information is adequate for an accurate estimate.

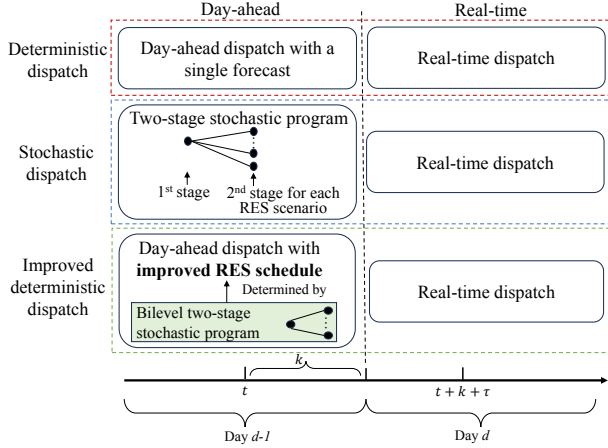approach the performance of stochastic dispatch. This schedule is determined by a stochastic bilevel program.

Fig. 1. The illustration of deterministic, stochastic, and improved deterministic dispatch, where k is the time interval between t and 1 am on day d.

A. Deterministic Dispatch

Day-ahead Dispatch with a Point Forecast Predicting Expected Renewable Realization: Let $y_{d,\tau}$ denote the RES generation at time τ on day d. As it is not known ex-ante, we often use a forecast $\hat{y}_{d,\tau}$ for decision planning. We use $\boldsymbol{x}_{d,\tau}$ to denote the decision variables regarding the power generation of slow-start generators at the day-ahead stage. Since the RES has zero marginal cost and we assume that its capacity is less than the minimum load, the operator uses all its forecast $\hat{y}_{d,\tau}$ to balance the load. The problem at the day-ahead stage, minimizing the linear operating cost of slow-start generators in a day, is given by

$$\min_{\{\boldsymbol{x}_{d,\tau}\}_{\tau=1}^T} \sum_{\tau=1}^T \boldsymbol{\rho}^\top \boldsymbol{x}_{d,\tau} \tag{1a}$$

$$\text{s.t.} \quad f_i^D(\boldsymbol{x}_{d,\tau}) \le 0, \forall i \in I^D, \forall \tau = 1,...,T \tag{1b}$$

$$-\boldsymbol{r} \le \boldsymbol{x}_{d,\tau+1} - \boldsymbol{x}_{d,\tau} \le \boldsymbol{r}, \forall \tau = 1,...,T-1 \tag{1c}$$

$$\mathbf{1}^\top \boldsymbol{x}_{d,\tau} + \hat{y}_{d,\tau} = l_{d,\tau}, \forall \tau = 1,...,T. \tag{1d}$$

where $\boldsymbol{\rho}$ is the marginal generation cost. We express the optimal solution of the decision $\boldsymbol{x}_{d,\tau}$ as $\boldsymbol{x}_{d,\tau}^*(\hat{y}_{d,\tau})$, since it depends on the forecast $\hat{y}_{d,\tau}$. (1b) is the constraint regarding the operational limits, such as limiting the output power of the generators, and I^D gives the set of those constraints. (1c) limits the up- and down-ramping of the generators between adjacent time steps within the limit $\boldsymbol{r}$. The power balance under the load $l_{d,\tau}$ is ensured by (1d). We assume the load forecast is rather accurate, and the uncertainty is only from RES. Additionally, we consider a single-node system. The total power generation equals the total demand.

Real-time Dispatch: As the errors between $\hat{y}_{d,\tau}$ and $y_{d,\tau}$ are unavoidable, it requires the operators to dispatch other assets to compensate the errors as time close to delivery. After the RES realization $y_{d,\tau}$ is revealed at time τ on day d, the energy imbalance caused by RES is addressed independently

at each time τ in real-time. The energy imbalance is then given by $y_{d,\tau} - \hat{y}_{d,\tau}$, which represents a surplus of generation, if positive, or a shortage, if negative. We group the decisions taken to cope with the energy shortage into the vector $\boldsymbol{z}_{d,\tau}^+, \forall \tau = 1,...,T$ and the decisions taken to cope with the energy surplus into the vector $\boldsymbol{z}_{d,\tau}^-, \forall \tau = 1,...,T$. Minimizing the cost of the flexible resources outputs, the problem solved at time τ for compensating the deviation $y_{d,\tau} - \hat{y}_{d,\tau}$ is,

$$\min_{\boldsymbol{z}_{d,\tau}^+, \boldsymbol{z}_{d,\tau}^-} \quad \boldsymbol{\rho}^{+\top} \boldsymbol{z}_{d,\tau}^+ - \boldsymbol{\rho}^{-\top} \boldsymbol{z}_{d,\tau}^- \tag{2a}$$

$$\text{s.t.} \quad f_i^{R+}(\boldsymbol{z}_{d,\tau}^+) \le 0, \forall i \in I^{R+} \tag{2b}$$

$$f_i^{R-}(\boldsymbol{z}_{d,\tau}^-) \le 0, \forall i \in I^{R-} \tag{2c}$$

$$\mathbf{1}^\top(\boldsymbol{z}_{d,\tau}^+ - \boldsymbol{z}_{d,\tau}^-) + y_{d,\tau} - \hat{y}_{d,\tau} = 0. \tag{2d}$$

where $\boldsymbol{\rho}^+$ and $\boldsymbol{\rho}^-$ are marginal cost and utility. We express the optimal solution of the decision $\boldsymbol{z}_{d,\tau}^+, \boldsymbol{z}_{d,\tau}^-$ as $\boldsymbol{z}_{d,\tau}^{+*}(\hat{y}_{d,\tau}, y_{d,\tau}), \boldsymbol{z}_{d,\tau}^{-*}(\hat{y}_{d,\tau}, y_{d,\tau})$, since it depends on the forecast $\hat{y}_{d,\tau}$ and the realization $y_{d,\tau}$. (2b) and (2c) comprise upper and lower bounds on the flexible resources output, and I^{R+}, I^{R-} give the set of those constraints. (2d) ensures that the deviation between RES forecast and realization is addressed, and the system remains in balance.

Forecasting Products Requirement: At the day-ahead stage, operators are required to predict a single value for the RES generation the next day, and then schedule the assets. The forecasts are issued at time t in day-ahead, by a model g with the estimated parameters $\hat{\Theta}_Q$ (the subscript Q stands for quality-oriented forecasts), based on the available information $\boldsymbol{s}_d$. Here, $\boldsymbol{s}_d$ is a set formed by the realization $\boldsymbol{s}_{d,\tau}$ of the random variable $S_{d,\tau}$ (denoting the contextual information at time slot τ), i.e., $\boldsymbol{s}_d = \{\boldsymbol{s}_{d,1},...,\boldsymbol{s}_{d,T}\}$. Take day-ahead wind power forecasting for instance, such information mainly contains numerical weather predictions (NWPs). Denote the forecast for RES at time slot τ as $\hat{y}_{d,\tau}$, which is obtained by

$$\hat{y}_{d,\tau} = g(\boldsymbol{s}_{d,\tau}; \hat{\Theta}_Q). \tag{3}$$

One of the most widely accepted practices is issuing $\hat{y}_{d,\tau}$ to predict the "most likely realization" of RES random variable $Y_{d,\tau}$. The expected value of $Y_{d,\tau}$ is a natural candidate to take this role. In this way, the parameter Θ_Q is usually estimated via data-driven methods and particularly by minimizing the mean squared error (MSE) at the training phase.

B. Two-stage Stochastic Program based Dispatch

Day-ahead Dispatch with Probabilistic Forecasts and Stochastic Optimization: As an input parameter, the forecast $\hat{y}_{d,\tau}$ influences the day-ahead and real-time problems. Concretely, $\hat{y}_{d,\tau}$ affects the generation schedule of slow-start generators, and in turn, affects the day-ahead operating cost. Its over- and down-prediction results in different operating decisions in real-time. In this way, the forecast $\hat{y}_{d,\tau}$ couples the day-ahead and real-time problems. However, the separate operation in (1) and (2) fails to take such coupling into account, thereby leading to higher overall operating costs. Therefore, a stochastic program for co-optimizing the two stages is advocated, which considers the temporal depen-

3

dency of the day-ahead and real-time operation and treats the RES schedule as a decision variable. Such a two-stage stochastic program is solved in day-ahead for all time slots $\tau, \forall \tau = 1, ..., T$ on the next day d. It determines the RES schedule $\tilde{y}_{d,\tau}$ and generation schedule $\boldsymbol{x}_{d,\tau}$, anticipating their influence on the expected real-time cost given any potential realization of the RES. For that, $\tilde{y}_{d,\tau}$ and $\boldsymbol{x}_{d,\tau}$ form the first-stage decisions, and the schedule of $\boldsymbol{x}_{d,\tau}$ incurs the first-stage cost $\boldsymbol{\rho}^\top \boldsymbol{x}_{d,\tau}$. The flexible resources output $\tilde{\boldsymbol{z}}_{d,\tau}^+, \tilde{\boldsymbol{z}}_{d,\tau}^-$ is the second-stage decision, which depends on each realization of RES forecasting error $Y_{d,\tau} - \tilde{y}_{d,\tau}$. Its dispatch incurs the second-stage cost $\boldsymbol{\rho}^{+\top} \tilde{\boldsymbol{z}}_{d,\tau}^+ - \boldsymbol{\rho}^{-\top} \tilde{\boldsymbol{z}}_{d,\tau}^-$. The stochastic program for minimizing the summation of the first-stage cost and the expected second-stage cost is formulated as,

$$\min_{\Xi} \quad \sum_{\tau=1}^{T} \{\boldsymbol{\rho}^\top \boldsymbol{x}_{d,\tau} + \mathbb{E}_{Y_{d,\tau}} \left[\boldsymbol{\rho}^{+\top} \tilde{\boldsymbol{z}}_{d,\tau}^+ - \boldsymbol{\rho}^{-\top} \tilde{\boldsymbol{z}}_{d,\tau}^- \right]\} \quad (4a)$$

$$\text{s.t.} \quad f_i^D(\boldsymbol{x}_{d,\tau}) \leq 0, \forall i \in I^D, \forall \tau = 1, ..., T \quad (4b)$$

$$-\boldsymbol{r} \leq \boldsymbol{x}_{d,\tau+1} - \boldsymbol{x}_{d,\tau} \leq \boldsymbol{r}, \forall \tau = 1, ..., T-1 \quad (4c)$$

$$\mathbf{1}^\top \boldsymbol{x}_{d,\tau} + \tilde{y}_{d,\tau} = l_{d,\tau}, \forall \tau = 1, ..., T \quad (4d)$$

$$0 \leq \tilde{y}_{d,\tau} \leq \bar{y}, \forall \tau = 1, ..., T \quad (4e)$$

$$f_i^{R+}(\tilde{\boldsymbol{z}}_{d,\tau}^+) \leq 0, \forall i \in I^{R+}, \forall \tau = 1, ..., T \quad (4f)$$

$$f_i^{R-}(\tilde{\boldsymbol{z}}_{d,\tau}^-) \leq 0, \forall i \in I^{R-}, \forall \tau = 1, ..., T \quad (4g)$$

$$\mathbf{1}^\top (\tilde{\boldsymbol{z}}_{d,\tau}^+ - \tilde{\boldsymbol{z}}_{d,\tau}^-) + Y_{d,\tau} - \tilde{y}_{d,\tau} = 0, \forall \tau = 1, ..., T, \quad (4h)$$

where $\Xi = \{\boldsymbol{x}_{d,\tau}, \tilde{\boldsymbol{z}}_{d,\tau}^+, \tilde{\boldsymbol{z}}_{d,\tau}^-, \tilde{y}_{d,\tau}\}_{\tau=1}^T$. $\bar{y}$ is the capacity of RES and serves as the upper bound for the RES schedule. $Y_{d,\tau}$ follows the conditional distribution $F_{Y_{d,\tau}|S_{d,\tau}=\boldsymbol{s}_{d,\tau}}$. One way to estimate the distribution is to predict the independent and identically distributed (i.i.d.) scenarios, with the same probability of occurrence. However, it usually requires a large number of i.i.d. scenarios to accurately model the distribution, which poses a huge computational burden to solving (4).

Notice that by anticipating the balancing operation with the constraints (4f),(4g) and (4h) and considering the expected balance cost in the objective (4a), the stochastic program (4) empowers the first-stage decisions to account for the impact of the energy imbalance on the second-stage. After solving (4), the solution $\boldsymbol{x}_{d,\tau}^*$ will be used for scheduling slow-start generators in day-ahead.

Real-time Dispatch: Similar to the deterministic approach, after the RES realization $y_{d,\tau}$ is revealed, the real-time problem in (2) is solved for settling the deviation between the realization and the optimal RES schedule, i.e., $y_{d,\tau} - \tilde{y}_{d,\tau}^*$.

Forecasting Products Requirement: Operators are required to predict the conditional distribution $F_{Y_{d,\tau}|S_{d,\tau}=\boldsymbol{s}_{d,\tau}}$.

C. Improved Deterministic Dispatch

Day-ahead Dispatch with Improved RES Schedule: The two-stage stochastic program (4) has the objective of minimizing the overall first-stage and second-stage costs. In an effort to enhance the performance of separate operation models (1) and (2), we optimize the RES schedule with the same objective as in (4), thereby ensuring that this value accounts for the the second-stage cost resulting from RES uncertainty.

Meanwhile, we need to ensure that such a schedule aligns with the deterministic dispatch framework. For that, a bilevel program is formulated [28], where the upper-level determines the schedule $\tilde{y}_{d,\tau}$, and the lower-level solves the first-stage and the second-stage problems separately which are equivalent to the conventional dispatch in (1) and (2),

$$\min_{\{\tilde{y}_{d,\tau}\}_{\tau=1}^T} \sum_{\tau=1}^{T} \{\boldsymbol{\rho}^\top \boldsymbol{x}_{d,\tau}^*(\tilde{y}_{d,\tau})+$$
$$\mathbb{E}_{Y_{d,\tau}} \left[\boldsymbol{\rho}^{+\top} \tilde{\boldsymbol{z}}_{d,\tau}^{+*}(\tilde{y}_{d,\tau}, Y_{d,\tau}) - \boldsymbol{\rho}^{-\top} \tilde{\boldsymbol{z}}_{d,\tau}^{-*}(\tilde{y}_{d,\tau}, Y_{d,\tau}) \right]\} \quad (5a)$$

$$\text{s.t.} \quad 0 \leq \tilde{y}_{d,\tau} \leq \bar{y}, \forall \tau = 1, ..., T \quad (5b)$$

$$\{\boldsymbol{x}_{d,\tau}^*(\tilde{y}_{d,\tau})\}_{\tau=1}^T = \arg\min_{\{\boldsymbol{x}_{d,\tau}\}_{\tau=1}^T} \sum_{\tau=1}^{T} \boldsymbol{\rho}^\top \boldsymbol{x}_{d,\tau} \quad (5c)$$

$$\text{s.t.} \quad f_i^D(\boldsymbol{x}_{d,\tau}) \leq 0, \forall i \in I^D, \forall \tau = 1, ..., T \quad (5d)$$

$$-\boldsymbol{r} \leq \boldsymbol{x}_{d,\tau+1} - \boldsymbol{x}_{d,\tau} \leq \boldsymbol{r}, \forall \tau = 1, ..., T-1 \quad (5e)$$

$$\mathbf{1}^\top \boldsymbol{x}_{d,\tau} + \tilde{y}_{d,\tau} = l_{d,\tau}, \forall \tau = 1, ..., T \quad (5f)$$

$$\{\tilde{\boldsymbol{z}}_{d,\tau}^{+*}(\tilde{y}_{d,\tau}, Y_{d,\tau}), \tilde{\boldsymbol{z}}_{d,\tau}^{-*}(\tilde{y}_{d,\tau}, Y_{d,\tau}) =$$
$$\arg\min_{\tilde{\boldsymbol{z}}_{d,\tau}^+, \tilde{\boldsymbol{z}}_{d,\tau}^-} \boldsymbol{\rho}^{+\top} \tilde{\boldsymbol{z}}_{d,\tau}^+ - \boldsymbol{\rho}^{-\top} \tilde{\boldsymbol{z}}_{d,\tau}^- \quad (5g)$$

$$\text{s.t.} \quad f_i^{R+}(\tilde{\boldsymbol{z}}_{d,\tau}^+) \leq 0, \forall i \in I^{R+} \quad (5h)$$

$$f_i^{R-}(\tilde{\boldsymbol{z}}_{d,\tau}^-) \leq 0, \forall i \in I^{R-} \quad (5i)$$

$$\mathbf{1}^\top (\tilde{\boldsymbol{z}}_{d,\tau}^+ - \tilde{\boldsymbol{z}}_{d,\tau}^-) + Y_{d,\tau} - \tilde{y}_{d,\tau} = 0\},$$
$$\forall \tau = 1, ..., T, \quad (5j)$$

where the random variable $Y_{d,\tau}$ follows the conditional distribution $F_{Y_{d,\tau}|S_{d,\tau}=\boldsymbol{s}_{d,\tau}}$. The upper-level decision $\tilde{y}_{d,\tau}$ serves as the parameter in the lower-level problems, and affects the lower-level decisions $\boldsymbol{x}_{d,\tau}^*(\tilde{y}_{d,\tau})$ and $\boldsymbol{z}_{d,\tau}^{+*}(\tilde{y}_{d,\tau}, Y_{d,\tau}), \boldsymbol{z}_{d,\tau}^{-*}(\tilde{y}_{d,\tau}, Y_{d,\tau})$, which form the upper-level objective in (5). After solving (5), the improved RES schedule $\tilde{y}_{d,\tau}^*, \forall \tau = 1, ..., T$ will enter the day-ahead problem in (1) to take the role of RES forecast $\hat{y}_{d,\tau}$.

Real-time Dispatch: Real-time dispatch is the same as that in Section II-B.

Forecasting Products Requirement: To solve the stochastic program in (5), scenarios for approximating the conditional distribution is needed. Therefore, forecasting products is the same as that in Section II-B.

Although the program (5) can determine the RES schedule by accounting for the future balancing cost caused by RES uncertainty, it requires resolving the stochastic bilevel program every time for determining the RES schedule on a new day, which is not computationally efficient. To save the computational burden, we will learn a forecasting model in the next section, which maps the realization of contextual information to the improved RES schedule.

III. METHODOLOGY

In this section, we propose to train a forecasting model g to issue the strategic RES schedule, given the contextual information $\boldsymbol{s}_{d,\tau}$. To this end, we align the objective of the

parameter estimation with the same objective of (5), i.e., minimizing the expected operating costs of the two stages, and term such an approach as value-oriented forecasting. Similar to the commonly used forecasts defined in (3), we still model forecasts with a function g with parameters Θ_V (to make a difference with the model (3), the subscript V stands for the proposed value-oriented forecasts). However, the goal here is not to forecast the expectation of the random variable $Y_{d,\tau}$, but a *strategic* quantity to be used at the day-ahead stage. Given the contextual information $s_{d,\tau}$, the value-oriented RES forecast is derived as

$$\tilde{y}_{d,\tau} = g(s_{d,\tau}; \Theta_V). \tag{6}$$

Usually, the parameters Θ_V are estimated via a training set formed by historical contextual information and RES realization in D days, i.e., $\{\{s_{d,\tau}, y_{d,\tau}\}_{\tau=1}^T\}_{d=1}^D$. The data in the training set are assumed to be independently and identically drawn from a joint distribution $F_{Y_{d,\tau}, S_{d,\tau}}$ over $Y_{d,\tau}$ and $S_{d,\tau}$. In this way, for any fixed contextual information $S_{d,\tau} = s_{d,\tau}$, the conditional distribution $F_{Y_{d,\tau}|S_{d,\tau}=s_{d,\tau}}$ is approximated by a single support $y_{d,\tau}$ [29]. The ultimate estimate for the parameters Θ_V is denoted as $\hat{\Theta}_V$. Particularly, at the training phase and given the context $s_{d,\tau}$, we denote the forecasts given by g as $\tilde{y}_{d,\tau}$, with Θ_V given by any value. We have the following assumption:

Assumption 1. *The day-ahead operation in (1) and the real-time operation in (2) have unique primal and dual solutions.*

The above assumption is reasonable. The primal solutions are interpreted as the scheduled energy, while the dual solutions represent the corresponding prices. Typically, these solutions are unique in the operation. In what follows, we formulate the parameter estimation problem at the training phase in Section III-A, followed by the operation phase in Section III-B.
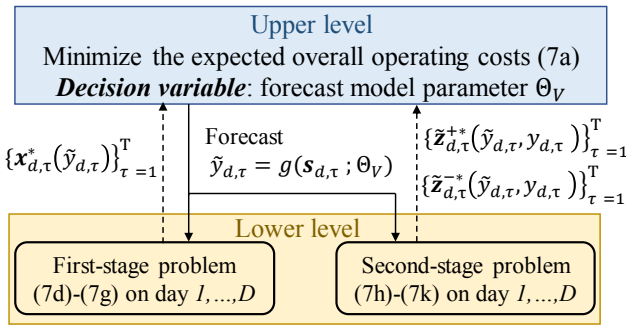

Fig. 2. Illustration of the proposed bilevel program (7) for model training of value-oriented forecasting.

A. Training Phase

For parameter estimation at the training phase, we propose formulating a bilevel program, adopting a similar bilevel form as (5). The upper-level problem determines the parameter Θ_V, for minimizing the average of the overall operating cost over D days. Specifically, since the conditional distribution $F_{Y_{d,\tau}|S_{d,\tau}=s_{d,\tau}}$ is supported only by $y_{d,\tau}$, the expected second-stage cost $\mathbb{E}_{Y_{d,\tau}}[\rho^{+\top} \tilde{z}_{d,\tau}^{+*}(\tilde{y}_{d,\tau}, Y_{d,\tau}) -$

$\rho^{-\top} \tilde{z}_{d,\tau}^{-*}(\tilde{y}_{d,\tau}, Y_{d,\tau})]$ is approximated by the cost under this support, i.e., $\rho^{+\top} \tilde{z}_{d,\tau}^{+*}(\tilde{y}_{d,\tau}, y_{d,\tau}) - \rho^{-\top} \tilde{z}_{d,\tau}^{-*}(\tilde{y}_{d,\tau}, y_{d,\tau})$. The lower-level repeats first- and second-stage problems for each day, which amounts to the day-ahead and real-time problems in (1) and (2). The optimal solutions parameterized by the forecast $\tilde{y}_{d,\tau}$ are obtained from the lower level. The bilevel program is,

$$\min_{\Theta_V} \quad \frac{1}{D \cdot T} \sum_{d=1}^D \sum_{\tau=1}^T \{\rho^\top x_{d,\tau}^*(\tilde{y}_{d,\tau}) +$$
$$\rho^{+\top} \tilde{z}_{d,\tau}^{+*}(\tilde{y}_{d,\tau}, y_{d,\tau}) - \rho^{-\top} \tilde{z}_{d,\tau}^{-*}(\tilde{y}_{d,\tau}, y_{d,\tau})\} \tag{7a}$$

$$\text{s.t.} \quad \tilde{y}_{d,\tau} = g(s_{d,\tau}; \Theta_V), \forall \tau = 1, ..., T, \forall d = 1, ..., D \tag{7b}$$

$$0 \le \tilde{y}_{d,\tau} \le \bar{y}, \forall \tau = 1, ..., T, \forall d = 1, ..., D \tag{7c}$$

$$\{\{x_{d,\tau}^*(\tilde{y}_{d,\tau})\}_{\tau=1}^T = \arg\min_{\{x_{d,\tau}\}_{\tau=1}^T} \sum_{\tau=1}^T \rho^\top x_{d,\tau} \tag{7d}$$

$$\text{s.t.} \quad f_i^D(x_{d,\tau}) \le 0 : \delta_{i,d,\tau}, \forall i \in I^D, \forall \tau = 1, ..., T \tag{7e}$$

$$-r \le x_{d,\tau+1} - x_{d,\tau} \le r : \underline{\eta_{d,\tau}}, \overline{\eta_{d,\tau}},$$
$$\forall \tau = 1, ..., T-1 \tag{7f}$$

$$\mathbf{1}^\top x_{d,\tau} + \tilde{y}_{d,\tau} = l_{d,\tau} : \lambda_{d,\tau}, \forall \tau = 1, ..., T\},$$
$$\forall d = 1, ..., D \tag{7g}$$

$$\{\tilde{z}_{d,\tau}^{+*}(\tilde{y}_{d,\tau}, y_{d,\tau}), \tilde{z}_{d,\tau}^{-*}(\tilde{y}_{d,\tau}, y_{d,\tau}) =$$
$$\arg\min_{\tilde{z}_{d,\tau}^+, \tilde{z}_{d,\tau}^-} \rho^{+\top} \tilde{z}_{d,\tau}^+ - \rho^{-\top} \tilde{z}_{d,\tau}^- \tag{7h}$$

$$\text{s.t.} \quad f_i^{R+}(\tilde{z}_{d,\tau}) \le 0 : \mu_{i,d,\tau}, \forall i \in I^{R+} \tag{7i}$$

$$f_i^{R-}(\tilde{z}_{d,\tau}) \le 0 : \zeta_{i,d,\tau}, \forall i \in I^{R-} \tag{7j}$$

$$\mathbf{1}^\top(\tilde{z}_{d,\tau}^+ - \tilde{z}_{d,\tau}^-) + y_{d,\tau} - \tilde{y}_{d,\tau} = 0 : \nu_{d,\tau}\},$$
$$\forall \tau = 1, ..., T, \forall d = 1, ..., D, \tag{7k}$$

where the term inside the curly bracket in (7a) is the expected day-ahead and real-time overall operating cost at the time slot τ on day d. The model parameter Θ_V, which is the upper-level variable, has an influence on the forecast $\tilde{y}_{d,\tau}$, and also the parameterized solutions $x_{d,\tau}^*(\tilde{y}_{d,\tau}), \tilde{z}_{d,\tau}^{+*}(\tilde{y}_{d,\tau}, y_{d,\tau}), \tilde{z}_{d,\tau}^{-*}(\tilde{y}_{d,\tau}, y_{d,\tau})$ as well as the average operating cost over D days through (7a). An illustration of such a bilevel program is shown in Fig. 2.

Remark 1. *As one of the inspiring works in the literature regarding the application of bilevel program for parameter estimation, [24] used a simplified first-stage dispatch model. We go beyond [24]. Our model (7) can address the temporal correlated constraints, such as ramping constraints in (7f), which aligns with the current operation practice.*

The optimal dual solutions of (7d)-(7g) parameterized by $\tilde{y}_{d,\tau}$ are denoted as $\delta_{i,d,\tau}^*(\tilde{y}_{d,\tau})$, $\underline{\eta_{d,\tau}}^*(\tilde{y}_{d,\tau})$, $\overline{\eta_{d,\tau}}^*(\tilde{y}_{d,\tau})$, and $\lambda_{d,\tau}^*(\tilde{y}_{d,\tau})$. The optimal dual solutions of (7h)-(7k) parameterized by $\tilde{y}_{d,\tau}, y_{d,\tau}$ are denoted as $\mu_{i,d,\tau}^*(\tilde{y}_{d,\tau}, y_{d,\tau})$, $\zeta_{i,d,\tau}^*(\tilde{y}_{d,\tau}, y_{d,\tau})$ and $\nu_{d,\tau}^*(\tilde{y}_{d,\tau}, y_{d,\tau})$. Usually, the operational constraints are linear, and the problems with linear operating cost in (7d)-(7g) and (7h)-(7k) are linear programs. In this way, we resort to the dual problems for explicitly rewriting (7a) as a function of the forecasting model output $\tilde{y}_{d,\tau}$. We

have the following proposition for it.

Proposition 1. *Consider the convex optimization problems* (7d)-(7g) *and* (7h)-(7k) *in the form of linear programs (LPs), with the parameterized dual solutions associated with the constraints* (7g) *and* (7k) *are* $\lambda_{d,\tau}^*(\tilde{y}_{d,\tau})$ *and* $\nu_{d,\tau}^*(\tilde{y}_{d,\tau}, y_{d,\tau})$, *respectively. The optimal primal objectives of* (7d) *and* (7h) *at time* τ *on day* d *are* $\boldsymbol{\rho}^\top \boldsymbol{x}_{d,\tau}^*(\tilde{y}_{d,\tau})$ *and* $\boldsymbol{\rho}^{+\top}\tilde{\boldsymbol{z}}_{d,\tau}^{+*}(\tilde{y}_{d,\tau}, y_{d,\tau}) - \boldsymbol{\rho}^{-\top}\tilde{\boldsymbol{z}}_{d,\tau}^{-*}(\tilde{y}_{d,\tau}, y_{d,\tau})$. *As the strong duality holds, the optimal primal objectives equal the respective optimal dual objectives, expressed by the optimal dual solutions. The term in the curly bracket of* (7a) *is equivalent as,*

$$-\lambda_{d,\tau}^*(\tilde{y}_{d,\tau})\tilde{y}_{d,\tau} - \nu_{d,\tau}^*(\tilde{y}_{d,\tau}, y_{d,\tau})(y_{d,\tau} - \tilde{y}_{d,\tau}) + \psi_{d,\tau}^{D,*} + \psi_{d,\tau}^{R,*}, \quad (8)$$

where the notations $\psi_{d,\tau}^{D,*}$ *and* $\psi_{d,\tau}^{R,*}$ *denote the remaining terms of the optimal dual objectives, which no longer involve* $\tilde{y}_{d,\tau}$.

Proof. See Appendix A. $\qquad\square$

Equation (8) shows that the model parameter estimation at the upper-level problem is associated with the parameterized dual solutions provided by the lower-level problem. Therefore, the primal problems at the lower level are replaced with their dual problems, denoted as $\mathrm{DP}^D(\tilde{y}_{d,\tau}), \mathrm{DP}^R(\tilde{y}_{d,\tau}, y_{d,\tau})$, such that the parameterized dual solutions can be obtained. Therefore, the primal problems in (7d)-(7g) and (7h)-(7k) are replaced with

$$\delta_{i,d,\tau}^*(\tilde{y}_{d,\tau}), \underline{\boldsymbol{\eta}_{d,\tau}}^*(\tilde{y}_{d,\tau}), \overline{\boldsymbol{\eta}_{d,\tau}}^*(\tilde{y}_{d,\tau}), \lambda_{d,\tau}^*(\tilde{y}_{d,\tau}) =$$
$$\underset{\delta_{i,d,\tau}\geq 0, \underline{\boldsymbol{\eta}_{d,\tau}}\geq 0, \overline{\boldsymbol{\eta}_{d,\tau}}\geq 0, \lambda_{d,\tau}}{\arg\max} \mathrm{DP}^D(\tilde{y}_{d,\tau}),$$
$$\mu_{i,d,\tau}^*(\tilde{y}_{d,\tau}, y_{d,\tau}), \zeta_{i,d,\tau}^*(\tilde{y}_{d,\tau}, y_{d,\tau}), \nu_{d,\tau}^*(\tilde{y}_{d,\tau}, y_{d,\tau}) =$$
$$\underset{\mu_{i,d,\tau}\geq 0, \zeta_{i,d,\tau}\geq 0, \nu_{d,\tau}}{\arg\max} \mathrm{DP}^R(\tilde{y}_{d,\tau}, y_{d,\tau}). \quad (9)$$

With the upper-level problem defined in (8) and the lower-level problem defined in (9). The bilevel program in (7) can be equivalently written as,

$$\min_{\Theta_V} \quad \frac{1}{D \cdot T} \sum_{d=1}^D \sum_{\tau=1}^T (8) \quad (10a)$$

$$\text{s.t.} \quad \tilde{y}_{d,\tau} = g(\boldsymbol{s}_{d,\tau}; \Theta_V) \quad (10b)$$

$$0 \leq \tilde{y}_{d,\tau} \leq \bar{y} \quad (10c)$$

$$(9). \quad (10d)$$

As can be seen from the structure of the problem, where the lower-level problem (the follower) in (9) optimizes its objective treating the forecast $\tilde{y}_{d,\tau}$ as the parameter, and the goal of the upper-level problem (the leader) is to determine $\tilde{y}_{d,\tau}$ that minimizes the objective which involves the parameterized solutions returned by the lower level. A natural idea is to resort to an iterative solution for model parameter estimation. The details are given in Section IV.

B. Operational Forecasting Phase

With the estimated parameter $\hat{\Theta}_V$, the forecasting model is ready for use during operation. Usually at 12 pm on day $d-1$ when the decisions are made at the day-ahead stage, with the

available information $\boldsymbol{s}_{d,\tau}, \forall \tau = 1, ..., T$, we forecast the RES generation for each time slot τ, i.e.,

$$\tilde{y}_{d,\tau} = g(\boldsymbol{s}_{d,\tau}; \hat{\Theta}_V), \ \tau = 1, \cdots, T.$$

Then, $\forall \tau = 1, ..., T, \tilde{y}_{d,\tau}$ is fed into the day-ahead problem (1) as a parameter. The real-time problem in (2) is solved when the realization $y_{d,\tau}$ is available at time slot τ on day d.

IV. ITERATIVE ALGORITHM FOR PARAMETER ESTIMATION

We present an iterative learning strategy for the parameter estimation problem (10) in Section IV-A. In Section IV-B, we explain why value-oriented forecasts are more preferable to VPP operators via the derived loss function.

A. Iterative Learning Algorithm

In the iterative learning strategy, the upper level issues the forecast $\tilde{y}_{d,\tau}^{e-1} = g(\boldsymbol{s}_{d,\tau}; \Theta_V^{e-1})$ at the epoch e, where the model parameter is denoted as Θ_V^{e-1}. Then, the forecast $\tilde{y}_{d,\tau}^{e-1}$ is passed to the lower level. The lower level solves the dual problems (9) and passes the optimal dual solutions parameterized by the forecast $\tilde{y}_{d,\tau}^{e-1}$ to the upper level. Such dual solutions update the coefficients in upper-level objectives. The upper level updates the model parameter with the updated objective. We denote the updated parameter at the epoch e as Θ_V^e. The forecast $\tilde{y}_{d,\tau}^e$ is then redetermined with the updated parameter Θ_V^e, and passed to the lower level. Such a process repeats for the epochs $e = 1, 2, ...,$. The parameter Θ_V^0 in the first epoch is randomly initialied. We denote the updated parameter at the last epoch as $\hat{\Theta}_V$, which will be used in the operational forecasting phase. Next, we detail the upper-level problem in the iterative scheme.

The parameterized dual solutions provided by the lower-level problem at the epoch e are denoted as $\delta_{i,d,\tau}^*(\tilde{y}_{d,\tau}^{e-1})$, $\underline{\boldsymbol{\eta}_{d,\tau}}^*(\tilde{y}_{d,\tau}^{e-1})$, $\overline{\boldsymbol{\eta}_{d,\tau}}^*(\tilde{y}_{d,\tau}^{e-1})$, $\lambda_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1})$, $\mu_{i,d,\tau}^*(\tilde{y}_{d,\tau}^{e-1}, y_{d,\tau})$, $\zeta_{i,d,\tau}^*(\tilde{y}_{d,\tau}^{e-1}, y_{d,\tau})$, and $\nu_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1}, y_{d,\tau})$. These dual solutions are treated as the constants at the upper-level problem under the iterative scheme, since the relationship between them and the forecasts is a constant function in the neighborhood of the forecast $\tilde{y}_{d,\tau}^{e-1}$ [30], [31]. Therefore, the last two terms $\psi_{d,\tau}^{D,*}$ and $\psi_{d,\tau}^{R,*}$ in (8) are constants as they only include the parameterized dual solutions. Thus, the upper-level objective in (8) can be further simplified by removing those terms, and the upper-level problem is reformulated as,

$$\min_{\Theta_V} \quad \frac{1}{D \cdot T} \sum_{d=1}^D \sum_{\tau=1}^T -\lambda_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1})\tilde{y}_{d,\tau} -$$
$$\nu_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1}, y_{d,\tau})(y_{d,\tau} - \tilde{y}_{d,\tau}) \quad (11)$$
$$\text{s.t.} \quad \tilde{y}_{d,\tau} = g(\boldsymbol{s}_{d,\tau}; \Theta_V)$$
$$0 \leq \tilde{y}_{d,\tau} \leq \bar{y},$$

whose objective is a linear function of $\tilde{y}_{d,\tau}$ in the neighborhood where the coefficients $\lambda_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1})$, $\nu_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1}, y_{d,\tau})$ remain unchanged. Such a locally linear objective acts as the value-oriented loss function for guiding the parameter
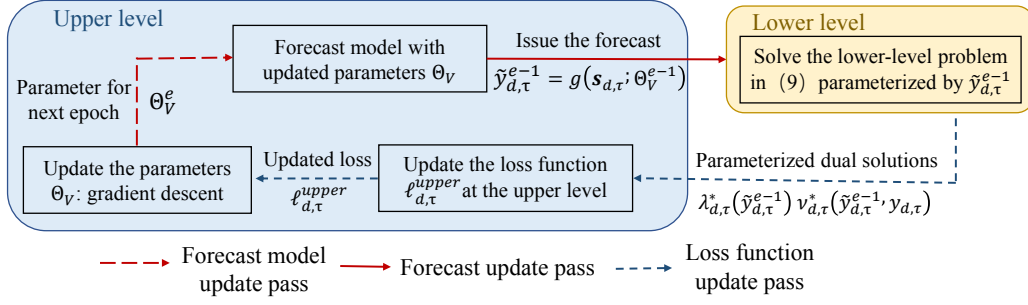
Fig. 3. The illustration of iterative scheme at the epoch e.

estimation. For a sample in the training set, we denote the value-oriented loss function as $\ell_{d,\tau}^{upper}$, i.e.,

$$\ell_{d,\tau}^{upper} := -\lambda_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1})\tilde{y}_{d,\tau} - \nu_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1}, y_{d,\tau})(y_{d,\tau} - \tilde{y}_{d,\tau}) \quad (12)$$

Therefore, the upper-level objective can be rewritten as $\frac{1}{D \cdot T}\sum_{d=1}^{D}\sum_{\tau=1}^{T}\ell_{d,\tau}^{upper}$.

With the defined upper- and lower-level problems, the training under the iterative scheme can be performed in a fixed number of epochs. Although various types of regression models can be used here as a forecasting model, we use neural network (NN) with batch optimization as an example, and its training paradigm is summarized in Algorithm 1. During training, with the defined loss function $\ell_{d,\tau}^{upper}$, the forecasting model parameter Θ_V is updated via gradient descent at the upper level in epoch e, i.e.,

$$\Theta_V^e \leftarrow \Theta_V^{e-1} - \alpha \bigtriangledown_{\Theta_V} \frac{1}{B \cdot T} \sum_{d=1}^{B}\sum_{\tau=1}^{T}\ell_{d,\tau}^{upper} \quad (13)$$

The forecasts are issued by the updated forecasting model and passed to the lower level. Then, the dual solutions are obtained from the lower level and update the loss function at the upper level. Such a process repeats for several epochs. An illustration is shown in Fig. 3.

B. Interpretating Value-oriented Forecasts

We would like to interpret the forecasts with the designed loss function (12). Its coefficients $\lambda_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1})$, $\nu_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1}, y_{d,\tau})$ are the respective optimal dual solutions associated with the equality constraint in day-ahead and real-time problems. They can be interpreted as electricity prices and are associated with the day-ahead marginal cost and real-time marginal cost/utility. In the following, we omit the parentheses and denote them as $\lambda_{d,\tau}^*, \nu_{d,\tau}^*$ for simplicity.

The loss function in (12) can be equivalently written as,

$$-\lambda_{d,\tau}^* y_{d,\tau} + (\lambda_{d,\tau}^* - \nu_{d,\tau}^*)(y_{d,\tau} - \tilde{y}_{d,\tau}) \quad (14)$$

Since the first term $-\lambda_{d,\tau}^* y_{d,\tau}$ is a constant, minimizing (14) is equivalent with minimizing the second term,

$$(\lambda_{d,\tau}^* - \nu_{d,\tau}^*)(y_{d,\tau} - \tilde{y}_{d,\tau}) \quad (15)$$

where $\lambda_{d,\tau}^* - \nu_{d,\tau}^*$ can be interpreted as the difference between day-ahead marginal cost and real-time marginal cost/utility. The asymmetric real-time marginal cost/utility can cause the

value $\lambda_{d,\tau}^* - \nu_{d,\tau}^*$ different in the cases of energy deficit and energy surplus, even if the day-ahead marginal cost takes the same value. For instance, when the real-time marginal cost of flexible resources addressing energy shortage (whose resulted real-time dual solution is denoted as $\nu_{d,\tau}^{+*}$) is larger than the real-time marginal utility of flexible resources addressing energy surplus (whose resulted real-time dual solution is denoted as $\nu_{d,\tau}^{-*}$) and satisfying $\nu_{d,\tau}^{+*} - \lambda_{d,\tau}^* \geq \lambda_{d,\tau}^* - \nu_{d,\tau}^{-*} \geq 0$, the proposed approach tends to forecast less renewable energy. The reason is that per forecast deviation in the case of energy shortage ($y_{d,\tau} < \tilde{y}_{d,\tau}$) results in more loss than in the case of energy surplus ($y_{d,\tau} > \tilde{y}_{d,\tau}$). The quality-oriented forecasts predicting the expected realization treat the two cases equally, ignoring the different impacts on the cost. Therefore, the quality-oriented forecasts result in more cost and are less preferable to the operator.

Algorithm 1 Training Phase of a NN-based forecasting model for Value-oriented Forecasting

---

Require: Learning rate α, batch size B, and initialized forecasting model parameters Θ_V^0

1: **for** epoch $e = 1, 2, \dots$ **do**
2:      Sample a batch $\{\{s_{d,\tau}, y_{d,\tau}\}_{\tau=1}^T\}_{d=1}^B$ from the training set
3:      Issue forecasts $\tilde{y}_{d,\tau}^{e-1} = g(s_{d,\tau}; \Theta_V^{e-1})$, $\forall \tau = 1, \cdots, T$, $d = 1, \cdots, B$ via the forecasting model g.
4:      **for** batch $d = 1, 2, \dots, B$ **do**
5:          Obtain dual solutions $\{\lambda_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1})\}_{\tau=1}^T$ and $\{\nu_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1}, y_{d,\tau})\}_{\tau=1}^T$ as the formula (9)
6:      **end for**
7:      Update the loss function with $\{\{\lambda_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1})\}_{\tau=1}^T\}_{d=1}^B$ and $\{\{\nu_{d,\tau}^*(\tilde{y}_{d,\tau}^{e-1}, y_{d,\tau})\}_{\tau=1}^T\}_{d=1}^B$ as the formula (12).
8:      Update model parameters via gradient descent (13).
9: **end for**

V. EXPERIEMENT SETUP

This section considers the sequential deterministic operational problem faced by a VPP operator managing wind power and analyzes the effectiveness of value-oriented forecasting in this context. A thorough examination of performance in the test set is carried out, focusing on five key aspects: 1) We will demonstrate that our proposed approach leads to lower average operating costs compared to the quality-oriented forecasting

approach, which employs MSE or pinball loss as the loss functions during training; 2) We will uncover the operational advantages under large wind power penetrations by investigating the performances at different levels of wind power capacities; 3) The computational efficiency of the proposed approach will be demonstrated by discussing the computation time in the test set; 4) We will discuss the potential of extending the proposed approach to operational problems involving binary variables; and 5) We will discuss the potential for the proposed approach to adapt to operational changes. Concretely, the changes are mainly related to parameter alterations rather than changes in the structure or type of the problem.

A. Description of Operation Simulations and Data

In the context of the energy dispatch problem, we consider the operation of a centralized VPP operator with two slow-start distributed generators (DGs) and two flexible resources. DGs and wind power generators are considered at the day-ahead stage. During operation, the operator first solves the day-ahead problem (1). After the ground truth of wind power generation is revealed, the real-time operation problem (2) is solved. The detailed model and parameters are given in [32]. In the real-time operation problem, if an energy surplus occurs (where the realization of wind power is greater than the forecast), the centralized operator will use flexible resources to absorb excessive power and obtain utility from the usage of electricity. Therefore, the real-time operating cost will be negative. If an energy shortage occurs (where wind power production is less than forecast), the centralized operator will leverage flexible resources to produce power that solves the shortage. Therefore, the real-time operating cost will be positive.

The hourly wind power production in 2012 from GEFCom 2014 is used, along with the aggregated annual load data from the Low Carbon London dataset, which has the valley demand of 50 kW, the average demand of 56 kW, and the peak demand of 70 kW. Wind and load data are provided in [32]. 80% data is divided into the training set, while the rest forms the test set. The wind data are scaled by multiplying a constant to fit the parameter setting of the case study. The contextual information consists of the predicted wind speed and direction at altitudes of 10 meters and 100 meters.

B. Benchmark Models

In general, we use six benchmark models for comparison: two quality-oriented (**Qua-E** and **Qua-Q**), two value-oriented (**Val-L** and **Val-O**) forecasting models, perfect forecasts (**Per-F**), and a stochastic program (**Sto-OPT**).

1) Qua-E: The forecasting model is trained under MSE and predicts the expected wind power realization. Qua-E corresponds to the forecasting product in Section II-A.
2) Qua-Q: The forecasting model is trained under pinball loss (asymmetric loss function) and predicts quantile.
3) Val-L: The value-oriented forecasting approach proposed by [24] is used as a benchmark, which requires the forecasting model to be linear.
4) Val-O: The forecasting model trained via OptNet [17] is used as another benchmark for value-oriented forecasting.

5) Per-F: The forecasts perfectly match the wind power realization. Since the forecasting error is inevitable, this is a very ideal benchmark.
6) Sto-OPT: We randomly generate 200 wind power scenarios by k-nearest-neighbors [14]. For each sample in the test set, the VPP operator first solves the two-stage stochastic program in (4) based on the scenarios, and then the real-time problem (2). Sto-OPT corresponds to the approach described in Section II-B.

We don't take the approach in Section II-C as the comparison candidate since the performance of this method cannot be better than the approach Sto-OPT in Section II-B; see the results in [28]. Therefore, we only compare the proposed approach with the best candidate. Since a specific type of forecasting model is not the main focus of the work, multi-layer perceptron (MLP) is used as the forecasting model for the proposed approach and benchmarks (Qua-E, Qua-Q, Val-O). The model hyper-parameters are summarized in Table I.

TABLE I
SUMMARY OF MLP HYPER-PARAMETERS

Item	Value
No. of hidden layers	2
No. of neurons in hidden layer	256
Dim. of contextual information	4
Optimizer	Adam
Learning rate	1e-3

C. Verification Metrics

We use the root mean square error (RMSE) and the average operating cost to measure the quality and value of the forecasts, respectively, both of which are negative oriented.

1) Root Mean Square Error: It is defined as

$$\text{RMSE} = \sqrt{\frac{1}{D_{test} \cdot T} \sum_{d=1}^{D_{test}} \sum_{\tau=1}^{T} (y_{d,\tau} - \hat{y}_{d,\tau})^2}, \qquad (16)$$

where D_{test} is the number of days in the test set.

2) Average Operating Cost: It is defined as,

$$\frac{1}{D_{test} \cdot T} \sum_{d=1}^{D_{test}} \sum_{\tau=1}^{T} \boldsymbol{\rho}^\top \boldsymbol{x}_{d,\tau}^* + \boldsymbol{\rho}^{+\top} \boldsymbol{z}_{d,\tau}^{+*} - \boldsymbol{\rho}^{-\top} \boldsymbol{z}_{d,\tau}^{-*}, \qquad (17)$$

where for the deterministic dispatch in Section II-A, $\boldsymbol{x}_{d,\tau}^*$ is the optimal solution of (1), i.e., $\boldsymbol{x}_{d,\tau}^*(\hat{y}_{d,\tau})$, with parentheses omitted. For the two-stage stochastic program in Section II-B, $\boldsymbol{x}_{d,\tau}^*$ is the optimal first-stage solution of (4). For both the deterministic and two-stage stochastic dispatches, the real-time optimal solutions $\boldsymbol{z}_{d,\tau}^{+*}, \boldsymbol{z}_{d,\tau}^{-*}$ are the ones from (2), with parentheses omitted.

VI. RESULTS AND DISCUSSION

A. The Operational Advantage

In this analysis, the capacity of wind power is scaled to 40 kW, whose capacity is 57% of that of the maximum demand.

In the test set, we use RMSE and average operating cost, calculated by (17) (which is the summation of the average day-ahead operating cost and average real-time operating cost), as the evaluation metrics for quality and value, respectively. Here, we consider two quality-oriented forecasting approaches, i.e., Qua-E and Qua-Q. The nominal level of the quantile, which is an input to the pinball loss, is set as $\frac{2}{9}$. Such nominal level is calculated by $\frac{\lambda^* - \nu^{-*}}{\nu^{+*} - \nu^{-*}}$ [22], where $\lambda^*, \nu^{-*}, \nu^{+*}$ are the parameterized dual solutions with the subscript and the parentheses omitted, and are given as an ex-ante "oracle. Specifically, the values are chosen as the marginal generation cost of one of the slow-start generators and flexible resources, i.e., 30 \$/kW, 10 \$/kW, and 100 \$/kW, respectively. The results of the value- and quality-oriented forecasting are reported in Table II. It clearly shows that accurate forecasting does not always benefit the operation. Qua-E achieves a lower RMSE score, which means the forecast shows more correspondence with the realization. Crafted to prioritize value, the proposed approach yields a significant 9.5% operating cost reduction, compared to Qua-E. Also, Qua-Q is trained under pinball loss, and can partially consider the impact of operating costs and is suitable for the specific case where no operation constraints are considered [22]. Therefore, it results in smaller operating costs, compared to Qua-E. Since it is unsuitable for the more general operation problem discussed in this work, its operating cost is still higher than the proposed one. The results manifest the operational advantage of our approach.

TABLE II
RMSE AND AVERAGE OPERATING COST IN THE TEST SET.

	RMSE /kW	Average operating cost/$	Average day-ahead cost/$	Average real-time cost/$
Qua-E	7.2	39344	33191	6153
Qua-Q	9.2	36009	37221	-1212
Proposed	11.1	35591	38671	-3080

Fig. 4 displays the 4-day wind power forecast profiles of the value- and quality-oriented forecasting approaches. The real-time problem has a clear influence on value-oriented forecasting. Due to the higher loss for energy shortage than energy surplus, the proposed approach and Qua-Q tend to forecast less wind power production than Qua-E, to avoid the less profitable situation of underproduction, such that the energy shortage is less likely to happen. This point can be further demonstrated by the last two columns of Table II, which show that on average, the proposed approach has lower real-time operating cost than Qua-E and Qua-Q. Since the proposed approach tends to forecast less wind power (which has zero marginal cost in the day-ahead problem), the proposed approach has a larger day-ahead operating cost. However, thanks to co-minimizing the overall day-ahead and the real-time costs at the training phase, the proposed approach achieves lower costs for the overall operation.

Furthermore, to demonstrate the advantage of the proposed iterative solution method, we compare it with Val-L [24]
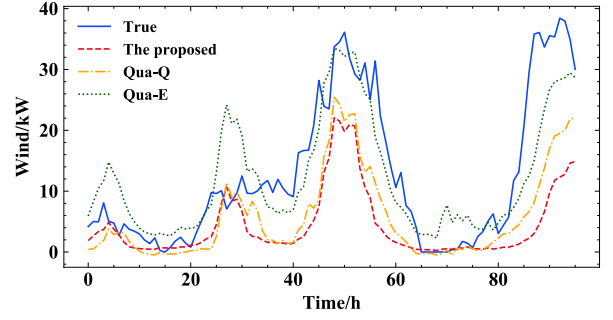

Fig. 4. 4-day wind power forecast profiles issued by the value- and quality-oriented forecasting approaches. Qua-E and Qua-Q stand for quality-oriented forecasts that issue the expectation and $\frac{2}{9}$ quantile of the wind power.

and Val-O [17]. To accommodate the setting in [24] where the ramping constraints are not considered in day-ahead, we omit the ramping constraint in (1c) for this comparison. The average operating cost in the test set is shown in Table III. Additionally, we present the average operating cost resulting from Qua-E. It shows that, due to the integration of more advanced forecasting models (NN v.s. linear model), our approach and Val-O achieve lower average cost, compared to Val-L. Additionally, despite employing a more sophisticated model, Qua-E has worse performance compared to Val-L, which employs a linear model. This underscores the necessity of aligning the forecasting model training with the decision value. The training time of neural network-based models is also reported in Table III. As solving the dual problems at the lower level and estimating the forecasting model parameters are iteratively performed, the proposed approach has a longer training time than Qua-E but much shorter than that of the Val-O based on OptNet. Since Val-O needs to calculate the inverse of the Jacobian matrix for the KKT conditions at each epoch, computational cost can be very high.

TABLE III
AVERAGE OPERATING COST IN THE TEST SET AND TRAINING TIME OF NEURAL NETWORKS. THE RAMPING CONSTRAINT (1c) HAS BEEN DROPPED FROM THE DAY-AHEAD DISPATCH MODEL IN (1).

	Proposed	Val-L [24]	Val-O [17]	Qua-E
Average operating cost	35598 \$	35781 \$	35628 \$	38667 \$
Training time	3 min 50s	-	45 min	8.26 s

To make a more nuanced analysis, we identify the sensitive periods when our approach has more obvious improvement compared to Qua-E. The hourly cost reduction of the proposed approach compared to Qua-E is calculated. Also, via the derived loss function (15), we identify the factors that affect the hourly cost reduction. The coefficient $\lambda^*_{d,\tau} - \nu^*_{d,\tau}$ is the marginal contribution of the forecasting error $y_{d,\tau} - \tilde{y}_{d,\tau}$ to the loss function. Therefore, in the following, we analyze the impact of $\lambda^*_{d,\tau} - \nu^*_{d,\tau}$ on the hourly cost reduction. We categorize the hourly $\lambda^*_{d,\tau} - \nu^*_{d,\tau}$ under the proposed forecasts into four intervals whose average values range from low

to high. The average hourly cost reduction within the four intervals is calculated and provided in Fig. 5. It shows that the smaller the $\lambda_{d,\tau}^* - \nu_{d,\tau}^*$ is, the more cost reduction our approach can lead to. This is because, in the setting that the real-time marginal cost (the marginal cost of flexible resources addressing energy deficit) is larger than the real-time marginal utility (the marginal utility of flexible resources addressing energy surplus), our approach tends to forecast less wind power to avoid the undesirable energy deficit situation. That is, our approach tends to have $y_{d,\tau} - \tilde{y}_{d,\tau} \geq 0$. When $\lambda_{d,\tau}^* - \nu_{d,\tau}^*$ is negative, such as the first three points, the term in (15) is negative in average under the proposed forecasts. This leads to a reduction in the loss and a corresponding reduction in the operating cost, compared to Qua-E. Also, the smaller the $\lambda_{d,\tau}^* - \nu_{d,\tau}^*$ is, the more cost reduction can be expected.
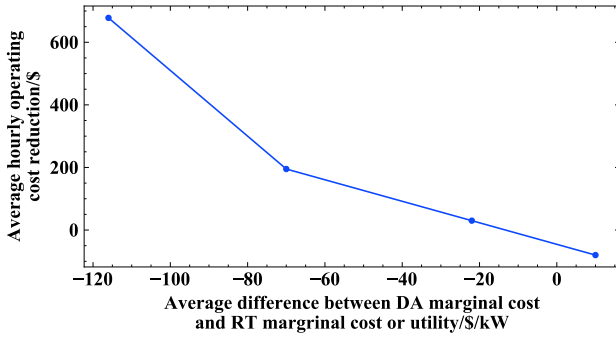

Fig. 5. Average hourly cost reduction of the proposed approach compared to Qua-E under different levels of difference between DA marginal cost and RT marginal cost/utility.

B. Sensitivity Analysis: Different Wind Power Penetrations

This study investigates the performance of value-oriented forecasting under different levels of wind power capacities: 20, 30, and 40 kW. Under these penetration levels, the average operating cost of the proposed approach, Qua-E, and Per-F are summarized in Fig. 6. It shows with the increase in wind power capacity, the average operating cost decreases as wind power, which has zero marginal cost, gradually contributes a larger share toward balancing the load. Also, the trend of cost reduction is more evident in the proposed approach, compared to Qua-E. Furthermore, under different levels of wind power penetration, the proposed value-oriented forecasting has lower operating costs compared to Qua-E, and has comparable performance with Per-F. The cost reduction is more evident under large wind power capacity. The results show that the benefit of value-oriented forecasting is more significant under the higher penetration of renewable energy resources.

C. Computational Complexity Comparison

Furthermore, to show the proposed approach is more computationally efficient than the two-stage stochastic program Sto-OPT, we compare the computation time of the two approaches in the test set. The scale of the Sto-OPT increases tremendously with the increase of the number of scenarios and the number of flexible resources whose dispatch is determined
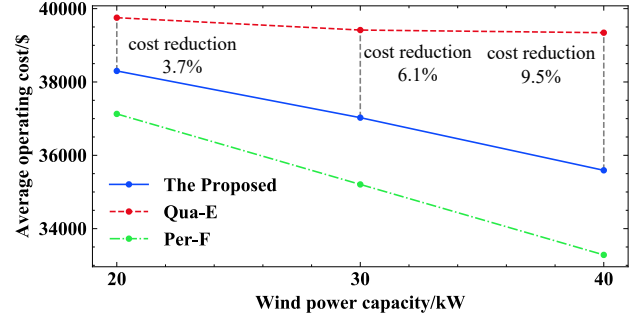

Fig. 6. Average operating cost under different wind power capacities. Qua-E stands for quality-oriented forecasts predicting expected wind power. Per-F stands for perfect forecasts and therefore results in the lowest operating cost.

at the second stage. Therefore, to show the advantage in a larger-scale problem, we consider an operation problem with a greater number of flexible resources. Concretely, there are 10 flexible resources addressing energy shortage (whose capacity is 6 kW each and marginal operating cost uniformly ranges from 90 $/kW to 120 $/kW), and 10 flexible resources addressing energy surplus (whose capacity is 6 kW each and marginal operating utility uniformly ranges from 10 $/kW to 20 $/kW). The wind power capacity is set as 40 kW.

We compare the computation time and average operating cost of the two approaches in the test set. The results are in Table IV. It shows that the proposed approach is much faster than the two-stage stochastic program while achieving a similar average operating cost.

TABLE IV
COMPUTATION TIME IN THE TEST SET AND AVERAGE OPERATING COST OF THE PROPOSED APPROACH AND STO-OPT.

	Computation time/s	Average operating cost/$
Proposed	3.39	35846
Sto-OPT	419	35837

D. Extending to Operational Problems with Binary Variables

We show the possibility of applying the proposed approach to the operational problem with binary variables. The day-ahead operation is a unit commitment problem where the binary variables indicate the on/off status of generators. The real-time problem dispatches flexible resources to address the forecast deviation, whose formulation is in (2). At the training phase, the day-ahead unit commitment problem is relaxed to a linear program [33], where the binary variables are relaxed to the continuous ones in the range of 0 and 1. We incorporate the relaxed day-ahead unit commitment and the real-time problem into the proposed approach for training a forecasting model in a value-oriented way. At the operational phase, we apply the issued forecast to serve as the input parameter to the unit commitment with binary variables and use the average overall cost in the test set as the evaluation metric. The mathematical programs of the unit commitment problem and the relaxed one are in Appendix B.

Considering different wind power capacities, we compare the average operating cost of the proposed approach and Qua-E. The results are given in Fig. 7. It shows that the proposed approach can result in a lower average operating cost compared to Qua-E. The cost reduction is more obvious under the large penetration of wind power. Therefore, the proposed approach is effective for the unit commitment problem.
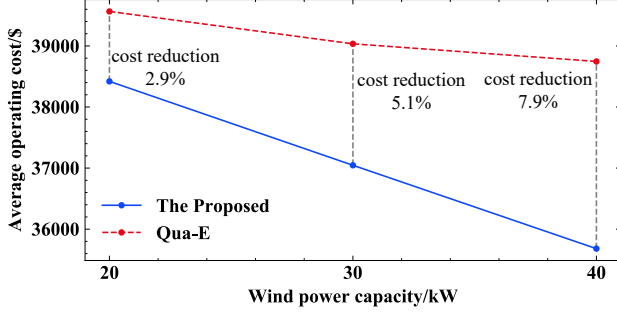

Fig. 7. Average operating cost of day-ahead unit commitment and real-time energy dispatch problems under different wind power capacities. Qua-E stands for quality-oriented forecasts, predicting the expected realization.

Furthermore, we note that the performance can be improved by relaxing the mixed integer program into a linear program in a more precise manner. For instance, [34] proposes to generate an exact continuous surrogate for the original discrete optimization problem by adding cutting planes. In this way, the mixed integer program is equivalently transformed into a linear program, which fits the proposed approach. We would like to leave this as the future work.

E. Adaptability to Operational Change

In this section, we test our approach under operational changes. Specifically, we consider the case where the parameters in the operational problems are fixed during the training phase. After training, a different set of parameters is used in the operational problems to test the forecasting performance. We provide an example of training the forecasting model with fixed marginal costs in day-ahead and in real-time, but using different real-time marginal costs at the operational phase.

Two different settings of real-time marginal costs $\rho^+ = [\rho_1^+, \rho_2^+]^\top$ are considered at the operational phase. In the first setting, the real-time marginal costs ρ_1^+, ρ_2^+ are still larger than the real-time marginal utility ρ_1^-, but take different values from the training phase. In the second setting, the real-time marginal costs ρ_1^+, ρ_2^+ are smaller than the real-time marginal utility ρ_1^-. We note that only the real-time marginal costs ρ_1^+, ρ_2^+ change at the test phase, and other parameters remain the same as those at the training phase. In Table V, we list the real-time marginal costs ρ_1^+, ρ_2^+ of the two settings. In these two settings, we compare the average operating cost in the test set of the proposed approach and Qua-E. The results are in Table VI.

It shows that in the high-cost setting, although the values of the marginal costs at the test phase are different from those at the training phase, our approach can adapt to the change, and achieves a lower average cost compared to Qua-E. However, in the low-cost setting, the proposed approach

	ρ_1^+	ρ_2^+
High-cost setting	90 $/kW	100 $/kW
Low-cost setting	5 $/kW	15 $/kW

	Proposed	Qua-E
High-cost setting	35511 $	37905 $
Low-cost setting	27578 $	24881 $

doesn't perform well and has an even larger cost compared to Qua-E. Therefore, in this case, the adaptability of the proposed approach highly depends on how the parameters change. There is no guarantee that the proposed approach can consistently adapt to these changes effectively. A potential solution is considering the cost change at the training phase. We would like to leave this as future work.

VII. CONCLUSIONS

This paper introduces a novel value-oriented point forecasting methodology tailored for VPP operators who possess renewable energy sources and solve the energy dispatch problems in sequential order. With the forecasts, operators are allowed to solve a computationally efficient deterministic problem during the day-ahead operation, while still attaining low expected operating costs of the day-ahead and real-time stages. To achieve this, we formulate the parameter estimation for the forecasting model as a bilevel program at the training phase. An iterative learning strategy is proposed to solve it, where the loss function aligning with the decision value is derived. For the operation problems in the form of LP, such a function exhibits local linearity with regard to forecasts.

Numerical outcomes from the test set reveal that, in comparison to the quality-oriented forecasting methods, the proposed approach yields a larger forecasting error measured by RMSE. However, it manages to attain a reduced average overall operating cost. This advantage becomes more sound, particularly under high levels of wind power integration. The computational efficiency is also demonstrated through a comparison of computation time with a stochastic program. Additionally, we demonstrate the potential of applying the proposed approach to operational problems involving binary variables and adapting to operational changes.

We note that the goal of this study is not to replace the quality-oriented forecasting approach but rather to provide tools for enhanced RES forecasts tailored to specific operational problems to achieve lower operational costs. In the future, we plan to develop value-oriented forecasts that can better adapt to operational changes and applicable to operation problems with network constraints.

11

APPENDIX A
PROOF OF PROPOSITION 1

Consider the linear programs (7d)-(7g) and (7h)-(7k). We derive the dual problems for them, respectively. Taking the problem (7d)-(7g) for instance, we rewrite it as

$$\min_{\{\boldsymbol{x}_{d,\tau}\}_{\tau=1}^{T}} \quad \sum_{\tau=1}^{T} \boldsymbol{\rho}^{\top}\boldsymbol{x}_{d,\tau} \tag{18a}$$

$$\text{s.t.} \quad \boldsymbol{A}_D \boldsymbol{x}_d \leq \boldsymbol{b}_D : \boldsymbol{\chi}_d \tag{18b}$$

$$\mathbf{1}^{\top}\boldsymbol{x}_{d,\tau} + \tilde{y}_{d,\tau} = l_{d,\tau} : \lambda_{d,\tau}, \forall \tau = 1, ..., T, \tag{18c}$$

where $\boldsymbol{\rho}, \boldsymbol{A}_D, \boldsymbol{b}_D$ are the coefficients, $\boldsymbol{x}_d = [\boldsymbol{x}_{d,\tau}]_{\tau=1,...,T}$, $\boldsymbol{\chi}_d = [[\delta_{i,d,\tau}]_{i\in I^D, \tau=1,...,T}, [\boldsymbol{\eta}_{d,\tau}, \overline{\boldsymbol{\eta}_{d,\tau}}]_{\tau=1,...,T-1}]$. The Lagrangian function of (18) can be derived as

$$\boldsymbol{\chi}_d^{\top}(\boldsymbol{A}_D\boldsymbol{x}_d - \boldsymbol{b}_D) + \sum_{\tau=1}^{T} \boldsymbol{\rho}^{\top}\boldsymbol{x}_{d,\tau} + \lambda_{d,\tau}(l_{d,\tau} - \mathbf{1}^{\top}\boldsymbol{x}_{d,\tau} - \tilde{y}_{d,\tau}). \tag{19}$$

Based on the Lagrangian function (19), the objective of the dual problem is

$$-\boldsymbol{\chi}_d^{\top}\boldsymbol{b}_D + \sum_{\tau=1}^{T} -\lambda_{d,\tau}\tilde{y}_{d,\tau} + \lambda_{d,\tau}l_{d,\tau}. \tag{20}$$

Likewise, we rewrite the linear program (7h)-(7k) as

$$\{\min_{\tilde{\boldsymbol{z}}_{d,\tau}} \quad \boldsymbol{c}_R^{\top}\tilde{\boldsymbol{z}}_{d,\tau} \tag{21a}$$

$$\text{s.t.} \quad \boldsymbol{A}_R \tilde{\boldsymbol{z}}_{d,\tau} \leq \boldsymbol{b}_R : \boldsymbol{\iota}_{d,\tau} \tag{21b}$$

$$\boldsymbol{d}_R^{\top}\tilde{\boldsymbol{z}}_{d,\tau} + y_{d,\tau} - \tilde{y}_{d,\tau} = 0 : \nu_{d,\tau}\}, \forall \tau = 1, ..., T, \tag{21c}$$

where $\boldsymbol{c}_R, \boldsymbol{A}_R, \boldsymbol{b}_R, \boldsymbol{d}_R$ are the coefficients. $\tilde{\boldsymbol{z}}_{d,\tau} = [\tilde{\boldsymbol{z}}_{d,\tau}^{+}; \tilde{\boldsymbol{z}}_{d,\tau}^{-}]$, $\boldsymbol{c}_R = [\boldsymbol{\rho}^{+}; -\boldsymbol{\rho}^{-}]$ $\boldsymbol{d}_R = [\mathbf{1}; -\mathbf{1}]$ and $\boldsymbol{\iota}_{d,\tau} = [[\mu_{i,d,\tau}]_{i\in I^{R+}}; [\zeta_{i,d,\tau}]_{i\in I^{R-}}]$. The objective of dual problem for (7h)-(7k) can be derived in a similar way

$$-\nu_{d,\tau}(y_{d,\tau} - \tilde{y}_{d,\tau}) - \boldsymbol{\iota}_{d,\tau}^{\top}\boldsymbol{b}_R, \forall \tau = 1, ..., T. \tag{22}$$

The summation of the dual objectives in (20) and (22) over T hours consists of the term $\sum_{\tau=1}^{T} -\lambda_{d,\tau}\tilde{y}_{d,\tau} - \nu_{d,\tau}(y_{d,\tau} - \tilde{y}_{d,\tau})$, with the remaining term only regarding the dual variables $\lambda_{d,\tau}, \nu_{d,\tau}, \boldsymbol{\chi}_d, \boldsymbol{\iota}_{d,\tau}$. For each sample indexed by d, τ, by plugging the parameterized dual solutions $\lambda_{d,\tau}^{*}(\tilde{y}_{d,\tau})$, $\nu_{d,\tau}^{*}(\tilde{y}_{d,\tau}, y_{d,\tau})$, $\delta_{i,d,\tau}^{*}(\tilde{y}_{d,\tau}), \boldsymbol{\eta}_{d,\tau}^{*}(\tilde{y}_{d,\tau}), \overline{\boldsymbol{\eta}_{d,\tau}}^{*}(\tilde{y}_{d,\tau})$, $\mu_{i,d,\tau}^{*}(\tilde{y}_{d,\tau}, y_{d,\tau})$ and $\zeta_{i,d,\tau}^{*}(\tilde{y}_{d,\tau}, \overline{y_{d,\tau}})$ into the objective of the dual problems in (20),(22), the summation of the optimal dual objectives for the sample indexed by d, τ is

$$-\lambda_{d,\tau}^{*}(\tilde{y}_{d,\tau})\tilde{y}_{d,\tau} - \nu_{d,\tau}^{*}(\tilde{y}_{d,\tau}, y_{d,\tau})(y_{d,\tau} - \tilde{y}_{d,\tau}) + \psi_{d,\tau}^{D,*} + \psi_{d,\tau}^{R,*}. \tag{23}$$

where $\psi_{d,\tau}^{D,*}$ and $\psi_{d,\tau}^{R,*}$ represent the remaining terms related with the dual solutions. (23) is equivalent with the term in the

curly bracket of (7a) by the strong duality theory, which ends the proof.

APPENDIX B
MATHEMATICAL FORMULATION OF UNIT COMMITMENT PROBLEMS IN SECTION VI-D

We give the unit commitment problem in day-ahead,

$$\min_{\{\boldsymbol{x}_{d,\tau}, \boldsymbol{u}_{d,\tau}\}_{\tau=1}^{T}} \quad \sum_{\tau=1}^{T} \boldsymbol{\rho}^{\top}\boldsymbol{x}_{d,\tau} + \boldsymbol{\rho}_0^{\top}\boldsymbol{u}_{d,\tau} \tag{24a}$$

$$\text{s.t.} \quad \mathbf{0} \leq \boldsymbol{x}_{d,\tau} \leq \overline{\boldsymbol{x}} \circ \boldsymbol{u}_{d,\tau}, \forall \tau = 1, ..., T \tag{24b}$$

$$\boldsymbol{x}_{d,\tau+1} - \boldsymbol{x}_{d,\tau} \leq \boldsymbol{r} \circ \boldsymbol{u}_{d,\tau+1}, \forall \tau = 1, ..., T-1 \tag{24c}$$

$$-\boldsymbol{r} \circ \boldsymbol{u}_{d,\tau} \leq \boldsymbol{x}_{d,\tau+1} - \boldsymbol{x}_{d,\tau}, \forall \tau = 1, ..., T-1 \tag{24d}$$

$$\mathbf{1}^{\top}\boldsymbol{x}_{d,\tau} + \hat{y}_{d,\tau} = l_{d,\tau}, \forall \tau = 1, ..., T \tag{24e}$$

where $\circ$ is the elementwise multiplication. $\boldsymbol{\rho}$ is the marginal generation cost and $\boldsymbol{\rho}_0$ is the start-up cost. $\boldsymbol{x}_{d,\tau}$ is the generation schedule at time τ on day d. $\boldsymbol{u}_{d,\tau}$ is the binary variable indicating on/off status. $\boldsymbol{r}$ is the ramping limit. $\hat{y}_{d,\tau}$ and $l_{d,\tau}$ are the wind power forecast and load demand. By relaxing the binary variable $\boldsymbol{u}_{d,\tau}$ to be a continuous variable in the range of 0 and 1, the relaxed unit commitment problem is as follows,

$$\min_{\{\boldsymbol{x}_{d,\tau}, \boldsymbol{u}_{d,\tau}\}_{\tau=1}^{T}} \quad \sum_{\tau=1}^{T} \boldsymbol{\rho}^{\top}\boldsymbol{x}_{d,\tau} + \boldsymbol{\rho}_0^{\top}\boldsymbol{u}_{d,\tau} \tag{25a}$$

$$\text{s.t.} \quad \text{(24b), (24c), (24d), (24e)} \tag{25b}$$

$$\mathbf{0} \leq \boldsymbol{u}_{d,\tau} \leq \mathbf{1} \tag{25c}$$

which is a linear program.

REFERENCES

[1] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating renewables in electricity markets: operational problems.* Springer Science & Business Media, 2013, vol. 205.

[2] Y. Wang, N. Zhang, Y. Tan, T. Hong, D. S. Kirschen, and C. Kang, "Combining probabilistic load forecasts," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3664–3674, 2018.

[3] Y. Zhang, H. Wen, Q. Wu, and Q. Ai, "Optimal adaptive prediction intervals for electricity load forecasting in distribution systems via reinforcement learning," *IEEE Transactions on Smart Grid*, 2022.

[4] H. Wen, P. Pinson, J. Ma, J. Gu, and Z. Jin, "Continuous and distribution-free probabilistic wind power forecasting: A conditional normalizing flow approach," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 4, pp. 2250–2263, 2022.

[5] Y. Chen, Y. Wang, D. Kirschen, and B. Zhang, "Model-free renewable scenario generation using generative adversarial networks," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3265–3275, 2018.

[6] T. Gneiting and M. Katzfuss, "Probabilistic forecasting," *Annual Review of Statistics and Its Application*, vol. 1, pp. 125–151, 2014.

[7] A. H. Murphy, "What is a good forecast? an essay on the nature of goodness in weather forecasting," *Weather and forecasting*, vol. 8, no. 2, pp. 281–293, 1993.

[8] T. Carriere and G. Kariniotakis, "An integrated approach for value-oriented energy forecasting and data-driven decision-making application to renewable energy trading," *IEEE transactions on smart grid*, vol. 10, no. 6, pp. 6933–6944, 2019.

[9] C. Zhao, C. Wan, and Y. Song, "Cost-oriented prediction intervals: On bridging the gap between forecasting and decision," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 3048–3062, 2021.

[10] X. Chen, Y. Yang, Y. Liu, and L. Wu, "Feature-driven economic improvement for network-constrained unit commitment: A closed-loop predict-and-optimize framework," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 3104–3118, 2021.

[11] A. Stratigakos, S. Camal, A. Michiorri, and G. Kariniotakis, "Prescriptive trees for integrated forecasting and optimization applied in trading of renewable energy," *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4696–4708, 2022.

[12] Y. Zhang, H. Wen, and Q. Wu, "A contextual bandit approach for value-oriented prediction interval forecasting," *IEEE Transactions on Smart Grid*, 2023.

[13] J. Zhang, Y. Wang, and G. Hug, "Cost-oriented load forecasting," *Electric Power Systems Research*, vol. 205, p. 107723, 2022.

[14] D. Bertsimas and N. Kallus, "From predictive to prescriptive analytics," *Management Science*, vol. 66, no. 3, pp. 1025–1044, 2020.

[15] A. N. Elmachtoub and P. Grigas, "Smart "predict, then optimize"," *Management Science*, vol. 68, no. 1, pp. 9–26, 2022.

[16] J. Mandi, J. Kotary, S. Berden, M. Mulamba, V. Bucarey, T. Guns, and F. Fioretto, "Decision-focused learning: Foundations, state of the art, benchmark and future opportunities," *arXiv preprint arXiv:2307.13565*, 2023.

[17] P. Donti, B. Amos, and J. Z. Kolter, "Task-based end-to-end model learning in stochastic optimization," *Advances in neural information processing systems*, vol. 30, 2017.

[18] D. Wahdany, C. Schmitt, and J. L. Cremer, "More than accuracy: end-to-end wind power forecasting that optimises the energy system," *Electric Power Systems Research*, vol. 221, p. 109384, 2023.

[19] S. Shah, K. Wang, B. Wilder, A. Perrault, and M. Tambe, "Decision-focused learning without decision-making: Learning locally optimized decision losses," *Advances in Neural Information Processing Systems*, vol. 35, pp. 1320–1332, 2022.

[20] D. S. Kirschen and G. Strbac, *Fundamentals of power system economics*. John Wiley & Sons, 2018.

[21] E. Y. Bitar, R. Rajagopal, P. P. Khargonekar, K. Poolla, and P. Varaiya, "Bringing wind energy to market," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1225–1235, 2012.

[22] P. Pinson, "Wind Energy: Forecasting Challenges for Its Operational Management," *Statistical Science*, vol. 28, no. 4, pp. 564 – 585, 2013.

[23] K. J. Arrow, T. Harris, and J. Marschak, "Optimal inventory policy," *Econometrica: Journal of the Econometric Society*, pp. 250–272, 1951.

[24] J. M. Morales, M. Muñoz, and S. Pineda, "Prescribing net demand for two-stage electricity generation scheduling," *Operations Research Perspectives*, vol. 10, p. 100268, 2023.

[25] J. D. Garcia, A. Street, T. Homem-de Mello, and F. D. Muñoz, "Application-driven learning: A closed-loop prediction and optimization approach applied to dynamic reserves and demand forecasting," *arXiv preprint arXiv:2102.13273*, 2021.

[26] A. Gomez-Exposito, A. J. Conejo, and C. Cañizares, *Electric energy systems: analysis and operation*. CRC press, 2018.

[27] D. Fan, S. Zhang, H. Huang, L. Zhou, Y. Wang, and X. Xiao, "Three-stage day-ahead scheduling strategy for regional thermostatically controlled load aggregators," *Protection and Control of Modern Power Systems*, vol. 8, no. 2, pp. 1–11, 2023.

[28] J. M. Morales, M. Zugno, S. Pineda, and P. Pinson, "Electricity market clearing with improved scheduling of stochastic production," *European Journal of Operational Research*, vol. 235, no. 3, pp. 765–774, 2014.

[29] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[30] X. Zheng, M. E. Khodayar, J. Wang, M. Yue, and A. Zhou, "Distributionally robust multistage dispatch with discrete recourse of energy storage systems," *IEEE Transactions on Power Systems*, 2024.

[31] F. Borrelli, A. Bemporad, and M. Morari, "Geometric algorithm for multiparametric linear programming," *Journal of optimization theory and applications*, vol. 118, pp. 515–540, 2003.

[32] "Descriptions of day-ahead and real-time energy dispatch problems." https://github.com/yufan0157/An-iterative-solution-approach-for-value-oriented-forecasting/tree/main, 2024.

[33] D. Zhao, V. Dvorkin, S. Delikaraoglou, A. Botterud *et al.*, "Uncertainty-informed renewable energy scheduling: A scalable bilevel framework," *IEEE Transactions on Energy Markets, Policy and Regulation*, 2023.

[34] A. Ferber, B. Wilder, B. Dilkina, and M. Tambe, "Mipaal: Mixed integer program as a layer," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 02, 2020, pp. 1504–1511.

13