Distributionally Robust Unit Commitment with Flexible Generation Resources Considering Renewable Energy Uncertainty

Siyuan Wang, Chaoyue Zhao, *Member, IEEE*, Lei Fan, *Senior Member, IEEE*, Rui Bo, *Senior Member, IEEE*

***Abstract*—As the penetration of intermittent renewable energy increases in bulk power systems, flexible generation resources, such as quick-start gas units, become important tools for system operators to address the power imbalance problem. To better capture their flexibility, we proposed a two-stage distributionally robust unit commitment framework with both regular and flexible generation resources, in which the unit commitment decisions for flexible generation resources can be adjusted in the second stage to accommodate the renewable energy intermittency. In order to tackle this challenging two-stage distributionally robust mixed-binary model, to which traditional separation algorithms won't apply, we designed a revised integer L-shaped algorithm with lift-and-project cutting plane techniques. In comparison to the traditional distributionally robust unit commitment, the proposed approach can reduce the system cost through an improved flexible resource quantification in the modeling.**

***Index Terms*—Unit commitment, renewable energy uncertainty, flexible generation resources, distributionally robust optimization, two-stage mixed-binary linear program, system flexibility.**

NOMENCLATURE

Indices and Sets

t, b, i	Index for time periods, buses, and units/transmission lines.
j	Index for pieces in piece-wise approximation.
$\mathcal{T}, \mathcal{B}, \mathcal{L}$	Set of time periods, buses, and transmission lines.
$\mathcal{W}, \mathcal{D}$	Set of renewable energy resources and loads.
$\mathcal{G}_r, \mathcal{G}_f, \mathcal{G}$	Set of regular, flexible, and all generation resources, i.e., $\mathcal{G} = \mathcal{G}_r \bigcup \mathcal{G}_f$.
$\cdot^b$	Set of devices $\cdot$ at bus b.

Parameters

N_t	Number of time periods, i.e., $N_t = \lvert\mathcal{T}\rvert$.
$D_{i,t}$	Power of load i in time period t.
$SF_{b,i}$	Shift factor that represents power flow change on branch i due to power injection change at bus b.
$\overline{F}_i$	Capacity of transmission line i.
SU_i	Start-up cost for unit i.
J	Number of pieces in piece-wise approximation.
β_i^j, γ_i^j	Per-unit and constant cost terms for unit i in the j-th piece of piecewise linear cost function.
RU_i, RD_i	Upward/downward ramp rate limits for unit i.
$\overline{RU}_i, \overline{RD}_i$	Start-up and shut-down ramp rate limits for unit i.
$\underline{P}_i, \overline{P}_i$	Minimum and maximum outputs for unit i.
UT_i, DT_i	Minimum on and off time periods for unit i.

Decision Variables

$u_{i,t}$	Binary commitment status variable for unit i in time period t.
$v_{i,t}$	Binary variable that indicates if unit i starts up at the beginning of time period t.
$\phi_{i,t}$	Approximated piece-wise linear cost for unit i in time period t.
$p_{i,t}$	Power output from unit i in time period t.
ϑ	Ancillary variables for second-stage objective.

Symbols for Confidence Set

$\xi, \hat{\xi}$	Real and empirical random variables for the day-ahead forecast error of renewable power.
$\mathbb{D}$	Confidence set for day-ahead forecast error distributions of renewable energy resources.
$\mathbb{P}, \hat{\mathbb{P}}$	True and empirical probability distributions for random variables ξ and $\hat{\xi}$, respectively.
$p^m, \hat{p}^n$	Probabilities for scenario m in true distribution $\mathbb{P}$ and scenario n in empirical distribution $\hat{\mathbb{P}}$.
$\mathbb{Q}$	Joint distribution of random variables ξ and $\hat{\xi}$ with marginal distributions $\mathbb{P}$ and $\hat{\mathbb{P}}$.
$q^{n,m}$	Probability in joint distribution $\mathbb{Q}$.
α	Confidence level for confidence set $\mathbb{D}$.
θ	Tolerance level for the distance between random variables ξ and $\hat{\xi}$.
N_h	Size of the historical data.
N	Number of bins.
δ	Diameter of supporting space.
$\mathrm{d}(\cdot,\cdot)$	Function to calculate the distance between random variables.

Compact Representation

y, x	Vector of first-stage and second-stage variables.
σ	Vector of slack variables.
a, b, c, d	Coefficient vectors for abstract formulation.
A, B, C, D	Coefficient matrices for abstract formulation.
$\Psi(\Phi), \omega$	Coefficient matrix or vector for parametric cuts.
μ, λ	Vector of dual variables.

This work was partially supported by National Science Foundation (NSF) under Grants 2045978 and 2046243. Paper no. TPWRS-00145-2021. *(Corresponding Author: Lei Fan.)*

Siyuan Wang and Rui Bo are with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, MO 65409 USA (e-mail: siyuanwang@mst.edu; rbo@mst.edu).

Chaoyue Zhao is with the Department of Industrial & Systems Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: cyzhao@uw.edu).

Lei Fan is with the Department of Engineering Technology, University of Houston, Houston, TX 77204 USA (e-mail: lfan8@central.uh.edu).

This article has been accepted for publication in IEEE Transactions on Power Systems. DOI: 10.1109/TPWRS.2022.3149506.

I. Introduction

THE penetration of renewable energy, typically wind and solar energy, keeps increasing in the power systems. The intermittent and unpredictable nature of renewable energy brings significant challenges to system operators. Flexible generation resources, such as quick-start units, can be quickly turned on to mitigate the shortage of energy supply that is caused by the intermittent renewable energy output in the near real-time. Due to this, more and more flexible generation resources become important tools for operators to enhance power system flexibility and security. Therefore, unit commitment (UC) models need to be robust to manage the uncertainty from renewable energy resources, and be powerful to reflect the capability of flexible generation resources.

To hedge against the risk of the intermittent renewable energy, stochastic programming (SP) [1]-[4] and robust optimization (RO) [5]-[7] approaches have been extensively studied for power system operation and planning applications. However, SP might result in unreliable decisions due to the blind assumption of probability distributions, and RO is too pessimistic because the worst-case scenario is typically unlikely to happen. By being able to efficiently utilize a large amount of historical data and address the probability distribution uncertainty with partial information of it, distributionally robust optimization (DRO) approaches can offer reliable while less conservative decisions, thus have recently been applied to UC problems [8]-[11]. In these DRO based works, the regular and flexible generation resources are treated in the same manner, i.e., the UC decisions for flexible generation resources are modeled as first-stage variables (also known as *here-and-now* variables). In fact, this neglects the flexible start-up and shut-down capabilities of flexible generation resources in near real-time operations, which can potentially cause an increase in cost (as shown in section IV). In this study, we propose a novel model to better quantify the flexibility of flexible generation resources in distributionally robust unit comment problems (denote as '*improved DRUC*' hereafter). The proposed model is formulated as a two-stage distributionally robust model with commitment variables for regular units in the first stage and commitment variables for flexible generation resources in the second stage. In contrast to traditional distributionally robust UC (denote as '*traditional DRUC*' hereafter) models in the literature, our improved DRUC model treats decision variables for flexible generation resources as second-stage variables (also known as *wait-and-see* variables), thus can better characterize the start-up and shut-down flexibility of flexible generation resources and reduce the system cost.

L-shaped method (or Benders decomposition) has been used to solve the two-stage distributionally robust optimization model with continuous recourse decisions in the second stage [8], [9], while we need to address the challenging mixed-binary recourse decisions in our improved DRUC model. To tackle this issue, we deploy the decomposition algorithmic framework proposed by [12], which can finitely converge with moment or Wasserstein metric based confidence sets.

However, in [12], the first-stage mixed-binary linear program (MBLP) is supposed to be relatively small, and thus can be solved purely by cutting plane methods to obtain its linear program (LP) basis matrix. This is numerically difficult for our first-stage UC problem, which contains a large number of commitment decision variables and feasibility cuts. Thus, a revised approach is needed to generate cuts that are valid for any first-stage solution, to convexify the second-stage LP relaxation. We note the parametric cut proposed in [13], which is used to solve stochastic UC with quick-start units, is valid for the second-stage problems given any first-stage solution. Authors of [14] further applied the parametric cut in [13] to a robust UC problem with quick-start units to guarantee the feasibility of second-stage problems. We first use the parametric cut in [13], and found it might stop improving the integrality gap after several iterations, thus cannot guarantee the tightness of the second-stage relaxation. To tackle this, we customize the lift-and-project cut generation approach in [15], [16] for our specific UC problem. In comparison to only using the parametric cut in [13], our cut generation approach can obtain a tighter second-stage LP relaxation.

In addition, the algorithm in [12] requires the recourse problem relatively complete, but first-stage UC decisions from initial iterations are most likely to make the recourse problem infeasible. In this work, we use a scenario filtering based feasibility cut approach to keep the recourse problem feasible.

We summarize our contributions in the following.

- Compared to traditional two-stage DRUC methods, our approach allows commitment decisions for flexible generation resources to be adjustable according to the near real-time realization of renewable energy uncertainty. We find modeling this feature has benefits in reducing expected operation cost, and potentially avoiding infeasibility caused by traditional DRUC modeling when feasible commitment schemes exist for the physical system.
- A revised integer L-shaped algorithm is proposed to solve our formulated two-stage distributionally robust mixed-binary program. In addition to the cutting plane method in [13], a customized lift-and-project cut generation method is used to strengthen the LP relaxation of the second-stage mixed-binary program, and thus enhance the performance of the solution strategy.

II. Problem Formulation

In this section, we introduce our improved DRUC formulation, and define a confidence set for the day-ahead renewable energy forecast error distribution by using Wasserstein metric.

A. Improved DRUC Formulation

In this work, our improved DRUC problem is formulated based on a two-stage distributionally robust optimization framework. UC decisions for regular units are modeled in the first stage, while UC decisions for flexible generation resources and economic dispatch decisions for all units are modeled in the second stage. In contrast to traditional two-stage stochastic programming models, the probability distribution $\mathbb{P}$ of renewable energy output in our model is assumed to be ambiguous and running within confidence set $\mathbb{D}$, which is constructed by using historical data. The robustness of UC solution is achieved by minimizing the overall cost under the worst-case probability distribution in the confidence set $\mathbb{D}$.

$$\min \quad \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{G}_r}(SU_i\cdot v_{i,t})+\max_{\mathbb{P}\in\mathbb{D}}E_{\mathbb{P}}\left[Q\left(u_{\mathcal{G}_r},v_{\mathcal{G}_r},\xi\right)\right] \quad (1a)$$

$$\text{s.t.} \quad u_{i,t}-u_{i,t-1}\le v_{i,t} \qquad \forall i\in\mathcal{G}_r, t\in\mathcal{T} \quad (1b)$$

$$\sum_{k=t-UT_i+1}^{t} v_{i,k}\le u_{i,t} \qquad \forall i\in\mathcal{G}_r, t\in[UT_i,N_t] \quad (1c)$$

$$\sum_{k=t-DT_i+1}^{t} v_{i,k}\le 1-u_{i,t-DT_i} \qquad \forall i\in\mathcal{G}_r, t\in[DT_i,N_t] \quad (1d)$$

$$u_{i,t},v_{i,t}\in\{0,1\} \qquad \forall i\in\mathcal{G}_r, t\in\mathcal{T} \quad (1e)$$

where $Q\left(u_{\mathcal{G}_r},v_{\mathcal{G}_r},\xi\right)$ is equal to,

$$\min \quad \sum_{t\in\mathcal{T}}\left(\sum_{i\in\mathcal{G}_r}\phi_{i,t}(\xi)+\sum_{i\in\mathcal{G}_f}(SU_i\cdot v_{i,t}(\xi)+\phi_{i,t}(\xi))\right) \quad (2a)$$

$$\text{s.t.} \quad \sum_{i\in\mathcal{G}}p_{i,t}(\xi)=\sum_{i\in\mathcal{D}}D_{i,t}-\sum_{i\in\mathcal{W}}W_{i,t}(\xi) \qquad \forall t\in\mathcal{T} \quad (2b)$$

$$-\overline{F}_i\le\sum_{b\in\mathcal{B}}SF_{b,i}\cdot\left(\sum_{i'\in\mathcal{G}^b}p_{i',t}(\xi)-\sum_{i'\in\mathcal{D}^b}D_{i',t}+\sum_{i'\in\mathcal{W}^b}W_{i',t}(\xi)\right)\le\overline{F}_i \qquad \forall i\in\mathcal{L}, t\in\mathcal{T} \quad (2c)$$

$$\beta_i^j\cdot p_{i,t}(\xi)+\gamma_i^j\cdot u_{i,t}\le\phi_{i,t}(\xi) \qquad \forall i\in\mathcal{G}_r, t\in\mathcal{T}, j\in[1,J] \quad (2d)$$

$$\underline{P}_i\cdot u_{i,t}\le p_{i,t}(\xi)\le\overline{P}_i\cdot u_{i,t} \qquad \forall i\in\mathcal{G}_r, t\in\mathcal{T} \quad (2e)$$

$$p_{i,t}(\xi)-p_{i,t-1}(\xi)\le RU_i\cdot u_{i,t-1}+\overline{RU}_i\cdot(1-u_{i,t-1}) \qquad \forall i\in\mathcal{G}_r, t\in\mathcal{T} \quad (2f)$$

$$p_{i,t-1}(\xi)-p_{i,t}(\xi)\le RD_i\cdot u_{i,t}+\overline{RD}_i\cdot(1-u_{i,t}) \qquad \forall i\in\mathcal{G}_r, t\in\mathcal{T} \quad (2g)$$

$$\beta_i^j\cdot p_{i,t}(\xi)+\gamma_i^j\cdot u_{i,t}(\xi)\le\phi_{i,t}(\xi) \qquad \forall i\in\mathcal{G}_f, t\in\mathcal{T}, r\in[1,J] \quad (2h)$$

$$\underline{P}_i\cdot u_{i,t}(\xi)\le p_{i,t}(\xi)\le\overline{P}_i\cdot u_{i,t}(\xi) \qquad \forall i\in\mathcal{G}_f, t\in\mathcal{T} \quad (2i)$$

$$u_{i,t}(\xi)-u_{i,t-1}(\xi)\le v_{i,t}(\xi) \qquad \forall i\in\mathcal{G}_f, t\in\mathcal{T} \quad (2j)$$

$$v_{i,t}(\xi)\le u_{i,t}(\xi) \qquad \forall i\in\mathcal{G}_f, t\in\mathcal{T} \quad (2k)$$

$$v_{t,k}(\xi)\le 1-u_{i,t-1}(\xi) \qquad \forall i\in\mathcal{G}_f, t\in\mathcal{T} \quad (2l)$$

$$u_{i,t}(\xi),v_{i,t}(\xi)\in\{0,1\} \qquad \forall i\in\mathcal{G}_f, t\in\mathcal{T} \quad (2m)$$

The objective function (1a) of improved DRUC formulation is the total cost that includes the first-stage UC cost of the regular units and the expected second-stage dispatch cost under the worst-case probability distribution. Here we use $v_{\mathcal{G}_r}$ and $u_{\mathcal{G}_r}$ to represent all the start-up and commitment variables for regular units. The lower bounds for start-up variables are modeled in (1b). Constraints (1c) and (1d) describe the minimum on-line and off-line time constraints, respectively. Binary variables are declared in (1e).

The second-stage objective function (2a) is the expected future dispatch cost, which includes UC costs for flexible generation resources. System power balance and branch flow limits are modeled in (2b) and (2c), respectively. Piece-wise linear constraints with R pieces are modeled in (2d) and (2h) to approximate quadratic cost functions for regular and flexible generation resources, respectively. Constraints (2e) and (2i) enforce the bounds for output power. Constraints (2f) and (2g) are ramp-up and ramp-down constraints for regular generation resources, respectively; corresponding constraints for flexible generation resources are not modeled due to their fast-ramp capabilities. In a similar manner to (1b)-(1e), commitment constraints for flexible generation resources are modeled in (2j)-(2m). Note constraints (2h)-(2l) together with corresponding trivial non-negative constraints for binary variables are the convex hull for quick-start units from a single-unit perspective [17]. However, with (2b)-(2g) included, the whole set of constraints is not a perfect formulation, thus needs further convexification.

The improved DRUC formulation is a two-stage distributionally robust optimization with mixed-binary recourse since binary commitment variables for flexible generation resources are modeled in the second-stage problem. In contrast to traditional DRUC formulations in the literature, our formulation can properly model the flexible adjustment capabilities of flexible generation resources to address the renewable energy power uncertainty, thus enabling a more accurate flexible resource quantification in UC problems.

B. Confidence Set

To construct a confidence set for the ambiguous distribution, several approaches are studied, such as moment information based sets [18], [19], probability metric based sets including L_1, $L_{\inf}$ metrics [8], [20] and Wasserstein metric [21], [22]. Among these, Wasserstein metric based confidence set is extensively studied in recent years, due to its good property on convergence and full utilization of historical data. As the distributionally robust integer L-shaped algorithm in [12] is proved to be finitely convergent under Wasserstein metric based confidence set, in this work, we use Wasserstein metric defined in (3) [23] to construct a confidence set with empirical data of day-ahead renewable power forecast error.

$$\mathfrak{D}_{\mathrm{w}}(\mathbb{P},\hat{\mathbb{P}}):=\inf_{\mathbb{Q}}\left\{E_{\mathbb{Q}}[\mathrm{d}(\xi,\hat{\xi})]:\mathbb{P}=\varphi(\xi),\hat{\mathbb{P}}=\varphi(\hat{\xi})\right\} \quad (3)$$

where ξ and $\hat{\xi}$ are random variables for day-ahead renewable power forecast error, which are associated with true distribution $\mathbb{P}$ and empirical distribution $\hat{\mathbb{P}}$, respectively. $\mathrm{d}(\xi,\hat{\xi})$ is a predefined distance between random variables ξ and $\hat{\xi}$, e.g., $\mathrm{d}(\xi,\hat{\xi})=|\xi-\hat{\xi}|$. $\mathbb{Q}$ denotes the joint distribution of ξ and $\hat{\xi}$ with marginal distributions $\mathbb{P}$ and $\hat{\mathbb{P}}$. The probability function is represented by $\varphi(\cdot)$.

Using Wasserstain metric, we construct a distribution-based confidence set $\mathbb{D}=\{\mathbb{P}\in\mathfrak{P}_+:\mathfrak{D}_{\mathrm{w}}(\mathbb{P},\hat{\mathbb{P}})\le\theta\}$. The tolerance level of the distance θ is determined by a given confidence level α, the number of bins N, the diameter δ of the supporting space, and the size of historical data N_h, as shown in (4) [24].

$$\theta=\frac{N\delta}{4N_h}\log\left(\frac{2N}{1-\alpha}\right) \quad (4)$$

We denote $\xi^1,\xi^2,\cdots,\xi^N$ as the discretized scenarios of ξ, and $\hat{\xi}^1,\hat{\xi}^2,\cdots,\hat{\xi}^N$ as the discretized scenarios of $\hat{\xi}$. Based on the definition of Wasserstein metric and the construction of $\mathfrak{D}_{\mathrm{w}}(\mathrm{P}, \hat{\mathrm{P}})$ in (3), we can reformulate the confidence set $\mathbb{D}$ as the constraints in (5).

$$\sum_{n=1}^{N} \sum_{m=1}^{N} q^{nm} \cdot \mathrm{d}(\xi^m, \hat{\xi}^n) \leq \theta \tag{5a}$$

$$\sum_{n=1}^{N} q^{nm} = p^m \qquad \forall m = 1, \ldots, N \tag{5b}$$

$$\sum_{m=1}^{N} q^{nm} = \hat{p}^n \qquad \forall n = 1, \ldots, N \tag{5c}$$

$$\sum_{m=1}^{N} p^m = 1 \tag{5d}$$

where constraint (5a) represents the expectation of distance between ξ and $\hat{\xi}$ over the joint distribution $\mathbb{Q}$. Constraints (5b) and (5c) represent that $\mathbb{P}$ and $\hat{\mathbb{P}}$ are the marginal distributions of $\mathbb{Q}$ respectively. Constraint (5d) ensures that the distribution $\mathbb{P}$ is indeed a distribution.

Although only renewable energy uncertainty is taken into account, the proposed method can be easily extend to considering load uncertainty. This can be achieved in a net-load manner. In equations (2b) and (2c), the term $\sum_{i\in\mathcal{D}^b} D_{i,t} - \sum_{i\in\mathcal{W}^b} W_{i,t}(\xi)$ can be represented as uncertain net-load $D^{\text{net}}_{b,t}(\xi)$. The corresponding confidence set can be constructed similarly with empirical distribution of net-load forecast error.

C. Abstract Formulation

The improved DRUC model in (1) can be presented in an abstract form, as shown in (6).

$$\min_{y} \quad a^\top y + \vartheta(y) \tag{6a}$$

$$\text{s.t.} \quad Ay \geq b,\ y \in \{0,1\}^{2|\mathcal{G}_r||\mathcal{T}|} \tag{6b}$$

where (6a) corresponds to the objective function in (1a) with $y = [u_{\mathcal{G}_r}^\top, v_{\mathcal{G}_r}^\top]^\top$, and (6b) corresponds to constraint (1b)-(1e). In (1a), $a^\top y$ is start-up cost for regular units; $\vartheta(y)$ denotes the second-stage objective function, which considers the expected cost under the worst-case distribution in confidence set $\mathbb{D}$, i.e., $\vartheta(y) = \max_{\mathbb{P}\in\mathbb{D}} E_{\mathbb{P}}[Q(y)]$.

The second-stage problem can be reformulated in (7) with discretized scenarios. This is referred to as distribution separation problem, and the algorithm to solve it is referred to as distribution separation algorithm [12].

$$DS(Q(y)) = \max_{\mathbb{P}\in\mathbb{D}} \sum_{n=1}^{N} p_n Q_n(y) \tag{7}$$

where,

$$Q_n(y) = \min_{x_n} \ c^\top x_n \tag{8a}$$

$$\text{s.t.} \quad Bx_n + Cy \geq d_n \ : \ \mu_n \tag{8b}$$

$$x_n \in \{0,1\}^{2|\mathcal{G}_f||\mathcal{T}|} \times \mathbb{R}^{V-2|\mathcal{G}_f||\mathcal{T}|} \tag{8c}$$

Corresponding to (2a), the objective function (8a) is to minimize the economic dispatch cost for scenario n including start-up cost for flexible generation resources. Here $x_n = (u_{\mathcal{G}_f}(\xi^n), v_{\mathcal{G}_f}(\xi^n), \phi(\xi^n), p(\xi^n))$, which contains commitment variables for flexible generation resources, as well as power output and cost variables for all generators. Constraints (8b)-(8c) corresponds to (2b)-(2m). V is the number of second-stage variables in $Q_n(y)$.

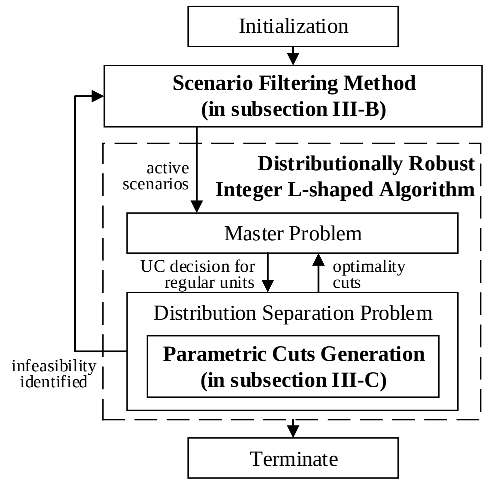

Fig. 1. Algorithmic framework.

III. Revised Integer L-shaped Algorithm

In this section, we describe a revised integer L-shaped algorithm to tackle the two-stage distributionally robust mixed-binary model. The algorithmic framework is presented first. A scenario filtering method is then introduced to address the first-stage feasibility issue. To offer high-quality optimally cuts, the second-stage convexification is achieved by parameter cut generation methods. Finally, we provide detailed steps for the whole algorithm.

A. Distributionally Robust Integer L-shaped Algorithm

We demonstrate the overall algorithmic framework to provide a big picture for this section, and then elaborate on a distribution separation algorithm to solve the max-min problem in (7).

1) Decomposition Framework: The algorithm follows a decomposition framework, as shown in Fig. 1. The first and second stage problems shown in (6) and (7) correspond to master and distribution separation problems, respectively. Given UC decisions for regular generation resources from the first stage, the distribution separation problem is solved to obtain the expected cost under the worst-case distribution. A sequential convexification procedure (in subsection III-C) for the second stage is also introduced by using parameter cut generation methods to offer high-quality optimally cuts for the master problem. In addition, to ensure the recourse problem relatively complete as mentioned in [12], an iterative scenario filtering method (in subsection III-B) is used when infeasibility is identified in the second stage.

2) Distribution Separation Problem: The distribution separation problem in (7) is solved as follows: $Q_n(\hat{y}^k)$ in (8) is first solved for each bin n given fixed $\hat{y}^k$; then solve $DS(Q(\hat{y}^k))$ in (7) with fixed optimal $Q_n^*(\hat{y}^k)$, i.e., $\max_{\mathbb{P}\in\mathbb{D}} \sum_{n=1}^{N} p_n Q_n^*(\hat{y}^k)$ to obtain the worst-case p_n^{k*}. This distribution separation algorithm associated with Wasserstein metric based confidence sets is finitely convergent, according to [12]. Optimally cuts for the first stage are further created with parameter cut generation methods in subsection III-C. The detailed steps that integrate distribution separation algorithm and optimally cut generation are summarized in subsection III-D.

B. 1st-Stage Feasibility Cuts: Scenario Filtering

A scenario filtering method is used to meet the relatively complete recourse requirement in [12]. We notice that the formulation has a relatively complete recourse if operational constraints (8b)-(8c) for all the scenarios are included in the first-stage problem (6). However, with more scenarios included, the size of (6) would become large. In light of the transmission constraint filtering methods proposed in [25], [26], we found constraints (8b) for most scenarios are not binding. In this work, a heuristic scenario filtering method is proposed to detect active scenarios and add corresponding operational constraints to (6). The method is described in the following steps.

F1 Calculate the peak net-load $\max_{t\in\mathcal{T}}\{\sum_{i\in\mathcal{D}} D_{i,t} - \sum_{i\in\mathcal{W}} W_{i,t}(\xi^n)\}$ for each scenario $n=1,...,N$. Denote the scenario with maximum peak net-load as $n_{\max}$.

F2 Include operational constraints (8b)-(8c) that correspond to $n_{\max}$ into the first-stage problem (6).

F3 Run the first-stage problem (6), and obtain a solution $\hat{y}$.

F4 Check the feasibility for each scenario n by adding slack variables σ_n:

$$\min_{x_n,\sigma_n} \quad \mathbf{1}^\top \sigma_n \tag{9a}$$
$$\text{s.t.} \quad Bx_n + D\sigma_n \geq d_n - C\hat{y} \tag{9b}$$
$$x_n \in \{0,1\}^{2|\mathcal{G}_f||\mathcal{T}|} \times \mathbb{R}^{V-2|\mathcal{G}_f||\mathcal{T}|} \tag{9c}$$

F5 If $\sigma_n = \mathbf{0}$ for all the n scenarios, recourse problems are feasible under the first-stage decision $\hat{y}$; If $\mathbf{1}^\top\sigma_n > 0$, add constraints (8b)-(8c) that correspond to scenario(s) with maximum objective into (6), and return to F3.

It should be noted the initial scenario identification in F1 is only a warm-start strategy, which doesn't necessarily guarantee the worst-case is identified. More scenarios might be included iteratively if infeasibility is found in subsequent steps.

C. 2nd-Stage Mixed-Binary Cuts: Sequential Convexification

To solve the two-stage distributionally robust optimization model, Benders decomposition is a traditional approach. However, given the second stage is a mixed-binary problem, duality theory cannot be directly applied to generate high-quality valid Benders cuts. In this work, we use two types of parametric cuts to strengthen the second stage. The idea is to sequentially convexify the second-stage relaxation by adding parametric cuts in each iteration. The parametric cuts in [13] and lift-and-project cuts are added sequentially. As the parametric cuts in [13] can be fast calculated, they are added first. Lift-and-project cuts are then used to further tighten the model. As the second-stage formulation becomes tighter, the quality of Benders cuts generated for (6) could be potentially improved.

1) Parametric Cuts in [13]: The parametric cut that is initially proposed by authors of [13] for stochastic unit commitment is valid for the second stage given any first-stage solution. We first deploy it to convexify our second-stage LP relaxation. The detailed implementation is provided in the appendix.

2) Lift-and-Project Cuts: We find the parametric cut in [13] cannot always guarantee the tightness of the second-stage relaxation, as it might stop improving the integrality gap after several iterations (as shown in subsection IV-C). Lift-and-project cut has been applied to two-stage distributionally robust mixed-binary problems in [12]. However, for our two-stage UC problem which contains a large number of binary variables in the first stage, it is difficult to obtain LP basis after using cutting plane algorithms to solve the first-stage problem. To address this issue, we revise the method in [12], and propose a customized lift-and-project cut generation process for our particular problem.

To ensure the generated lift-and-project cuts are valid for the second-stage problem in (8) given any first-stage solution from (6), we apply a lift-and-project process to optimization problem (10) for scenario n. Note regardless of objective parameter values $\tilde{a}^\top$ and $\tilde{c}^\top$, the generated cuts are valid for the feasible region formed by (10b)-(10d).

$$\min_{y,x_n} \quad \tilde{a}^\top y + \tilde{c}^\top x_n \tag{10a}$$
$$\text{s.t.} \quad Ay \geq b,\ Bx_n + Cy \geq d_n \tag{10b}$$
$$y \in \{0,1\}^{2|\mathcal{G}_r||\mathcal{T}|} \tag{10c}$$
$$x_n \in \{0,1\}^{2|\mathcal{G}_f||\mathcal{T}|} \times \mathbb{R}^{V-2|\mathcal{G}_f||\mathcal{T}|} \tag{10d}$$

Given optimal solution $z^* = (y^*, x_n^*)$ for an LP relaxation of (10) (probably with previous generated cuts) and a binary variable index l, we use the LP in (11) to generate a lift-and-project cut [15], [16]. With the index l, inequalities $z_l \leq 0$ and $z_l \geq 1$ split the feasible region of LP relaxation of (10). Lift-and-project cuts are obtained from the disjunction of the splitted regions [16]. Here (11f) is a normalization constraint. Readers can refer to [15], [16] for more details of the split-cut generation LP in (11). To distinguish the matrices/vectors in (10) and (12), primes are marked in the corresponding terms in (12) as trivial bound constraints for binary variables are included in it.

$$\min_{\kappa,\zeta,g,g_0,h,h_0} \quad z^{*\top}\kappa - \zeta \tag{11a}$$
$$\text{s.t.} \quad \kappa - \tilde{A}^\top g + g_0 \cdot \boldsymbol{e}_l \geq 0 \tag{11b}$$
$$\kappa - \tilde{A}^\top h - h_0 \cdot \boldsymbol{e}_l \geq 0 \tag{11c}$$
$$-\zeta + \tilde{b}^\top g = 0 \tag{11d}$$
$$-\zeta + \tilde{b}^\top h + h_0 = 0 \tag{11e}$$
$$\mathbf{1}^\top g + g_0 + \mathbf{1}^\top h + h_0 = 1 \tag{11f}$$
$$\kappa \in \mathbb{R}^{n^{\text{var}}},\ \zeta \in \mathbb{R} \tag{11g}$$
$$g, h \in \mathbb{R}_+^{n^{\text{con}}},\ g_0, h_0 \in \mathbb{R}_+ \tag{11h}$$

where, $\boldsymbol{e}_l$ is the l-th unit vector.

$$\tilde{A} = \begin{pmatrix} A' & \mathbf{0} \\ C' & B' \end{pmatrix}_{n^{\text{con}}\times n^{\text{var}}}, \quad \tilde{b} = \begin{pmatrix} b' \\ d'_n \end{pmatrix}_{n^{\text{con}}\times 1} \tag{12}$$

and n^{con}, n^{var} are the number of constraints and variables in the LP relaxation of problem (10), respectively, as indicated in (12).

With the optimal solution of (11), i.e., κ^* and ζ^*, a lift-and-project cut can be obtained in the form of (13).

$$(\hat{y}^\top, x_n^\top) \cdot \kappa^* \geq \zeta^* \tag{13}$$

The lift-and-project cut generation process is summarized in the following.

L1 Initialization. Set the counter for cut number $n_{\text{L\&P}}$ as 0, maximum cut number as $n_{\text{L\&P}}^{\max}$, and tolerance for integer solutions as ϵ_{int}.

L2 Solve the LP relaxation of (10) with previous generated cuts, obtain an optimal solution $z = (y^*, x_n^*)$.

L3 If all the binary variables in y^* and x_n^* are close to integral values within an ϵ_{int} tolerance, or the cut number reaches the predefined maximum number (i.e., $n_{\text{L\&P}} \geq n_{\text{L\&P}}^{\max}$), terminate the cut generation process and return the generated cuts; otherwise, go to Step L4.

L4 Given a predefined priority list, let l be the first binary variables with non-integral value that exceed ϵ_{int}-distance to integer values. Solve the cut generation LP (11) that splits the l-th binary variable, to generate a lift-and-project cut in the form of (13). Assign $n_{\text{L\&P}} \leftarrow n_{\text{L\&P}} + 1$, and return to Step L2.

In comparison to the algorithm in [15], which is designed for general purposes, some customized adaptions and heuristic rules are made for our particular UC problem.

a) In this work, we limit the maximum cut number as $n_{\text{L\&P}}^{\max}$ in each iteration. The authors of [15] aim to solve an MBLP, while we are trying to obtain a tight LP relaxation. Thus, in our problem, it may not be necessary to keep adding cuts until an integral solution is obtained.

b) A largest-index policy is used in [15], in fact, the sequence of adding split inequalities can affect the convexification performance. Due to the cost minimization objective of UC problems, $u_{i,t}$ values that should be 1 in the MBLP tend to approach $p_{i,t}^*/\overline{P}_i$ in the relaxed LP. A cost-based heuristic priority index (PI), as defined in (14), is proposed to estimate potential objective value decrease in LP relaxation that is caused by non-integral value of $u_{i,t}$. Higher priorities are assigned to variables with higher PI values. We denote this heuristic rule as 'priority-index policy' hereafter.

$$\text{PI}_{i,t} = \left(1 - p_{i,t}^*/\overline{P}_i\right)\left(a_i u_{i,t}^* + SU_i v_{i,t}^*\right) \;\; \forall i \in \mathcal{G}_f, t \in \mathcal{T} \quad (14)$$

where, a_i is the no-load cost of unit i; $p_{i,t}^*$, $u_{i,t}^*$ and $v_{i,t}^*$ are the optimal solution from the second-stage MBLP (8) (when performing Step 4 of distributionally robust integer L-shaped algorithm in subsection III-D). Note although on-the-shelf solvers can quickly solve MBLP problems, they may not provide perfect formulations, as pure cutting plane approaches are usually not used. However, the solutions can be leveraged in our cut generation process.

c) It is important to assign appropriate values for $\tilde{a}$ and $\tilde{c}$ in the objective of (10) to guide the direction for convexification. In our implementation, for dispatched units in the MBLP solution, we scale down the objective terms that correspond to $\phi_{i,t}$ and $v_{i,t}$. Thus, commitment variables $u_{i,t}$ have higher objective coefficients, which intuitively would drive the convexification direction along with these variables. Note these objective coefficients are only used for cut generation. We denote this heuristic rule as 'objective-scaling policy' hereafter.

Finally, for the k-th iteration in distributionally robust integer L-shaped algorithm as shown later in subsection III-D, we present the parametric cuts in (20) and (13), which are generated from the two aforementioned cutting plane methods, in a compact formulation in (15).

$$\Psi_{n,k}^\top x_n \geq \omega_{n,k} - \Phi_{n,k}^\top \hat{y}^k \;\;:\;\; \lambda_{n,k} \quad (15)$$

D. Steps of the Algorithm

We integrate the scenario filtering method, and two kinds of parametric cuts into the distributionally robust integer L-shaped algorithm in [12] to solve our improved DRUC model. Here a set X_n is used to represent the feasible region formed by second-stage constraints (8b)-(8c) for scenario n. The LP relaxation of $Q_n(y)$ in (8) is denoted as $RQ_n(y)$. The whole algorithm is summarized in the following steps.

1. Initialization. Set iteration counter $k \leftarrow 1$, scenario counter $n \in \{1, \cdots, N\}$, and a relative gap ϵ.
2. Obtain a first-stage decision $\hat{y}^k$ and a lower bound LB^k for the original problem through:
   - If $k = 1$, obtain $\hat{y}^k$ and LB^k using proposed scenario filtering method, i.e., Steps F1-F5 in subsection III-B.
   - If $k > 1$, obtain $\hat{y}^k$ and LB^k by solving the master problem (6).
3. For every renewable energy scenario $\hat{\xi}^n$, solve $RQ_n(\hat{y}^k)$ with all parametric cuts (15) generated in previous iterations. Obtain optimal solution $\tilde{x}_n^*$, optimal objective $RQ_n^*(\hat{y}^k)$, and the associated optimal dual multipliers.
4. Check for every renewable energy scenario $\hat{\xi}^n$ whether $\tilde{x}_n^* \in X_n$. For each n such that $\tilde{x}_n^* \in X_n$, set $x_n^* \leftarrow \tilde{x}_n^*$, and $Q_n^*(\hat{y}^k) \leftarrow RQ_n^*(\hat{y}^k)$. For each n such that $\tilde{x}_n^* \notin X_n$:
   - Solve MBLP $Q_n(\hat{y}^k)$ to get x_n^* and $Q_n^*(\hat{y}^k)$.
   - Generate parametric cuts in (15) by using Steps Z1-Z3 in the appendix and Steps L1-L4 in subsection III-C, then add them to the corresponding second-stage relaxation problem (8).
   - Solve the second stage relaxation problem with parametric cuts to update $\tilde{x}_n^*$, $RQ_n^*(\hat{y}^k)$, and the associated optimal dual multipliers.
5. If any of problems $RQ_n(\hat{y}^k)$ in Step 3 or $Q_n(\hat{y}^k)$ in Step 4 is infeasible, set $k \leftarrow k + 1$, and execute scenario filtering method in subsection III-B from step F4. After obtaining an updated $\hat{y}^k$ and LB^k, then go to Step 3. Otherwise, for each n, use $Q_n^*(\hat{y}^k)$ to solve $DS(Q(\hat{y}^k))$ in (7). Obtain optimal solution p_n^{k*} and optimal objective DS^{k*}. Set $UB^k \leftarrow a^\top \hat{y}^k + DS^{k*}$ as an upper bound for the original problem.
6. If $(UB^k - LB^k)/UB^k \leq \epsilon$, terminate and output the results. Otherwise, add optimality cut in (16) to master problem (6).

$$\vartheta \geq \sum\nolimits_{n=1}^{N} p_n^{k*} \cdot \Big\{ \mu_{n,k}^{*\top} \cdot (d - Cy) + \sum\nolimits_{j=1}^{k} \lambda_{n,j,k}^{*\top} \cdot (\omega_{n,j} - \Phi_{n,j}^\top y) \Big\} \quad (16)$$

where, $\mu_{n,k}^*$ and $\lambda_{n,j,k}^*$ are optimal dual multipliers associated with constraints (8b) and parametric cuts in (15), respectively, obtained by solving $RQ_n(\hat{y}^k)$ in Steps 3-4. Set $k \leftarrow k + 1$ and go to Step 2.

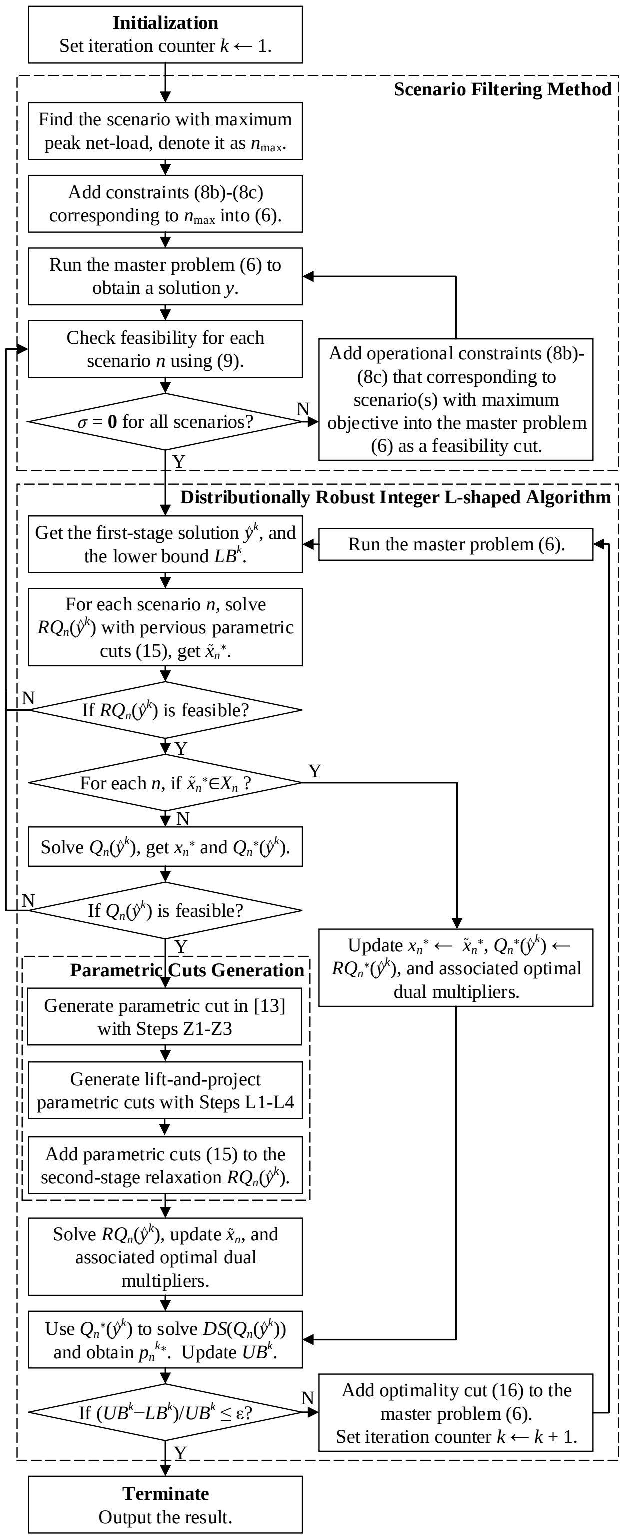

Fig. 2. Flow chart of the proposed algorithm.

It should be noted that parametric cuts in (15) are valid for any first-stage decision $\hat{y}$, thus cuts from previous iterations will be remained in each iteration.

To clearly illustrate the process of the proposed algorithm, a flow chart is provided in Fig. 2.

IV. Case Study

In this section, we test the performance of our approach on a 6-bus system and modified IEEE 118-bus system. CPLEX 12.10 [27] is used to solve LP and MBLP problems on a computer with Intel Core i7-9700 CPU and 64 GB RAM.

A. 6-Bus System

This test system contains 6 buses and 8 transmission lines. It has three regular thermal units (G1, G2, and G4), one quick-start unit (G3) as a flexible generation resource, and one wind farm. The wind farm is located at bus 2 with 100 MW installed capacity. The system diagram and detailed data are available online in [28]. The normalized day-ahead forecast error of wind power is calculated by data in the year 2019 from California ISO (CAISO) open access same-time information system (OASIS) [29].

1) Analysis of Flexibility Benefits: In comparison to existing DRUC models in the literature, the key feature of our improved DRUC lies in modeling the flexible start-up and shut-down capabilities of flexible generation resources in the near real time. To demonstrate the benefits of our proposed formulation, a traditional DRUC approach [8], which doesn't model the commitment capability of flexible generation resources in the recourse problem, is used for comparison. In our numerical experiments, we set the number of bins as 20, and the confidence levels α as 0.99 [8] for both improved and traditional DRUC cases.

Table I and Table II list the unit commitment solutions obtained from improved and traditional DRUC, respectively, with 365-day historical data. As the improved DRUC solution in Table I indicated, the regular generator G2 is dispatched in the day-ahead to satisfy the base load, while the flexible generation resource G3 can be scheduled depending on the realization of wind power (i.e, G3 starts up for scenarios S1-S3, and is not committed for others). In this case, either G1 or G3 could possibly be committed to address the peak net-load in S1-S3. Although the fuel cost of regular unit G1 is cheaper than that of G3, the quick-start unit G3 is supposed to be

TABLE I
Improved DRUC Solution in 6-Bus System (with 365-day Historical Data)

	hour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
u_{G1}, u_{G4}		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
u_{G2}		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
u_{G3}	S1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0
	S2	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
	S3	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	S4-20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

TABLE II
Traditional DRUC Solution in 6-Bus System (with 365-day Historical Data)

hour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
u_{G1}	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1
u_{G2}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
u_{G3}, u_{G4}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

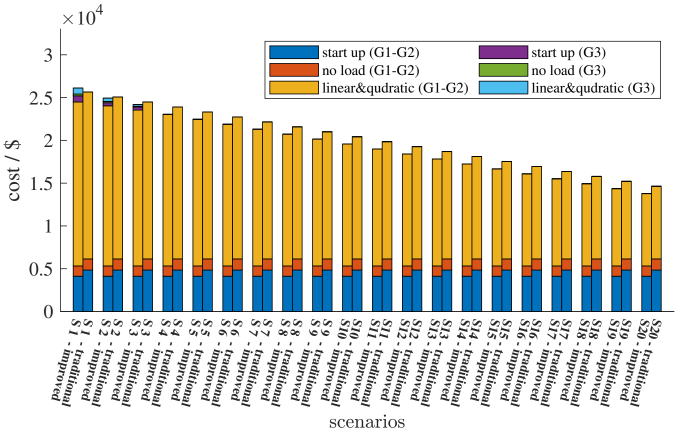

Fig. 3. In-sample cost comparison for improved and traditional DRUC approaches in 6-bus system (with 365-day historical data).

TABLE III
AN ILLUSTRATIVE CASE FOR TRADITIONAL DRUC INFEASIBILITY

scenario	probability	net-load [load, wind] (MW)		
		$t = 1$	$t = 2$	$t = 3$
S1	0.5	80 [110, 30]	125 [140, 15]	125 [140, 15]
S2	0.5	42 [47, 5]	42 [42, 0]	80 [88, 8]

TABLE IV
COST FOR 6-BUS SYSTEM UNDER DIFFERENT SIZE OF HISTORICAL DATA

# of days for historical data	in-sample			out-of-sample		
	improved DRUC ($)	traditional DRUC ($)	VF	improved DRUC ($)	traditional DRUC ($)	VF
10	21609.2	21933.6	1.48%	21616.2	21920.8	1.39%
50	19865.4	20585.7	3.50%	19857.0	20582.9	3.53%
100	19653.3	20427.5	3.79%	19675.8	20423.6	3.66%
365	19566.3	20390.6	4.04%	19583.6	20413.3	4.06%

more economically committed for possible further scenarios, as only a few scenarios need an additional unit to start up. On the contrary, given G1 is not flexible to decide commitment status in the near real time, having G1 committed in the day-ahead may not be preferred as it is not needed for most future scenarios. The solution of improved DRUC in fact takes advantage of the flexibility of the quick-start unit G3 to hedge against the risk of renewable uncertainty. However, from the traditional DRUC solution in Table II, one can observe that regular units G1 and G2 are dispatched, while quick-start unit G3 is not committed. As traditional DRUC does not appropriately model the flexibility of the quick-start unit G3 in the near real time, the commitment decisions for all units are supposed to be made in the first stage (i.e., day ahead). Consequently, the regular unit G1 is committed due to its lower fuel price, and has to stay online for most scenarios in which it is not needed.

For overall costs under the worst-case distribution, in contrast to $\$20390.6$ from the traditional DRUC solution, cost from the improved DRUC solution $\$19566.3$ is reduced by 4.04%. Fig. 3 shows the detailed cost comparison of improved and traditional DRUC for each scenario. As indicated, the quick-start unit G3 only needs to be committed in 3 scenarios. In S1, traditional DRUC has cost benefits over improved DRUC as the fuel cost of G1 is cheaper. In S2 and S3, the cost of the improved DRUC solution is lower even when G3 is committed, as the quick-start unit G3 is dispatched for a few hours in the improved DRUC solution, while the regular unit G2 is committed for 8 hours in the day ahead to handle all the scenario due to its inflexibility in the near real time. For the rest of the scenarios, i.e., from S4 to S20, the improved DRUC solution has cost benefits. The reason is that the quick-start unit G3 is not supposed to be committed in S4-S20, while the regular unit G2 still has to be online. As a result, with our improved DRUC model, the cost under the worst-case distribution can be reduced by leveraging the flexibility of quick-start units.

2) A Toy Example for Traditional DRUC Infeasibility: We design a toy illustrative case to show the traditional DRUC formulation might be infeasible in some extreme cases, while our improved DRUC can find a solution that physically makes sense. Assume we only have two generators G2 and G3, and the data settings in Table III is used in this 3 time-period example. The formulation is extended to consider net-load uncertainty. From improved DRUC, we obtain an optimal solution $u_{G2} = (1, 1, 1)$, $u_{G3} = (0, 1, 1)$ for S1, and $u_{G3} = (0, 0, 0)$ for S2. However, the traditional DRUC formulation is infeasible. The master problem (6) is reported to be infeasible by the CPLEX solver in the first iteration of our algorithm after S1 and S2 are added as feasibility cuts, which indicates S1 and S2 are conflicting in the traditional DRUC formulation under a no-load-shedding assumption. In detail, G3 should be online in hour 2 of S1 as the net-load 125 MW is larger than the capacity of G2 (i.e., $\overline{P}_{G2} = 120$ MW), while G3 should be offline in the same time period of S2 as the net-load 42 MW is smaller than the sum of minimum stable levels of G2 and G3 (i.e., $\underline{P}_{G2} + \underline{P}_{G3} = 44$ MW). This conflict also cannot be resolved even if wind curtailment is considered. As the unit commitment decisions for quick-start unit G3 are modeled in the first stage of the traditional DRUC formulation, the aforementioned conflict causes infeasibility. Although this is a conceptual example, it shows the advantage of the improved DRUC formulation on modeling the near real-time adjustment flexibility of flexible generation resources.

3) Statistical Results for Flexibility Benefits: Cost reduction of the improved DRUC formulation from the traditional DRUC formulation under the worst-case distribution is tested with different sizes of historical data. To quantify the cost benefits, we propose a metric, namely, value of flexibility (VF), to measure the impact of our flexible generation resources modeling on the system cost, as shown in (17).

$$\text{VF} = \frac{\text{cost}_{\text{traditional DRUC}} - \text{cost}_{\text{improved DRUC}}}{\text{cost}_{\text{traditional DRUC}}} \times 100\% \quad (17)$$

where $\text{cost}_{\text{traditional DRUC}}$ and $\text{cost}_{\text{improved DRUC}}$ are costs from traditional and improved DRUC, respectively.

We employ Monte-Carlo method to generate 5000 samples for the worst-case distribution that corresponds to each test. These samples are then used to estimate out-of-sample costs for traditional and improved DRUC approaches. In Table IV, we noticed the out-of-sample cost is close to corresponding in-sample cost. As indicated, both in-sample and out-of-sample costs obtained from traditional DRUC are greater than the costs from improved DRUC under the same size of historical data. VF increases as the size of historical data increases. This verifies our proposed DRUC formulation can reduce the system cost through properly modeling the fast adjustment capability of flexible generation resources.

Furthermore, 500 distributions with different standard deviations are generated in the corresponding confidence set of each case. For each distribution, 5000 samples are used to estimate the expected cost. As shown in Fig. 4, in this case, the improved DRUC can generally reduce the system cost in comparison to traditional DRUC under various distributions. This benefit becomes more significant if more historical data is used.

4) Comparison to Robust and Stochastic Approaches: The improved DRUC approach is also compared to robust optimization and stochastic programming approaches. Note we use robust UC (denote as 'RUC' hereafter) and stochastic UC (denote as 'SUC' hereafter) that consider the near real-time adjustable capability of flexible generation resources to facilitate a fair comparison. In fact, RUC can be regarded as a special case of the proposed framework by setting the confidence level α as 1. When $\alpha = 1$, θ tends towards $+\inf$, and the constraint (5a) in the confidence set is not enforced. Thus, the probability for the worst-case scenario will be 1 when solving $DS(Q(\hat{y}^k))$ in (7). Then, the proposed formulation is equivalent to RUC formulation. We implement RUC in this way. SUC is implemented in a scenario-based manner, and solved by Benders decomposition with our the

TABLE V
OUT-OF-SAMPLE COST COMPARISON OF DRUC, RUC AND SUC FOR 6-BUS SYSTEM UNDER WORST-CASE DISTRIBUTION

# of days for historical data	DRUC		RUC		SUC	
	cost ($)	infeasible samples	cost ($)	infeasible samples	cost ($)	infeasible samples
10	21616.2	0	21937.7	0	–	2014
50	19857.0	0	20578.2	0	–	516
100	19675.8	0	20443.9	0	19728.9	0
365	19583.6	0	20404.1	0	19583.6	0

cutting plane approaches in subsection III.

The out-of-sample cost comparisons of SUC, RUC, and our proposed DRUC (taking $\alpha = 0.99$ as an example) are shown in Table V with 5000 generated samples for each worst-case distribution from DRUC. As indicated, SUC suffers from infeasibility issues when the number of historical scenarios is relatively small. In such cases, as distributional uncertainties are considered in SUC, scenarios with zero occurrences may not have an accurate assessment of probability. Therefore, decisions from DRUC are more reliable than those from SUC. On the other hand, RUC aims to minimize the cost for the worst-case scenario, which may result in conservative commitment decisions. In this case, to optimize the cost for the worst-case scenario S1, regular unit G1 will be committed as analyzed before. In comparison to RUC, our proposed DRUC can avoid this conservative solution. As the size of historical data increases, DRUC cost under the worst-case distribution converges to the risk-neutral SUC cost.

We also evaluate the performance of DRUC, RUC, and SUC under the aforementioned 500 distributions. Note Fig. 5a only shows feasible cases for SUC. This indicates SUC suffers from

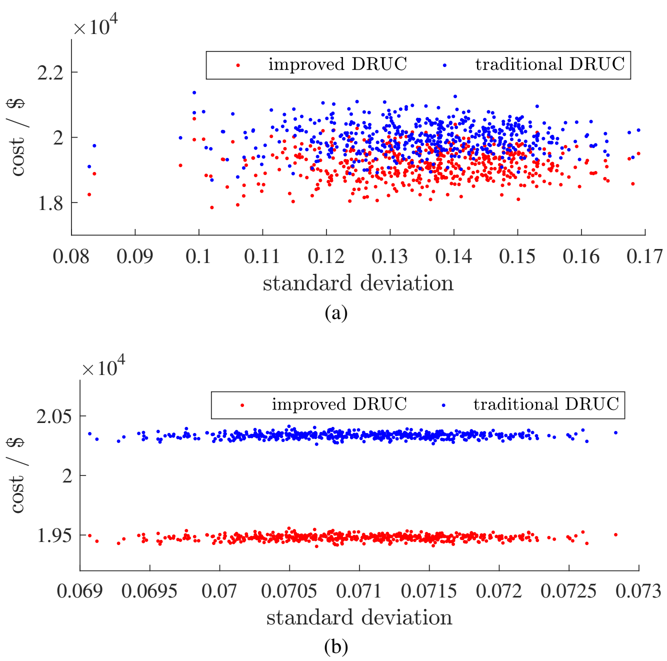

Fig. 4. Expected cost comparisons for improved and traditional DRUC under 500 distributions in 6-bus system. The evaluated UC decisions in subplots (a) and (b) are from models with 10-day and 365-day historical data. Each point corresponds to expected cost under a distribution.

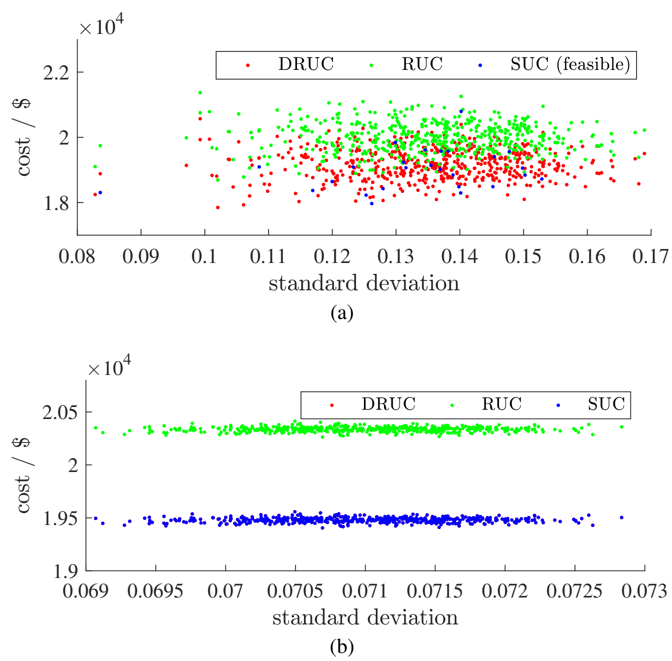

Fig. 5. Expected cost comparisons for DRUC, RUC and SUC under 500 distributions in 6-bus system. The evaluated UC decisions in subplots (a) and (b) are from models with 10-day and 365-day historical data. Points of DRUC and SUC may coincide partially in (b).

TABLE VI
RELAXED LP SOLUTION COMPARISON FOR CASES WITH AND WITHOUT LIFT-AND-PROJECT CUTS IN 6-BUS SYSTEM

settings		RLP solution of u_{G3} for 24 hours																								RLP ($)	MBLP ($)	IGap
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24			
S1	w/o L&P	0	0	0	0	0	0	0.02	0.25	0.44	0.46	0.41	0.19	0	0	0	0	0	0	0	0	0	0	0.02	0	21222	21993	3.5%
	w/ L&P	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	21993	21993	0.0%
S2	w/o L&P	0	0	0	0	0	0	0	0.07	0.26	0.29	0.24	0.01	0	0	0	0	0	0	0	0	0	0	0	0	20372	20811	2.1%
	w/ L&P	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	20811	20811	0.0%
S3	w/o L&P	0	0	0	0	0	0	0	0	0.09	0.11	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0	19596	20060	2.3%
	w/ L&P	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	20060	20060	0.0%

TABLE VII
RELAXED LP SOLUTION COMPARISON FOR CASES WITH AND WITHOUT HEURISTICS FOR LIFT-AND-PROJECT CUT GENERATION IN 6-BUS SYSTEM

settings		iteration 1				iteration 2				iteration 3				iteration 4				$n_{\text{L\&P}}^{\text{total}}$
		$n_{\text{L\&P}}$	RLP ($)	MBLP ($)	IGap	$n_{\text{L\&P}}$	RLP ($)	MBLP ($)	IGap	$n_{\text{L\&P}}$	RLP ($)	MBLP ($)	IGap	$n_{\text{L\&P}}$	RLP ($)	MBLP ($)	IGap	
S1	w/o heuristics	100	21443	21993	2.5%	100	21443	21993	2.5%	100	21443	21993	2.5%	100	21443	21993	2.5%	> 400
	w/ heuristics	23	21993	21993	0.0%	100	24595	24595	0.0%	92	20797	20797	0.0%	23	21993	21993	0.0%	238
S2	w/o heuristics	100	20372	20811	2.1%	100	20372	20811	2.1%	100	20372	20811	2.1%	100	20372	20811	2.1%	> 400
	w/ heuristics	100	20811	20811	0.0%	100	23259	23280	0.1%	44	20216	20216	0.0%	100	20811	20811	0.0%	344
S3	w/o heuristics	100	19601	20060	2.3%	100	19601	20060	2.3%	100	19601	20060	2.3%	100	19601	20060	2.3%	> 400
	w/ heuristics	47	20060	20060	0.0%	100	21911	22100	0.9%	54	19636	19636	0.0%	47	20060	20060	0.0%	248

infeasibility issues when the size of historical data is small. The cost of DRUC solutions converges to the cost of SUC solutions as the size of historical data increases, as shown in Fig. 5b. The UC solutions from DRUC generally have cost advantages over those from RUC.

5) Value of Lift-and-Project Cuts: To demonstrate the value of lift-and-project cuts for our problem, we perform numerical experiments in two case settings: the first one follows our proposed algorithm, which incorporates lift-and-project cuts; the second one only uses parametric cut in [13]. The maximum cut number limit $n_{\text{L\&P}}^{\max}$ for each scenario in each iteration is set as 100. Taking our test on 6-bus system with 365-day historical data as an example, the distributionally robust integer L-shaped algorithm converges in 4 iterations with lift-and-project cuts (denote as 'w/ L&P' in Table VI), while it doesn't converge with cuts in [13] only (denote as 'w/o L&P' in Table VI). In fact, LP relaxation solutions remain unchanged for further iterations in the case without lift-and-project cuts. Table VI shows the detailed LP relaxation solutions of the second-stage problem for S1-S3 in iteration 4. As indicated, an integral solution can be found with lift-and-project cuts in this small test case, so that the objective value of relaxed linear program (RLP) is the same as that of the original MBLP. However, for the case without lift-and-project cuts, a larger integral gap (as defined in (18)) for each scenario appears. Thus, compared to only using the parametric cut in [13], the second-stage problem can be further strengthened by using our proposed lift-and-project cut.

$$\text{IGap} = \frac{\text{obj}_{\text{MBLP}} - \text{obj}_{\text{RLP}}}{\text{obj}_{\text{MBLP}}} \times 100\% \tag{18}$$

We also propose heuristic rules, i.e., priority-index policy and objective-scaling policy in subsection III-C, to accelerate the sequential convexification process. Lift-and-project cut generation methods with and without heuristic rules (denote as 'w/ heuristics' and 'w/o heuristics', respectively) are also compared. Again, we use 6-bus case with 365-day historical data as an example. Using our proposed heuristic rules, the algorithm converges in 4 iterations. However, it doesn't converge in 10 iterations without these heuristics. In Table VII, we collect the numbers of generated lift-and-project cut $n_{\text{L\&P}}$, RLP objectives, and integral gaps for scenarios S1-S3 in the first four iterations. As indicated, oftentimes the maximum cut number limit $n_{\text{L\&P}}^{\max}$ (which is 100 in this case) is reached in the case without the proposed heuristic rules. After incorporating these heuristic rules, a smaller integral gap is obtained with fewer lift-and-projection cuts.

B. IEEE 118-Bus System

A modified IEEE 118-bus system [30] is used to test the scalability of the proposed approach. This system has 118 buses, 188 transmission lines, 54 generators, and 3 wind farms. Among these generators, there are 11 gas-fired quick-start units as flexible generation resources. Wind farms are located at buses 36, 69, and 77, with 400 MW, 800 MW, and 650 MW installed capacities respectively. Transmission lines 23-32, 34-36, and 77-78 are added to relieve congestion issues caused by including the wind farms.

In the numerical tests, we set the gap tolerance ϵ as 0.05%, and the confidence levels α for all the cases as 0.99. We also generate 5000 Monte-Carlo samples for the worst-case distribution that corresponds to each test, then use these samples to evaluate out-of-sample costs. As shown in Table VIII, it can be observed that the cost from improved DRUC is less than the cost from traditional DRUC for all the in-sample and out-of-sample cases. This verifies the cost can be reduced in our improved DRUC model by taking advantage of the flexibility of quick-start units. Given the expectation that the installed capacity of flexible generation resources will

TABLE VIII
COST FOR IEEE 118-BUS SYSTEM UNDER WORST-CASE DISTRIBUTION

# of days for historical data	in-sample			out-of-sample		
	improved DRUC ($)	traditional DRUC ($)	VF	improved DRUC ($)	traditional DRUC ($)	VF
10	2111554	2116187	0.22%	2107793	2120323	0.59%
50	2033177	2040069	0.34%	2033108	2041434	0.41%
100	2035504	2042953	0.36%	2036826	2044321	0.37%
365	2024093	2031945	0.39%	2024619	2031685	0.35%

TABLE IX
COST FOR FUTURE 118-BUS SYSTEM WITH 19 QUICK-STARTS UNDER WORST-CASE DISTRIBUTION

# of days for historical data	in-sample			out-of-sample		
	improved DRUC ($)	traditional DRUC ($)	VF	improved DRUC ($)	traditional DRUC ($)	VF
10	2110612	2118401	0.37%	2112978	2117318	0.20%
50	2030721	2042206	0.56%	2028418	2042194	0.67%
100	2033567	2044883	0.55%	2032145	2043175	0.54%
365	2021790	2033813	0.59%	2023434	2034088	0.52%

gradually increase in the near future, we also test a system with 19 quick-start units. In this system, 8 regular units with less than or equal to 100 MW capacities are assumed to retire in the near future, and 8 quick-start units with the same capacities are connected to the same buses. The results are reported in Table IX. We use the same approach to evaluate out-of-sample costs. Similar observations to those from the previous system with 11 quick-start units can be obtained. In addition, we can observe that the cost reduction from improved DRUC with 19 quick-starts is more significant, which indicates flexible operations of quick-start units can play an important role in reducing the system cost.

C. Discussions

1) On the Solution Time: We found although high-quality second-stage relaxations can be obtained through the proposed approach, which improves the convergence performance in comparison to the methods in the literature, the convexification process takes more time for large systems. For example, the IEEE 118 bus system takes $11,668.7$ seconds to solve. On the other hand, we notice that the computation process of Steps 3-4 (in the distributionally robust integer L-shaped algorithm) can be parallelized. Thus, finer-grained parallel computing methods can be further explored for larger-scale systems, which is beyond the scope of this paper.

2) On the Convergence: The distributionally robust integer L-shaped algorithm is proved to have finite convergence in [12]. The employed lift-and-project cut is also proved to be a finite-step cutting plane algorithm for MBLP in [15]. In our numerical experience, the iterations for revised integer L-shaped algorithm are relatively low. For tests with 365-day historical data, the 6-bus and two 118-bus cases use 4, 4, and 5 iterations to converge, respectively. It should be noted, although the cutting plane algorithm is finitely convergent, its converge rate depends on the heuristic rule as shown in our case study. Moreover, the converge rate of the cutting plane algorithm also affects that of the distributionally robust integer L-shaped algorithm.

V. CONCLUSION

In this work, we propose a novel DRUC model that addresses UC problems with flexible generation resources such as quick-start units. As binary variables appear in the second stage so that traditional separation algorithms won't apply, we propose a distributionally robust integer L-shaped algorithm to solve this two-stage mixed-binary model. Furthermore, revised lift-and-project cut generation method is used to strengthen the formulation of the second-stage problem. The numerical experiment verifies that our improved DRUC approach yields less cost in comparison to the traditional DRUC approach, due to the appropriate flexibility modeling of flexible generation resources in our approach.

APPENDIX: DETAILED IMPLEMENTATION OF PARAMETRIC CUT IN [13]

For demonstration brevity, we present the discrete and continuous parts of x_n as x_n^{d} and x_n^{c}, respectively. Accordingly, $c = [c^{\mathrm{d}}, c^{\mathrm{c}}]$, and $B = [B^{\mathrm{d}}, B^{\mathrm{c}}]$. The cut can be generated in three steps as shown in the following.

Z1 Given a first-stage decision $\hat{y}$, for each scenario n, the second-stage MBLP in (8) is solved, and the optimal binary variables $x_n^{\mathrm{d}*}$ and continuous variables $x_n^{\mathrm{c}*}$ are obtained.

Z2 We fix the binary variables $x_n^{\mathrm{d}*}$ to get an LP problem from $Q_n(\hat{y})$, as shown in (19a)-(19c).

$$\min_{x_n^{\mathrm{c}}} \quad c^{\mathrm{c}\top} x_n^{\mathrm{c}} \tag{19a}$$

$$\text{s.t.} \quad B^{\mathrm{c}} x_n^{\mathrm{c}} \geq d_n - C\hat{y} - B^{\mathrm{d}} x_n^{\mathrm{d}*} \ : \ \eta_n \tag{19b}$$

$$x_n^{\mathrm{c}} \in \mathbb{R}^{V - 2|\mathcal{G}_f||\mathcal{T}|} \tag{19c}$$

Z3 After the optimal dual multipliers η_n^* from (19b) are obtained, the parametric cuts can be generated as in (20) and added to the relaxation of the second stage problem $RQ_n(\hat{y})$ iteratively.

$$c^{\mathrm{c}\top} x_n^{\mathrm{c}} \geq \eta_n^{*\top} (d_n - C\hat{y} - B^{\mathrm{d}} x_n^{\mathrm{d}}) \tag{20}$$

REFERENCES

[1] F. Bouffard and F. D. Galiana, "Stochastic security for operations planning with significant wind power generation," in *Proc. IEEE Power Energy Soc. General Meeting*, 2008.

[2] P. A. Ruiz, C. R. Philbrick, E. Zak, K. W. Cheung, and P. W. Sauer, "Uncertainty management in the unit commitment problem," *IEEE Trans. on Power Syst.*, vol. 24, no. 2, pp. 642–651, 2009.

[3] A. Papavasiliou, S. S. Oren, and R. P. O'Neill, "Reserve requirements for wind power integration: A scenario-based stochastic programming framework," *IEEE Trans. on Power Syst.*, vol. 26, no. 4, pp. 2197–2206, 2011.

[4] H. Park, R. Baldick, and D. P. Morton, "A stochastic transmission planning model with dependent load and wind forecasts," *IEEE Trans. on Power Syst.*, vol. 30, no. 6, pp. 3003–3011, 2015.

[5] R. Jiang, J. Wang, and Y. Guan, "Robust unit commitment with wind power and pumped storage hydro," *IEEE Trans. on Power Syst.*, vol. 27, no. 2, pp. 800–810, 2011.

[6] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. on Power Syst.*, vol. 28, no. 1, pp. 52–63, 2012.

[7] S. Wang, G. Geng, and Q. Jiang, "Robust co-planning of energy storage and transmission line with mixed integer recourse," *IEEE Trans. on Power Syst.*, vol. 34, no. 6, pp. 4728–4738, 2019.
[8] C. Zhao and Y. Guan, "Data-driven stochastic unit commitment for integrating wind generation," *IEEE Trans. on Power Syst.*, vol. 31, no. 4, pp. 2587–2596, 2015.
[9] C. Zhao and R. Jiang, "Distributionally robust contingency-constrained unit commitment," *IEEE Trans. on Power Syst.*, vol. 33, no. 1, pp. 94–102, 2017.
[10] P. Xiong, P. Jirutitijaroen, and C. Singh, "A distributionally robust optimization model for unit commitment considering uncertain wind power generation," *IEEE Trans. on Power Syst.*, vol. 32, no. 1, pp. 39–49, 2016.
[11] C. Duan, L. Jiang, W. Fang, and J. Liu, "Data-driven affinely adjustable distributionally robust unit commitment," *IEEE Trans. on Power Syst.*, vol. 33, no. 2, pp. 1385–1398, 2017.
[12] M. Bansal, K. L. Huang, and S. Mehrotra, "Decomposition algorithms for two-stage distributionally robust mixed binary programs," *SIAM J. Optim.*, vol. 28, no. 3, pp. 2360–2383, 2017.
[13] Q. P. Zheng, J. Wang, P. M. Pardalos, and Y. Guan, "A decomposition approach to the two-stage stochastic unit commitment problem," *Ann. Oper. Res.*, vol. 210, no. 1, pp. 387–410, 2013.
[14] 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 Trans. on Power Syst.*, vol. 31, no. 2, pp. 1407–1419, 2015.
[15] E. Balas, S. Ceria, and G. Cornuéjols, "A lift-and-project cutting plane algorithm for mixed 0-1 programs," *Math. Program.*, vol. 58, no. 1-3, pp. 295–324, 1993.
[16] E. Balas and M. Perregaard, "A precise correspondence between lift-and-project cuts, simple disjunctive cuts, and mixed integer gomory cuts for 0-1 programming," *Math. Program.*, vol. 94, no. 2-3, pp. 221–245, 2003.
[17] B. Hua and R. Baldick, "A convex primal formulation for convex hull pricing," *IEEE Trans. on Power Syst.*, vol. 32, no. 5, pp. 3814–3823, 2016.
[18] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Oper. Res.*, vol. 58, no. 3, pp. 595–612, 2010.
[19] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Math. Program.*, vol. 137, no. 1-2, pp. 167–198, 2013.
[20] A. M.-C. So, J. Zhang, and Y. Ye, "Stochastic combinatorial optimization with controllable risk aversion level," *Math. Oper. Res.*, vol. 34, no. 3, pp. 522–537, 2009.
[21] G. C. Pflug and A. Pichler, *Multistage stochastic optimization*. Springer, 2014.
[22] G. C. Pflug, A. Pichler, and D. Wozabal, "The 1/N investment strategy is optimal under high model ambiguity," *J. Bank. Financ.*, vol. 36, no. 2, pp. 410–417, 2012.
[23] L. Kantorovich and G. Rubinshtein, "On a space of totally additive functions," *Vestn. Leningr. Univ.*, vol. 13, pp. 52–59, 1958.
[24] A. Bagheri, C. Zhao, F. Qiu, and J. Wang, "Resilient transmission hardening planning in a high renewable penetration era," *IEEE Trans. on Power Syst.*, vol. 34, no. 2, pp. 873–882, 2018.
[25] Y. Chen, A. Casto, F. Wang, Q. Wang, X. Wang, and J. Wan, "Improving large scale day-ahead security constrained unit commitment performance," *IEEE Trans. on Power Syst.*, vol. 31, no. 6, pp. 4732–4743, 2016.
[26] A. S. Xavier, F. Qiu, F. Wang, and P. R. Thimmapuram, "Transmission constraint filtering in large-scale security-constrained unit commitment," *IEEE Trans. on Power Syst.*, vol. 34, no. 3, pp. 2457–2460, 2019.
[27] IBM. ILOG CPLEX Homepage. Armonk, NY, USA. [Online]. Available: http://www.ilog.com
[28] S. Wang, C. Zhao, L. Fan, and R. Bo. Supplementary material for "Distributionally robust unit commitment with flexible generation resources considering renewable energy uncertainty". [Online]. Available: https://github.com/ee-swang/supplementary-material/blob/main/SUPPL-TPWRS-00145-2021.pdf
[29] California ISO. Open Access Same-time Information System. Folsom, CA, USA. [Online]. Available: http://oasis.caiso.com/mrioasis/logon.do
[30] IIT. Data for IEEE 118-bus system. Chicago, IL, USA. [Online]. Available: http://motor.ece.iit.edu/data/SCUC_118/SCUC_118.xls

Siyuan Wang received the B.S. and Ph.D. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2013 and 2019, respectively. He is currently a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology (formerly University of Missouri-Rolla), Rolla, MO, USA. His research interests include power system planning and operation, renewable energy integration, and applications of energy storage technology in power systems.

Lei Fan (Senior Member, IEEE) is an Assistant Professor with the Engineering Technology Department at the University of Houston. Before this position, he worked in the electricity energy industry for several years. He received the Ph.D. degree in operations research from the Industrial and System Engineering Department at the University of Florida. His research includes optimization methods, complex system operations, power system operations and planning.

Chaoyue Zhao (Member, IEEE) received the B.S. degree in information and computing sciences from Fudan University, Shanghai, China, in 2010, and the Ph.D. degree in industrial and systems engineering from the University of Florida, Gainesville, FL, USA, in 2014. She is currently an Assistant Professor in Industrial and Systems Engineering with the University of Washington, Seattle, WA, USA and previously was an Assistant Professor in Industrial Engineering and Management with Oklahoma State University. Her research interests include distributionally robust optimization and reinforcement learning with their applications in power system scheduling, planning, and resilience. She was with the Pacific Gas and Electric Company in 2013.

Rui Bo (Senior Member, IEEE) received the BSEE and MSEE degrees in electric power engineering from Southeast University (China) in 2000 and 2003, respectively, and received the Ph.D. degree in electrical engineering from the University of Tennessee, Knoxville (UTK) in 2009. He is currently an Assistant Professor of the Electrical and Computer Engineering Department with the Missouri University of Science and Technology (formerly University of Missouri-Rolla). He worked as a Principal Engineer and Project Manager at Midcontinent Independent System Operator (MISO) from 2009 to 2017. His research interests include computation, optimization and economics in power system operation and planning, high performance computing, electricity market simulation, evaluation and design.