

Day-ahead Network-constrained Unit Commitment Considering Distributional Robustness and Intraday Discreteness: A Sparse Solution Approach

Xiaodong Zheng, Baorong Zhou, Xiuli Wang, Bo Zeng, Jizhong Zhu, Haoyong Chen, and Waisheng Zheng

Abstract—Quick-start generation units are critical devices and flexible resources to ensure a high penetration level of renewable energy in power systems. By considering the wind uncertainty and both binary and continuous decisions of quick-start generation units within the intraday dispatch, we develop a Wasserstein-metric-based distributionally robust optimization model for the day-ahead network-constrained unit commitment (NCUC) problem with mixed-integer recourse. We propose two feasible frameworks for solving the optimization problem. One approximates the continuous support of random wind power with a finite number of events, and the other leverages the extremal distributions instead. Both solution frameworks rely on the classic nested column-and-constraint generation (C&CG) method. It is shown that due to the sparsity of L_1 -norm Wasserstein metric, the continuous support of wind power generation could be represented by a discrete one with a small number of events, and the rendered extremal distributions are sparse as well. With this reduction, the distributionally robust NCUC model with complicated mixed-integer recourse problems can be efficiently handled by both solution frameworks. Numerical studies are carried out, demonstrating that the model considering quick-start generation units ensures unit commitment (UC) schedules to be more robust and cost-effective, and the distributionally robust optimization method captures the wind uncertainty well in terms of out-of-sample tests.

Index Terms—Unit commitment, distributional robustness, mixed-integer recourse, nested column-and-constraint generation (C&CG), sparsity, Wasserstein metric.

NOMENCLATURE

A. Sets

$\mathcal{G}^1, \mathcal{G}^2, \mathcal{G}$	Sets of slow-acting generation units (coal-fired units), quick-start generation units (gas turbines), and all generation units
$\mathcal{L}, \mathcal{D}, \mathcal{W}$	Sets of transmission lines, loads, and wind farms
\mathcal{T}	Set of dispatch time periods, i.e., $\{1, 2, \dots, 24\}$

B. Parameters and Functions

$\kappa_{g,l}^G, \kappa_{d,l}^D, \kappa_{w,l}^W$	Power flow distribution factors from unit g , load d , and wind farm w to line l
$C_g(\cdot)$	Variable cost of unit g
$C_{d,t}^{LS}(\cdot)$	Load shedding cost for load d at time t
$C_w^{WC}(\cdot)$	Wind curtailment cost for wind farm w
$C^{Pen}(\cdot)$	Penalty factor of slack variables
$D_{d,t}$	Demand level of load d at time t
$D_{d,t}^{\max}$	The maximum value of load d that could be interrupted at time t
F_l	Thermal power rating of transmission line l
MU_g, MD_g	The minimum up and down time of unit g
NL_g	No-load cost of unit g
P_g^{\min}, P_g^{\max}	The minimum and maximum production levels of unit g
R_g^+, R_g^-	Ramp-up and ramp-down limits of unit g
SU_g, SD_g	Start-up and shut-down costs of unit g
$W_{w,t}$	Day-ahead forecasting generation of wind farm w at time t

C. Decision Variables and Random Variables

$\delta_{d,t}^{LS}$	Load shedding scheduled for load d at time t
$\delta_{w,t}^{WC}$	Wind curtailment of wind farm w at time t
$\zeta_{w,t}$	Forecasting error of wind farm w at time t
$p_{g,t}$	Scheduled production level of unit g at time t
$S_{g,t}^G$	Positive slack variable added to unit g at time t

Manuscript received: June 26, 2021; revised: October 27, 2021; accepted: February 2, 2022. Date of CrossCheck: February 2, 2022. Date of online publication: February 11, 2022.

The authors thank Dr. Z. Li and Dr. J. Lyu for their valuable discussions. The work of X. Zheng and B. Zhou was supported by the Guangdong R&D Program in Key Areas (No. 2021B0101230004); the work of B. Zeng was supported in part by the U.S. National Science Foundation (No. CMMI-1635472); and the work of H. Chen was supported by the Key Program of National Natural Science Foundation of China (No. 51937005).

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

X. Zheng (corresponding author) is with the School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China, and he is also with the Electric Power Research Institute of China Southern Power Grid Co., Ltd., Guangzhou 510663, China (e-mail: eezhengxd@xjtu.edu.cn).

B. Zhou is with the Electric Power Research Institute of China Southern Power Grid Co., Ltd., Guangzhou 510663, China (e-mail: zhoubr@csg.cn).

X. Wang is with the Shaanxi Key Laboratory of Smart Grid, School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: xuliw@xjtu.edu.cn).

B. Zeng is with the Department of Industrial Engineering, and the Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15260, USA (e-mail: bzeng@pitt.edu).

J. Zhu and H. Chen are with the School of Electric Power Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: zhujz@scut.edu.cn; eehychen@scut.edu.cn).

W. Zheng is with China Southern Power Grid Co., Ltd., Guangzhou 510663, China (e-mail: ws-zheng@163.com).

DOI: 10.35833/MPCE.2021.000413



$x_{g,t}$ $u_{g,t}$ $v_{g,t}$ Binary variables indicating whether unit g is on, start-up, or shut-down at time t

I. INTRODUCTION

THE past decade has witnessed a rapid increase of wind power generation worldwide. Although wind power plays a central role in a sustainable power system, its unpredictability generates a huge demand on the real-time flexibility and generation capacity of power systems. In order to develop a modern power system with dominant renewable energies like wind power, the randomness should be carefully hedged against. In this context, a sufficient amount of flexible dispatchable resources should be installed such as gas turbines, combined-cycle units, pumped hydro storage, and bulk energy storage [1]-[4].

In addition to the resource adequacy from the hardware perspective, a powerful decision-making model is another key factor from the software perspective to ensure the efficiency and reliability of system operation subject to wind uncertainty. The stochastic programming is a classic method for handling the uncertainty in scheduling models [5]. However, the inevitable errors in the probability distribution approximation and the scenario sampling process may render the unit commitment (UC) schedule unreliable. In this regard, the robust optimization (RO) and distributionally robust optimization (DRO) have been introduced to support the network-constrained unit commitment (NCUC) model to handle the uncertainty of variable renewable generation as well as the ambiguity on their probability distributions [1], [3], [6] - [8]. The robustness of RO or DRO decision is achieved by simultaneously considering all (usually infinite) scenarios or probability distributions defined in an uncertainty set or an ambiguity set.

Due to the strong modeling capacity and the rational modeling philosophy, DRO attracts more attention from the power and energy society in recent years. The applicability and generality of DRO to the UC problem basically rely on two facts. First, the concept of distributional robustness enables system operators to deal with the distribution shift of randomness in the systems in a data-driven fashion, and thus to make UC schedules in a more accurate and quantitative manner. Second, the two-stage DRO framework admits general modeling components like discrete randomness and discrete recourse, yet remains problem tractability. Over the last half decade, many research works have been devoted to the application of two-stage DRO to the UC problem. In a nutshell, these research works can be divided into density-based [9], [10], distance-based [7], [11], [12], and moment-based [8], [13] - [15] methods, depending on the information used to characterize the ambiguous probability distribution and the method used to describe the distributional robustness. They can also be categorized as linear-decision-rule-based [10], [12] - [14], Benders-decomposition-based [9], [11], and delayed-constraint-generation-based [7] - [11] methods, depending on the type of algorithms used to compute the optimal solution.

It is noted that the above-mentioned works have not considered the intraday and even real-time discrete behavior fea-

tured by a majority of flexible resources, and the algorithmic frameworks therein would be incompatible if such behavior is considered in the day-ahead UC. Actually, the importance of including quick-start generation units and bulk storage with discrete decisions within the recourse, i.e., the intraday operation, has been recognized in the RO-based day-ahead NCUC models [1]-[3], where the resulting formulations are solved by the nested column-and-constraint generation (C&CG) method [16]. But existing literature in electrical engineering does not cover the topic of computing a DRO model with discrete recourse decisions, which is considered computationally intractable.

To the best of our knowledge, [17] is the most relevant literature in electrical engineering that deals with a two-stage DRO problem with discrete or mixed-integer recourse. Nevertheless, the second-stage mixed-integer linear program (MILP) therein is immediately relaxed into a continuous optimization problem before the solution method is developed [17]. References [18]-[20] are representative publications addressing this kind of problem in the operation research society. But the models in [18] and [19] have restricted the support for the random vector to be a scenario-based discrete set. The support adopted in the model of [20] is continuous as in this paper, but in essence, an *aforehand* discretization of this support is used. It is worth noting that a predefined finite support does not cover all extreme scenarios as the continuous one does, and it may fail to render a solution that is truly robust in some battlefields. For example, in the scheduling of renewable power systems, the random parameter is usually in high dimensions, in which case the sampled or historical scenarios are less representative unless the scenario set is extremely huge [21]. Hence, the *a priori* discrete support is unlikely to provide a robust schedule. Also, we believe that the two-stage DRO with continuous support is more general yet more challenging to handle.

The NCUC problems modeled by the two-stage DRO are not standard mathematical programs, and they are more complicated in structure than those modeled by the two-stage RO. Therefore, the tractability of the DRO-based NCUC models is of serious concern, especially the subset of these models that has a mixed-integer recourse problem. Even if the problems have been successfully recast as mixed-integer programs (MIPs) by leveraging some effective decomposition methods, the resulting mathematical programs are generally too heavy to handle due to the massive dual/primal variables and constraints that have been augmented. Another concern is about the ambiguity set. Despite the wide use of the aforementioned ambiguity sets, it is still difficult to figure out which one should be used on a specific problem, since the critical properties of these ambiguity sets have rarely been investigated. To promote the development and applications of DRO in power systems, the following two contributions are made in this paper.

1) Two feasible frameworks are proposed for solving the two-stage distributionally robust NCUC problem with quick-start generation units, which leverage the extreme points and the extremal distributions, respectively. Since the DRO-based NCUC problems considering mixed-integer recourse

and the solution methods have not been studied before, the efficiencies of these two solution methods are compared as well.

2) It is demonstrated that the L_1 -norm Wasserstein metric reduces the number of extreme points needed to represent the full support, and produces an extremal distribution with less events of wind power generation. This sparsity effect is beneficial in producing an equivalent MIP formulation of the two-stage distributionally robust NCUC problem, which is smaller in size and easier to handle.

II. MODEL FORMULATION

A. Deterministic NCUC Model with Quick-start Generation Units

We first present the deterministic NCUC model used in this paper. Noting that the minimum up and down time of most quick-start generation units, especially steam turbines and combined-cycle units, is more than one hour [22]-[24], we adopt the 1-hour resolution for the deterministic NCUC model. The 1-hour resolution is widely used in the clearing models of system operators and relative literature [1], [3], [25]. Besides, only the quick-start generation units are modeled, although other flexible resources like bulk energy storages can be easily included. We note that the solution frameworks and algorithms proposed in this paper are apparently compatible with different time resolutions and other discretely behaved devices. The deterministic NCUC model is formulated as:

$$\min \left\{ \sum_{t \in \mathcal{T}} \left(\sum_{g \in \mathcal{G}} x_{g,t} \cdot NL_g + u_{g,t} \cdot SU_g + v_{g,t} \cdot SD_g + C_g(p_{g,t}) \right) + \sum_{t \in \mathcal{T}} \left(\sum_{d \in \mathcal{D}} C_{d,t}^{LS}(\delta_{d,t}^{LS}) + \sum_{w \in \mathcal{W}} C_w^{WC}(\delta_{w,t}^{WC}) \right) + \sum_{t \in \mathcal{T}} \sum_{g \in \mathcal{G}} C^{\text{Pen}}(s_{g,t}^G) \right\} \quad (1)$$

s.t.

$$x_{g,t} - x_{g,t-1} = u_{g,t} - v_{g,t} \quad \forall g \in \mathcal{G}, t \in \mathcal{T} \setminus \{1\} \quad (2)$$

$$\sum_{\tau=\max\{1, t-MU_g+1\}}^t u_{g,\tau} \leq x_{g,t} \quad \forall g \in \mathcal{G}, t \in \mathcal{T} \setminus \{1\} \quad (3)$$

$$\sum_{\tau=\max\{1, t-MD_g+1\}}^t v_{g,\tau} \leq 1 - x_{g,t} \quad \forall g \in \mathcal{G}, t \in \mathcal{T} \quad (4)$$

$$P_g^{\min} x_{g,t} \leq p_{g,t} \leq P_g^{\max} x_{g,t} \quad \forall g \in \mathcal{G}, t \in \mathcal{T} \quad (5)$$

$$-R_g^- \leq p_{g,t} - p_{g,t-1} \leq R_g^+ \quad \forall g \in \mathcal{G}, t \in \mathcal{T} \setminus \{1\} \quad (6)$$

$$0 \leq \delta_{d,t}^{LS} \leq D_{d,t}^{\max} \quad \forall d \in \mathcal{D}, t \in \mathcal{T} \quad (7)$$

$$0 \leq \delta_{w,t}^{WC} \leq W_{w,t} + \zeta_{w,t} \quad \forall w \in \mathcal{W}, t \in \mathcal{T} \quad (8)$$

$$-F_l \leq \sum_{g \in \mathcal{G}} \kappa_{g,l}^G(p_{g,t} - s_{g,t}^G) + \sum_{w \in \mathcal{W}} \kappa_{w,l}^W(W_{w,t} + \zeta_{w,t} - \delta_{w,t}^{WC}) - \sum_{d \in \mathcal{D}} \kappa_{d,l}^D(D_{d,t} - \delta_{d,t}^{LS}) \leq F_l \quad \forall l \in \mathcal{L}, t \in \mathcal{T} \quad (9)$$

$$\sum_{g \in \mathcal{G}} (p_{g,t} - s_{g,t}^G) + \sum_{w \in \mathcal{W}} (W_{w,t} + \zeta_{w,t} - \delta_{w,t}^{WC}) = \sum_{d \in \mathcal{D}} (D_{d,t} - \delta_{d,t}^{LS}) \quad \forall t \in \mathcal{T} \quad (10)$$

The objective function as shown in (1) considers the no-load cost, the start-up and shut-down costs, the variable

costs of units, the costs of load shedding and wind curtailment, as well as the penalty factor of slack variables. The linear cost functions are used in this paper. Constraints (2)-(4) are the state transition equation, and the minimum up and down time limits of both non-quick-start and quick-start generation units, respectively. Constraint (5) is the production limit of units. Constraint (6) denotes the ramp-up and ramp-down limits of units. Constraints (7) and (8) impose limits on the amount of load shedding and wind curtailment, respectively. Constraint (9) denotes the flow limits of transmission lines. Constraint (10) denotes the power balance condition.

To ensure that the economic dispatch problem is surely feasible given any UC solution, slack variables are attached to the generators, as shown in (9) and (10). The slack variables are introduced to hedge against the minimum output requirement of units, which can be shown to guarantee the feasibility of the economic dispatch problem with the load shedding and wind curtailment variables. According to [26], a decomposition algorithm for two-level programs often relies on some feasibility cuts, which are identified from an infinite second-level problem. But many solvers have difficulty in returning a solution when the objective value is infinity. Therefore, we develop a surely feasible dispatch to alleviate the difficulty in searching the feasibility cuts. Note that the slack variables should be penalized in the objective function with a big penalty factor. Moreover, no nonzero slack variables are allowed in the final solution, as the generators practically cannot operate below the minimum output level. In another word, a nonzero slack variable in the final solution indicates that the power balance condition does not hold.

For ease of illustration, we use the compact matrix formulation of (1)-(10), which is expressed as:

$$\min_{\mathbf{x}, \mathbf{z}, \mathbf{y}} \{ \mathbf{c}_x^T \mathbf{x} + \mathbf{c}_z^T \mathbf{z} + \mathbf{c}_y^T \mathbf{y} \} \quad (11)$$

s.t.

$$\mathbf{F}\mathbf{x} \geq \mathbf{f} \quad (12)$$

$$\mathbf{G}\mathbf{z} \geq \mathbf{g} \quad (13)$$

$$\mathbf{H}\mathbf{y} \geq \mathbf{h} \quad (14)$$

$$\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} + \mathbf{C}\mathbf{y} \geq \mathbf{d} \quad (15)$$

$$\mathbf{U}\mathbf{y} + \mathbf{V}\boldsymbol{\zeta} = \mathbf{w} \quad (16)$$

The problem data of (11)-(16) are acquired from the original NCUC model (1)-(10), which include the cost vectors \mathbf{c}_x , \mathbf{c}_z , and \mathbf{c}_y , the matrixes \mathbf{F} , \mathbf{G} , \mathbf{H} , \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{U} , and \mathbf{V} , and the right-hand-side vectors \mathbf{f} , \mathbf{g} , \mathbf{h} , \mathbf{d} , and \mathbf{w} . The vectors \mathbf{x} , \mathbf{z} , and \mathbf{y} are the decision variables, while the vector $\boldsymbol{\zeta}$ is the random vector. Note that the vector \mathbf{x} denotes the statuses of slow-acting generation units, which should be determined day-ahead; the vectors \mathbf{z} and \mathbf{y} are “wait-and-see” decision variables, which are adjustable in intraday operation; and the random vector $\boldsymbol{\zeta}$ represents the wind power. The correspondences between the variables in (1)-(10) and (11)-(16) are given as: $\{x_{g,t}, u_{g,t}, v_{g,t} | g \in \mathcal{G}, t \in \mathcal{T}\} \Rightarrow \mathbf{x}$, $\{x_{g,t}, u_{g,t}, v_{g,t} | g \in \mathcal{G}, t \in \mathcal{T}\} \Rightarrow \mathbf{z}$, $\{p_{g,t}, s_{g,t}^G, \delta_{d,t}^{LS}, \delta_{w,t}^{WC} | g \in \mathcal{G}, d \in \mathcal{D}, w \in \mathcal{W}, t \in \mathcal{T}\} \Rightarrow \mathbf{y}$, and $\{W_{w,t} + \zeta_{w,t} | w \in \mathcal{W}, t \in \mathcal{T}\} \Rightarrow \boldsymbol{\zeta}$.

B. Distributionally Robust Counterpart

To be precise, the distributionally robust counterpart of the NCUC model will be a multi-stage DRO model. However, the multi-stage problem is extremely complicated and computationally unaffordable. Hence, we focus on the two-stage DRO counterpart, which is still challenging nowadays. There is another fact that encourages us to attack the two-stage problem, that is, the multi-stage NCUC problem can be approximately formulated as a two-stage one using some modeling skills [2].

Adopting the two-stage setting, the distributionally robust counterpart of the NCUC model (11)-(16) can be formulated as [7], [8]:

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P} \in \mathcal{P}} \left\{ \mathbf{c}_x^T \mathbf{x} + \mathbb{E}_{\mathbb{P}}(Q(\mathbf{x}, \xi)) \right\} \quad (17)$$

where \mathcal{X} is the feasible region for \mathbf{x} defined as $\{\mathbf{x} | \mathbf{F}\mathbf{x} \geq \mathbf{f}\}$; $\mathbb{P} \in \mathcal{P}$ is a probability distribution that belongs to the set \mathcal{P} ; $\mathbb{E}_{\mathbb{P}}(\cdot)$ is the expectation operator regarding the distribution \mathbb{P} ; and $Q(\mathbf{x}, \xi)$ is the optimal value function defined as:

$$\begin{cases} Q(\mathbf{x}, \xi) = \min_{\{\mathbf{z}, \mathbf{y}\} \in \mathcal{Z} \times \mathcal{Y}} \left\{ \mathbf{c}_z^T \mathbf{z} + \mathbf{c}_y^T \mathbf{y} \right\} \\ \text{s.t. } \mathcal{Z} = \{\mathbf{z} | \mathbf{G}\mathbf{z} \geq \mathbf{g}\} \\ \mathcal{Y} = \{\mathbf{y} | \mathbf{H}\mathbf{y} \geq \mathbf{h}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} + \mathbf{C}\mathbf{y} \geq \mathbf{d}, \mathbf{U}\mathbf{y} + \mathbf{V}\xi = \mathbf{w}\} \end{cases} \quad (18)$$

We adopt the L_1 -norm Wasserstein metric, which has a sparsity property that facilitates the solution procedure of the resulting DRO-based NCUC problem. Therefore, we have the ambiguity set \mathcal{P} for (17) [7], [27]:

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) \mid \mathbb{E}_{\mathbb{P} \times \hat{\mathbb{P}}} (d(\xi, \hat{\xi})) \leq \varepsilon; \xi \sim \mathbb{P}, \hat{\xi} \sim \hat{\mathbb{P}} \right\} \quad (19)$$

where Ξ is the support of ξ defined as $\{\xi | \xi^{\min} \leq \xi \leq \xi^{\max}\}$; \mathcal{P}_0 is the set of all distributions supported on Ξ ; $\hat{\mathbb{P}}$ is the empirical distribution constructed by historical data; $\mathbb{P} \times \hat{\mathbb{P}}$ is a joint distribution; $d(\xi, \hat{\xi})$ is the L_1 -norm distance function returning $\|\xi - \hat{\xi}\|_1$; and ε is the distance parameter controlling the size of the ambiguity set. As can be observed from (19), the ambiguity set \mathcal{P} contains all probability distributions that are no more than ε away from the empirical distribution, measured by the L_1 -norm Wasserstein metric.

III. SOLUTION METHODS

The DRO-based NCUC model (17) is novel in that the discrete behavior of quick-start generation units has been precisely considered. In this section, we propose two solution methods that can compute the globally optimal solution of the model, so as to facilitate the application of this model and other models that fall into this model family.

A. Reformulating the Second-stage Max-min Problem

As suggested by [7], N historical observations of ξ are aggregated into S data points before being utilized to construct the empirical distribution. Therefore, we have $\hat{\mathbb{P}} = \sum_{s=1}^S n^s / (N \delta_{\xi^s})$, where n^s / N is the probability mass of the s^{th} data point ξ^s ; and δ_{ξ^s} is the Dirac distribution.

By leveraging the conditional distribution interpretation of

$\mathbb{P} \times \hat{\mathbb{P}}$, the second stage of (17) can be written as:

$$\begin{cases} Z^{\text{P,INF}}(\mathbf{x}) = \max_{f_{\xi^s}^s \geq 0} \int_{\Xi} \frac{n^s}{N} Q(\mathbf{x}, \xi) f_{\xi^s}^s d\xi \\ \text{s.t. } \int_{\Xi} f_{\xi^s}^s d\xi = 1; \alpha^s \quad s = 1, 2, \dots, S \\ \sum_{s=1}^S \int_{\Xi} \frac{n^s}{N} \|\xi - \hat{\xi}^s\|_1 f_{\xi^s}^s d\xi \leq \varepsilon; \beta \end{cases} \quad (20)$$

where $f_{\xi^s}^s$ is the probability density function (PDF) for the s^{th} conditional distribution; and α^s and β are the dual variables.

Similar to the standard Lagrangian duality in convex optimization, the semi-infinite program (20) can be dualized into:

$$\begin{cases} Z^{\text{D,INF}}(\mathbf{x}) = \min_{\alpha^s, \beta \geq 0} \left\{ \sum_{s=1}^S \alpha^s + \varepsilon \beta \right\} \\ \text{s.t. } \frac{n^s}{N} Q(\mathbf{x}, \xi) - \alpha^s - \frac{n^s}{N} \|\xi - \hat{\xi}^s\|_1 \beta \leq 0 \\ \forall \xi \in \Xi, s = 1, 2, \dots, S \end{cases} \quad (21)$$

Please refer to [7] and [27] for the detail of the reformulation procedure. Note that the strong duality holds between (20) and (21) due to two facts: ① (20) has at least one feasible solution, which imitates the empirical distribution $\hat{\mathbb{P}}$, hence it is a relative interior point [27], [28]; ② the economic dispatch problem is surely feasible, and thus $Q(\mathbf{x}, \xi)$ is bounded.

It can be observed that the constraints of (21) are actually S independent robust constraints. But the standard technique addressing this kind of constraints (i.e., dualizing the left-hand side of the constraint [7], [8]) is not applied here, since $Q(\mathbf{x}, \xi)$ is the value function of an MILP. Fortunately, (21) can be recast as a two-stage robust optimization problem with mixed-integer recourse [26], [29], which is shown as:

$$Z^{\text{D,RO}}(\mathbf{x}) = \min_{\beta \geq 0} \varepsilon \beta + \sum_{s=1}^S \max_{\xi^s \in \Xi} \frac{n^s}{N} \left(-\beta \|\xi^s - \hat{\xi}^s\|_1 + Q(\mathbf{x}, \xi^s) \right) \quad (22)$$

The second term of (22) equals the first term of the objective of (21), because the optimal solution of α^s satisfies $\alpha^s = \max_{\xi^s \in \Xi} \frac{n^s}{N} \left(-\beta \|\xi^s - \hat{\xi}^s\|_1 + Q(\mathbf{x}, \xi^s) \right)$. Note that the second stage of (22) includes S independent max-min sub-problems, and they share the first-stage decision variable β in the objective function.

Problem (22) is solvable with a nested C&CG method according to a seminal paper [16]. It is anticipated that, by using the nested C&CG method, a finite number of ξ values could be found to sufficiently characterize the infinitely dimensional constraints in (21). Assuming the nested C&CG method terminates at the J^{th} iteration, (21) is representable with the MILP as:

$$\begin{cases} Z^{\text{D,FINT}}(\mathbf{x}) = \min_{\beta \geq 0, \alpha^s} \left\{ \varepsilon \beta + \sum_{s=1}^S \alpha^s \right\} \\ \text{s.t. } \alpha^s \geq \frac{n^s}{N} \left(-\beta \|\xi_j^{s*} - \hat{\xi}^s\|_1 + Q(\mathbf{x}, \xi_j^{s*}) \right); m_{\xi_j^{s*}} \\ s = 1, 2, \dots, S, j = 1, 2, \dots, J \end{cases} \quad (23)$$

where ξ_j^{s*} is the j^{th} extreme point for the s^{th} sub-problem of

(22) generated by the nested C&CG method; and m_{ξ^s} is the dual variable.

Remark 1: it is noted that the infinite support in (21) is approximated by (23) with a finite discrete support refined by C&CG method. This technique, which is also a delayed constraint generation method, is an alternative to addressing the semi-infinite program (21) aside from the one used in [7], [8], dualizing the left-hand side of the infinitely dimensional constraints and refining the second-stage value function instead. The alternative is adopted in this paper as it enables us to deal with the mixed-integer recourse problems. However, it can be observed that, when solving the semi-infinite programming problem, this technique could suffer from slow convergence in problems with a moment-based ambiguity set [30]. The slow convergence means that a massive number of events are needed in order to equivalently characterize the continuous support. In the following, this technique is examined on the DRO-based NCUC problem equipped with an L_1 -norm Wasserstein ambiguity set. Owing to the sparsity rendered from the L_1 norm, the technique turns out to be effective in handling this kind of problems.

B. Solution Framework 1: Applying Nested C&CG Method

It is straightforward to combine the first stage of (17) with (22), and then solve the problem as a whole using the nested C&CG method. Therefore, the DRO-based NCUC problem becomes a two-stage robust optimization problem with mixed-integer recourse:

$$\min_{x \in \mathcal{X}, \beta \geq 0} \left\{ c_x^T x + \varepsilon \beta + \sum_{s=1}^S \max_{\xi^s \in \Xi} \frac{n^s}{N} \left(-\beta \|\xi^s - \hat{\xi}^s\|_1 + Q(x, \xi^s) \right) \right\} \quad (24)$$

We can observe that the first term of the second-stage objective function in (24), i.e., $-\beta \|\xi^s - \hat{\xi}^s\|_1$, could lead to a sparse solution [31]. The sparsity effect can be illustrated by Fig. 1.

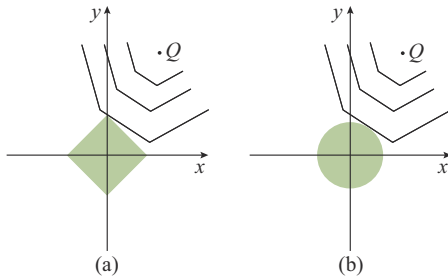


Fig. 1. Illustration of sparsity effect. (a) Graph of an L_1 -norm case. (b) Graph of an L_2 -norm case.

First, the main component of the second-stage problem of (24) is written as $\max_{\xi^s \in \Xi, \|\xi^s - \hat{\xi}^s\|_1 \leq \psi} Q(x, \xi^s)$ with a suitable constant ψ [32]. Then, the contour of $Q(x, \xi^s)$, i.e., a piecewise affine function of ξ^s , is plotted in a coordinate plane on each sub-figure of Fig. 1. Also, the shaded areas representing the constant ψ are plotted. Figure 1(a) corresponds to the L_1 -norm case, where the shaded area is a rectangular shape. Since the rectangle is sharp, the optimal solution is much

more likely to be attained at the corner. In this case, the optimal ξ^s takes zero at the first coordinate, hence it is a sparse solution. In other cases, like the L_2 -norm one shown by Fig. 1(b), a sparse solution can not be easily attained.

Basically, the sparsity effect becomes more salient as the regularization parameter β increases. When β takes zero, the solution sparsity vanishes. It can be observed from the second constraint of (20) that when the dual variable β takes zero, the distance constraint is loose, so the extreme event can be chosen more casually. We mention that the sparsity effect or the L_1 -regularization has been studied in the random convex programming. For example, [33] uses the L_1 -regularization to alleviate a critical obstacle that the number of samples needed may dramatically increase in order to achieve the solution robustness.

The sparsity effect could reduce the number of extreme points needed to characterize the continuous support, so one can expect the outer loop (main loop) of nested C&CG method to converge fast. Hereinafter, the standard nested C&CG method that solves (24) is shown in Algorithm 1. Since the nested C&CG method has not yet been implemented into a DRO-based NCUC problem before, especially one that uses the Wasserstein distance, Algorithm 1 is detailed in this paper.

The main loop of Algorithm 1 goes from Line 3 to Line 16, which iteratively solves a master problem with augmented extreme events, and a sub-problem that generates a new extreme event regarding each sample ξ^s . Line 5 to Line 14 elaborate the procedure of solving the sub-problem, which is known as the inner-level C&CG algorithm [16]. Note that the sub-problem has S parallelizable instances, each of which corresponds to a data point that constructs the ambiguity set. To solve each instance of the sub-problem, a bilinear program (BLP) will be solved repeatedly given some recourse decisions from quick-start generation units, and the critical wind power generation event yielded from the BLP will be fed to an MILP to explore a new recourse decision. When there is no better recourse decision, the inner-level C&CG procedure terminates and the latest extreme event is augmented to the master problem of the next main loop. Algorithm 1 terminates when the gap between the lower and upper bounds is closed, and it returns a day-ahead UC decision for slow-acting generation units.

It should be noted that the master problem makes use of the primal formulation of $Q(x, \xi_j^{s*})$, which is similar to (18). Due to the “ \geq ” relationship, the minimization operator in $Q(x, \xi_j^{s*})$ can be discarded, so the master problem is an MILP. It is also worth mentioning that the empirical distribution is included in the master problem, i.e., $\xi_1^{s*} = \hat{\xi}^s$. The empirical distribution is necessary to ensure the feasibility of the master problem. As for the BLP, the dual formulation of $Q(x, \xi^s)$ is adopted, where λ , μ , and ν are dual variables concerning constraints (14)-(16), respectively.

C. Solution Framework 2: Incorporating Extremal Distributions

When (21) is addressed by refining the second-stage value function $Q(x, \xi)$, as reported in [7], [8], the resulting prob-

lem cannot be efficiently solved by delayed constraint generation if we combine the first-stage problem together. To tackle such an issue, [7] and [8] propose an efficient approach, which updates the first-stage UC decision in a separate routine via extremal distribution generation (EDG). Given a first-stage decision in each main loop, the EDG procedure will simultaneously identify a set of events that weight more than zero in the worst-case distribution, and this is where the gain of efficiency mainly relies [7], [8]. Facing this unexplored DRO-based NCUC problem, it is of interest to compare the performance of an EDG-based algorithm with that of the classic nested C&CG method. Hence, we are about to derive the EDG-based algorithm for (17).

Algorithm 1: nested C&CG method for DRO-based NCUC problem

- 1: Choose a convergence tolerance ϵ ; denote \mathbf{x}^* as the optimum of a variable \mathbf{x} ; let $UB = \infty$, $LB = -\infty$, $J = 1$, $K = 0$, $\xi_1^{ss} = \hat{\xi}^s$; let \mathbf{z}_0^s be an initial value of \mathbf{z}^s
 - 2: **while** $(UB - LB)/UB > \epsilon$, **do**
 - 3: Solve the master problem

$$\begin{cases} \underline{g} = \min_{\mathbf{x} \in \mathcal{X}, \beta \geq 0, \alpha^s} \mathbf{c}_x^T \mathbf{x} + \epsilon \beta + \sum_{s=1}^S \alpha^s \\ \text{s.t. } \alpha^s \geq \frac{n^s}{N} \left(-\beta \left\| \xi_j^{ss} - \hat{\xi}^s \right\|_1 + Q(\mathbf{x}, \xi_j^{ss}) \right) \\ s = 1, 2, \dots, S, j = 1, 2, \dots, J \end{cases}$$
 - 4: $J \leftarrow J + 1$, $\mathbf{x}_J^* \leftarrow \mathbf{x}^*$, $\beta_J^* \leftarrow \beta^*$, $LB \leftarrow \underline{g}$
 - 5: **for** $s = 1, 2, \dots, S$ **do**
 - 6: A subroutine to obtain an extreme value of ξ^s with \mathbf{x}_J^* and β_J^*
 - 7: **repeat**
 - 8: Obtain an extremum of ξ^s by solving

$$\begin{cases} \max_{\xi^s \in \Xi, (\lambda, \mu, \mathbf{v}), \gamma^s} \gamma^s \\ \text{s.t. } \gamma^s \leq \frac{n^s}{N} \left[-\beta_J^* \left\| \xi^s - \hat{\xi}^s \right\|_1 + \mathbf{c}_z^T \mathbf{z}_k^{ss} + \lambda^T \mathbf{h} + \mu^T (\mathbf{d} - \mathbf{A} \mathbf{x}_J^* - \mathbf{B} \mathbf{z}_k^{ss}) + \mathbf{v}^T (\mathbf{w} - \mathbf{V} \xi_k^{ss}) \right] \\ \mathbf{c}_y^T - \lambda^T \mathbf{H} - \mu^T \mathbf{C} - \mathbf{v}^T \mathbf{U} = \mathbf{0} \end{cases}$$
 - 9: $K \leftarrow K + 1$, $\xi_K^{ss} \leftarrow \xi^{ss}$
 - 10: Obtain a recourse of \mathbf{z}^s by solving

$$\begin{cases} \min_{\mathbf{z}^s, \mathbf{y}} \{ \mathbf{c}_z^T \mathbf{z}^s + \mathbf{c}_y^T \mathbf{y} \} \\ \text{s.t. (13)-(15)} \\ \mathbf{U} \mathbf{y} + \mathbf{V} \xi_K^{ss} = \mathbf{w} \end{cases}$$
 - 11: $K \leftarrow K + 1$, $\mathbf{z}_K^{ss} \leftarrow \mathbf{z}^{ss}$
 - 12: **until** \mathbf{z}^{ss} duplicates \mathbf{z}_k^{ss} for $k \in [1, K - 1]$
 - 13: $K \leftarrow 0$
 - 14: **end for**
 - 15: $UB \leftarrow \mathbf{c}_x^T \mathbf{x}_J^* + \epsilon \beta_J^* + \sum_{s=1}^S \gamma^{ss}$
 - 16: **end while**
 - 17: Return \mathbf{x}_J^*
-

Herein, in order to implement the EDG, (22) will be solved instead of (24), i.e., the variable \mathbf{x} is fixed and the first-stage problem is bypassed at this stage. Then, as (22) is solved, the byproduct (23) is dualized again into:

$$\begin{cases} Z^{\text{P.FINT}}(\mathbf{x}) = \max_{m_{\xi_j^{ss}} \geq 0} \sum_{s=1}^S \sum_{j=1}^J \frac{n^s}{N} Q^*(\mathbf{x}, \xi_j^{ss}) m_{\xi_j^{ss}} \\ \text{s.t. } \sum_{j=1}^J m_{\xi_j^{ss}} = 1 \quad s = 1, 2, \dots, S \\ \sum_{s=1}^S \sum_{j=1}^J \frac{n^s}{N} \left\| \xi_j^{ss} - \hat{\xi}^s \right\|_1 m_{\xi_j^{ss}} \leq \epsilon \end{cases} \quad (25)$$

where $Q^*(\mathbf{x}, \xi_j^{ss})$ is the optimal value of $Q(\mathbf{x}, \xi_j^{ss})$, i.e., a scalar; and $m_{\xi_j^{ss}}$ is the conditional probability of the event ξ_j^{ss} . From (20) and (25), it has been evident that the set of optimal values $m_{\xi_j^{ss}}$ establishes a discrete case of the distribution \mathbb{P} , i.e., $\mathbb{P}^* = \sum_{s=1}^S \sum_{j=1}^J \frac{n^s}{N} m_{\xi_j^{ss}}^*$. Please refer to [7] and [8] for the detailed procedure of recovering a discrete distribution.

The distribution \mathbb{P}^* is an extremal distribution as it attains the maximum of the primal semi-infinite program (20). This can be observed from:

$$Z^{\text{P.INF}}(\mathbf{x}) = Z^{\text{D.INF}}(\mathbf{x}) = Z^{\text{D.FINT}}(\mathbf{x}) = Z^{\text{P.FINT}}(\mathbf{x}) \quad (26)$$

where the first and third equalities hold due to strong duality, and the second equality holds because the nested C&CG method solves (22) globally.

The extremal distribution introduced above can be used to update the first-stage decision. As a result, we have a variant of the solution framework, denoted as Algorithm 2.

Algorithm 2: EDG method with nested C&CG subroutine

- 1: Choose a convergence tolerance ϵ ; denote \mathbf{x}^* as the optimum of a variable \mathbf{x} ; let $UB = \infty$, $LB = -\infty$, $I = 1$, $J_1 = 1$, $\xi_{1,1}^{ss} = \hat{\xi}^s$, and $m_{\xi_{1,1}^{ss}}^* = 1$.
 - 2: **while** $(UB - LB)/UB > \epsilon$ **do**
 - 3: Solve the master problem:

$$\begin{cases} \underline{g} = \min_{\mathbf{x} \in \mathcal{X}, \eta \geq 0} \mathbf{c}_x^T \mathbf{x} + \eta \\ \text{s.t. } \eta \geq \sum_{s=1}^S \sum_{j=1}^{J_I} \frac{n^s}{N} m_{\xi_{ij}^{ss}}^* Q(\mathbf{x}, \xi_{ij}^{ss}) \quad i = 1, 2, \dots, I \end{cases}$$
 - 4: $I \leftarrow I + 1$, $\mathbf{x}_I^* \leftarrow \mathbf{x}^*$, $LB \leftarrow \underline{g}$, $J^I \leftarrow 0$
 - 5: Solve (22) with \mathbf{x}_I^* to obtain $\xi_{I,J}^{ss}$
 - 6: Solve (25) to extract $m_{\xi_{I,J}^{ss}}^*$
 - 7: $UB \leftarrow \mathbf{c}_x^T \mathbf{x}_I^* + Z^{\text{D.FINT}}(\mathbf{x}_I^*)$, $J_I \leftarrow$ number of $\xi_{I,J}^{ss}$
 - 8: **end while**
 - 9: Return \mathbf{x}_I^*
-

In Algorithm 2, the main loop from line 3 to line 8 is the EDG procedure, in which the master problem utilizes the primal formulation of the extremal distributions, i.e., the weighted sum of $Q(\mathbf{x}, \xi_{i,j}^{ss})$. The subroutine in line 5 invokes the nested C&CG method to refine the discrete support and yields (23) for establishing an extremal distribution. Note that in line 5, (22) is solved in a fashion similar to Algorithm 1, excepting that the first-stage variable \mathbf{x} will be fixed as \mathbf{x}_I^* . The distinct difference between these two algorithms is that Algorithm 2 adds multiple support points in each main loop (with nonzero probability certificated by the extremal distribution) to the master problem, instead of adding one support point as performed by Algorithm 1.

IV. CASE STUDIES

Numerical experiments are carried out in this section on two modified IEEE systems to validate the proposed model and the solution methods. The main information of the test systems is reported in Table I. With some mild modifications, the 6-bus system contains one 100 MW quick-start generation unit, while the 118-bus system contains seven quick-start generation units with relatively smaller capacities. The minimum up and down time of quick-start generation

units is set to be one hour without loss of generality. The locations of these units have been reported for reproducibility. As can be observed from Table I, wind farms provide 25%

and 20% of the total generation capacity in the 6-bus system and the 118-bus system, respectively.

TABLE I
CONFIGURATION OF TEST SYSTEMS

System	All units		Quick-start generation units			Wind farms		
	Number	Capacity (MW)	Number	Capacity (MW)	Location (bus No.)	Number	Capacity (MW)	Location (bus No.)
6-bus	4	630	1	100	6	1	210	6
118-bus	54	7220	7	270	4, 6, 90, 91, 105, 107	3	1805	25, 37, 66

The whole data over a three-year time span are divided into two sets. The first set, namely the in-sample data set, is reduced to an empirical distribution [7]. The rest is used for simulations, which is the out-of-sample data set. In Fig. 2, we can visualize the load demand profile, the empirical distribution of wind power (with 5 events as depicted by five-black lines in Fig. 2, or $S=5$) as well as the out-of-sample data set for the 6-bus system.

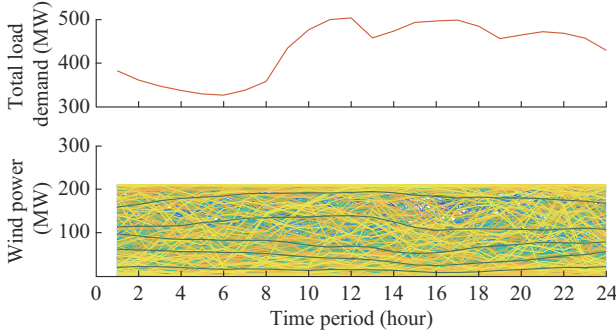


Fig. 2. Profiles of total load demand and wind power.

The models and algorithms are implemented in MATLAB using the YALMIP toolbox. MILPs and linear programs (LPs) are solved with GUROBI 9 on an Intel i5-8250U CPU personal computer running at 3 GHz. The BLPs are handled by sequentially solving some LPs with multiple initial points [8]. The optimality gaps are set to be 10^{-4} and 10^{-3} for the 6-bus system and the 118-bus system, respectively. Also, only the 14 transmission lines connected with wind farms are imposed with power flow limits in the 118-bus system.

A. Cost Efficiency

In this subsection, we compare the out-of-sample cost of the proposed DRO-based NCUC model in (3), which is denoted as Model 1, with those of other three models (Models 2-4). Model 2 is a simplified DRO model, which replaces the mixed-integer recourse problem with an LP by fixing the intraday statuses of quick-start generation units as optimized day-ahead statuses [1]. Definitely, Model 2 is less complex, whose optimal solution can be efficiently computed using the method proposed in [7]. For both Model 1 and Model 2, the historical scenarios are aggregated to obtain an empirical distribution with 5 events [7], [20]. Besides, we denote the per-unit distance parameter as $\varepsilon^* = \varepsilon / \|\xi^{\max} - \xi^{\min}\|_1$ and employ the cross-validation to select a best value ε^* out of a se-

ries of candidates for each model [7], [27].

In order to demonstrate the advantage of the DRO method, a data-driven RO method is also tested on our problem. As suggested by [34], a data-driven uncertainty set represented by the convex hull of the in-sample data is used. The model equipped with such an uncertainty set is denoted as Model 3.

It is also of interest to evaluate how much is lost by making the decision upon a set of ambiguous distributions instead of a perfect one. In this light, another model equipped with a perfect distribution, i.e., the out-of-sample data set, is solved. The model is denoted as Model 4, and its out-of-sample cost would be in accordance with the scheduling cost since perfect information of wind power has been assumed. The main features of the above-mentioned optimization models are collected in Table II. Note that whether the quick-start generation units allowed to be started up/shut down in intraday operation are defined in the sense of decision making, whereas in the simulation phase, they are freely adjustable according to the technique advantage.

TABLE II
DEFINITION OF OPTIMIZATION MODELS

Model	Status of quick-start generation units	Wind power information
1	Adjustable (intraday)	Ambiguous distribution
2	Determined (day-ahead)	Ambiguous distribution
3	Adjustable (intraday)	Data-driven uncertainty set
4	Adjustable (intraday)	Explicitly known

Since these models are mainly used to determine the on/off statuses for slow-acting generation units, the day-ahead UC schedules of 6-bus and 108-bus systems are presented by polar plots, as shown in Fig. 3 and Fig. 4, respectively. The polar plot with 0-1 values naturally visualizes the off/on statuses in a compact manner. Each polar plot starts at the rightmost at Hour 1 of the first generation unit G1, and ends at Hour 24 of the last generator. We set the perfect UC solution derived by Model 4 as the benchmark, which is shown in Figs. 3(d) and 4(d). The UC schedules derived from other three models are presented in independent polar plots, with colored boxes on the circumference of the circle marking the entries, where the on/off statuses are distinct from those of Model 4.

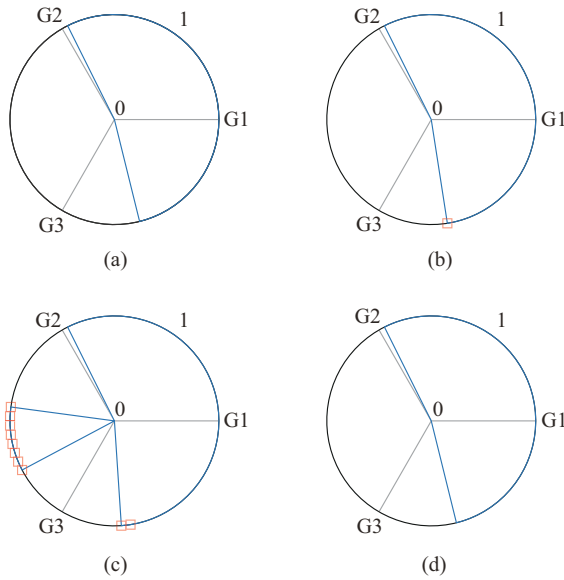


Fig. 3. Day-ahead UC schedules with different models for a problem instance in 6-bus system (each circle contains statuses of three generators). (a) Model 1. (b) Model 2. (c) Model 3. (d) Model 4.

As can be observed from Fig. 3, in the 6-bus system, Model 1 achieves a same day-ahead UC solution as Model 4, and Model 2 achieves a slightly different solution, which turns on G3 at Hour 8 instead of Hour 9. As optimized by Model 3, however, G2 is scheduled on from Hour 11 to Hour 18, and G3 is turned on even earlier at Hour 7. In the 118-bus system, the solutions of Model 1 and Model 2 are quite different from the perfect solution. The UC schedule derived by Model 3 is distinctly different from the others, which schedules more units on with frequent start-up and shut-down and seems more conservative.

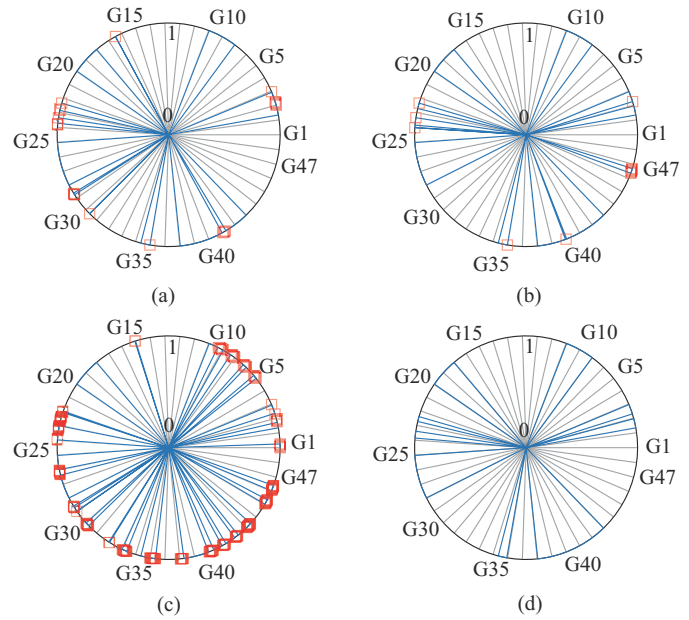


Fig. 4. Day-ahead UC schedules with different models for a problem instance in 118-bus system (each circle contains statuses of 47 generators). (a) Model 1. (b) Model 2. (c) Model 3. (d) Model 4.

Table III summarizes some important indices of these models, including the distance parameters ε^* used by Model 1 and Model 2, the scheduling cost and out-of-sample cost as well as the wind curtailment and load shedding. It can be observed that the best distance parameter ε^* basically varies with the model formulation and the system configuration. In Table III, the scheduling costs are derived from the optimization models, while the out-of-sample costs are yielded from simulations with a UC solution, and they reflect the “real” expected costs in real-time operations.

TABLE III
COST EFFICIENCY AND ROBUSTNESS OF DIFFERENT MODELS

System	Model	ε^* (p.u.)	Scheduling cost (k\$)	Out-of-sample cost (k\$)	Wind curtailment (MWh)		Load shedding (MWh)		Slack power (MWh)	
					Mean value	The maximum value	Mean value	The maximum value	Mean value	The maximum value
6-bus	1	10^{-3}	96.45	103.98	5.11	65.82	0.23	1.85	0	0
	2	$10^{-1.5}$	107.53	104.57	7.69	95.31	0.23	1.85	0	0
	3		126.52	105.94	8.98	119.51	0.01	0.40	0	0
	4		103.98	103.98	5.11	65.82	0.23	1.85	0	0
118-bus	1	$10^{-1.5}$	944.97	974.66	0.82	18.20	0	0	0	0
	2	$10^{-2.25}$	931.95	975.75	0.82	18.20	0	0	0	0
	3		1157.32	985.41	54.76	628.91	0	0	0	0
	4		974.52	974.52	0.82	18.20	0	0.14	0	0

The out-of-sample costs obtained from Model 1 are lower than those from Model 2 in both systems, while the expected amount of wind curtailment and load shedding obtained from Model 1 are smaller than or equal to those from Model 2. This demonstrates that it is of high necessity to precisely account for the discrete behavior of quick-start generation units. Regarding Model 3, even though the robust solutions provide UC schedules with the least load shedding, the solu-

tions are least favorable in two systems in terms of both the cost efficiency and the wind curtailment. It is worth noting that the performance of Model 3 is not improved and even degraded when the convex hull is constructed with partial in-sample data. According to the scheduling costs of Model 3, it is asserted that the data-driven uncertainty set is actually quite conservative. This phenomenon can be explained by the so-called curse of dimensionality: in high dimensions,

most of the samples are concentrated near the edge of a hyperrectangle [21], and thus the convex hull of in-sample data could be unsuitable for robust optimization. The situation is quite different in Model 1. Even though Model 1 protects the system against all scenarios contained in the full hyperrectangular support, the worst-case distribution approach that it takes has made the UC schedule less conservative.

We further analyze the gap between Model 1 and Model 4. The out-of-sample cost of Model 1 is equivalent with that of Model 4 in the 6-bus system, which coincides with the UC results shown in Fig. 3. As for the 118-bus system, although the UC solution of Model 1 is quite different from that of Model 4, the out-of-sample cost is extremely close to the cost under perfect information, i. e., \$974660 v. s. \$974520. Basically, this attractive performance is due to the precise modeling, that is, the intraday start-up/shut-down behavior of quick-start generation units has been considered in the scheduling phase. The performance should also be attributed to the DRO approach with a fine-tuned distance parameter, which renders the system being not only robust against extreme random events, bus also cost-effective under realistic distributions.

Aside from the day-ahead UC schedules, the intraday dispatch schemes also manifest the performance of decision-making models. In Fig. 5, the intraday adjustments of on/off statuses for quick-start generation units in the 118-bus system under the day-ahead UC schedules determined with different models are compared, and for conciseness, only the result on a typical day with the maximum load shedding is presented.

It can be observed from Fig. 5(a) that the real-time power outputs of three wind farms are simultaneously low during the second half of Day 671. As a result, the quick-start generation units are started up within this time period to support the power balance, i. e., the quick-start generation unit G1 of Model 2 under the day-ahead UC schedule, all quick-start generation units of Model 3, and the quick-start generation units G4 and G6 of Model 4. It is noticed that no quick-start generation units are invoked on this simulating day, given the day-ahead UC schedule determined by the proposed DRO-based model. The operating costs are also reported on this day, as shown in Fig. 5(b). The cost of Model 1 is exactly the lowest.

B. Performance of Solution Methods

To analyze the performance of the solution methods, we start with an extremal distribution yielded from the 6-bus system. As shown in Fig. 6(a), the upper panel shows the empirical distribution, while the lower panel shows the extremal one. In the extreme distribution, the original Event 5 in the upper panel is split into two events, denoted as Event 5 and Event 6 with the possibilities of 69% and 31%, respectively (i. e., extreme points $\xi_{i,1}^{5*}$ and $\xi_{i,2}^{5*}$), so the number of extreme points needed to equivalently represent the constraints in (21) is quite small indeed. We can observe that the sparsity of extremal distributions always happens in the test cases. Moreover, the number of events in an extremal distribution is highly related to that of the empirical one, i. e., $\hat{\mathbb{P}}$. This

suggests that the size of the master problems is controllable. If the ambiguity set in (19) is constructed with a highly aggregated distribution $\hat{\mathbb{P}}$, the size of the MILP master problems in Algorithm 2 will be effectively reduced. When using a highly aggregated empirical distribution, unlike the stochastic programming approach, the DRO-based model can still hedge the bias and shift of probability distributions, through protecting the system against all wind generation events within the support \mathcal{E} (the robustness), and quantitatively finding a solution regarding a set of distributions centered at the empirical one from a suitable distance ε^* (the cost efficiency).

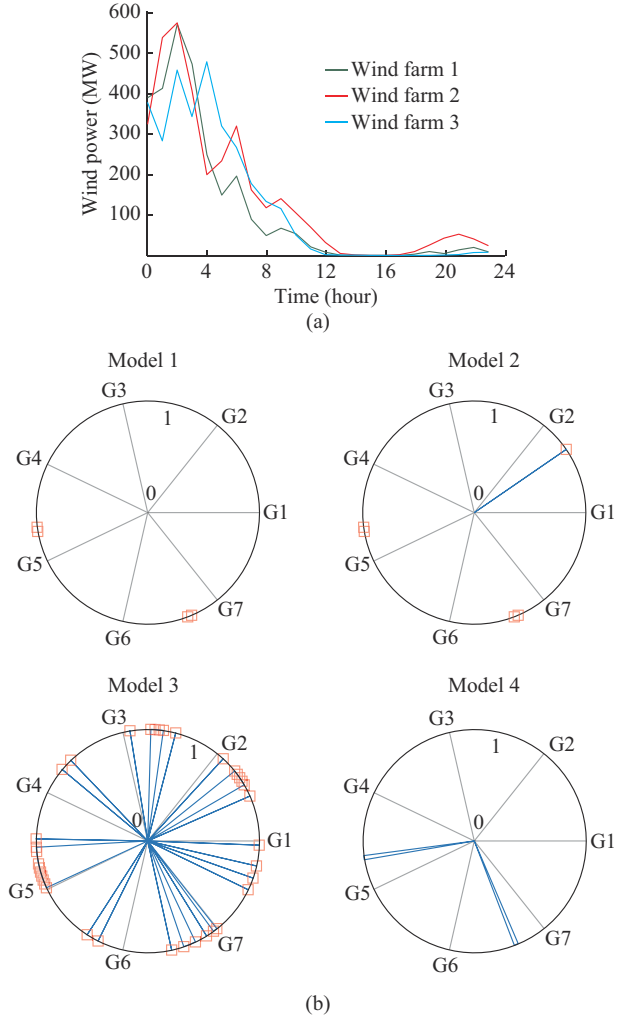


Fig. 5. Simulation results of 118-bus system on with the maximum load shedding. (a) Real-time wind power. (b) Intraday statuses of quick-start generation units.

The L_1 -norm also has a significant effect on the events. The entries of ξ_j^{s*} in Fig. 6(a) mostly take the same values as those of $\hat{\xi}^s$, i. e., the wind power is perturbed in only a few hours. Since the vector $\xi - \hat{\xi}^s$ is sparse, reasonably, the set of extreme points for the final master problem in Algorithm 2 will be small-sized as well. The sparsity of events could alleviate the computational burden as it reduces the scale of the master problems and the number of main loops.

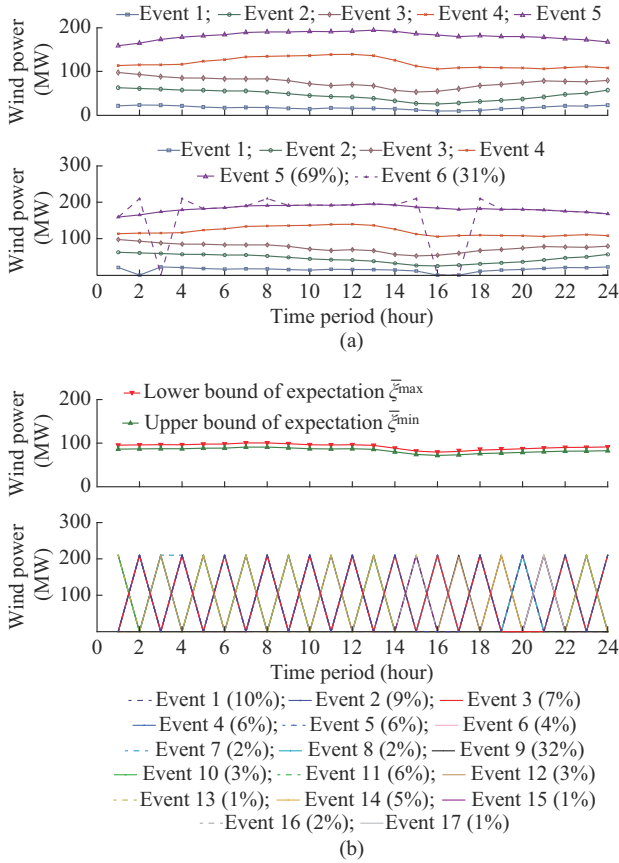


Fig. 6. Visualization of empirical and extremal distributions. (a) Events in empirical distribution (upper panel) and extremal distribution (lower panel) of L_1 -norm Wasserstein ambiguity set. (b) Expectation band (upper panel) and extremal distribution (lower panel) of moment-based ambiguity set.

To better demonstrate the sparsity of L_1 -norm Wasserstein ambiguity set, an extremal distribution obtained from a first-moment ambiguity set is presented for comparison. The first-moment ambiguity set contains all distributions that share a predefined expectation (band) [8], [13], i.e., $\mathbb{P} =$

$\{\mathbb{P} \in \mathcal{P}_0(\Xi) \mid \xi^{\min} \leq \mathbb{E}_{\mathbb{P}}(\xi) \leq \xi^{\max}\}$. The result is visualized in Fig. 6(b), where the lower panel shows that the extremal distribution has split the expectation band into 17 events, the probabilities of which add up to 1. Typically, the extremal distribution yielded from the first-moment-based ambiguity set is a dense one with many ramp events. The extremal distribution could become even denser if the expectation band is squeezed to a profile. It can be observed that the dense distributions make the DRO-based NCUC problem much more computationally expensive.

As illustrated in Fig. 1, the sparsity of solutions is mainly due to the L_1 -norm design in the Wasserstein distance, so other models such as the moment-based distributionally robust NCUC models studied in [8] can hardly achieve a sparse solution. Please also see Fig. 3 in [8] for the extremal distribution of a second-moment-based case.

In order to figure out which solution framework is more suitable, several instances of the distributionally robust NCUC problem (17) are solved for each test system. Basically, Algorithm 2 takes fewer main loops to converge in both test systems, as shown in Table IV. The number of main loops in 6-bus and 118-bus systems are summarized by the range, mean, and standard deviation (std.), where the mean and std. are fixed to integers. The results validate that the cutting plane provided by the extremal distribution is stronger. As shown in Figs. 7 and 8, the strong cutting plane is illustrated. Although both algorithms have a high convergence rate, Algorithm 2 always achieves a smaller gap at the second iteration, as shown by Fig. 7(b) and Fig. 8(b). It is also noted that Algorithm 2 returns a smaller upper bound at the first iteration, which is due to the aforementioned setup, i.e., the master problem has to include the empirical distribution to ensure the feasibility when using Algorithm 2. Note that the results delivered by these algorithms are highly matched given the convergence errors. Although the sub-optimality induced by the way of solving the BLPs may affect the consistency of UC schedules obtained from two algorithms, the deviation is rather small in our numerical experiments.

TABLE IV
PERFORMANCES OF DIFFERENT SOLUTION ALGORITHMS IN 6-AND 118-BUS SYSTEMS

Algorithm	Item	6-bus system			118-bus system		
		Number of main loops	Number of solutions	Runtime (s)	Number of main loops	Number of solutions	Runtime (s)
Algorithm 1	Range	[2, 21]	[2, 9]	[8, 2090]	[2, 8]	[2, 7]	[161, 4315]
	Mean	6	4	234	4	3	1014
	Std.	4	2	461	2	1	1208
Algorithm 2	Range	[2, 18]	[2, 7]	[20, 1876]	[2, 5]	[2, 3]	[110, 2220]
	Mean	5	3	282	3	2	835
	Std.	4	2	510	1	0	698

Other main indices, including the number of solutions generated by the inner C&CG subroutine and the total runtime spent in solving a problem instance, are shown in Table IV. The nested C&CG method in two algorithms handles problems in different structures with the first-stage decision variable x unfixed or fixed, respectively. Therefore, the average numbers of solutions rendered are different. The runtime

should be the most crucial index. Concerning the problem addressed in this paper, it is found that the strong cutting plane unnecessarily leads to less runtime in total. Since each main loop of Algorithm 2 involves S nested C&CG subroutines that solve the second-stage problem, more computational effort is spent for many cases. Due to the extra burden involved in solving the second-stage problem, the average run-

time turns out to be longer in the 6-bus system. In the system where the runtime consumed by solving the master problem dominates, namely the 118-bus system, the advantage of Algorithm 2 reveals due to the time saved in solving fewer master problems. Since the real-world TCUC problems are mostly large-scale, an algorithm is suggested with stronger cutting plane and less main loops, hence Algorithm 2 is chosen. In specific cases where Algorithm 2 has been proven to be more time-consuming, it is necessary to introduce the EDG procedure to this problem only if the system operators care more about the extremal distributions of wind power rather than the solution time.

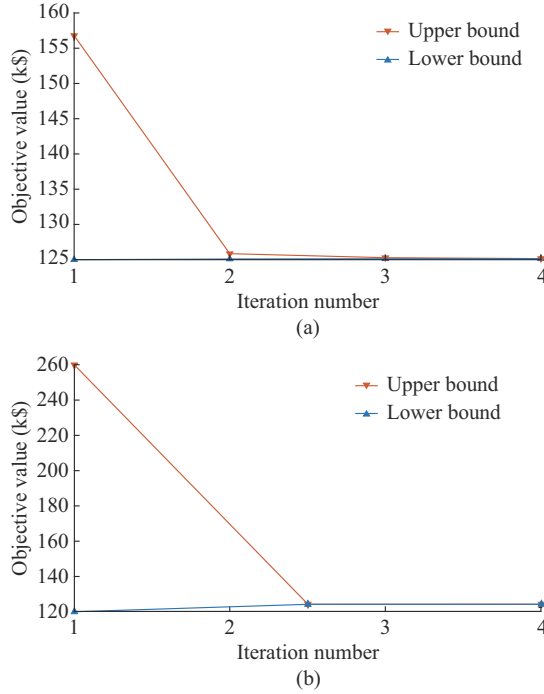


Fig. 7. Convergence profiles of two algorithms on 6-bus system. (a) Algorithm 1. (b) Algorithm 2.

To sum up, the sparse solution approach is beneficial in that the number of iterations as well as the size of problems solved within each iteration could be effectively reduced. However, it does not mean that any DRO-based NCUC problem can be solved efficiently, e.g., within a 2-hour market clearing time window. There are two important noteworthy facts: ① the complexity of the DRO-based NCUC problem is basically dictated by the deterministic counterpart; ② the complexity can be influenced by many other factors, which are hard to quantify, such as the distance parameter, the share of wind power capacity, and the penalty factor of wind curtailment.

Finally, we would like to mention that although the EDG framework has been reported in a recent paper dealing with the L_1 -norm Wasserstein distributionally robust UC problem [7], some novel contributions make this paper distinctly different from [7]. First, the model herein considers the intraday start-up/shut-down behavior of quick-start generation units. Second, in order to solve the complicated DRO-based NCUC problem with mixed-integer recourse, the solution ap-

proach is quite different in how the second-stage problem is handled. Finally, this is the first time that the sparse solution resulting from the L_1 -norm DRO modeling in NCUC problems has been revealed.

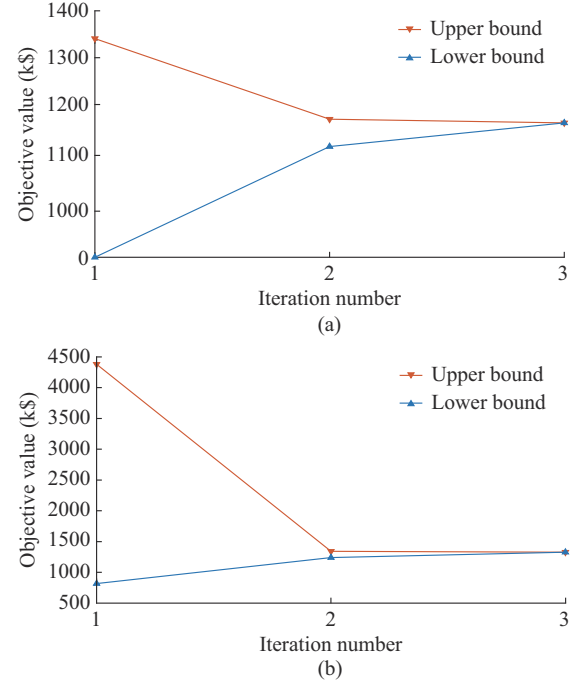


Fig. 8. Convergence profiles of two algorithms on 118-bus system. (a) Algorithm 1. (b) Algorithm 2.

V. CONCLUSIONS

The Wasserstein-metric-based distributionally robust NCUC problem with quick-start generation units can be efficiently solved by leveraging the nested C&CG method. Due to the L_1 -norm design, the infinitely dimensional optimization problem could be represented with a reasonable number of extreme events, or several sparse extremal distributions. It has been observed that the events identified by both algorithms are analogous to the events of the empirical distribution that constructs the ambiguity set, and specifically, they are the variants obtained by perturbing a few values. Although the extreme events and the extremal distributions are sparse, they are more representative than a prior discrete supports, and by construction they render the same degree of robustness as the original continuous support does. This sparse effect results in a reduction in both the problem size and the iteration number, and therefore allows the nested C&CG method to perform well on solving the complex problem. The sparse solution approach is believed to facilitate the DRO method in real-world applications, as the concern of problem tractability can be alleviated. Nevertheless, the runtime of algorithms can be affected by the problem data, such as the distance parameter, wind power data and penalty factor of wind curtailment.

This paper shows that the cost efficiency of the UC solution can be significantly increased by precisely considering the intraday start-up/shut-down behavior of quick-start generation units. Under the proposed modeling and solution frame-

work, the UC solution yielded is not only robust, but also cost-effective under realistic situations. However, it has also been demonstrated that the good out-of-sample performance of the Wasserstein-metric-based distributionally robust NCUC model depends on a fine-tuned distance parameter, which generally varies with the configuration and operating condition of a system, and is time-consuming to obtain.

Regarding the aforementioned limitation, a meaningful focus in the future work is thus to develop a quick estimation of the best distance parameter. Statistical learning methods could be promising approaches to making the quick estimation. Some aspects on modeling are also of interest to study such as addressing the nonparticipative issue in the DRO-based NCUC model, where integer recourse variables are time-coupled, and applying the proposed solution framework to multi-stage distributionally robust NCUC problems.

REFERENCES

- [1] B. Hu and L. Wu, "Robust SCUC considering continuous/discrete uncertainties and quick-start units: a two-stage robust optimization with mixed-integer recourse," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1407-1419, Mar. 2016.
- [2] N. G. Cobos, J. M. Arroyo, N. Alguacil *et al.*, "Robust energy and reserve scheduling considering bulk energy storage units and wind uncertainty," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5206-5216, Sept. 2018.
- [3] N. G. Cobos, J. M. Arroyo, N. Alguacil *et al.*, "Robust energy and reserve scheduling under wind uncertainty considering fast-acting generators," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 4, pp. 2142-2151, Oct. 2019.
- [4] N. Jia, C. Wang, W. Wei *et al.*, "Decentralized Robust energy management of multi-area integrated electricity-gas systems," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 6, pp. 1478-1489, Nov. 2021.
- [5] H. Chen, P. Xuan, Y. Wang *et al.*, "Key technologies for integration of multitype renewable energy sources—research on multi-timeframe robust scheduling/dispatch," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 471-480, Jan. 2016.
- [6] C. Liu, C. Lee, H. Chen *et al.*, "Stochastic robust mathematical programming model for power system optimization," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 821-822, Jan. 2016.
- [7] X. Zheng and H. Chen, "Data-driven distributionally robust unit commitment with Wasserstein metric: tractable formulation and efficient solution method," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4940-4943, Nov. 2020.
- [8] X. Zheng, K. Qu, J. Lv *et al.*, "Addressing the conditional and correlated wind power forecast errors in unit commitment by distributionally robust optimization," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 944-954, Apr. 2021.
- [9] C. Zhao and Y. Guan, "Data-driven stochastic unit commitment for integrating wind generation," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2587-2596, Jul. 2015.
- [10] C. Duan, L. Jiang, W. Fang *et al.*, "Data-driven affinely adjustable distributionally robust unit commitment," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1385-1398, Mar. 2017.
- [11] Y. Chen, Q. Guo, H. Sun *et al.*, "A distributionally robust optimization model for unit commitment based on Kullback-Leibler divergence," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5147-5160, Sept. 2018.
- [12] R. Zhu, H. Wei, and X. Bai, "Wasserstein metric based distributionally robust approximate framework for unit commitment," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2991-3001, Jul. 2019.
- [13] P. Xiong, P. Jirutitijaroen, and C. Singh, "A distributionally robust optimization model for unit commitment considering uncertain wind power generation," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 39-49, Jan. 2017.
- [14] S. Babaei, C. Zhao, and L. Fan, "A data-driven model of virtual powerplants in day-ahead unit commitment," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 5125-5135, Nov. 2019.
- [15] X. Zheng, H. Chen, Y. Xu *et al.*, "A mixed-integer SDP solution approach to distributionally robust unit commitment with second order moment constraints," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 2, pp. 374-383, Jun. 2020.
- [16] L. Zhao and B. Zeng, "An exact algorithm for two-stage robust optimization with mixed integer recourse problems," Technical Report, University of South Florida, 2012.
- [17] F. Pourahmadi, J. Kazempour, C. Ordoudis *et al.*, "Distributionally robust chance-constrained generation expansion planning," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2888-2903, Jul. 2019.
- [18] M. Bansal, K.-L. Huang, and S. Mehrotra, "Decomposition algorithms for two-stage distributionally robust mixed binary programs," *SIAM Journal on Optimization*, vol. 28, no. 3, pp. 2360-2383, Aug. 2018.
- [19] F. Luo and S. Mehrotra, "A decomposition method for distributionally robust two-stage stochastic mixed-integer conic programs," *Mathematical Programming*, vol. 2021, pp. 1-45, Jun. 2021.
- [20] K. Kim, "Dual decomposition of two-stage distributionally robust mixed-integer programming under the Wasserstein ambiguity set," Preprint manuscript, 2020, pp. 1-21.
- [21] J. H. Friedman, "Local methods in high dimensions," in *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer Open, 2017, pp. 22-27.
- [22] I. Rossi, A. Sorce, and A. Traverso, "Gas turbine combined cycle start-up and stress evaluation: a simplified dynamic approach," *Applied Energy*, vol. 190, pp. 880-890, Mar. 2017.
- [23] H. Li, Z. Lu, Y. Qiao *et al.*, "The flexibility test system for studies of variable renewable energy resources," *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 1526-1536, Mar. 2021.
- [24] U.S. Energy Information Administration. (2020, Nov.). About 25% of U.S. power plants can start up within an hour. [Online]. Available: <https://www.eia.gov/todayinenergy/detail.php?id=45956>
- [25] L. Nan, Y. Liu, L. Wu *et al.*, "Graph theory based N-1 transmission contingency selection and its application in security-constrained unit commitment," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 6, pp. 1458-1467, Nov. 2021.
- [26] 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.
- [27] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, no. 1-2, pp. 115-166, Jul. 2018.
- [28] C. Zhao and Y. Guan, "Data-driven risk-averse stochastic optimization with Wasserstein metric," *Operations Research Letters*, vol. 46, no. 2, pp. 262-267, Sept. 2018.
- [29] S. Babaei, R. Jiang, and C. Zhao, "Distributionally robust distribution network configuration under random contingency," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3332-3341, Sept. 2020.
- [30] A. Velloso, D. Pozo, and A. Street, "Distributionally robust transmission expansion planning: a multi-scale uncertainty approach," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3353-3365, Sept. 2020.
- [31] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.
- [32] J. Lv, X. Zheng, M. Pawlak *et al.*, "Very short-term probabilistic wind power prediction using sparse machine learning and nonparametric density estimation algorithms," *Renewable Energy*, vol. 177, pp. 181-192, May 2021.
- [33] M. C. Campi and A. Car'è, "Random convex programs with L_1 -regularization: sparsity and generalization," *SIAM Journal on Control and Optimization*, vol. 51, no. 5, pp. 3532-3557, Sept. 2013.
- [34] A. Velloso, A. Street, D. Pozo *et al.*, "Two-stage robust unit commitment for co-optimized electricity markets: an adaptive data-driven approach for scenario-based uncertainty sets," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 958-969, Apr. 2019.

Xiaodong Zheng received the B.S. degree and the Ph.D. degree from South China University of Technology, Guangzhou, China, in 2015 and 2020, respectively, both in electrical engineering. From 2018 to 2019, he was a Research Assistant with Nanyang Technological University, Singapore. Prior to that, he was a joint Graduate Student with the Power Dispatching and Control Center of China Southern Power Grid Co., Ltd., Guangzhou, China, from 2015 to 2017. He is currently a Postdoctoral Scholar with the Electric Power Research Institute of China Southern Power Grid Co., Ltd., Guangzhou, China, and the School of Electrical Engineering of Xi'an Jiaotong University, Xi'an, China. His research interests include robust optimization, distributed optimization, game theory, statistical learning, and their applica-

tions in power system operation, renewable energy integration, and electricity markets.

Baorong Zhou received the B.S. degree from Wuhan University, Wuhan, China, in 1996, and the Ph.D. degree from Tianjin University, Tianjin, China, in 2004. From 2004 to 2010, he was a Power Planning Engineer with China Southern Power Grid (CSG) Co., Ltd., Guangzhou, China. Since 2011, he has been a Senior Power Expert with the Electric Power Research Institute of CSG. He is a Senior Engineer at Professor Grade. His research interests include power system planning, power system analysis and control, renewable energy resources, etc.

Xiuli Wang received the B.S., M.S., and Ph.D. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1982, 1985, and 1997, respectively. She is currently a Professor with the School of Electrical Engineering, Xi'an Jiaotong University. Her research interests include power system planning, integration of renewable energy, and electricity market.

Bo Zeng received the Ph.D. degree in operation research from the School of Industrial Engineering, Purdue University, West Lafayette, USA, in 2007. He is currently an Associate Professor with the Department of Industrial Engineering, and the Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, USA. He is a Member of the Institute of Industrial and Systems Engineers and the Institute for Operations Research and the Management Sciences. His research interests include theory and algorithm development for stochastic, robust, and multilevel mixed-integer

programs, with applications in power, logistics, healthcare, and cyber-physical systems.

Jizhong Zhu received the B.S., M.S., and Ph.D. degrees in electrical engineering from Chongqing University, Chongqing, China, in 1985, 1987, and 1990, respectively. He is an IEEE Fellow and a Professor at the School of Electric Power Engineering, South China University of Technology, Guangzhou, China. His research interests include power system operation and control, smart grid, microgrid, virtual power plant, electric vehicle, renewable energy application, and integrated energy system.

Haoyong Chen received the B.S., M.S., and Ph.D. degrees in electrical engineering all from Xi'an Jiaotong University, Xi'an, China, in 1995, 1997, and 2000, respectively. He is currently a Professor with the Department of Electrical Engineering, South China University of Technology, Guangzhou, China. He is the Executive Director and Chief Scientist of the Guangdong Key Laboratory of Clean Energy Technology, the National-Local Joint Engineering Laboratory for Wind Power Control and Integration Technology and Guangdong Engineering Laboratory for Wind Power Control and Integration Technology. His current research interests include renewable energy & smart grid, power system planning, operation and control, power economics & electricity market, and integrated energy system, etc.

Waisheng Zheng is currently with China Southern Power Grid Co., Ltd., Guangzhou, China. His research interests include power system planning, scheduling, and smart grid.