Robust Generation Dispatch with Purchase of Renewable Power and Load Predictions

Rui Xie, *Member, IEEE,* Pierre Pinson, *Fellow, IEEE,* Yin Xu, *Senior Member, IEEE,* Yue Chen, *Member, IEEE*

***Abstract*—The increasing use of renewable energy sources (RESs) and responsive loads has made power systems more uncertain. Meanwhile, thanks to the development of advanced metering and forecasting technologies, predictions by RES and load owners are now attainable. Many recent studies have revealed that pooling the predictions from RESs and loads can help the operators predict more accurately and make better dispatch decisions. However, how the prediction purchase decisions are made during the dispatch processes needs further investigation. This paper fills the research gap by proposing a novel robust generation dispatch model considering the purchase and use of predictions from RESs and loads. The prediction purchase decisions are made in the first stage, which influence the accuracy of predictions from RESs and loads, and further the uncertainty set and the worst-case second-stage dispatch performance. This two-stage procedure is essentially a robust optimization problem with decision-dependent uncertainty (DDU). A mapping-based column-and-constraint generation (C&CG) algorithm is developed to overcome the potential failures of traditional solution methods in detecting feasibility, guaranteeing convergence, and reaching optimal strategies under DDU. Case studies demonstrate the effectiveness, necessity, and scalability of the proposed model and algorithm.**

***Index Terms*—Robust generation dispatch, prediction purchase, decision-dependent uncertainty, mapping-based C&CG**

NOMENCLATURE

A. Abbreviations

C&CG	Column-and-constraint generation
DDU	Decision-dependent uncertainty
DIU	Decision-independent uncertainty
RES	Renewable energy resource
RGD	Robust generation dispatch
RO	Robust optimization
RUC	Robust unit commitment

B. Indices and Sets

$i \in \mathcal{I}_r, \mathcal{I}_d, \mathcal{I}$	Index and set of RESs/loads/agents
$j \in \mathcal{J}$	Index and set of controllable generators
$l \in \mathcal{L}$	Index and set of transmission lines
$t \in \mathcal{T}$	Index and set of periods
$\mathcal{X}$	Feasible set of the first-stage variable
$\mathcal{X}_R$	Robust feasible set
$\mathcal{U}$	Uncertainty set
$\mathcal{Y}(x,u)$	Feasible set of the second-stage variable
$V(\cdot)$	Vertex set

C. Parameters

I_r, I_d, I	Number of RESs/loads/agents
J	Number of generators.
L	Number of transmission lines
T	Number of periods
$\overline{u}_{it}^r, \overline{u}_{it}^d$	Expected value of the maximum power output of RES i or power demand of load i in period t
$\sigma_{U_i}^2$	Variance of operator's estimation for the uncertainty of agent i
δ, ξ	Probability parameters of uncertainty set
m	Prediction cost parameter
$\theta_{jt}, \theta_{jt}^U, \theta_{jt}^D$	Binary parameter for the on/startup/shutdown state of generator j in period t
ρ_j	Output cost coefficient of generator j
ρ_j^+, ρ_j^-	Upward/downward regulation cost coefficient of generator j
ρ^c	Penalty coefficient of real-time RES power curtailment
γ_j^+, γ_j^-	Upward/downward reserve cost coefficient of generator j
R_j^+, R_j^-	Maximum upward/downward reserve of generator j
$\mathcal{R}_j^+, \mathcal{R}_j^-$	Maximum upward/downward ramping of generator j
P_j^{min}, P_j^{max}	Minimum/maximum output of generator j
F_l	Capacity of transmission line l
π_{jl}, π_{il}	Power transfer distribution factors

D. Decision Variables

u_{it}	Uncertain power of agent i in period t
τ_i	Prediction accuracy of the uncertainty of agent i
C_i	Operator's payment to agent i for predictions
p_{jt}	Contemporary output of generator j in t
r_{jt}^+, r_{it}^-	Upward/downward reserve of generator j in period t
p_{jt}^+, p_{jt}^-	Upward/downward power adjustment of generator j in period t
p_{it}^c	Real-time power curtailment of RES i in period t

I. INTRODUCTION

This work was supported by the National Natural Science Foundation of China under Grant No. 52307144 and the Shun Hing Institute of Advanced Engineering, the Chinese University of Hong Kong, through Project RNE-p2-23 (Grant No. 8115071). (Corresponding to Y. Chen)

R. Xie and Y. Chen are with the Department of Mechanical and Automation Engineering, the Chinese University of Hong Kong, HKSAR. (email: ruixie@cuhk.edu.hk; yuechen@mae.cuhk.edu.hk)

P. Pinson is with the Dyson School of Design Engineering, Imperial College London, UK. (email: p.pinson@imperial.ac.uk)

Y. Xu is with the School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China. (email: xuyin@bjtu.edu.cn)

GEOGRAPHICALLY distributed renewable energy sources (RESs) and responsive loads have flourished in recent years, posing great challenges on power system operations including higher risks of power imbalance and inadequate ramping capacities [1]. Robust generation dispatch (RGD) is an essential way to cope with the rising uncertainty [2]. It allows the transmission grid operator to minimize the operation cost under the worst-case uncertainty realizations. The performance of RGD largely depends on the quality of the uncertainty set it employs. A vast literature has been devoted to building more accurate uncertainty sets [3].

The parameters of an uncertainty set were determined to achieve the best trade-off between security and conservativeness [4]. A method for building polyhedral uncertainty sets based on the theory of coherent risk measures was introduced [5]. In addition to a better selection of parameters of the uncertainty set, enhancing the accuracy of uncertainty predictions is another important way. An improved wind forecasting framework considering the spatio-temporal correlation in wind speed was developed and used to build a more accurate uncertainty set [6]. A dynamic uncertainty set was proposed in [7], which improves the forecasts for the current period based on uncertainty realizations in the past few periods. In the studies above, the uncertainty sets were estimated by the operator through processing their own data and forecasts. Thanks to the development of advanced metering and forecasting technologies, there is an emerging opportunity to further improve the forecast and uncertainty set. That is, through pooling of predictions from RES and load owners [8].

To aggregate and make use of the local predictions, data/information markets have been introduced. For example, regression markets were developed to aggregate local data for energy forecasting with proper incentives based on cooperative game theory [9], [10]. A data-sharing mechanism was designed for electricity retailers to improve their profits in the wholesale market [11]. A blockchain-based data transmission framework was developed for energy imbalance market [12]. Due to the high communication burdens and the risk of private data leakage, sharing data to perform a central prediction may not always be a good way. An alternative approach is to build an information market for aggregating predictions [8], [13]. A binary prediction market was proposed in [14] to forecast RESs. The day-ahead trading and valuation of load forecast were studied in [15]. This paper chooses to focus on the latter approach, i.e., to help the operator improve the uncertainty set by purchasing predictions from RESs and loads.

The fruitful works above focused on the forecasting tasks. But in fact, the forecast improvement via prediction purchase and the power system dispatch have mutual influences. Therefore, a holistic model that integrates them is necessary, which remains unexplored. This paper takes an initial step by proposing a RGD model that allows the operator to purchase and use predictions from RESs and loads. The proposed model turns out to be a case of robust optimization (RO) with decision-dependent uncertainty (DDU). DDU appears in various optimization models under uncertain environments and brings new solution challenges.

In decision-dependent stochastic programming (SP) problems, the current decision affects the underlying probability distribution or the realization time of the future uncertainty [16]. These problems were solved by hybrid mixed-integer disjunctive programming [16], quasi-exact solution approach based on discretization and linearization [17], Benders decomposition-based method [18], genetic algorithm [19], and scenario reduction method [20].

For RO problems with DDU, a reformulation method by exploiting the uncertainty set structure was proposed [21] and extended to a more general uncertainty set [22]. These two studies focused on static robust models, and the adjustable two-stage robust model with DDU is even more challenging. Aiming at solving general two-stage RO with polyhedral decision-dependent uncertainty set, modified Benders decomposition [23], Benders C&CG [24] and parametric C&CG [24], [25] were proposed, which guarantees to find the optimal solution if the master problem and the second-stage problems can be solved effectively. However, if the left-hand-side coefficient of the uncertainty set is decision-dependent, the master problem will have bilinear terms and become nonconvex, making the master problem hard to solve and hindering the solution of the RO with DDU. Adaptive C&CG was another exact solution method [26], which uses projection to keep the critical vertex in the decision-dependent uncertainty set. However, this method is not the most efficient one when the uncertainty set has a natural mapping structure. Multi-parametric programming method [27] was also established for RO with DDU, but may easily become time-consuming because of the curse of dimensionality. The uncertainty set studied in this paper has decision-dependent left-hand-side coefficients and possesses a natural vertex mapping structure. Therefore, a more efficient solution method can be designed.

In distributionally robust optimization (DRO) problems with DDU, ambiguity sets of probability distributions are decision-dependent. Tractable reformulations of two-stage DRO problems with DDU were derived using duality theory in [28] for various ambiguity sets in the finite support case. Multi-stage mixed-integer DRO with DDU under a moment-based ambiguity set was solved by the Stochastic Dual Dynamic integer Programming (SDDiP) [29]. The total variation ambiguity set case was solved by Benders decomposition [30]. A scenario-based formulation was handled by a customized C&CG method in [31]. The two-stage decision-dependent DRO under Wasserstein-distance ambiguity set with a continuous support was solved by an approximated SP based on Lipschitz constants in [32].

Our core objective is to provide a holistic model that allows the operator to exploit local predictions to build an improved uncertainty set and considers the tradeoff between prediction purchase cost and operation cost in the RGD. A mapping-based C&CG algorithm is developed to solve the problem. Our contributions are two-fold:

(1) *Robust Generation Dispatch Model with Purchase of Local Predictions*. A novel RGD model is developed to help the operator make better dispatch decisions by exploiting predictions purchased from agents (RESs and loads). Distinct from previous research that estimates the uncertainty set by the operator's own data/forecast, this paper builds an improved

uncertainty set based on conditional expectations and variances derived by combining forecasts from the operator and agents. The proposed model turns out to be an RO with DDU. It is worth noting that the proposed model is generic and not limited to the RGD problem in this paper.

(2) *Solution Algorithm*. A mapping-based C&CG algorithm is developed to solve the proposed model by returning mapping constraints rather than the worst-case scenarios directly. The proposed algorithm is proven to converge to the optimal solution within a finite number of iterations, while the traditional RO algorithms may fail to detect feasibility and guarantee optimality when dealing with DDU. Several comparisons are conducted to demonstrate its effectiveness.

The rest of this paper is organized as follows. Section II builds an improved uncertainty set based on the purchased predictions. The RGD model and its solution algorithm are developed in Section III and IV, respectively. Numerical experiments are carried out in Section V. Section VI concludes the paper.

II. Improved Uncertainty Set with Purchase of Predictions

We start with a general form of two-stage RO and develop approaches to improving its uncertainty set via purchased predictions from agents. Later in Section III, the RGD problem will be introduced in detail. The two-stage RO can be generally formulated as

$$\min_{x} \quad \left\{ f(x) + \max_{u\in\mathcal{U}} \min_{y\in\mathcal{Y}(x,u)} g(y) \right\}, \tag{1}$$
$$\text{s.t.} \quad x \in \mathcal{X} \cap \mathcal{X}_R,$$

with

$$\mathcal{X}_R = \{x \mid \mathcal{Y}(x,u) \neq \emptyset, \forall u \in \mathcal{U}\}, \tag{2}$$

where x and y are the first- and second-stage decision variables, respectively. In parallel, u represents the potential uncertainty realization varying within an uncertainty set $\mathcal{U}$. $\mathcal{X}$ and $\mathcal{X}_R$ are the feasible and robust feasible sets of x, respectively. According to (2), a first-stage decision x is robust feasible if and only if for any realization of uncertainty u, we can find a feasible second-stage solution $y \in \mathcal{Y}(x,u)$.

Traditionally, the uncertainty set $\mathcal{U}$ is estimated by the operator based on their own forecasts. With the recent advance in metering and data analytic technologies, the agents (RESs and loads) can also produce their predictions [14], [15], which may be used to improve the operator's forecast. To predict more precisely, the operator can buy predictions from the forecasting agents. For example, the operator may buy RESs' predictions about their future outputs. It is worth noting that there can also be independent forecasting agents, such as an agent that has advanced technologies for weather forecasting to support RES forecasting. The operator pays the forecasting agents according to their prediction accuracy. In addition, the agent does not need to disclose the specific prediction technology and there is no restriction on the forecasting methods, except that the prediction should be accompanied by a variance value to indicate its accuracy. In the following, we first investigate how, if provided with predictions from agents, the operator can improve their forecasts and obtain an improved uncertainty set. To make it easier to follow, the procedures are summarized in Fig. 1.

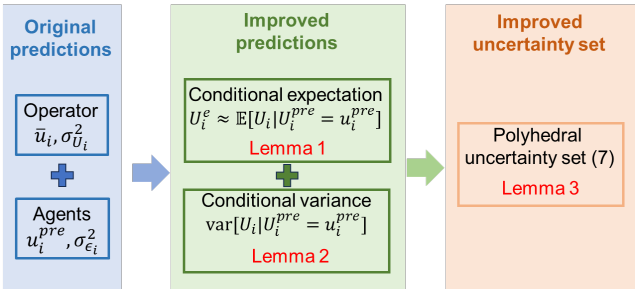

Fig. 1. Overview of Section II.

A. Improved Forecasts

Suppose there are I agents indexed by $i \in \mathcal{I} = \{1, 2, ..., I\}$. The uncertainty of agent $i \in \mathcal{I}$ is represented as a random variable U_i in $\mathbb{R}$ with an unknown distribution, and u_i is its realization. In this paper, we focus on the polyhedral uncertainty set [2] that is commonly used in RO. The two key parameters to determine such a set are the expectation and variance of the uncertain factor. Traditionally, the operator estimates the expectation and variance of U_i using their own historical data, denoted by $\bar{u}_i := \mathbb{E}[U_i]$ and $\sigma^2_{U_i} := \text{var}[U_i]$, respectively. These two estimates may not be accurate enough due to the limited data of the operator. To predict more precisely, the operator can buy predictions from the agents. Suppose agent i's forecast of U_i is U_i^{pre}, then we have $U_i = U_i^{pre} + \epsilon_i$, where ϵ_i is a random noise. The prediction U_i^{pre} is also a random variable and let u_i^{pre} be its realization. Denote by $\sigma^2_{\epsilon_i} := \text{var}[\epsilon_i]$ the variance of ϵ_i. The higher the agent's prediction accuracy, the smaller the $\sigma^2_{\epsilon_i}$. Agents are heterogeneous and may have different prediction errors reflected in $\sigma^2_{\epsilon_i}$. Throughout the paper, we adopt the following assumption, which is a commonly adopted assumption about multiple uncorrelated predictions [33]:

A1: $\{\epsilon_i, \forall i \in \mathcal{I}\}$ are independent. Each ϵ_i has zero expectation, i.e., $\mathbb{E}[\epsilon_i] = 0$, and ϵ_i is independent of U_i.

First, let us see how the operator can improve their estimation of U_i with the help of prediction u_i^{pre}. We propose to use the conditional probability $\mathcal{P}(U_i|U_i^{pre} = u_i^{pre})$ as an approximation of the actual probability $\mathcal{P}(U_i)$ of U_i. Then the uncertainty set can be constructed based on the conditional expectation $\mathbb{E}[U_i|U_i^{pre} = u_i^{pre}]$ and variance $\text{var}[U_i|U_i^{pre} = u_i^{pre}]$. Generally, these two parameters can be complicated nonlinear functions of u_i^{pre}. For simplicity, in this paper, we adopt the *best linear predictor* of U_i, $U_i^e := \alpha_i + \beta_i U_i^{pre}$, that minimizes the squared error expectation as follows,

$$\min_{\alpha_i,\beta_i} \ \mathbb{E}\left[(U_i - (\alpha_i + \beta_i U_i^{pre}))^2\right], \tag{3}$$

where α_i and β_i are parameters to be determined. Denote the realization of U_i^e by u_i^e.

Lemma 1: When **A1** holds, the two parameters of the best linear predictor $U_i^e = \alpha_i + \beta_i U_i^{pre}$ are

$$\beta_i = \frac{\sigma_{U_i}^2}{\sigma_{U_i}^2 + \sigma_{\epsilon_i}^2},\ \alpha_i = (1-\beta_i)\overline{u}_i. \tag{4}$$

Moreover, $\mathbb{E}[U_i - U_i^e] = 0$ and $\mathrm{cov}(U_i - U_i^e, U_i^{pre}) = 0$.

The proof of Lemma 1 can be found in Appendix A. When the error is very small ($\sigma_{\epsilon_i}^2 \to 0$), we have $\alpha_i = 0, \beta_i = 1$, and thus $u_i^e = u_i^{pre}$. It means that the prediction u_i^{pre} is accurate so the operator just uses it. When the error is very high ($\sigma_{\epsilon_i}^2 \to \infty$), we have $\alpha_i = \bar{u}_i, \beta_i = 0$, and thus $u_i^e = \overline{u}_i$. It means that the prediction u_i^{pre} is so inaccurate that the operator cannot get a better estimation than the original one $\overline{u}_i$. Let $\eta_i := U_i - U_i^e$. By Lemma 1, we have $\mathrm{cov}(\eta_i, U_i^{pre}) = 0$, so it is reasonable to make a stronger assumption that the improved forecast and its forecast error are independent. Moreover, this assumption can be proven in the case of Gaussian random variables [34].

A2: η_i and U_i^{pre} are independent.

Based on the best linear predictor, the conditional variance $\mathrm{var}[U_i | U_i^{pre} = u_i^{pre}]$ can be calculated by Lemma 2.

Lemma 2: When **A1** and **A2** hold, we have

$$\mathrm{var}[U_i | U_i^{pre} = u_i^{pre}] = (1-\beta_i)^2\sigma_{U_i}^2 + \beta_i^2\sigma_{\epsilon_i}^2 \tag{5}$$

and $\mathrm{var}[U_i | U_i^{pre} = u_i^{pre}] \le \sigma_{U_i}^2$.

The proof of Lemma 2 can be found in Appendix B. From Lemma 1, we have observed that the more accurate the prediction (i.e., the smaller the $\sigma_{\epsilon_i}^2$), the larger the β_i (weight on u_i^{pre}). In an extreme case when $\sigma_{\epsilon_i}^2$ is zero, we have $u_i^e = u_i^{pre}$ and $\mathrm{var}[U_i | U_i^{pre} = u_i^{pre}] = 0$. This indicates that with the prediction from agent i, the operator can know the exact value of u_i, and so there is no uncertainty. On the contrary, when $\sigma_{\epsilon_i}^2 \to \infty$, we have $u_i^e = \overline{u}_i$, and $\mathrm{var}[U_i | U_i^{pre} = u_i^{pre}] = \sigma_{U_i}^2$, meaning that the agent's prediction is so inaccurate that the operator still uses the same estimation as if there were no purchased prediction. From the analysis above, we find that $\beta_i \in [0,1]$ can be a good indicator of prediction accuracy. We borrow similar concepts from economics and define the prediction accuracy as follows.

Definition 1: (Prediction Accuracy [34]) The parameter β_i in (4) can be formally defined as the prediction accuracy. To differentiate, we use τ_i to denote the prediction accuracy, i.e.,

$$\tau_i := \frac{\sigma_{U_i}^2}{\sigma_{U_i}^2 + \sigma_{\epsilon_i}^2} \in [0,1]. \tag{6}$$

The accuracy of each agent's prediction to the operator is influenced both by the agent's forecasting technology and the incentive paid by the operator (will be explained later).

B. Improved Uncertainty Set

With the improved forecasts above, the operator can then construct an improved uncertainty set. Suppose there are T periods indexed by $t \in \mathcal{T} = \{1, ..., T\}$, then we have the best linear predictor u_{it}^e and variance $\mathrm{var}[U_{it} | U_{it}^{pre} = u_{it}^{pre}]$ for all $i \in \mathcal{I}$ and $t \in \mathcal{T}$. We adopt a polyhedral uncertainty set with the following form.

$$\begin{aligned}\mathcal{U}(\tau) = \ \{u_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T} \mid \\ u_{it}^0 - u_{it}^h \le u_{it} \le u_{it}^0 + u_{it}^h, \forall i \in \mathcal{I}, \forall t \in \mathcal{T} \\ \sum\nolimits_i \frac{|u_{it} - u_{it}^0|}{u_{it}^h} \le \Gamma_S,\ \forall t \in \mathcal{T} \\ \sum\nolimits_t \frac{|u_{it} - u_{it}^0|}{u_{it}^h} \le \Gamma_T,\ \forall i \in \mathcal{I}\},\end{aligned} \tag{7}$$

where Γ_S and Γ_T are the uncertainty budgets to restrain the spatial and temporal deviations from the forecast u_{it}^0. $u_{it}^0, u_{it}^h, \forall i, \forall t$ and Γ_S, Γ_T are parameters to be determined based on the best linear predictor u_{it}^e and variance $\mathrm{var}[U_{it} | U_{it}^{pre} = u_{it}^{pre}]$. Their values are given in Lemma 3. Denote $v_{it} := |u_{it} - u_{it}^0| / u_{it}^h, \forall i, \forall t$, so $v_{it} = |\eta_{it}/u_{it}^h|$ when $u_{it}^0 = u_{it}^e$. For simplicity, we assume that:

A3: $\{\eta_{it}/\sqrt{\mathrm{var}[\eta_{it}]}, \forall i, \forall t\}$ are independent and identically distributed (i.i.d.).

Assumption **A3** means that the errors of improved forecasts are independent, and they follow the same type of distribution except that the variances may be different. This assumption will hold if the predictions are obtained independently using the same type of forecasting method. We adopt this assumption for simplicity and derive theoretical confidence results for the constraints of the uncertainty set in Lemma 3, which provides guidance for constructing the uncertainty set and choosing the parameters in general cases.

Lemma 3: When **A1**-**A3** hold, if the parameters of the uncertainty set (7) are chosen as

$$u_{it}^0 = \ u_{it}^e = (1-\tau_{it})\overline{u}_{it} + \tau_{it}u_{it}^{pre},\ \forall i, \forall t, \tag{8a}$$

$$\begin{aligned} u_{it}^h &= \sqrt{\mathrm{var}[U_{it} | U_{it}^{pre} = u_{it}^{pre}]/(1-\delta)} = \sqrt{\mathrm{var}[\eta_{it}]/(1-\delta)} \\ &= \sqrt{\left[(1-\tau_{it})^2\sigma_{U_i}^2 + \tau_{it}^2\sigma_{\epsilon_i}^2\right]/(1-\delta)},\ \forall i, \forall t, \end{aligned} \tag{8b}$$

$$\Gamma_S = \sqrt{\frac{I(1-\delta)(1+I-I\xi)}{1-\xi}}, \tag{8c}$$

$$\Gamma_T = \sqrt{\frac{T(1-\delta)(1+T-T\xi)}{1-\xi}}. \tag{8d}$$

Then, we can ensure that $\mathcal{P}(v_{it} \ge 1) \le 1-\delta, \forall i, \forall t$, $\mathcal{P}\left(\sum_i v_{it} \ge \Gamma_S\right) \le 1-\xi, \forall t$ and $\mathcal{P}\left(\sum_t v_{it} \ge \Gamma_T\right) \le 1-\xi, \forall i$.

The proof of Lemma 3 can be found in Appendix C. With the help of purchased predictions, a smaller and more accurate uncertainty set can be obtained. The improved uncertainty set (7) depends on the prediction accuracy $\tau_{it}, \forall i, \forall t$ and the confidence thresholds δ and ξ. τ_i is influenced by the payment of the operator to agent i for buying the agent's prediction (denoted by $C_i, \forall i$). The higher the payments, the more accurate the predictions. δ and ξ are chosen according to the operator's preferences. The relationship between the operator's payment and the prediction accuracy is given in the next section. The improved uncertainty set is impacted by the payment $C_i, \forall i$ determined in the first stage, and thus, is decision-dependent.

Remark: The proposed method allows the operator to purchase and use predictions in RGD. The participation of RESs and loads in the day-ahead electricity market is not considered. In the proposed model, we assume that the forecasting agents are trustworthy, which means that they will provide

the operator with the predictions of the claimed quality. This can be enforced by a penalty mechanism based on statistical hypothesis testing after the real values are observed. The hypothesis testing may combine multiple days' forecasts for a high enough level of significance. As long as the agents are trustworthy, the operator's optimal strategy can be found using the proposed method. Since the operator optimizes toward the best trade-off between prediction cost and operation cost, the agents are encouraged to improve their forecasting technologies to the most efficient degree.

III. Robust Generation Dispatch Model With Purchase of Predictions

With the improved uncertainty set above, in the following, we develop the RGD model integrating the purchase and use of predictions from agents (RESs and loads).

A. General Form

To integrate the prediction purchase and RGD processes, we need the relationship between the operator's payment C_i and the prediction accuracy τ_{it}. The agent i assesses the prediction cost $h_i(\tau_i)$ as a function of the prediction accuracy $\tau_i := \{\tau_{it}, \forall t\}$ according to the forecasting method and then offers it to the operator. The operator's payment C_i to agent i should cover the prediction cost, i.e., $C_i \geq h_i(\tau_i)$. For the sake of testing the proposed RGD framework with prediction purchase, we use the prediction cost function in (9). The function in (9) has been widely applied to market information aggregation [34] and virtual power plants [13] to model the prediction costs. It is worth noting that the proposed method can be adapted to other forms of prediction cost function.

$$h_i(\tau_i) = \sum_{t=1}^{T} \frac{m}{\sigma_{\epsilon_i}^2} = \sum_{t=1}^{T} \frac{m}{\sigma_{U_{it}}^2} \frac{\tau_{it}}{1-\tau_{it}}, \tag{9}$$

where m is a cost parameter estimated by the forecasting agent. The prediction cost $h_i(\tau_i)$ covers the costs in T periods. In each period, the prediction cost is inversely proportional to the variance, which indicates that the more accurate the prediction (the smaller the $\sigma_{\epsilon_i}^2$), the higher the cost $h_i(\tau_i)$.

For notation conciseness, we further assume that for each agent, the prediction accuracy values in different periods are the same, based on which we obtain the formulation in (10). Nonetheless, the proposed model and algorithm can also be applied to the case with heterogeneous prediction accuracy.

$$C_i \geq h_i(\tau_i) = \frac{\hat{m}}{\sigma_{U_i}^2} \frac{\tau_i}{1-\tau_i}, \forall i \in \mathcal{I}, \tag{10}$$

where $\hat{m} = Tm$.

Definition 2: (Value of Prediction) The operator's payment C_i to agent i can be formally defined as the value of prediction from agent i.

Remark: Prediction trading methods can be divided into two categories, depending on whether the payments are determined before or after the provision of predictions. In references [8], [11], [15], [35], the payments for the forecasts are determined based on the actual uncertainty realizations and benefits afterwards. In the proposed model, we adopt the other method, where the agents offer the cost-accuracy function, and then the operator decides on whether to buy or not and the level of accuracy. The adopted method is more flexible in that the prediction accuracy is adaptive.

The total payment for buying predictions from agents is also a cost of the operator in the first stage. Therefore, the two-stage RO model considering the purchase and use of predictions can be formulated as

$$\min_{x,C,\tau} \left\{ f(x) + \sum_{i\in\mathcal{I}} C_i + \max_{u\in\mathcal{U}(\tau)} \min_{y\in\mathcal{Y}(x,u)} g(y) \right\}, \tag{11}$$
$$\text{s.t. } x \in \mathcal{X} \cap \tilde{\mathcal{X}}_R,$$
$$(10),\ 0 \leq \tau_i \leq 1, \forall i \in \mathcal{I},$$

where

$$\tilde{\mathcal{X}}_R = \{x \mid \mathcal{Y}(x,u) \neq \emptyset, \forall u \in \mathcal{U}(\tau)\}. \tag{12}$$

As the focus of this paper is the operator's decision-making in RGD, the predictive information market design between the operator and agents will be left for future study. The model (11) is an RO with DDU since the uncertainty set $\mathcal{U}(\tau)$ is influenced by the first-stage decision τ. The details of the objective function and constraints are given below.

B. Detailed Robust Generation Dispatch Model

There are J controllable generators indexed by $j \in \mathcal{J} = \{1, ..., J\}$, L lines indexed by $l \in \mathcal{L} = \{1, ..., L\}$, I_r RESs indexed by $i \in \mathcal{I}_r = \{1, ..., I_r\}$, and I_d loads indexed by $i \in \mathcal{I}_d = \{I_r + 1, ..., I_r + I_d\}$ in a transmission grid. Let $\mathcal{I} = \mathcal{I}_r \cup \mathcal{I}_d$ denote the set of agents whose power outputs/demands $u_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}$ are uncertain. In particular, for $i \in \mathcal{I}_r$, u_{it} represents the uncertain maximum power output of RES i in period t; for $i \in \mathcal{I}_d$, u_{it} represents the uncertain power demand of load i in period t. The operator may buy predictions from the RESs and loads to improve their predictions. The best linear predictors u_{it}^e of u_{it} can be obtained by (8a).

In the first stage (day-ahead pre-dispatch stage), the transmission grid operator decides on the reference output and reserve capacity of the controllable generators and the payments for buying predictions from RESs and loads. In the second stage (re-dispatch stage), knowing the exact RES power outputs and exact demands, the operator adjusts the output of controllable generators within their reserve capacity or curtails RES power to maintain power balance. The RGD problem can be formulated as

$$\min_{\substack{p,r^{\pm},\\ C,\tau}} \underbrace{\sum_{t\in\mathcal{T}} \sum_{j\in\mathcal{J}} (\varrho_j p_{jt} + \gamma_j^+ r_{jt}^+ + \gamma_j^- r_{jt}^-)}_{f(x)} + \sum_{i\in\mathcal{I}} C_i$$
$$+ \max_{u\in\mathcal{U}(\tau)} \min_{\substack{(p^{\pm},p^c)\in\\ \mathcal{Y}(p,r^{\pm},u)}} \underbrace{\sum_{t\in\mathcal{T}} \left(\sum_{j\in\mathcal{J}} (\varrho_j^+ p_{jt}^+ + \varrho_j^- p_{jt}^-) + \sum_{i\in\mathcal{I}_r} \rho^c p_{it}^c \right)}_{g(y)}, \tag{13a}$$
$$\text{s.t. } (p, r^{\pm}) \in \mathcal{X} \cap \tilde{\mathcal{X}}_R, \tag{13b}$$
$$C_i \geq \frac{m}{\sigma_{U_i}^2} \frac{\tau_i}{1-\tau_i},\ 0 \leq \tau_i \leq 1,\ \forall i \in \mathcal{I}, \tag{13c}$$

where

$$\mathcal{X} = \Big\{(p, r^{\pm}) \mid$$
$$\sum\nolimits_{j\in\mathcal{J}} p_{jt} + \sum\nolimits_{i\in\mathcal{I}_r} u^e_{it} = \sum\nolimits_{i\in\mathcal{I}_d} u^e_{it}, \forall t \in \mathcal{T}, \tag{14a}$$
$$0 \le r^+_{jt} \le R^+_j \theta_{jt}, 0 \le r^-_{jt} \le R^-_j \theta_{jt}, \forall j \in \mathcal{J}, \forall t \in \mathcal{T}, \tag{14b}$$
$$P^{min}_j \theta_{jt} + r^-_{jt} \le p_{jt} \le P^{max}_j \theta_{jt} - r^+_{jt}, \forall j \in \mathcal{J}, \forall t \in \mathcal{T}, \tag{14c}$$
$$(p_{jt} + r^+_{jt}) - (p_{j(t-1)} - r^-_{j(t-1)}) \le \mathcal{R}^+_j \theta_{j(t-1)} + \mathcal{R}^U_j \theta^U_{jt}, \forall j \in \mathcal{J}, \forall t = 2, ..., T, \tag{14d}$$
$$-(p_{jt} - r^-_{jt}) + (p_{j(t-1)} + r^+_{j(t-1)}) \le \mathcal{R}^-_j \theta_{jt} + \mathcal{R}^D_j \theta^D_{jt}, \forall j \in \mathcal{J}, \forall t = 2, ..., T, \tag{14e}$$
$$-F_l \le \sum_{j\in\mathcal{J}} \pi_{jl} p_{jt} + \sum_{i\in\mathcal{I}_r} \pi_{il} u^e_{it} - \sum_{i\in\mathcal{I}_d} \pi_{il} u^e_{it} \le F_l, \forall l, \forall t \Big\}, \tag{14f}$$

and

$$\mathcal{Y}(p, r^{\pm}, u) = \Big\{p^{\pm}, p^c \mid$$
$$0 \le p^+_{jt} \le r^+_{jt}, 0 \le p^-_{jt} \le r^-_{jt}, \forall j \in \mathcal{J}, \forall t \in \mathcal{T}, \tag{15a}$$
$$\sum_{j\in\mathcal{J}} (p_{jt} + p^+_{jt} - p^-_{jt}) + \sum_{i\in\mathcal{I}_r} (u_{it} - p^c_{it}) = \sum_{i\in\mathcal{I}_d} u_{it}, \forall t \in \mathcal{T}, \tag{15b}$$
$$0 \le p^c_{it} \le u_{it}, \forall i \in \mathcal{I}_r, \forall t \in \mathcal{T}, \tag{15c}$$
$$-F_l \le \sum_{j\in\mathcal{J}} \pi_{jl}(p_{jt} + p^+_{jt} - p^-_{jt}) + \sum_{i\in\mathcal{I}_r} \pi_{il}(u_{it} - p^c_{it}) - \sum_{i\in\mathcal{I}_d} \pi_{il} u_{it} \le F_l, \forall l \in \mathcal{L}, \forall t \in \mathcal{T} \Big\}. \tag{15d}$$

The objective function (13a) minimizes the total cost under the worst-case scenario, i.e., the total generation-related cost and prediction purchase payment in the first stage plus the total generation adjustment cost and curtailment penalty in the second stage. In the first stage, the decision variable x consists of the reference output $\{p_{jt}, \forall j, \forall t\}$ and upward/downward reserve capacity $\{r^+_{jt}, r^-_{jt}, \forall j, \forall t\}$ of controllable generators. $C_i, \forall i$ is the payment for buying information from the agents. The second-stage decision variable y includes the upward/downward power output adjustment $\{p^+_{jt}, p^-_{jt}, \forall j, \forall t\}$ of controllable units and the real-time RES power curtailment $\{p^c_{it}, \forall i \in \mathcal{I}_r, \forall t\}$. θ_{jt}, θ^U_{jt}, θ^D_{jt} are binary parameters for the on, startup, and shutdown states of generator j in period t, respectively [25]. ϱ_j, $\gamma^{\pm}_j$, and $\varrho^{\pm}_j$ are the cost coefficients of power output, upward/downward reserve, and upward/downward regulation, respectively. ρ^c is the real-time curtailment penalty coefficient.

Constraints (14a)-(15d) stipulate the operational limits. Constraints (14a) and (15b) are the power balance conditions. The upward/downward reserve capacity should not exceed the bounds $R^{\pm}_j$ as in (14b). The upper/lower power limits of controllable generators considering reserve requirements are given in (14c), where P^{min}_j / P^{max}_j is the minimum/maximum power output. (14d)-(14e) ensure the satisfaction of ramping limits when offering reserves [26]. The upper/lower power limits of RES curtailment are given in (15c). The network capacity limits are imposed in (14f) and (15d); F_l is the power flow limit of line l and π_{il}, π_{jl} are the power transfer distribution factors (PTDFs) deduced from the DC power flow model, so constraints (14a), (14f), (15b), and (15d) constitute the network model. Constraint (15a) ensure that the power adjustment is within the reserve capacity.

Remark: Load shedding is not allowed in the proposed model (13)-(15). This is because in some countries such as China, load shedding is viewed as an operation failure and is prohibited [2]. But it is worth noting that the proposed model can also accommodate load shedding by adding slack variables to (15b), (15d) and a penalty term in the objective function.

As mentioned earlier, the proposed model (13)-(15) is an RO with DDU. The traditional algorithms such as Benders decomposition and C&CG cannot be directly applied since they may fail to converge or lead to suboptimal solutions. In the next section, a mapping-based C&CG algorithm will be developed to overcome this difficulty.

IV. Solution Algorithm

In this section, a mapping-based C&CG algorithm is developed to solve problem (13), an RO with DDU. Noticing that the re-dispatch problem (15) is a linear program, $g(y)$ and $\mathcal{Y}(x, u)$ can be expressed by

$$g(y) = c^\top y, \tag{16}$$
$$\mathcal{Y}(x, u) = \{y \in \mathbb{R}^{n_y} \mid Ax + By + Du \le q\}. \tag{17}$$

A. Second-Stage Problem Transformation

Given the first-stage decision $x \in \mathcal{X}$ and $\tau_i \in [0, 1], \forall i \in \mathcal{I}$, the second-stage problem is a bilevel optimization:

$$S(x, \tau) = \max_{u\in\mathcal{U}(\tau)} \min_{y\in\mathcal{Y}(x,u)} c^\top y, \tag{18}$$

which is equivalent to the sub-problem (**SP**) (19).

$$\textbf{SP}: \max_{u\in\mathcal{U}(\tau), y, \nu} c^\top y, \tag{19a}$$
$$\text{s.t. } B^\top \nu = c, \tag{19b}$$
$$0 \le -\nu \perp [-(Ax + By + Du) + q] \ge 0. \tag{19c}$$

In **SP** (19), the decision variables are the uncertainty u, the second-stage decision variable y, and the second-stage dual variable ν. The constraints (19b) and (19c) constitute the KKT conditions of the inner "min" problem $\min_{y\in\mathcal{Y}(x,u)} c^\top y$. The complementary slackness condition (19c) can be linearized by the Big-M method [36].

Furthermore, for a given first-stage decision (x, τ), the problem (18) may be infeasible. Remember that we need to ensure x is robust feasible ($x \in \tilde{\mathcal{X}}_R$), so we construct the following relaxed problem for checking feasibility.

$$F(x, \tau) = \max_{u\in\mathcal{U}(\tau)} \min_{y, s} 1^\top s, \tag{20a}$$
$$\text{s.t. } Ax + By + Du - s \le q, s \ge 0. \tag{20b}$$

In the relaxed problem (20), a nonnegative slack variable s is added to the constraint $Ax + By + Du \le q$ so that the new constraint (20b) is always feasible. Moreover, the original

problem is feasible if the minimum $1^\top s = 0$ in the relaxed problem, which means the slack variable is not needed. By maximizing the minimum $1^\top s$ over $u \in \mathcal{U}(\tau)$, we can know whether the original second-stage problem is feasible for all possible realizations in the uncertainty set by checking whether the optimal value $F(x,\tau) = 0$. Similarly, the relaxed problem (20) is equivalent to

$$\textbf{FC}: \max_{\substack{u\in\mathcal{U}(\tau)\\ y,s,\nu,\mu}} 1^\top s, \tag{21a}$$

$$\text{s.t. } B^\top \nu = 0, \ -\nu + \mu = 1, \ 0 \le \mu \perp s \ge 0, \tag{21b}$$

$$0 \le -\nu \perp [-(Ax + By + Du - s) + q] \ge 0. \tag{21c}$$

We call (21) the feasibility-check (**FC**) problem. In **FC** (21), ν and μ are the dual variables of the two constraints in (20b) in the inner "min" problem, respectively, and the constraints (21b) and (21c) constitute the KKT conditions. In particular, the parameters of $\mathcal{U}(\tau)$ are fixed in (19)-(21), including u^e and u^h which can be determined by τ in the first-stage decisions.

Given a candidate first-stage decision, we first solve the **FC** problem to check whether $x \in \tilde{\mathcal{X}}_R$. If not, a feasibility cut will be returned; otherwise, we continue to solve the **SP** problem to identify an optimality cut.

Lemma 4: Suppose u^* is the optimal solution of **SP** or **FC**, then u^* can be reached at a vertex of $\mathcal{U}(\tau)$.

The proof of Lemma 4 is similar to that in [37] and is omitted here. The traditional RO algorithms return the worst-case scenario $\{u^*_{it}, \forall i, \forall t\}$ directly to the master problem to generate a feasibility/optimality cut. However, when dealing with DDU, a previously selected scenario may no longer be a vertex of the new uncertainty set when the first-stage decision changes ($\mathcal{U}(\tau)$ changes with τ). This causes the traditional algorithms to fail to find the optimal solutions.

To tackle this problem, instead of returning the scenario $u^*_{it}, \forall i, \forall t$ directly, we propose to map the worst-case scenario to an unchanged vertex set and return the mapping constraints. Define the set Φ by:

$$\begin{aligned}\Phi := \ \{&\phi_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T} \mid \\ &-\psi_{it} \le \phi_{it} \le \psi_{it}, \psi_{it} \le 1, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \\ &\textstyle\sum_{i\in\mathcal{I}} \psi_{it} \le \Gamma_S, \forall t \in \mathcal{T}, \sum_{t\in\mathcal{T}} \psi_{it} \le \Gamma_T, \forall i \in \mathcal{I}\}.\end{aligned}$$

Lemma 5: The set Φ and its vertex set $V(\Phi)$ are unchanged, i.e., they are independent of τ. For any τ subject to $\tau_i \in (0,1), \forall i \in \mathcal{I}$, the mapping $\pi : \Phi \to \mathcal{U}(\tau)$ defined by $\pi(\phi) = u$ with

$$u_{it} = u^e_{it}(\tau_i) + u^h_i(\tau_i)\phi_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T},$$

is bijective. Moreover, the restriction on the vertex set $\pi : V(\Phi) \to V(\mathcal{U}(\tau))$ is also bijective.

The proof of Lemma 5 can be found in Appendix D. Lemma 5 uses the assumption that $\tau_i \in (0,1), \forall i \in \mathcal{I}$, which naturally holds when the forecasting variances are not zero, i.e., the operator and the agents cannot predict exactly. Under this assumption, Lemma 5 shows that there is a bijection between the unchanged set Φ and uncertainty set $\mathcal{U}(\tau)$ that preserves the vertices. For a worst-case scenario u^* generated by **FC** or **SP**, we can get the corresponding $\phi^*_{it}, \forall i, \forall t$ by:

$$\phi^*_{it} = (u^*_{it} - u^e_{it})/u^h_i, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}.$$

Then, instead of returning u^* to the master problem, we return the following mapping constraints:

$$u_{it} = u^e_{it}(\tau_i) + u^h_i(\tau_i)\phi^*_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}. \tag{22}$$

Here, u_{it}, $u^e_{it}(\tau_i)$, and $u^h_i(\tau_i)$ are all variables in the master problem. When τ_i changes, the point calculated by (22) remains at a vertex of $\mathcal{U}(\tau)$, as illustrated in Fig. 2.

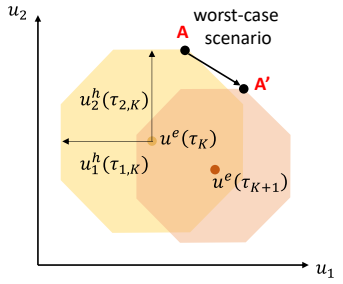

Fig. 2. Illustration of (22). The yellow and orange regions are the uncertainty sets in the K and $K+1$ iterations, respectively. Point A is the worst-case scenario in the K iteration. When u^e and u^h change with τ, (22) moves point A to point A′, which is a vertex of the new uncertainty set.

B. Mapping-Based C&CG Algorithm

With the mapping constraints returned, the master problem (**MP**) can be formulated as

$$\textbf{MP}: \min_{x,C,\tau,\zeta,y^k,u^k} f(x) + \sum_i C_i + \zeta, \tag{23a}$$

$$\text{s.t. } x \in \mathcal{X}, (10), \tau_i \in [0,1], \forall i \in \mathcal{I}, \tag{23b}$$

$$\zeta \ge c^\top y^k, \forall k \in [K], \tag{23c}$$

$$Ax + By^k + Du^k \le q, \forall k \in [K], \tag{23d}$$

$$u^k = u^e(\tau) + u^h(\tau)\phi_k, \forall k \in [K], \tag{23e}$$

where the symbols with superscript k are variables while the symbols with subscript k are given parameters. $[K]$ represents all positive integers not exceeding K.

The overall procedure of the proposed mapping-based C&CG algorithm is given in Algorithm 1. The proposed algorithm is different from the traditional C&CG algorithm [38] as it returns the mapping constraints instead of the worst-case scenarios $\{u^*_{it}, \forall i, \forall t\}$ to the master problem to generate new cuts. To be specific, in the traditional C&CG algorithm, Step 6 in Algorithm 1 is replaced by: "6: Create variables y^{K+1} and add the following constraints to **MP** (23):

$$\zeta \ge c^\top y^{K+1}, \ Ax + By^{K+1} + Cu^*_{K+1} \le q.$$

Update $K = K + 1$ and go to Step 2." Note that u^*_{K+1} is a constant vector obtained by solving **FC** (21) or **SP** (19).

Theorem 1: Let $n_U := |V(\Phi)|$ be the number of extreme points of Φ. The mapping-based C&CG algorithm generates the optimal solution to problem (11) within $\mathcal{O}(n_U)$ iterations.

The proof of Theorem 1 can be found in Appendix E. While the traditional algorithms fail to guarantee finite-step convergence and optimality of the obtained strategy, the proposed algorithm can overcome these limitations. Case studies in Section V-C also demonstrate the advantages and necessity of the proposed algorithm.

Algorithm 1 Mapping-Based C&CG Algorithm

1: **Initiation**: Error tolerance $\epsilon > 0$; $K = 0$; $UB_K = +\infty$.
2: **Solve the Master Problem**
Solve **MP** (23). Derive the optimal solution $(x^*_{K+1}, C^*_{K+1}\tau^*_{K+1}, \zeta^*_{K+1}, y^{1*}, ..., y^{K*}, u^{1*}, ..., u^{K*})$ and update $LB_{K+1} = f(x^*_{K+1}) + \sum_{i\in\mathcal{I}} C^*_{i,K+1} + \zeta^*_{K+1}$.
3: **Solve the Feasibility-check Problem**
Solve **FC** (21) with $(x^*_{K+1}, \tau^*_{K+1})$.
Let $(u^*_{K+1}, \phi^*_{K+1}, \nu^*_{K+1}, \mu^*_{K+1}, y^*_{K+1}, s^*_{K+1})$ be the optimal solution. If $1^\top s^*_{K+1} > 0$, let $UB_{K+1} = UB_K$ and go to Step 6. Otherwise, go to Step 4.
4: **Solve the Sub-problem**
Solve **SP** (19) with $(x^*_{K+1}, \tau^*_{K+1})$. Denote the optimal solution by $(u^*_{K+1}, \phi^*_{K+1}, \nu^*_{K+1}, y^*_{K+1})$. Let

$$UB_{K+1} = f(x^*_{K+1}) + \sum\nolimits_{i\in\mathcal{I}} C^*_{i,K+1} + c^\top y^*_{K+1}$$

5: If $|UB_{K+1} - LB_{K+1}| \le \epsilon$, terminate and output $(x^*_{K+1}, \tau^*_{K+1})$. Otherwise, go to Step 6.
6: Create variables (y^{K+1}, u^{K+1}) and add the following constraints to **MP** (23):

$$\zeta \ge c^\top y^{K+1},\ Ax + By^{K+1} + Cu^{K+1} \le q,$$
$$u^{K+1} = u^e(\tau) + u^h(\tau)\phi^*_{K+1}.$$

Update $K = K + 1$ and go to Step 2.

C. Transformation and Linearization

In Algorithm 1, **MP** (23) is highly nonlinear due to the term $u^e(\tau)$, $u^h(\tau)$, and the constraint (10). In the following, we show how to turn (23) into a solvable form.

First, it is easy to prove that at the robust optimum, we have $C_i = h_i(\tau_i)$. Otherwise, if $C_i > h_i(\tau_i)$, we can always reduce C_i a little bit without changing the value of the other variables, so that all constraints are still satisfied but the objective value decreases. This contradicts the definition of the robust optimum. Therefore, we can eliminate constraint (10) and replace $\sum_i C_i$ in the objective function with $\sum_i h_i(\tau_i)$.

Second, if we let $u^h_i, \forall i \in \mathcal{I}$ be the decision variables and use them to represent $\tau_i, \forall i \in \mathcal{I}$, then the prediction cost $\sum_i h_i(\tau_i)$ can be represented by

$$\sum\nolimits_{i\in\mathcal{I}} h_i(\tau_i) = \sum\nolimits_{i\in\mathcal{I}} \hat{m}\left(\frac{1}{(1-\delta)(u^h_i)^2} - \frac{1}{\sigma^2_{U_i}}\right). \quad (24)$$

Let $\tilde{h}_i(u^h_i)$ denote the term in the right-hand side of (24) for each $i \in \mathcal{I}$, where u^h_i is the decision variable. Then, $\tilde{h}_i(u^h_i)$ is a convex function. Similarly, $u^e_{it}(\tau)$ can be represented by u^h_i, which is

$$u^e_{it}(u^h_i) = \frac{(1-\delta)(u^h_i)^2}{\sigma^2_{U_i}}\bar{u}_{it} + \left(1 - \frac{(1-\delta)(u^h_i)^2}{\sigma^2_{U_i}}\right) u^{pre}_{it}, \forall i, \forall t. \quad (25)$$

Next, we introduce a new variable $\tilde{u}^e_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}$, use it to replace $u^e(\tau)$ in (23e), and add the following penalty function to the objective:

$$\mathcal{H}(\tilde{u}^e, u^h) = \iota \sum\nolimits_{i\in\mathcal{I}} \sum\nolimits_{t\in\mathcal{T}} \left(\tilde{u}^e_{it} - u^e_{it}(u^h_i)\right)^2, \quad (26)$$

where ι is a large constant. Then, the remaining nonlinear term in the objective function, $\sum_i \tilde{h}_i(u^h_i) + \mathcal{H}(\tilde{u}^e, u^h)$, can be linearized by a convex combination approach [39], after which **MP** (23) has been turned into a linear program that can be solved efficiently.

D. Comparison with Generalized C&CG Methods

In this subsection, we discuss generalized C&CG methods in literature designed for two-stage RO problems with DDU, namely, Benders C&CG [24], parametric C&CG [24], [25], and adaptive C&CG [26].

The main difficulty of using C&CG to solve two-stage RO with DDU is that the worst-case scenarios found in previous iterations may become inside or outside of the new uncertainty set if the first-stage decision changes. In Benders C&CG and parametric C&CG, the idea is to add the KKT conditions for the following optimization problem to **MP** in order to consider the uncertainty set vertices:

$$\max_{u\in\mathcal{U}(\tau)} \quad -\nu^\top Du, \quad (27)$$

where ν is a dual variable value obtained from the second-stage problems **SP** and **FC**. The proposed uncertainty set $\mathcal{U}(\tau)$ in (7) can be written as:

$$\begin{aligned}\mathcal{U}(\tau) = \ &\{u_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T} \mid \\ &-v_{it} \le u_{it} - u^0_{it} \le v_{it}, v_{it} \le u^h_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T} \\ &\sum_{i\in\mathcal{I}} \frac{v_{it}}{u^h_{it}} \le \Gamma_S, \forall t \in \mathcal{T},\ \sum_{t\in\mathcal{T}} \frac{v_{it}}{u^h_{it}} \le \Gamma_T, \forall i \in \mathcal{I}\},\end{aligned}$$

where the auxiliary variable v_{it} represents $|u_{it} - u^0_{it}|$. Since the parameters $u^0_{it} = u^e_{it}$ and u^h_{it} depend on τ, the uncertainty set $\mathcal{U}(\tau)$ has the following compact form:

$$\mathcal{U}(\tau) = \{u \mid G_1 u + G_2(\tau)v \le H(\tau)\},$$

where G_1, $G_2(\tau)$, and $H(\tau)$ are coefficients. Since $G_2(\tau)$ and $H(\tau)$ depend on τ, both the left-hand-side and right-hand-side coefficients are decision-dependent. Then the KKT conditions of problem (27) are:

$$G_1^\top \lambda = -D^\top \nu, G_2(\tau)^\top \lambda = 0, \quad (28a)$$
$$0 \le \lambda \perp [H(\tau) - G_1 u - G_2(\tau)v] \ge 0. \quad (28b)$$

When added in **MP**, the KKT condition constraint (28) contains decision variables τ, u, v, and λ. Although the complementarity relationship in (28b) can be linearized using binary variables and the Big-M method, the bilinear term $G_2(\tau)v$ still exists, which brings nonlinearity and nonconvexity to **MP** and makes it intractable. Since the effectiveness of the Benders C&CG and parametric C&CG methods relies on the optimal solution of **MP**, the two methods are not suitable for the current model. Further linearization may help to solve such a type of **MP** effectively, which is one of our future directions.

To deal with the problem of the changing uncertainty set, the adaptive C&CG method projects the previously found worst-case scenario to a vertex of the new uncertainty set, and the projection method works for the general polyhedral decision-dependent uncertainty set. The considered uncertainty set $\mathcal{U}(\tau)$ has a natural mapping to the unchanged polyhedron Φ and

we utilize this fact to track the uncertainty set vertex in the proposed solution method. To this end, the proposed method can be regarded as another implementation of the idea of adaptive C&CG that further exploits the specific uncertainty set structure and reduces the computation burden of vertex projection.

In summary, the considered uncertainty set has decision-dependent left-hand-side coefficients and a natural mapping structure, which makes the proposed method the most suitable one compared with the three generalized C&CG methods.

V. Case Studies

We first use a simple 5-bus system to verify the proposed method and reveal some interesting phenomena; then, larger systems (33-, 69-, and 123-bus systems) are tested to show the scalability. Detailed data can be found in [40]. The mapping-based C&CG algorithm is implemented in MATLAB with GUROBI 9.5. All the simulations are conducted on a laptop with Intel i7-12700H processor and 16 GB RAM.

A. Benchmark

A 5-bus system with 3 controllable generators, 5 agents (2 wind farms and 3 uncertain loads), and 3 fixed loads is tested, whose parameters are shown in Table I. The time interval is 1 h. In the case study, it is assumed that the agent's prediction equals the actual value plus an error term, whose size depends on the variance value. We call the expected values estimated by the operator without using the agents' prediction information by original forecasts. The operator's original forecasts $\bar{u}_i$ and the actual uncertainty realization u_i of agent i in hindsight are depicted in Fig. 3. The original forecasts of agents 3 and 4 are the same. Suppose in every period $\sigma_U^2 = [8000, 2000, 4000, 9000, 1000]$ MW^2.

TABLE I
Parameters

Parameter	Value
Generator cost coefficient ρ	[30,35,25] \$/MWh
Generator maximum output P^{max}	[700,700,800] MW
Generator minimum output P^{min}	[280,280,320] MW
Generator maximum reserve $R^{\pm}$	[350,350,400] MW
Generator maximum ramp $\mathcal{R}^{\pm}$	[350,350,400] MW/h
Period number T	24
Prediction cost parameter m	10^4 \$·MW^2
Curtailment penalty ρ^c	100 \$/MWh
Linearization penalty ι	10^4 \$/MW^2
Uncertainty set confidence δ, ξ	$0.95, 0.95$

The proposed algorithm converges after 14 iterations in 987 s, which is acceptable for the day-ahead scheduling. The total operation cost (sum of the first-stage and the worst-case second-stage operation costs) is $\$1.115 \times 10^6$, while the prediction payments C_i for the 5 agents are $\$[4.00, 4.28, 3.73, 3.55, 0] \times 10^3$, respectively. Hence, the total cost (13a) under the worst-case scenario is $\$1.131 \times 10^6$. We also test the performance of the obtained day-ahead predispatch strategy when dealing with the actual uncertainty realizations. A feasible real-time redispatch strategy still exists

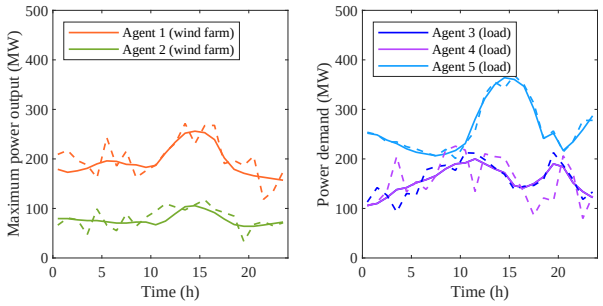

Fig. 3. Original forecasts (solid lines) and actual values (dashed lines).

but with a lower total cost ($\$1.071 \times 10^6$). Although the prediction cost is relatively small compared to the operation cost, the purchased predictions still play an important role in the dispatch, because they improve the forecasting accuracy, help to narrow the uncertainty set, decrease the conservativeness of the robust dispatch strategy, and therefore reduce the operation cost. The operation cost is $\$1.178 \times 10^6$ without prediction purchasing, whereas the total cost is $\$1.131 \times 10^6$ ($\$1.115 \times 10^6$ for operation and $\$1.556 \times 10^4$ for prediction) if it is allowed to purchase predictions. Therefore, the purchased predictions have a nonnegligible impact on the operation cost even if the prediction cost is small, and they can reduce the total cost by about 4% in the benchmark case.

The reference and worst-case power outputs of controllable generators are shown in Fig. 4. In the results, the cheapest generator (generator 3) has the highest output, and the most expensive generator (generator 2) has the lowest output.

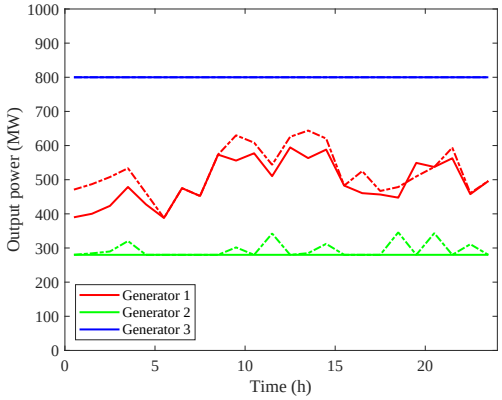

Fig. 4. The reference (solid lines) and worst-case (dash-dotted lines) power outputs of controllable generators.

To visualize the impact of predictions from RESs and loads on the forecast of the operator, the original (green area) and improved (blue area) uncertainty sets of agents 3 and 4 (i.e., loads 1 and 2) are shown in Fig. 5. Both sets have a confidence probability $\delta = 0.95$ in each period. The centers of the original uncertainty sets are the original forecasts $\bar{u}_{it}, \forall i, \forall t$. After purchasing the predictions from RESs and loads, the centers become the best linear predictors $u^e_{it}, \forall i, \forall t$, which are closer to the actual uncertainty realizations. The shaded areas show the variation ranges of the demand of loads 1 and 2.

Both uncertainty sets contain the actual load demand, but

the improved sets are much narrower so the operator is facing less uncertainty. The original forecasts of agents 3 and 4 are the same, but the uncertainty variance of agent 4 is larger, so agent 4's original uncertainty set is wider ($u_4^h = 134$ MW) than that of agent 3 ($u_3^h = 89$ MW). The widths of their improved sets are similar with $u_3^h = 33$ MW and $u_4^h = 35$ MW. Moreover, Fig. 6 shows how the uncertainty sets narrow as the prediction payments increase. Note that Fig. 6 is a semi-log plot and the prediction payment grows very quickly when the width of the uncertainty set is small. This is because the marginal prediction cost increases with a higher accuracy. In other words, it costs more to improve the accuracy of an already quite accurate prediction. The optimal payments and the corresponding widths $2u_i^{h*}, \forall i$ are also marked in Fig. 6. We can find that the values of $u_i^{h*}, \forall i$ are similar, which is due to the equal incremental principle, i.e., at the optimum, $\partial C_i / \partial u_i^h, \forall i$ are equal. The subtle difference between $u_i^{h*}, \forall i$ is caused by the linearization approximation errors.

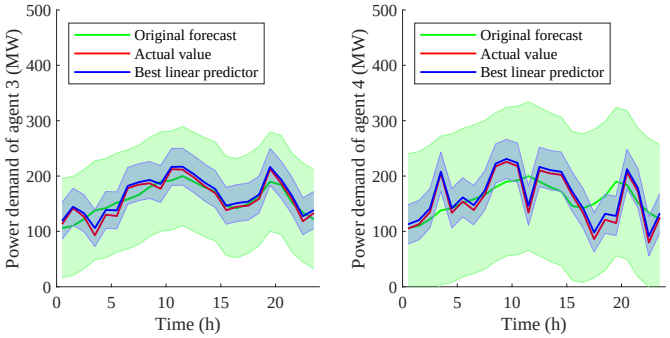

Fig. 5. Original and improved uncertainty sets of agents 3 and 4.

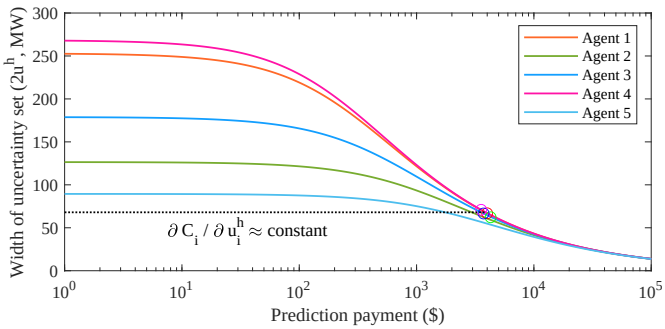

Fig. 6. Width of uncertainty sets under different prediction payments; circles represent the results at the optimum (the optimal payment for agent 5 is zero, which is outside of this figure).

B. Sensitivity Analysis

We further investigate the impacts of three different factors: the agent's prediction cost coefficient m, the probability parameters δ and ξ of the uncertainty set, and the variance of the uncertain factor σ_U^2.

1) Impact of Prediction Cost Coefficient: First, we test how the strategy of the operator changes with a rising agent prediction cost by changing m from 0 to 2×10^5 \$·MW^2. The total costs, operation costs, and prediction payments under different m are shown in Fig. 7. The change of prediction accuracy τ and the width of the improved uncertainty set are given in Fig. 8. We can find that when $m = 0$, the agents' prediction payments are zero since the operator can know the exact value of u without making any payment, and thus, there is no uncertainty ($\tau_i = 1, \forall i$). When m is very large, e.g., 2×10^5 \$·MW^2, the agents' prediction costs are extremely high, so the operator cannot afford to purchase predictions from the agents. Therefore, as shown in Fig. 8, the final prediction accuracy τ_i is zero for each agent i and the uncertainty sets are the widest. As m grows, from Fig. 7, the operation cost and the total cost are always less than the cost of the traditional model (1) without buying predictions from the agents. The lower the m, the higher the operation cost reduction, showing the potential of our model. According to Definition 2, the prediction payment C_i can be interpreted as the value of prediction from agents. This value is influenced by agent's prediction cost coefficient m and the system parameters. From Fig. 7, the value of prediction of all agents follows a similar trend (first increases and then declines) and the peak value of the agent with a larger uncertainty variance $\sigma_{U_i}^2$ tends to be higher. This indicates that predictions from agents will play an increasingly important role in future power systems with higher uncertainties.

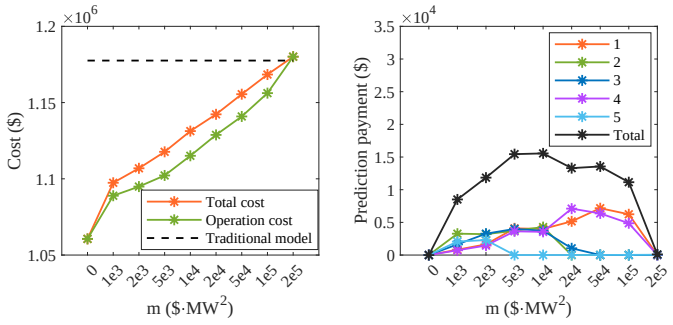

Fig. 7. Costs and prediction payments under different m.

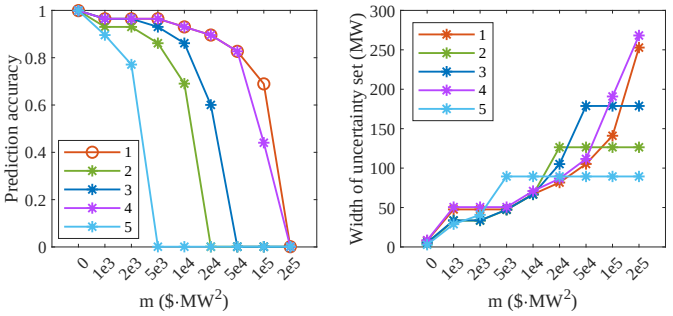

Fig. 8. Prediction accuracy and width of uncertainty set under different m.

2) Impact of Uncertainty Set Probability Parameters: We next change the probability parameters δ and ξ simultaneously, i.e., keeping $\delta = \xi$. The costs as well as the widths of uncertainty sets of agents 3 and 4 are shown in Fig. 9. When δ and ξ increase, the uncertainty sets expand, giving a more robust optimal predispatch strategy but also resulting in higher total and operation costs. Moreover, the total

cost of the proposed model is always less than that of the traditional model without predictions from the agents (when $\delta = \xi \geq 0.99$, the traditional model become infeasible). The improved uncertainty sets are much narrower than the original uncertainty sets. The widths of the improved uncertainty sets of agents 3 and 4 are similar under different δ and ξ due to the same reason as in Section V-A.

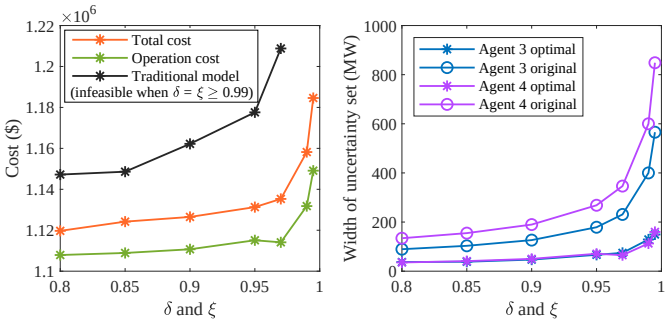

Fig. 9. Costs and widths of uncertainty sets under different δ and ξ.

3) Impact of Uncertainty Variance: We further investigate the impact of σ_U^2, the variance of the uncertain factor (or the operator's original estimate). To do this, we multiply σ_U^2 by a positive constant. The original forecasts and actual values are still the same as those in Fig. 3. The cost and prediction accuracy $\tau_i, \forall i$ are shown in Fig. 10. When the multiple of variance exceeds 2.0, the traditional model is infeasible because the uncertainty is too severe. The proposed model is still feasible because buying predictions from the agents enables the operator to effectively mitigate the uncertainty they face. This shows the advantage of the proposed model. When the variance is small, the operator already has a relatively good original estimate, so they tend to pay less for buying predictions from the agents. When the variance is large, at the optimum, the prediction accuracy is close to 1 for every agent, meaning that the operator relies on the predictions from the agents to make better dispatch decisions.

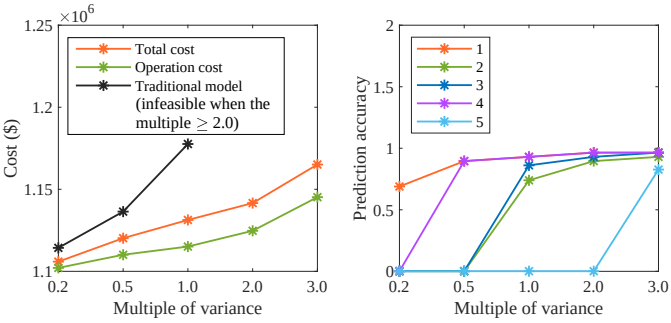

Fig. 10. Cost and prediction accuracy τ under different multiples of variance.

C. Comparison With the Traditional C&CG Algorithm

To show the necessity of the proposed mapping-based C&CG algorithm, we compare it with the traditional C&CG algorithm [38] using the benchmark case. The iteration processes are shown in Fig. 11. The mapping-based C&CG algorithm converges to the optimal solution given in Section V-A. The traditional algorithm stops in 30 iterations with the optimal objective value equals $\$1.178 \times 10^6$, which is higher than that of the proposed algorithm ($\$1.131 \times 10^6$). This is because a previously added worst-case scenario may lie outside of the uncertainty set when the first-stage decision changes. Therefore, the master problem is no longer a relaxation of the robust optimization, which may lead to over-conservative results. Moreover, the previous scenarios that are outside of the uncertainty set hinder the improvement by buying predictions from the agents. At the optimum of the traditional C&CG, we can find that the prediction payments are zero. Given the reasons above, the proposed algorithm is necessary.

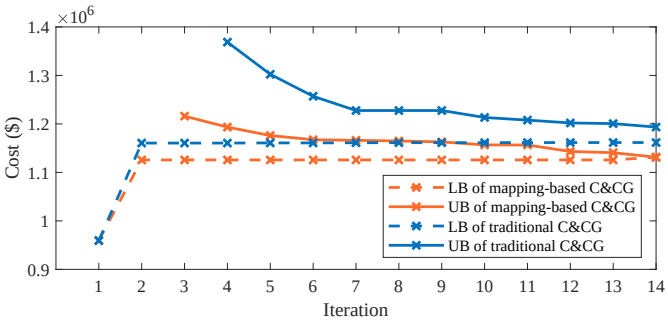

Fig. 11. Iteration processes of the proposed and traditional C&CG algorithms.

D. Out-of-Sample Test

To analyze the statistical performance of the obtained strategy, out-of-sample tests are conducted. To imitate various prediction errors, we randomly generate scenarios from a uniform and a Gaussian distribution with the same expectation u and standard deviation, respectively. We change the standard deviation via multiplying $\sqrt{\text{var}[U|U^{pre}]}$ by a constant from 0.5 to 2.5 and test the average total cost under the selected scenarios. Ten thousand (10000) scenarios are tested for each setting. The average total costs of the proposed algorithm and the traditional C&CG are compared in Table II. We can find that the proposed algorithm has lower costs.

TABLE II
OUT-OF-SAMPLE TEST OF THE PROPOSED ALGORITHM AND THE TRADITIONAL C&CG: AVERAGE TOTAL COST (10^6 $)

Multiple of standard deviation		0.5	1.0	1.5	2.0	2.5
Uniform	Proposed	1.0733	1.0741	1.0762	1.0796	1.0840
	Traditional	1.1237	1.1245	1.1257	1.1274	1.1294
Gaussian	Proposed	1.0733	1.0741	1.0762	1.0794	1.0830
	Traditional	1.1237	1.1245	1.1257	1.1272	1.1286

E. Scalability

The proposed RGD model and algorithm can be extended to the robust unit commitment (RUC) problem by adding the unit commitment decision variables, constraints, and costs to the first stage. The new decision variables θ_{jt}, θ_{jt}^U, and θ_{jt}^D are binary variables for the on, startup, and shutdown states

of the controllable generator j in period t, respectively. The first-stage feasible region $\mathcal{X}$ in the RGD model (14) already contains the power balance, reserve, ramp rates, and network capacity constraints. The following new constraints are added to further incorporate unit commitment [25]:

$$\theta_{jt}, \theta^U_{jt}, \theta^D_{jt} \in \{0,1\}, \forall j, \forall t, \tag{29a}$$

$$\sum\nolimits_{k=t}^{t+UT_j-1} \theta_{jk} \geq UT_j \theta^U_{jt}, \forall j, 1 \leq t \leq T - UT_j + 1, \tag{29b}$$

$$\sum\nolimits_{k=t}^{T} (\theta_{jk} - \theta^U_{jt}) \geq 0, \forall j, T - UT_j + 2 \leq t \leq T, \tag{29c}$$

$$\sum\nolimits_{k=t}^{t+DT_j-1} (1 - \theta_{jk}) \geq DT_j \theta^D_{jt}, \forall j, 1 \leq t \leq T - DT_j + 1, \tag{29d}$$

$$\sum\nolimits_{k=t}^{T} (1 - \theta_{jk} - \theta^D_{jt}) \geq 0, \forall j, T - DT_j + 2 \leq t \leq T, \tag{29e}$$

$$\theta_{jt} - \theta_{j(t-1)} = \theta^U_{jt} - \theta^D_{jt}, \forall j, \forall t, \tag{29f}$$

$$\theta^U_{jt} + \theta^D_{jt} \leq 1, \forall j, \forall t. \tag{29g}$$

Constraint (29a) defines the binary variables. Constraints (29b)-(29e) stipulate the minimum up time UT_j and minimum down time DT_j of the controllable generator j [41]. Constraint (29f) depicts the state change and constraint (29g) prohibits simultaneous startup and shutdown. The total startup and shutdown cost $\sum_{t\in\mathcal{T}} \sum_{j\in\mathcal{J}} (\rho^U_j \theta^U_{jt} + \rho^D_j \theta^D_{jt})$ is also added to the first-stage objective function $f(x)$, where ρ^U_j and ρ^D_j are cost coefficients.

To show the scalability of the proposed algorithm on models with and without unit commitment decisions, the computational time and the number of iterations under different settings are recorded in Table III. The 33-bus, 69-bus, and 123-bus cases have 4, 8, and 8 generators, respectively. UT_j and $DT_j, \forall j$ are set as 4 h.

TABLE III
COMPUTATIONAL TIME/NUMBER OF ITERATIONS UNDER DIFFERENT SETTINGS

Case	Model	Number of agents			
		4	8	12	16
33-bus	RGD	84 s / 3	116 s / 3	163 s / 3	1576 s / 15
69-bus	RGD	104 s / 3	329 s / 6	2614 s / 17	3609 s / 17
123-bus	RGD	542 s / 6	968 s / 5	2156 s / 8	6411 s / 11
33-bus	RUC	58 s / 2	243 s / 4	331 s / 4	741 s / 6
69-bus	RUC	123 s / 3	566 s / 8	2458 s / 13	5399 s / 9
123-bus	RUC	938 s / 6	1043 s / 6	3793 s / 8	4557 s / 4

As Table III shows, the computational time needed is less than 2 h in all test cases, which is acceptable for day-ahead generation dispatch and unit commitment. Generally, as the number of agents increases, the number of iterations usually increases due to the growing dimension of the uncertainty set. Apart from this, the number of iterations does not have a direct connection with the case complexity and whether unit commitment is considered or not. The number of iterations in C&CG types of methods is theoretically guaranteed to be no larger than the vertex number of the uncertainty set but is case-by-case. It is possible that we may need a large number of iterations to converge. In the test cases, most iteration numbers are below 10, whereas the maximum iteration number is 17, which is still much lower than the theoretical bound.

Compared with the RGD model, the RUC model generally requires longer computational time, but it is at most about two times as long in the test cases. The solution process of RGD and RUC differs in the master problem **MP**. RGD's **MP** is a linear program. In the RUC case, the total dimension of additional binary variables is $3JT$, and **MP** becomes a mixed-integer linear program (MILP). However, the uncertainty set and the second-stage MILP problems **SP** and **FC** are not changed. Therefore, the computational efficiency is still acceptable for day-ahead usage after incorporating unit commitment in the proposed model.

VI. CONCLUSION

This paper proposes a novel RGD model in which the operator can purchase predictions from the agents to obtain a more accurate uncertainty set and make better decisions. The proposed model renders a two-stage RO with DDU. A mapping-based C&CG algorithm with convergence guarantee is developed to solve the model. Some interesting findings are:

- Compared with the traditional model without buying predictions from the agents, the proposed model can help the operator greatly narrow the uncertainty set and reduce the total cost.
- The value of predictions from the agents grows with the variance of uncertainty, indicating that they will play an increasingly important role in future power systems with more volatile renewable generation.
- When dealing with DDU, the proposed algorithm outperforms the traditional C&CG algorithm in terms of solution optimality.

A detailed predictive information market design between the operator and agents and a more efficient solution algorithm will be our future research directions.

APPENDIX A
PROOF OF LEMMA 1

Set a function

$$\begin{aligned} g(\alpha_i, \beta_i) :=\ & \mathbb{E}\left[(U_i - (\alpha_i + \beta_i U^{pre}_i))^2\right] \\ =\ & \alpha_i^2 + \mathbb{E}[(U^{pre}_i)^2]\beta_i^2 + 2\mathbb{E}[U^{pre}_i]\alpha_i\beta_i \\ & - 2\mathbb{E}[U_i]\alpha_i - 2\mathbb{E}[U_i U^{pre}_i]\beta_i + \mathbb{E}[U_i^2]. \end{aligned} \tag{A.1}$$

Note that ϵ_i is independent of U_i, then

$$\begin{aligned} &\mathbb{E}[U^{pre}_i] = \mathbb{E}[U_i - \epsilon_i] = \mathbb{E}[U_i] = \overline{u}_i, \\ &\mathbb{E}[U_i^2] = \text{var}[U_i] + (\mathbb{E}[U_i])^2 = \sigma^2_{U_i} + \overline{u}_i^2, \\ &\mathbb{E}[U_i\epsilon_i] = \mathbb{E}[U_i]\mathbb{E}[\epsilon_i] = 0, \\ &\mathbb{E}[U_i U^{pre}_i] = \mathbb{E}[U_i^2] - \mathbb{E}[U_i\epsilon_i] = \mathbb{E}[U_i^2] = \sigma^2_{U_i} + \overline{u}_i^2, \\ &\mathbb{E}[(U^{pre}_i)^2] = \mathbb{E}[U_i^2] - 2\mathbb{E}[U_i\epsilon_i] + \mathbb{E}[\epsilon_i^2] = \sigma^2_{U_i} + \overline{u}_i^2 + \sigma^2_{\epsilon_i}. \end{aligned}$$

Since α_i and β_i minimizes $g(\alpha_i, \beta_i)$, we have

$$\frac{\partial g(\alpha_i, \beta_i)}{\partial \alpha_i} = 0, \frac{\partial g(\alpha_i, \beta_i)}{\partial \beta_i} = 0, \tag{A.2}$$

whose solution is

$$\beta_i = \frac{\sigma_{U_i}^2}{\sigma_{U_i}^2 + \sigma_{\epsilon_i}^2},\ \alpha_i = (1-\beta_i)\overline{u}_i. \tag{A.3}$$

Then

$$\mathbb{E}[U_i - U_i^e] = \overline{u}_i - \alpha_i - \beta_i \overline{u}_i = 0, \tag{A.4}$$

and

$$\begin{aligned} & \text{cov}(U_i - U_i^e, U_i^{pre}) \\ =\ & \mathbb{E}[(U_i - U_i^e)U_i^{pre}] - \mathbb{E}[U_i - U_i^e]\mathbb{E}[U_i^{pre}] \\ =\ & \mathbb{E}[(U_i - (\alpha_i + \beta_i U_i^{pre}))U_i^{pre}] \\ =\ & \mathbb{E}[U_i U_i^{pre}] - \alpha_i \mathbb{E}[U_i^{pre}] - \beta_i \mathbb{E}[(U_i^{pre})^2] = 0. \end{aligned} \tag{A.5}$$

This completes the proof.

APPENDIX B
PROOF OF LEMMA 2

Since $U_i^e = \alpha_i + \beta_i U_i^{pre}$ is a function of U_i^{pre}, for any random variable X with finite second moment we have $\mathbb{E}[XU_i^e|U_i^{pre}] = U_i^e \mathbb{E}[X|U_i^{pre}]$ [33, p. 348]. Therefore,

$$\begin{aligned} & \text{var}[\eta_i|U_i^{pre}] \\ =\ & \mathbb{E}[(U_i - U_i^e)^2|U_i^{pre}] - (\mathbb{E}[U_i - U_i^e|U_i^{pre}])^2 \\ =\ & (\mathbb{E}[U_i^2|U_i^{pre}] - 2U_i^e \mathbb{E}[U_i|U_i^{pre}] + (U_i^e)^2) \\ & - ((\mathbb{E}[U_i|U_i^{pre}])^2 - 2U_i^e \mathbb{E}[U_i|U_i^{pre}] + (U_i^e)^2) \\ =\ & \mathbb{E}[U_i^2|U_i^{pre}] - (\mathbb{E}[U_i|U_i^{pre}])^2 \\ =\ & \text{var}[U_i|U_i^{pre}]. \end{aligned} \tag{B.1}$$

Moreover, we have η_i and U_i^{pre} are independent random variables. Therefore,

$$\begin{aligned} \text{var}[U_i|U_i^{pre}] =\ & \text{var}[\eta_i|U_i^{pre}] = \text{var}[\eta_i] \\ =\ & \text{var}[U_i - (1-\beta_i)\overline{u}_i - \beta_i U_i^{pre}] \\ =\ & \text{var}[(1-\beta_i)U_i + \beta_i \epsilon_i] \\ =\ & (1-\beta_i)^2 \sigma_{U_i}^2 + \beta_i^2 \sigma_{\epsilon_i}^2. \end{aligned} \tag{B.2}$$

Moreover,

$$\text{var}[U_i|U_i^{pre}] = \sigma_{U_i}^2 \frac{\sigma_{\epsilon_i}^2}{\sigma_{\epsilon_i}^2 + \sigma_{U_i}^2} \le \sigma_{U_i}^2. \tag{B.3}$$

This completes the proof.

APPENDIX C
PROOF AND DISCUSSION OF LEMMA 3

A. Proof of Lemma 3

Recall that according to Lemma 1 we have $\mathbb{E}[\eta_{it}] = \mathbb{E}[U_{it} - U_{it}^e] = 0$. Combining the fact $\text{var}[\eta_{it}] = \text{var}[U_{it}|U_{it}^{pre}]$ shown in (B.2), we have $\mathbb{E}[\eta_{it}/u_{it}^h] = 0$ and $\text{var}[\eta_{it}/u_{it}^h] = 1 - \delta$. Then according to the Chebyshev inequality,

$$\begin{aligned} & \mathcal{P}\left(|\eta_{it}/u_{it}^h| \ge 1\right) \\ =\ & \mathcal{P}\left(|(\eta_{it}/u_{it}^h) - \mathbb{E}[\eta_{it}/u_{it}^h]| \ge \sqrt{\text{var}[\eta_{it}/u_{it}^h]/(1-\delta)}\right) \\ \le\ & 1/\left(\sqrt{1/(1-\delta)}\right)^2 = 1 - \delta. \end{aligned} \tag{C.1}$$

Note that $v_{it} = |\eta_{it}/u_{it}^h|$. Thus, we have $\mathcal{P}(v_{it} \ge 1) = \mathcal{P}(|\eta_{it}/u_{it}^h| \ge 1) \le 1 - \delta$.

Again by the Chebyshev inequality,

$$\begin{aligned} & \mathcal{P}\left(\sum\nolimits_i v_{it} \ge \mathbb{E}\left[\sum\nolimits_i v_{it}\right] + \sqrt{\text{var}\left[\sum\nolimits_i v_{it}\right]/(1-\xi)}\right) \\ =\ & \mathcal{P}\left(\sum\nolimits_i v_{it} - \mathbb{E}\left[\sum\nolimits_i v_{it}\right] \ge \sqrt{\text{var}\left[\sum\nolimits_i v_{it}\right]}\sqrt{1/(1-\xi)}\right) \\ \le\ & \{1/(1-\xi)\}^{-1} = 1 - \xi. \end{aligned} \tag{C.2}$$

We find an upper bound for $\mathbb{E}[\sum_i v_{it}] + \sqrt{\text{var}[\sum_i v_{it}]/(1-\xi)}$. $\{\eta_{it}/\text{var}[\eta_{it}], \forall i, \forall t\}$ are independent and identically distributed (i.i.d.), so $\{v_{it}, \forall i, \forall t\}$ are also i.i.d. random variables. Therefore, $\mathbb{E}[\sum_i v_{it}] = I\mathbb{E}[v_{it}]$ and $\text{var}[\sum_i v_{it}] = I\text{var}[v_{it}]$. Moreover,

$$\begin{aligned} \text{var}[v_{it}] =\ & \mathbb{E}[v_{it}^2] - (\mathbb{E}[v_{it}])^2 \\ =\ & \mathbb{E}[(\eta_{it}/u_{it}^h)^2] - (\mathbb{E}[v_{it}])^2 \\ =\ & \text{var}[\eta_{it}/u_{it}^h] + (\mathbb{E}[\eta_{it}/u_{it}^h])^2 - (\mathbb{E}[v_{it}])^2 \\ =\ & 1 - \delta - (\mathbb{E}[v_{it}])^2. \end{aligned} \tag{C.3}$$

Then

$$\begin{aligned} & \mathbb{E}\left[\sum\nolimits_i v_{it}\right] + \sqrt{\text{var}\left[\sum\nolimits_i v_{it}\right]/(1-\xi)} \\ =\ & I\mathbb{E}[v_{it}] + \sqrt{I(1-\delta-(\mathbb{E}[v_{it}])^2)/(1-\xi)}\ =: G(\mathbb{E}[v_{it}]) \end{aligned}$$

is a function of $\mathbb{E}[v_{it}]$ for $0 \le \mathbb{E}[v_{it}] \le \sqrt{1-\delta}$. By calculating its derivative, it is easy to show that $G(\mathbb{E}[v_{it}])$ first increases and then declines, whose unique maximum value is Γ_S. Therefore, we have $\mathcal{P}(\sum_i v_{it} \ge \Gamma_S) \le 1 - \xi, \forall t$. Similarly, we can prove that $\mathcal{P}(\sum_t v_{it} \ge \Gamma_T) \le 1 - \xi, \forall i$.

B. Discussion of Lemma 3

The bound $\mathcal{P}(v_{it} \ge 1) \le 1 - \delta$ is tight for $\eta_{it}/\sqrt{\text{var}[\eta_{it}]}$ with the following discrete probability distribution:

$$\begin{aligned} & \mathcal{P}(\eta_{it}/\sqrt{\text{var}[\eta_{it}]} = -1/\sqrt{1-\delta}) = (1-\delta)/2, \\ & \mathcal{P}(\eta_{it}/\sqrt{\text{var}[\eta_{it}]} = 0) = \delta, \\ & \mathcal{P}(\eta_{it}/\sqrt{\text{var}[\eta_{it}]} = 1/\sqrt{1-\delta}) = (1-\delta)/2, \end{aligned}$$

which satisfies

$$\mathbb{E}[\eta_{it}/\sqrt{\text{var}[\eta_{it}]}] = 0,\ \text{var}[\eta_{it}/\sqrt{\text{var}[\eta_{it}]}] = 1.$$

Now we focus on the bound $\mathcal{P}(\sum_i v_{it} \ge \Gamma_S) \le 1 - \xi$, where Γ_S is given by (8c). We consider some specific testing probability distributions of $\eta_{it}/\sqrt{\text{var}[\eta_{it}]}$ and calculate $\mathcal{P}(\sum_i v_{it} \ge \Gamma_S)$ under varying Γ_S.

First, consider the probability distribution with

$$\mathcal{P}(\eta_{it}/\sqrt{\text{var}[\eta_{it}]} = 1) = \mathcal{P}(\eta_{it}/\sqrt{\text{var}[\eta_{it}]} = -1) = 1/2.$$

Then $\mathcal{P}(v_{it} = \sqrt{1-\delta}) = 1, \forall i$ and $\mathcal{P}(\sum_i v_{it} = I\sqrt{1-\delta}) = 1$. Therefore, $\mathcal{P}(\sum_i v_{it} \ge \Gamma_S) = 1$ for $\Gamma_S \le I\sqrt{1-\delta}$ and $\mathcal{P}(\sum_i v_{it} \ge \Gamma_S) = 0$ for $\Gamma_S > I\sqrt{1-\delta}$.

Second, construct a discrete probability distribution as follows with parameter $\xi \in (0, 1)$.

$$\mathcal{P}(\eta_{it}/\sqrt{\text{var}[\eta_{it}]} = a) = \mathcal{P}(\eta_{it}/\sqrt{\text{var}[\eta_{it}]} = -a) = p,$$

$$\mathcal{P}(\eta_{it}/\sqrt{\text{var}[\eta_{it}]} = 0) = 1 - 2p,$$

where

$$a = \sqrt{\frac{1+I-I\xi}{I(1-\xi)}},\ p = \frac{I(1-\xi)}{2(1+I-I\xi)}.$$

Then $\mathcal{P}(v_{it} = a\sqrt{1-\delta}) = 2p$ and $\mathcal{P}(v_{it} = 0) = 1 - 2p, \forall i$, based on which $\mathcal{P}(\sum_i v_{it} \geq \Gamma_S)$ can be calculated under different Γ_S.

Bound (8c) is compared with the values of $\mathcal{P}(\sum_i v_{it} \geq \Gamma_S)$ under the two kinds of testing probability distributions in Fig. 12, where $I = 3$ or $I = 5$ and ξ varies in $(0, 1)$. All the curves of the testing probability distributions are above the curves of bound (8c), which verifies the effectiveness of the bound.

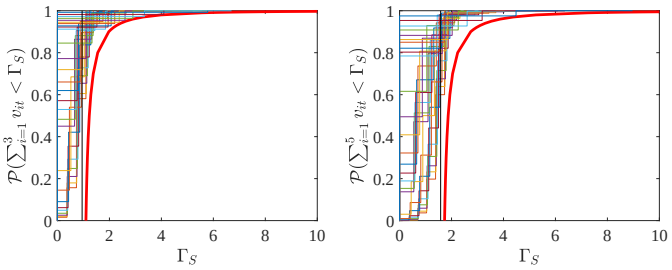

Fig. 12. Bounds (8c) (red thick curves) and testing values (other curves) of $\mathcal{P}(\sum_i v_{it} \geq \Gamma_S)$ when $I = 3$ (left) and $I = 5$ (right).

When I is large, by the central limit theorem, the probability distribution of $\sum_i v_{it}$ can be approximated by a normal distribution with mean $I\mathbb{E}[v_{it}]$ and variance $I\text{var}[v_{it}] = I(1 - \delta - (\mathbb{E}[v_{it}])^2)$. Therefore, with a high probability, $\sum_i v_{it}$ will not exceed $I\sqrt{1-\delta}$ much. Hence, $\Gamma_S \approx I\sqrt{1-\delta}$ can be used as the bound in $\mathcal{P}(\sum_i v_{it} \geq \Gamma_S) \approx 0$ when I is large.

APPENDIX D
PROOF OF LEMMA 5

In the definition of the set Φ, the coefficients Γ_S and Γ_T are independent of τ according to their definitions in Lemma 3. Then Φ is unchanged and thus so is the vertex set $V(\Phi)$.

Now fix some τ subject to $\tau_i \in (0, 1), \forall i \in \mathcal{I}$. By the definitions of τ_i and u_i^h in Definition 1 and Lemma 3, we have $u_i^h(\tau_i) > 0, \forall i \in \mathcal{I}$. For any $\phi \in \Phi$, it can be verified that $u = \pi(\phi)$ and $v_{it} = |u_{it} - u_{it}^e(\tau_i)|, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}$ satisfy the constraints of $\mathcal{U}(\tau)$. Therefore, $\pi(\phi) \in \mathcal{U}(\tau)$ and $\pi : \Phi \to \mathcal{U}(\tau)$ is well-defined. The inverse map $\pi^{-1}(u) = \phi$ is given by:

$$\phi_{it} = \frac{u_{it} - u_{it}^e(\tau)}{u_i^h(\tau)}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}. \tag{D.1}$$

Similarly, for any $u \in \mathcal{U}(\tau)$, $\phi = \pi^{-1}(u)$ and $\psi_{it} = |\phi_{it}|, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}$ satisfy the constraints of Φ. Therefore, $\pi^{-1}(u) \in \Phi$ and then $\pi : \Phi \to \mathcal{U}(\tau)$ is bijective.

Assume $u^* \in V(\mathcal{U}(\tau))$ is a vertex and $\phi^* = \pi^{-1}(u^*)$. Suppose $\phi^* = a\phi' + (1-a)\phi''$ for $a \in [0, 1]$ and $\phi', \phi'' \in \Phi$. Then $\pi(\phi'), \pi(\phi'') \in \mathcal{U}(\tau)$ and $u^* = a\pi(\phi') + (1-a)\pi(\phi'')$. Since $u^* \in V(\mathcal{U}(\tau))$, we have $\pi(\phi') = \pi(\phi'') = u^*$, which implies $\phi' = \phi'' = \phi^*$. Therefore, $\phi^* \in V(\Phi)$ and then $\pi^{-1}(V(\mathcal{U}(\tau))) \subset V(\Phi)$. Similarly, $\pi(V(\Phi)) \subset V(\mathcal{U}(\tau))$. Then the restriction on the vertex set $\pi : V(\Phi) \to V(\mathcal{U}(\tau))$ is also bijective.

APPENDIX E
PROOF OF THEOREM 1

Suppose the optimal solution of the two-stage RO model (11) is (x^*, τ^*) and the optimal objective value is

$$O^* := \min_{\substack{x \in \mathcal{X} \cap \tilde{\mathcal{X}}_R \\ \tau_i \in [0,1], (10), \forall i \in \mathcal{I}}} f(x) + \sum_{i \in \mathcal{I}} C_i(\tau_i) + \max_{u \in \mathcal{U}(\tau)} \min_{y \in \mathcal{Y}(x,u)} g(y).$$

We start the proof of Theorem 1 by providing the following claims. For any $K \in \mathbb{Z}^+$:

(1) **Claim 1**: $LB_K \leq O^* \leq UB_K$;

(2) **Claim 2**: If the algorithm does not terminate after K iterations, then for any $K_1, K_2 \in [K]$, we have $\phi_{K_1}^* \neq \phi_{K_2}^*$.

Proofs of claims:

(1) **Claim 1**: The master problem in the K-th iteration is equivalent to

$$LB_K = \min_{\substack{x \in \mathcal{X} \cap \mathcal{X}_{K-1} \\ \tau_i \in [0,1], (10), \forall i \in \mathcal{I}}} f(x) + \sum_{i \in \mathcal{I}} C_i(\tau_i) + \max_{u \in \mathcal{U}_{K-1}(\tau)} \min_{y \in \mathcal{Y}(x,u)} g(y),$$

where

$$\mathcal{U}_{K-1}(\tau) := \{u^k = u^e(\tau) + u^h(\tau)\phi_k^* \mid k \in [K-1]\}, \tag{E.1a}$$

$$\mathcal{X}_{K-1} := \{x \mid \mathcal{Y}(x, u) \neq \emptyset, \forall u \in \mathcal{U}_{K-1}(\tau)\}. \tag{E.1b}$$

Since $\mathcal{U}_{k-1}(\tau) \subset \mathcal{U}(\tau)$ and $\mathcal{X}_{K-1} \supset \tilde{\mathcal{X}}_R$, we have $LB_K \leq O^*$.

Next, we prove $UB_K \geq O^*$ by induction. First of all, $UB_0 = +\infty \geq O^*$. Suppose for the sake of induction that $UB_{K-1} \geq O^*$, then if $1^\top s_K^* > 0$, we have $UB_K = UB_{K-1} \geq O^*$; otherwise, (x_K^*, τ_K^*) is robust feasible and

$$\begin{aligned} UB_K &= f(x_K^*) + \sum_i C_{i,K}^* + g(y_K^*) \\ &= f(x_K^*) + \sum_i C_{i,K}^* + \max_{u \in \mathcal{U}(\tau_K^*)} \min_{y \in \mathcal{Y}(x_K^*, u)} g(y) \\ &\geq O^*. \end{aligned} \tag{E.2}$$

The last inequality is due to the optimality of (x^*, τ^*).

(2) **Claim 2**: Without loss of generality, we assume that $K_1 < K_2$. If we have $\phi_{K_1}^* = \phi_{K_2}^*$, then $u_{K_2}^* = u^e(\tau_{K_2}^*) + u^h(\tau_{K_2}^*)\phi_{K_2}^* = u^e(\tau_{K_2}^*) + u^h(\tau_{K_2}^*)\phi_{K_1}^* \in \mathcal{U}_{K_2-1}(\tau_{K_2}^*)$, so $x_{K_2}^*$ must be robust feasible. Moreover,

$$\begin{aligned} LB_{K_2} &= f(x_{K_2}^*) + \sum_i C_{i,K_2}^* + \max_{u \in \mathcal{U}_{K_2-1}(\tau_{K_2}^*)} \min_{y \in \mathcal{Y}(x_{K_2}^*, u)} g(y) \\ &\geq f(x_{K_2}^*) + \sum_i C_{i,K_2}^* + \min_{y \in \mathcal{Y}(x_{K_2}^*, u_{K_2}^*)} g(y) = UB_{K_2}. \end{aligned} \tag{E.3}$$

Together with $LB_{K_2} \leq UB_{K_2}$ from Claim 1, we have $LB_{K_2} = UB_{K_2}$. This contradicts to the assumption that the algorithm does not terminate after $K \geq K_2$ iterations.

Now the proof of **Theorem 1** is given below.

First, we prove that the algorithm converges in $\mathcal{O}(n_U)$ iterations. With Lemma 4, we know that the worst-case scenario u_k^* can be achieved at a vertex of $\mathcal{U}_k(\tau_k^*)$. A vertex of the set Φ

corresponds to a vertex of the set $\mathcal{U}_k(\tau_k^*)$. According to Claim 2, the same vertex of Φ will not be picked up twice. Moreover, the number of vertices of Φ is n_U. Hence, the algorithm stops in $\mathcal{O}(n_U)$ iterations.

Suppose the algorithm terminates after $K \leq n_U$ iterations.

Next, we show the robust feasibility of (x_K^*, τ_K^*). Obviously, we have $x_K^* \in \mathcal{X}$ and $\tau_K^* \in [0, 1]$. Moreover, $s_K^* = 0$ when the algorithm terminates, so (x_K^*, τ_K^*) is robust feasible.

Finally, we show the optimality of (x_K^*, τ_K^*). According to Claim 1, we have $LB_K \leq O^* \leq UB_K$. Together with the condition for termination $|UB_K - LB_K| \leq \epsilon$, we have

$$|UB_K - O^*| \leq |UB_K - LB_K| \leq \epsilon. \tag{E.4}$$

This completes the proof.

REFERENCES

[1] Y. Yang, W. Wu, B. Wang, and M. Li, "Chance-constrained economic dispatch considering curtailment strategy of renewable energy," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5792–5802, 2021.

[2] W. Wei, F. Liu, S. Mei, and Y. Hou, "Robust energy and reserve dispatch under variable renewable generation," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 369–380, 2014.

[3] Y. Guan and J. Wang, "Uncertainty sets for robust unit commitment," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1439–1440, 2013.

[4] D. Bertsimas and M. Sim, "The price of robustness," *Operations research*, vol. 52, no. 1, pp. 35–53, 2004.

[5] D. Bertsimas and D. B. Brown, "Constructing uncertainty sets for robust linear optimization," *Operations research*, vol. 57, no. 6, pp. 1483–1495, 2009.

[6] L. Xie, Y. Gu, X. Zhu, and M. G. Genton, "Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 511–520, 2013.

[7] A. Lorca and X. A. Sun, "Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1702–1713, 2014.

[8] A. A. Raja, P. Pinson, J. Kazempour, and S. Grammatico, "A market for trading forecasts: A wagering mechanism," *International Journal of Forecasting*, 2023.

[9] P. Pinson, L. Han, and J. Kazempour, "Regression markets and application to energy forecasting," *TOP*, pp. 1–41, 2022.

[10] L. Han, P. Pinson, and J. Kazempour, "Trading data for wind power forecasting: A regression market with lasso regularization," *Electric Power Systems Research*, vol. 212, p. 108442, 2022.

[11] B. Wang, Q. Guo, and Y. Yu, "Mechanism design for data sharing: An electricity retail perspective," *Applied Energy*, vol. 314, p. 118871, 2022.

[12] J. Cui, N. Gu, and C. Wu, "Blockchain enabled data transmission for energy imbalance market," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 1254–1266, 2021.

[13] Y. Chen, T. Li, C. Zhao, and W. Wei, "Decentralized provision of renewable predictions within a virtual power plant," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2652–2662, 2020.

[14] M. Shamsi and P. Cuffe, "Prediction markets for probabilistic forecasting of renewable energy sources," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 1244–1253, 2021.

[15] Z. Sun, L. Von Krannichfeldt, and Y. Wang, "Trading and valuation of day-ahead load forecasts in an ensemble model," *IEEE Transactions on Industry Applications*, 2023.

[16] V. Goel and I. E. Grossmann, "A class of stochastic programs with decision dependent uncertainty," *Mathematical programming*, vol. 108, no. 2-3, pp. 355–394, 2006.

[17] Y. Zhan, Q. P. Zheng, J. Wang, and P. Pinson, "Generation expansion planning with large amounts of wind power via decision-dependent stochastic programming," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3015–3026, 2016.

[18] S. Giannelos, I. Konstantelos, and G. Strbac, "Option value of demand-side response schemes under decision-dependent uncertainty," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5103–5113, 2018.

[19] B. Zeng, J. Feng, N. Liu, and Y. Liu, "Co-optimized parking lot placement and incentive design for promoting pev integration considering decision-dependent uncertainties," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1863–1872, 2020.

[20] W. Yin, S. Feng, J. Hou, C. Qian, and Y. Hou, "A decision-dependent stochastic approach for joint operation and maintenance of overhead transmission lines after sandstorms," *IEEE Systems Journal*, vol. 17, no. 1, pp. 1489–1500, 2022.

[21] O. Nohadani and K. Sharma, "Optimization under decision-dependent uncertainty," *SIAM Journal on Optimization*, vol. 28, no. 2, pp. 1773–1795, 2018.

[22] N. H. Lappas and C. E. Gounaris, "Robust optimization for decision-making under endogenous uncertainty," *Computers & Chemical Engineering*, vol. 111, pp. 252–266, 2018.

[23] Y. Zhang, F. Liu, Z. Wang, Y. Su, W. Wang, and S. Feng, "Robust scheduling of virtual power plant under exogenous and endogenous uncertainties," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1311–1325, 2021.

[24] B. Zeng and W. Wang, "Two-stage robust optimization with decision dependent uncertainty," *[Online]. Available: arXiv:2203.16484*, 2022.

[25] H. Haghighat, W. Wang, and B. Zeng, "Robust unit commitment with decision-dependent uncertain demand and time-of-use pricing," *IEEE Transactions on Power Systems*, 2023.

[26] Y. Chen and W. Wei, "Robust generation dispatch with strategic renewable power curtailment and decision-dependent uncertainty," *IEEE Transactions on Power Systems*, vol. 38, no. 5, pp. 4640–4654, 2023.

[27] S. Avraamidou and E. N. Pistikopoulos, "Adjustable robust optimization through multi-parametric programming," *Optimization Letters*, vol. 14, no. 4, pp. 873–887, 2020.

[28] F. Luo and S. Mehrotra, "Distributionally robust optimization with decision dependent ambiguity sets," *Optimization Letters*, vol. 14, pp. 2565–2594, 2020.

[29] X. Yu and S. Shen, "Multistage distributionally robust mixed-integer programming with decision-dependent moment-based ambiguity sets," *Mathematical Programming*, vol. 196, no. 1-2, pp. 1025–1064, 2022.

[30] L. Qiu, Y. Dan, X. Li, and Y. Cao, "Decision-dependent distributionally robust integrated generation, transmission, and storage expansion planning: An enhanced benders decomposition approach," *IET Renewable Power Generation*, 2023.

[31] Y. Li, S. Lei, W. Sun, C. Hu, and Y. Hou, "A distributionally robust resilience enhancement strategy for distribution networks considering decision-dependent contingencies," *IEEE Transactions on Smart Grid*, 2023.

[32] R. Xie, W. Wei, and Y. Chen, "Sizing grid-connected wind power generation and energy storage with wake effect and endogenous uncertainty: A distributionally robust method," *[Online]. Available: arXiv:2212.14665*, 2022.

[33] G. Grimmett and D. Stirzaker, *Probability and Random Processes*. Oxford University Press, 2001.

[34] X. Vives, *Information and learning in markets: the impact of market microstructure*. Princeton University Press, 2010.

[35] L. Han, J. Kazempour, and P. Pinson, "Monetizing customer load data for an energy retailer: A cooperative game approach," in *2021 IEEE Madrid PowerTech*. IEEE, 2021, pp. 1–6.

[36] S. Pineda and J. M. Morales, "Solving linear bilevel problems using big-ms: not all that glitters is gold," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 2469–2471, 2019.

[37] H. Konno, "A cutting plane algorithm for solving bilinear programs," *Mathematical Programming*, vol. 11, no. 1, pp. 14–27, 1976.

[38] B. Zeng and L. Zhao, "Solving two-stage robust optimization problems using a column-and-constraint generation method," *Operations Research Letters*, vol. 41, no. 5, pp. 457–461, 2013.

[39] L. Wu, "A tighter piecewise linear approximation of quadratic cost curves for unit commitment problems," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2581–2583, 2011.

[40] R. Xie, "Information valuation," https://github.com/xieruijx/Information-Valuation, 2022.

[41] J. M. Arroyo and A. J. Conejo, "Optimal response of a thermal unit to an electricity spot market," *IEEE Transactions on power systems*, vol. 15, no. 3, pp. 1098–1104, 2000.

Rui Xie (Member, IEEE) received the B.E. degree in electrical engineering and the B.Sc. degree in mathematics in 2017, and the Ph.D. degree in electrical engineering in 2022 from Tsinghua University, Beijing, China. She is currently a research associate with the Department of Mechanical and Automation Engineering and the Shun Hing Institute of Advanced Engineering, the Chinese University of Hong Kong, Hong Kong SAR. Her current research interests include stochastic optimization problems in power systems.

Pierre Pinson (Fellow, IEEE) received the M.Sc. degree in applied mathematics from the National Institute of Applied Sciences (INSA), Toulouse, France, in 2002 and the Ph.D. degree in energetics from Ecole des Mines de Paris, France, in 2006. He is the chair of data-centric design engineering at Imperial College London, United Kingdom, Dyson School of Design Engineering. He is also an affiliated professor of operations research and analytics with the Technical University of Denmark and a chief scientist at Halfspace (Denmark). He is the editor-in-chief of the International Journal of Forecasting. His research interests include analytics, forecasting, optimization and game theory, with application to energy systems mostly, but also logistics, weather-driven industries and business analytics. He is a Fellow of the IEEE, an INFORMS member and a director of the International Institute of Forecasters (IIF).

Yin Xu (Senior Member, IEEE) received the B.E. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2008 and 2013, respectively. From 2013 to 2016, he was a Research Assistant Professor with the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, USA. He is currently a Professor with the School of Electrical Engineering, Beijing Jiaotong University, Beijing, China. His research interests include power grid resilience, power system electromagnetic transient modeling and high-performance simulation.

Dr. Xu is currently serving as Chair of the Energy Internet Resilience Working Group under the IEEE PES Energy Internet Coordinating Committee and Secretary of the Distribution Test Feeder Working Group under the IEEE PES Distribution System Analysis Subcommittee. He is an Editorial Broad member of IEEE Transactions on Power Systems, IEEE Power Engineering Letters, IET Smart Grid, and Energy Conversion and Economics.

Yue Chen (Member, IEEE) received the B.E. degree in electrical engineering from Tsinghua University, Beijing, China, in 2015, the B.S. degree in economics from Peking University, Beijing, China, in 2017, and the Ph.D. degree in electrical engineering from Tsinghua University, in 2020. She is currently a Vice-Chancellor Assistant Professor with the Department of Mechanical and Automation Engineering, the Chinese University of Hong Kong, Hong Kong SAR. Her research interests include optimization, game theory, mathematical economics, with applications in smart grid and integrated energy systems. She is currently an Associate Editor for IEEE Transactions on Smart Grid, IEEE Power Engineering Letters, and IET Renewable Power Generation.