A Fully Adaptive DRO Multistage Framework Based on MDR for Generation Scheduling under Uncertainty

Ying Yang, Linfeng Yang, *Member, IEEE*, and Zhaoyang Dong, *Fellow, IEEE*

***Abstract*—The growing proliferation of wind power into the power grid achieves a low-cost sustainable electricity supply while introducing technical challenges with associated intermittency. This paper proposes a fully adaptive distributionally robust multistage framework based on mixed decision rules (MDR) for generation scheduling under uncertainty to adapt wind power respecting non-anticipativity in quick-start unit status decision and dispatch process. Compared with existing multistage models, the proposed framework introduces improved MDR to handle all decision variables to expand the feasible region. Therefore, our model can find a feasible solution to some problems that are not feasible in the traditional models while finding a better solution to feasible problems, so as to better exploit wind energy and accordingly fall consumption of fossil fuels. Besides, the proposed model is reformulated with advanced optimization methods and improved MDR to the mixed integer linear programming (MILP) to address computational intractability. The effectiveness and superiority of the proposed model have been validated with case studies using IEEE benchmark systems.**

***Index Terms*—non-anticipativity, wind power, mixed decision rules, multistage, distributionally robust optimization, uncertain generation scheduling.**

NOMENCLATURE

Indices:

i, j	Index for unit/bus/wind generator.
t	Index for time period.

Constants:

N, T, R	Total number of units/periods/wind generators.
S^{W}	Set of buses with wind generators, $\left\|S^{\mathrm{W}}\right\| = R$.
S^{B}	Set of buses, $\left\|S^{\mathrm{B}}\right\| = N$ and $S^{\mathrm{B}} = \{1, \dots, N\}$.
S^{G}	Set of buses with units. $\left\|S^{\mathrm{G}}\right\| = G$ and $S^{\mathrm{G}} \subseteq S^{\mathrm{B}}$.
S^{TP}	Set of time periods, and $\left\|S^{\mathrm{TP}}\right\| = T$.
S^{L}	Set of lines, and $\left\|S^{\mathrm{L}}\right\| = L$.
S^{ref}	Set of reference bus.
a_i, b_i, c_i	Coefficients of the cost function of unit i.
$C_{\mathrm{cold},i}, C_{\mathrm{hot},i}$	Cold and hot startup cost of unit i.
$\underline{P}_i$, $\overline{P}_i$	Minimum and maximum power output of unit i.
$P_{\mathrm{D},t}$	System load demand in period t.
$P_{\mathrm{up},i}, P_{\mathrm{down},i}$	Ramp up and ramp down limit of unit i.
$P_{\mathrm{start},i}, P_{\mathrm{shut},i}$	Startup and shutdown ramp limit of unit i.
$u_{i,0}$	Initial commitment state of unit i.
$T_{i,0}$	Number of periods unit i has been online $(+)$ or offline $(-)$ prior to the first period of the time span.
$\underline{T}_{\mathrm{on},i}, \underline{T}_{\mathrm{off},i}$	Minimum up and minimum down time of unit i.
U_i	$\left[\min\left[T, u_{i,0}\left(\underline{T}_{\mathrm{on},i} - T_{i,0}\right)\right]\right]^+$
L_i	$\left[\min\left[T, (1-u_{i,0})\left(\underline{T}_{\mathrm{off},i} + T_{i,0}\right)\right]\right]^+$
τ_i	$\max\left(t - \underline{T}_{\mathrm{off},i} - T_{\mathrm{cold},i}, 1\right)$
$f_{i,t}$	$\begin{cases} 1, (t-\tau_i \le 0) \,\&\, \left(\left[-T_{i,0}\right]^+ < \lvert t-\tau_i\rvert + 1\right) \\ 0, \text{otherwise} \end{cases}$
C^l	Value of load loss.
$\underline{l}^{\mathrm{D}}_{i,t}, \overline{l}^{\mathrm{D}}_{i,t}$	Minimum/maximum load loss on bus i in period t.

Variables:

$u_{i,t}$	On/off status of unit i in period t.
$s_{i,t}, d_{i,t}$	Startup/shutdown status of unit i in period t.
$l^{\mathrm{D}}_{i,t}$	Load loss at bus i during time period t.
$P^{\mathrm{G}}_{i,t}$	Power output of unit on bus i in time period t.
$\theta_{i,t}$	Phase angle on bus i in time period t.

Operator:

$\{\cdot\}^+$	$\max(0,\cdot)$
$\mathbb{I}(\cdot)$	The indicator function
$[\cdot]$	$\{1,2,\dots,\cdot\}$
$[\cdot]^\wedge$	$\{0,1,2,\dots,\cdot\}$

I. Introduction

Typical unit commitment (UC) aims to minimize the generator's fuel consumption or the operation cost of the whole system by scheduling generators [1]. Recently, higher and higher penetration of renewable energy resources of UC problem to address the energy crisis and impact of greenhouse gas emissions [2]-[4]; wind energy is the most widely used among them due to its free availability and environment friendliness [4], [5]. Even though there are the above advantages, integrating wind power into power system still presents a formidable challenge since its intermittency and volatility make the UC a large-scale uncertain problem [6], [7].

Fortunately, there are two kinds of practical ways to combat the challenges: wind power forecast [7], [8] and optimization methods for the uncertain UC problem (UUC). The forecasting errors will be appropriately modeled with the introduction of different approaches. J. An *et al*. proposed a short-term wind power prediction method based on multisource wind speed fusion [7]. An ensemble pruning and combination problem is formulated in [8], which can enhance short-term prediction. Notwithstanding, these deterministic methods cannot completely meet the decision-making requirements in an uncertain environment. By contrast, for the past few decades, optimization methods have been representative methods as their practical significance of UUC. There are three main approaches [6]: robust optimization (RO), stochastic optimization (SO), and distributionally robust optimization (DRO). RO can guarantee the robustness of the scheduling scheme by optimizing the

This work was supported by the Natural Science Foundation of Guangxi (2020GXNSFAA297173, 2020GXNSFDA238017, 2021GXNSFBA075012), the Natural Science Foundation of China (51767003, 71861002). (Corresponding author: Linfeng Yang)

Y. Yang and L.F. Yang are with the School of Computer Electronics and Information, Guangxi University, Nanning 530004, China. And L.F. Yang is also with the Guangxi Key Laboratory of Multimedia Communication and Network Technology, Guangxi University. (email: yingyoung1997@gmail.com, ylf@gxu.edu.cn).

Z.Y Dong is with School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore. (e-mail: zydong@ieee.org)

worst-case in the uncertainty set [9], whilst SO maximizes profits by assuming that its distribution is the real distribution [10]. However, the security constraints of SO may be breached because of the unreliable probability distribution, while RO always obtains an over-conservative scheme in real scenarios.

Therefore, DRO demonstrates significant superiority over other methods since it can weigh the radicalism of SO and the conservatism of RO, all probability distributions of DRO's ambiguity set have been immunized by finding and optimizing the worst-case in the possible distribution family [11]. After years of research and development, DRO, especially moment-based DRO [12], [13], has been used in many fields, such as energy systems and UC. Nevertheless, moment-based DRO also has drawbacks. Firstly, it cannot guarantee that the unknown distribution will converge to the true distribution. Furthermore, the only utilization of moment information is still conservative. Motivated by the above facts, references [11], [14]-[16] developed a method called distance-based DRO, where the ambiguity set consisted of the distributions within a fixed distance from the nominal distribution. [11] provides a Wasserstein-based DRO and its approximate framework, which can manage the risk from wind power forecasted errors and minimize the generating cost. The authors in [14] proposed a min-max-min KL-based DRO to control the expected cost in the worst case. A chance-constrained optimal power flow model for UUC was introduced by [15].

Meanwhile, a class of two-stage models was developed in [17], [18]. In the two-stage model, the day-ahead unit state decision is made at stage 1, and the real-time dispatch decision is made at stage 2 across the scheduling horizon. Stage 2's decisions are based on the complete knowledge of uncertain variables across the entire operating period. However, the decisions are made sequentially in practice because of the generation dispatch of each hour only depends on the information of the realized uncertain variables up to that hour. In order to improve the deficiency of two-stage, multistage models were introduced in [17]-[20], where the state of unit decisions are *here-and-now*, i.e., day-ahead, and the hourly dispatch decisions of the next day are made on *wait-and-see*, which demonstrates the non-anticipativity for the sequential observation of uncertainties. However, most of those models are generally intractable to solve. An effective approach to overcome this problem is decision rules, which choose some real-valued functions with pre-specified structures [12], [21]. Reference [22] first used the decision rules for real-valued functions in 1974. Linear/affine decision rule's real-valued functions are employed as linear functions of uncertain variables in [23]. These approaches are direct and crude. The nonlinear decision rule used in [24] enables the real-valued adaptive decisions parameterized by a nonlinear lifting operator to define the structure of the decision rule. [17] used simplified affine policies to construct a more solvable structure. Linear decision rules are utilized in multistage models to ensure computational tractability [18]. Buhan *et al*. [20] designed a multistage method for wind forecast of wind power plants.

The above multistage models still have some flaws, although these models use rolling optimization for dispatch generation to adapt to uncertain variables, the state of units is still fixed. Accordingly, no feasible scheme may adapt to the uncertainties in any given unit state when uncertainties fluctuate considerably. That is, no feasible solution can be found in this scenario. As we know, units can be divided into two types roughly: quick-start and slow-start units. Therefore, quick-start units also could be optimized in real-time while slow-start units remain day-ahead optimization, meanwhile, introducing distance-based DRO not only improves utilization of renewable but also maintains reliability, which motivates us to construct a fully adaptive DRO model. Consequently, we propose a framework that provides a larger feasible region and results in a better optimal solution than traditional deterministic and multistage models. The main contributions can be summarized as:

1) A fully adaptive multistage distributionally robust framework based on Wasserstein-metric and mixed decision rules (MDR) is proposed for the generation scheduling under uncertainty (DR-MDR), respecting non-anticipativity in quick-start unit decision and dispatch process. Such a fully adaptive framework can get a much more advanced model with a larger feasible region, which has two main merits: It can find the feasible solution to some problems that traditional models are not feasible and find a better solution to problems that traditional models are feasible.

2) We used improved binary and linear decision rules to handle most decision variables, including integers and continuities for further adapting decision variables to uncertainty. As we know, this is the first work that utilizes DRO with MDR in uncertain generation scheduling. It provides a sequential decision-making process in that the uncertain wind is observed at a given time, and the decision variables are made immediately after this observation. Therefore, our model can better exploit wind energy and accordingly fall consumption of fossil fuel.

3) We developed a mixed integer linear programming (MILP) reformulation for the DR-MDR based on optimization theory to address the computational intractability. We also improved the decision rules for the reformulation of equality constraints. The extensive computational study of the proposed model and existing multistage model on IEEE test systems demonstrate the superiority of our model.

The rest of the paper is organized as follows. Section II presents the typical UUC models and their limitations. DR-MDR and its reformulation is proposed in Sections III. Case studies using the proposed model for the IEEE test systems are given in Section IV. Finally, Section V concludes this article.

II. Typical UUC Models and Their Limitations

A. Description of uncertain UC

1) Objective function

The objective of the problem with wind power is to minimize the total cost, which includes the following three parts:

The generation cost of the thermal unit $C_{i,t}^{FG}$ is described as a quadratic function of power output $P_{i,t}^{\mathrm{G}}$:

$$C_{i,t}^{FG} = a_i + b_i P_{i,t}^{\mathrm{G}} + c_i \left(P_{i,t}^{\mathrm{G}}\right)^2, \tag{1}$$

Given parameter $\mathcal{L}$, let $p_{i,l} = \underline{P_i} + l\,(\overline{P}_i - \underline{P_i})/\mathcal{L}$ and $l =$

$0,1,2,\dots,\mathcal{L}-1$, and replacing $C_{i,t}^{FG}$ with auxiliary variable $z_{i,t}$, then replacing (1) with the following constraints:

$$z_{i,t} \geq \left(2c_i p_{i,l} + b_i\right) C_{i,t}^{FG} + \left(a_i - \gamma_i \left(p_{i,l}\right)^2\right) u_{i,t}, \quad (2)$$

The startup cost of thermal unit $C_{i,t}^{ST}$, including hot startup cost $C_{i,t}^{HS}$ and excess of hot startup cost $C_{i,t}^{ES}$, which is

$$C_{i,t}^{ST} = C_{i,t}^{HS} + C_{i,t}^{ES} \quad (3)$$

$$C_{i,t}^{HS} = C_{\text{hot},i} s_{i,t}, \quad (4)$$

$$C_{i,t}^{ES} \geq \left\{\left(C_{\text{cold},i} - C_{\text{hot},i}\right)\left[s_{i,t} - \sum_{t_i}^{t-1} d_{i,\tau} - f_{i,t}\right]\right\}^+, \quad (5)$$

And the cost of load loss C^{CL}:

$$C^{CL} = \sum_{t \in S^{\text{TP}}} \sum_{i \in S^{\text{B}}} C^l l_{i,t}^{\text{D}}, \quad (6)$$

Thus, the objective function is:

$$f(\cdot) = \sum_{t \in S^{\text{TP}}} \sum_{i \in S^{\text{G}}} \left(C_{i,t}^{FG} + C_{i,t}^{ST}\right) + C^{CL}. \quad (7)$$

2) Thermal unit constraints

$$u_{i,t} \underline{P}_i \leq P_{i,t}^{\text{G}} \leq u_{i,t} \overline{P}_i, \quad (8)$$

$$P_{i,t}^{\text{G}} - P_{i,t-1}^{\text{G}} \leq u_{i,t-1} P_{i,\text{up}} + s_{i,t} P_{i,\text{start}}, \quad (9)$$

$$P_{i,t-1}^{\text{G}} - P_{i,t}^{\text{G}} \leq u_{i,t} P_{i,\text{down}} + d_{i,t} P_{i,\text{shut}}. \quad (10)$$

where $\forall i \in S^{\text{G}}, \forall t \in S^{\text{TP}}$. (8) is the capacity limits of thermal units, (9)-(10) enforce ramp up/down limits of individual unit.

3) Binary variables constraints

$$u_{i,t} = u_{i,0}, \quad (11)$$

$$\sum_{\varpi=\left[t-\underline{T}_{\text{on},i}\right]^+ +1}^{t} s_{i,\varpi} \leq u_{i,t}, \quad (12)$$

$$\sum_{\varpi=\left[t-\underline{T}_{\text{off},i}\right]^+ +1}^{t} d_{i,\varpi} \leq 1 - u_{i,t}, \quad (13)$$

$$s_{i,t} - d_{i,t} = u_{i,t} - u_{i,t-1}, t \in S^{\text{TP}}. \quad (14)$$

where $\forall i \in S^{\text{G}}, t \in [L_i + 1, \dots, T]$. (11) is the initial status of units, (12) and (13) define minimum on/off time, and (14) denotes the state equality constraint of units.

4) System constraints

$$\underline{l}_{i,t}^{\text{D}} \leq l_{i,t}^{\text{D}} \leq \overline{l}_{i,t}^{\text{D}}, \forall i \in S^{\text{B}}, \quad (15)$$

$$P_{i,t}^{\text{G}} + P_{i,t}^{\text{W}} - P_{i,t}^{\text{D}} + l_{i,t}^{\text{D}} = \sum_{j \in S_{\text{B}}} B_{ij} \theta_{j,t}, \forall i \in S^{\text{B}}, \quad (16)$$

$$\theta_{i,t} = 0, \forall i \in S^{\text{ref}}, \quad (17)$$

$$\left|\frac{\theta_{i,t} - \theta_{j,t}}{X_{ij}}\right| \leq F_{ij}^{\max}, \forall ij \in S^{\text{L}}. \quad (18)$$

where $\forall t \in S^{\text{TP}}$. (15) and (18) are the limitation of load loss and transmission lines, respectively, (16) represents the DC network model, and (17) denotes the reference bus.

5) Wind generators constraints

In this paper, $P_{i,t}^{\text{W}}$ is used to denote the wind generation, which is equal to the expected value $\bar{P}_{i,t}^{\text{W}}$ plus a random term representing forecast error $v_{i,t}^{\text{W}}$:

$$P_{i,t}^{\text{W}} \leq \bar{P}_{i,t}^{\text{W}} + v_{i,t}^{\text{W}}, \forall i \in S^{\text{W}}, \forall t \in S^{\text{TP}}. \quad (19)$$

6) Formulation of UUC

Finally, we obtain the UUC problem:

$$\min f(\cdot) \quad (20)$$

$$s.t. \{(2)-(6), (8)-(19)\}.$$

B. Typical Two-stage UUC

The popular DRO or robust optimization UUC model [12], [17], [18] can be represented by the typical two-stage formula. The first-stage scheduling determines unit status $\boldsymbol{x}$ before wind power is realized in the future, and the second-stage determines the dispatch $\boldsymbol{y}$ after knowing wind power. Besides, taking infinite space of the adaptive recourse function $\boldsymbol{y}(\boldsymbol{v}^{\text{W}})$ to solve the problem is generally intractable to solve [21].

$$\min_{\boldsymbol{x} \in \mathcal{X}} \left\{\sum_{t=1}^{T} \left[\boldsymbol{c}_t^{\text{T}} \boldsymbol{x}_t\right] + \min_{\boldsymbol{y} \in \mathcal{Y}(\boldsymbol{x}, \boldsymbol{v}^{\text{W}})} \sup_{\mathbb{P}^{\text{W}} \in \Upsilon^{\text{W}}} \mathbb{E}_{\mathbb{P}^{\text{W}}}\left[\sum_{t=1}^{T} \left[\boldsymbol{d}_t^{\text{T}} \boldsymbol{y}_t\right]\right]\right\}, \quad (21)$$

where $\boldsymbol{x}_t = \left[u_{[G],t}, s_{[G],t}, d_{[G],t}\right]$ denotes binary decision variables, $\boldsymbol{c}_t^T \boldsymbol{x}_t = \sum_{i \in S^G} C_{i,t}^{HS}$, $\mathcal{X} = \{x | (11) - (14)\}$ is the feasible constraints of $\boldsymbol{x}$, and $\boldsymbol{y}_t = \left[P_{[G],t}^G, \theta_{[G],t}, z_{[G],t}, C_{[G],t}^{ES}, P_{[W],t}^W, l_{[N],t}^D\right]$ denotes continuous decision variables, $\boldsymbol{d}_t^{\text{T}} \boldsymbol{y}_t = \sum_{i \in S^{\text{G}}} (z_{i,t} + C_{i,t}^{ES}) + \sum_{i \in S^{\text{B}}} C_{i,t}^{CL}$, $\mathcal{Y}(\boldsymbol{x}, \boldsymbol{v}^{\text{W}}) = \{y | (2), (5) - (6), (8) - (10), (15) - (19)\}$ is the feasible constraints of $\boldsymbol{y}$ parameterized by commitment decisions $\boldsymbol{x}$ and forecast error $\boldsymbol{v}^{\text{W}}$, $\mathbb{P}^{\text{W}}$ is a distribution of $\boldsymbol{v}^{\text{W}}$ and Υ^{W} is the ambiguity set.

C. Adaptive DRO UUC

Similar to Claims 1 and 2 of the two-stage RO model in [17], the scheduling scheme obtained by model (21) can lead to infeasibility in real-time applications [18]. Because, in actual scenarios, dispatch decisions must be sequentially optimized real-time with uncertainty observations prior to the operating hour. To truly figure this process, the decision variables $\boldsymbol{y}_t$ should count on the history of wind power generation $\boldsymbol{v}_{[t]}^{\text{W}} := (\boldsymbol{v}_1^{\text{W}}, \dots, \boldsymbol{v}_t^{\text{W}})$, $\boldsymbol{v}_t^{\text{W}} = \boldsymbol{v}_{[R],t}^{\text{W}}$, i.e., the decision maker observes first wind power $\boldsymbol{\xi}_1$ and then makes the first decision $\boldsymbol{y}_1(\cdot)$. Subsequently, second wind power $\boldsymbol{\xi}_2$ is revealed, and in response, the decision maker makes second decision $\boldsymbol{y}_2(\cdot)$, this alternating sequence of observations and decisions extends to *T* periods. Because of the above reasons, [18] develops an adaptive multistage DRO model by substituting $\boldsymbol{y}(\boldsymbol{v}^{\text{W}})$ with $\left[\boldsymbol{y}_t\left(\boldsymbol{v}_{[t]}^{\text{W}}\right), \forall t \in S^{\text{TP}}\right]$ in (21), then it formulates the adaptive DRO model:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \left\{\sum_{t=1}^{T} \left[\boldsymbol{c}_t^{\text{T}} \boldsymbol{x}_t\right] + \min_{\boldsymbol{y}_t(\cdot) \in \mathcal{Y}} \sup_{\mathbb{P}^{\text{W}} \in \Upsilon^{\text{W}}} \mathbb{E}_{\mathbb{P}^{\text{W}}}\left[\sum_{t=1}^{T} \left[\boldsymbol{d}_t^{\text{T}} \boldsymbol{y}_t(\cdot)\right]\right]\right\}, \quad (22)$$

However, (22) still has drawbacks in some scenarios, and we will present a contrast flow chart and an example to epitomize the superiority of our framework over (22).

D. Typical model's limitations

Fig. 1 shows the difference in scheduling between the typical models and ours at a macro level, specially $\boldsymbol{x}_i^s$ and $\boldsymbol{x}_i^q$ denotes state of slow-start and quick-start units, respectively. Furthermore, Fig. 2 gives the one-bus system structure, the system only has one period and one load bus *A*, which has both quick-start thermal unit and wind unit. Then, we assume the ambiguity is a single point set, unit ramp constraints are ignored, and wind power can be abandoned while load loss is not allowed. Nodal load $d_A = 20\text{MW}$, wind generation $p_w \in [0,25]$ and thermal generation $p_A \in \{0, [15,20]\}$, the generation cost of thermal unit is 60/MW. $\xi \in [0, +\infty]$ is wind power.

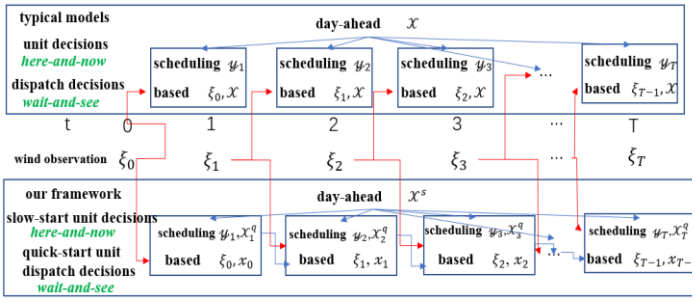

Fig. 1. Scheduling framework for typical models and ours.

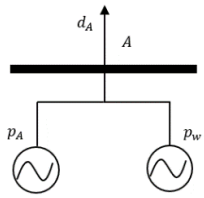

Fig. 2. Simple case to illustrate advantages of proposed model over typical multistage model (1-1 system).

Typical multistage model likes (22) have to fix unit status u before rolling scheduling, it is not difficult to find that system will be infeasible when $u=0$ since the load d_A can only be satisfied when $p_w \geq 20$. Therefore, (22) fixes $u=1$, in this case, thermal unit generation $p_A(\xi)=15+(5-\xi)^+$ and windy unit generation $p_w(\xi)=5-(5-\xi)^+$, the expectation of generation cost is $E_{obj}=(900+60\times(5-\xi)^+)\ \epsilon[900,1200]$.

By contrast, in our framework, quick-start unit status u does not need to be fixed in advance but making real-time decisions based on the wind power ξ. So, we have:

$$u(\boldsymbol{\xi})=\begin{cases}0, & \xi \geq 20\\ 1, & else\end{cases}$$

In this case, thermal unit generation is $p_A(\xi)=\begin{cases}0, & \xi \geq 20\\ 15+(5-\xi)^+, & else\end{cases}$ and windy unit generation is $p_w(\xi)=\begin{cases}\xi, & \xi \geq 20\\ 5-(5-\xi)^+, & else\end{cases}$, the expectation of generation cost is $E_{obj}=\left(\frac{20}{25}(900+60\times(5-\xi)^+)+\frac{5}{20}\times 0\right)\ \epsilon[720,960]$.

It can be seen from Fig. 1 and 2 that as the binary decision variables can also adapt to the wind power in real-time in our fully adaptive framework, the whole scheduling scheme is more willing to use the wind power and reduce the expectation of generation cost consequently. More advantages of our model will be delineated experimentally in Section IV.

III. The Proposed Fully Adaptive Framework

A. Fully Adaptive Framework

Here, we will introduce a sequential decision-making process in which uncertain wind power is observed at a given time, followed immediately (fully adaptive) by operational decisions (including $\boldsymbol{x^q}(\cdot)$ and $\boldsymbol{y}(\cdot)$) for a given period.

Therefore, our fully adaptive multistage DRO framework for uncertain generation scheduling can be expressed by the following formula:

$$\min\left\{\sum_{t=1}^{T}[\boldsymbol{c}_t^{\mathrm{T}}\boldsymbol{x}_t^s]+\min_{\boldsymbol{x^q},\boldsymbol{y}}\sup_{\mathbb{P}\in\Upsilon}\mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^{T}\left[\boldsymbol{c}_t^{\mathrm{T}}\boldsymbol{x}_t^q(\boldsymbol{\xi}_{[t_1]^\wedge})+\boldsymbol{d}_t^{\mathrm{T}}\boldsymbol{y}_t(\boldsymbol{\xi}_{[t_2]^\wedge})\right]\right]\right\} \quad (23)$$

$$\text{s.t.}\begin{cases}\sum_{\tau=1}^{t}\left[\boldsymbol{A}_\tau^t\boldsymbol{x}_\tau(\boldsymbol{\xi}_{[\tau_1]^\wedge})+\boldsymbol{B}_\tau^t\boldsymbol{y}_\tau(\boldsymbol{\xi}_{[\tau_2]^\wedge})\right]\leq \boldsymbol{D}^t\boldsymbol{\xi}_{[t]^\wedge}\\ \sum_{\tau=1}^{t}\left[\boldsymbol{\mathcal{A}}_\tau^t\boldsymbol{x}_\tau(\boldsymbol{\xi}_{[\tau_1]^\wedge})+\boldsymbol{\mathcal{B}}_\tau^t\boldsymbol{y}_\tau(\boldsymbol{\xi}_{[\tau_2]^\wedge})\right]=\boldsymbol{d}^t\end{cases}\quad \forall\boldsymbol{\xi}\in\Xi,\forall t\in[T]$$

where $\boldsymbol{x}_t^s$ and $\boldsymbol{x}_t^q$ is slow-start and quick-start unit, respectively, $\boldsymbol{\xi}:=\left[\xi_{0,0}=1;\xi_{i,t}=v_{i,t}^{\mathrm{w}}\right],\ t\in[T],\ t_1,t_2\leq t,\ \tau_1,\tau_2\leq\tau$. $\xi_{0,0}$ is introduced to represent affine functions of $\boldsymbol{v}^{\mathrm{w}}$[30], and $\boldsymbol{A}_\tau^t$, $\boldsymbol{B}_\tau^t$, $\boldsymbol{D}^t$, $\boldsymbol{\mathcal{A}}_\tau^t$, $\boldsymbol{\mathcal{B}}_\tau^t$, and $\boldsymbol{d}^t$ are the coefficient matrices/vectors associated with the constraints in time period t. $\mathbb{P}$ is the probability distribution of $\boldsymbol{\xi}$ which satisfied $\mathbb{P}(A):=\mathbb{P}^{\mathrm{w}}\left(\left\{\boldsymbol{\xi}_{[T]}\middle|\left[1;\boldsymbol{\xi}_{[T]}\right]\in A\right\}\right)$, $A\in\sigma(\Xi)$, Ξ is any closed convex set containing the support set of $\mathbb{P}$, Υ is the ambiguity set of $\mathbb{P}$. In this framework, we focus on finding binary functions $\boldsymbol{x}_t(\boldsymbol{\xi}_{[t_1]^\wedge})$ and real-valued functions $\boldsymbol{y}_t(\boldsymbol{\xi}_{[t_2]^\wedge})$ to solve the model.

Compared to (22), besides to continuous variables (dispatch decisions), the binary variables (quick-start unit status) in ours (23) also are modeled as functions of uncertain variables, which implies adaptivity and non-anticipativity in unit status since binary variables in t period only depend on uncertain variables in the previous periods without the later periods. Here, we provide a unified framework for typical DROs in the recent literature:

1) (22) is a special case of (23) when $t_1=0$ and $t_2=t$.

2) (21) is a special case of (23) when $t_1=0$ and $t_2=T$.

3) when $t_1\leq t, t_2=t$. It is the most important scene for our framework, which can be used to implement rolling scheduling. In this scene, with uncertain wind power are observed at a given time and the operational dispatch ($\boldsymbol{y}(\cdot)$) are made immediately, quick-start units can be also scheduled immediately ($t_1=t$), slow-start units can be scheduled several time periods after the observation ($t_1<t$).

Without loss of generality, for the convenience of discussion, this paper mainly discuss $t_1=t_2=t$, $\boldsymbol{x}_t$ includes $\boldsymbol{x}_t^s$ and $\boldsymbol{x}_t^q$, but everything discussed in this paper can be easily generalized to the general version of our framework (23).

B. Reformulation of the Proposed Model

In this subsection, we first use more flexibility *non-linear* mixed decision rules [24] to restricts the infinite space of the adaptive decisions $\boldsymbol{x}_t(\cdot)$ and $\boldsymbol{y}_t(\cdot)$. Secondly, we equivalently transform the DRO problem to MILP based on duality theory.

1) Data-driven uncertainty set

Given a set $\{\boldsymbol{P}^{\mathrm{w},j}\}_{j=1}^{M}$ of M samples, and $\bar{\boldsymbol{P}}^{\mathrm{w}}=\frac{1}{M}\sum_{j=1}^{M}\boldsymbol{P}^{\mathrm{w},j}$, $\boldsymbol{v}^{\mathrm{w},j}=\boldsymbol{P}^{\mathrm{w},j}-\bar{\boldsymbol{P}}^{\mathrm{w}}$, $\boldsymbol{\xi}^j:=[1;\boldsymbol{v}^{\mathrm{w},j}]$. Let $v_{i,t}^{\mathrm{w}-}=\min\{v_{i,t}^{\mathrm{w},j},j=1,\dots,M\}$, $v_{i,t}^{\mathrm{w}+}=\max\{v_{i,t}^{\mathrm{w},j},j=1,\dots,M\}$, $\boldsymbol{v}^{\mathrm{w}-}=\left[v_{[R],[T]}^{\mathrm{w}-}\right]^{\mathrm{T}}$, $\boldsymbol{v}^{\mathrm{w}+}=\left[v_{[R],[T]}^{\mathrm{w}+}\right]^{\mathrm{T}}$. Then, we have the data-driven support of $\mathbb{P}$:

$$\Xi=\left\{\boldsymbol{\xi}\middle|\boldsymbol{\xi}^-:=\begin{bmatrix}1\\ \boldsymbol{v}^{\mathrm{w}-}\end{bmatrix}\leq\boldsymbol{\xi}\leq\boldsymbol{\xi}^+:=\begin{bmatrix}1\\ \boldsymbol{v}^{\mathrm{w}+}\end{bmatrix}\right\} \quad (24)$$

Unbiased moment estimator [25] is used to estimate $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$.

2) Non-linear operator and lifted uncertainty set

Inserting $r_{i,t}-1$ breakpoints into the marginal support of each uncertain variable $\xi_{i,t}$ as follows:

$$v_{i,t}^{\mathrm{w}-}=p_{i,t}^0<p_{i,t}^1<\cdots<p_{i,t}^{r_{i,t}-1}<p_{i,t}^{r_{i,t}}=v_{i,t}^{\mathrm{w}+}. \quad (25)$$

Lifting operators $\bar{L}$ [24] and improved $\hat{L}$ are introduced to facilitate the construction of flexible mixed decision rules:

$$\bar{\xi}_{i,t}^j=\bar{L}_{i,t}^j(\xi_{i,t}):=\begin{cases}\xi_{i,t}, & \text{if } r_{i,t}=1, j=0\\ \left\{\min\left\{\xi_{i,t},p_{i,t}^j\right\}-p_{i,t}^{j-1}\right\}^+, & \text{if } r_{i,t}>1, j=2,\dots,r_{i,t}\end{cases} \quad (26)$$

$$\hat{\xi}_{i,t}^j=\hat{L}_{i,t}^j(\xi_{i,t}):=\begin{cases}1 & \text{if } r_{i,t}=1, j=0\\ \mathbb{I}(\xi_{i,t}\geq p_{i,t}^j) & \text{if } r_{i,t}>1, j=1,\dots,r_{i,t}-1\end{cases} \quad (27)$$

$$\bar{\xi}_{0,0}=\bar{L}_{0,0}(\xi_{0,0})=1 \quad (28)$$

$$\hat{\xi}_{0,0} = \hat{L}_{0,0}(\xi_{0,0}) = 1 \tag{29}$$

Then, we have the uncertainty set of $\tilde{\boldsymbol{\xi}}_{0,0}$, and $\tilde{\boldsymbol{\xi}}_{i,t}$:

$$\tilde{\Xi}_{0,0} \coloneqq \{\tilde{\boldsymbol{\xi}}_{0,0}\} \tag{30}$$

$$\tilde{\Xi}_{i,t} \coloneqq \{\tilde{\boldsymbol{\xi}}_{i,t} | \xi_{i,t} \in [\xi_{i,t}^-, \xi_{i,t}^+], \bar{\boldsymbol{\xi}}_{i,t}, \hat{\boldsymbol{\xi}}_{i,t}\} \tag{31}$$

$\tilde{\boldsymbol{\xi}}$ is separable in each component $\tilde{\boldsymbol{\xi}}_{i,t}$ and $\tilde{\boldsymbol{\xi}}_{0,0}$, consequently, the uncertainty set of $\tilde{\boldsymbol{\xi}}$ is

$$\tilde{\Xi} \coloneqq \{\tilde{\boldsymbol{\xi}} | \tilde{\boldsymbol{\xi}}_{0,0} \in \tilde{\Xi}_{0,0}, \tilde{\boldsymbol{\xi}}_{i,t} \in \tilde{\Xi}_{i,t}\} \tag{32}$$

$\tilde{\Xi}_{i,t}$ is an open set because of the discontinuity of $\hat{\boldsymbol{\xi}}_{i,t}$. For the sake of further reformulation, we're going to give the closed convex hull of $\tilde{\Xi}_{i,t}$ denote as $\mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi}_{i,t})\right)$ [19], [30].

$$\mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi}_{i,t})\right) = \left\{ \tilde{\boldsymbol{\xi}}_{i,t} \middle| \begin{matrix} \sum_{\widetilde{\boldsymbol{v}}_{i,t} \in \tilde{\mathcal{V}}_{i,t}^j} \lambda_{i,t}(\widetilde{\boldsymbol{v}}_{i,t}) = 1 \\ \tilde{\boldsymbol{\xi}}_{i,t} = \sum_{\widetilde{\boldsymbol{v}}_{i,t} \in \tilde{\mathcal{V}}_{i,t}^j} \lambda_{i,t}(\widetilde{\boldsymbol{v}}_{i,t}) \widetilde{\boldsymbol{v}}_{i,t} \\ \xi_{i,t} \in [\xi_{i,t}^-, \xi_{i,t}^+] \\ \lambda_{i,t}(\widetilde{\boldsymbol{v}}_{i,t}) \in \Re_+, \forall \widetilde{\boldsymbol{v}}_{i,t} \in \tilde{\mathcal{V}}_{i,t} \end{matrix} \right\} \tag{33}$$

$$\mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi}_{0,0})\right) = \{[1;1;1]\} \tag{34}$$

where $\lambda_{i,t}(\widetilde{\boldsymbol{v}}_{i,t})$ is the coefficient associated with vertex $\widetilde{\boldsymbol{v}}_{i,t}$

It can be seen from (30)-(32) that $\tilde{\Xi}$ is separable in each component $\tilde{\boldsymbol{\xi}}_{i,t}$ and $\tilde{\boldsymbol{\xi}}_{0,0}$. Therefore, we have

$$\mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi}_S)\right) = \left\{\tilde{\boldsymbol{\xi}} \middle| \tilde{\boldsymbol{\xi}}_{i,t} \in \mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi}_{i,t})\right), \forall (i,t) \in S\right\} \tag{35}$$

where $S \subseteq \{(0,0), (1,1), \dots, (i,t), \dots, (R,T)\}$.

3) Data-Driven Lifted ambiguity set

On the basis of uncertainty set, we construct the ambiguity set based on Wasserstein-metric through prior knowledge or historical data, which can provide a flexible framework for decision maker to model uncertainty through partial information. The Wasserstein-metric is defined as:

$$W(\mathbb{Q}, \mathbb{P}) = \inf_{\pi \in m(\tilde{\Xi} \times \tilde{\Xi})} \{\mathbb{E}_\pi[d(\xi^q, \xi^p)] : \xi^q \sim \mathbb{Q}, \xi^p \sim \mathbb{P}\} =$$

$$\inf_{\pi \in m(\tilde{\Xi} \times \tilde{\Xi})} \left\{ \int_{\tilde{\Xi} \times \tilde{\Xi}} d(\xi^q, \xi^p) \pi(\mathrm{d}\xi^q, \mathrm{d}\xi^p) : \pi\left(\hat{\Xi}, \mathrm{d}\xi^p\right) = \right.$$

$$\left. \mathbb{P}(\mathrm{d}\xi^p), \pi\left(\mathrm{d}\xi^q, \hat{\Xi}\right) = \mathbb{Q}(\mathrm{d}\xi^q) \right\} \tag{36}$$

$$d(\xi^q, \xi^p) = \|\xi^q - \xi^p\| \tag{37}$$

Where $\hat{\Xi}$ is a compact supporting space and $\mathbb{P}$, $\mathbb{Q}$ are two probability distribution. $m(\cdot)$ is the set of probability measures on a measurable space $(\cdot, \mathcal{F})$, and the distance between uncertain variables ξ^q and ξ^p is defined by $d(\xi^q, \xi^p)$. ξ^q and ξ^p are follow distribution $\mathbb{Q}$ and $\mathbb{P}$, respectively, and the infimum is taken over all joint distributions π with marginals $\mathbb{Q}$ and $\mathbb{P}$.

Given $\tilde{\boldsymbol{\xi}}$ and $\mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi})\right)$, the Wasserstein-metric on ambiguity set Υ can be expressed as [16]:

$$\mathbb{B}_\varepsilon(\widetilde{\mathbb{P}}_M) = \{\widetilde{\mathbb{Q}} | W(\widetilde{\mathbb{Q}}, \widetilde{\mathbb{P}}_M) \leq \varepsilon\} \tag{38}$$

Parameter ε plays an important role in the performance of ambiguity set, which can weigh the robustness and economy. Meanwhile, it can be calculated by statistical methods, such as the one given in reference [15]. For the sake of concision, we do not go into details of the calculation of ε here.

4) Mixed decision rules

We adopt mixed decision rules to restrict uncertain variables to be linear functions of the lifted parameters:

$$\boldsymbol{x}_t(\boldsymbol{\xi}_{[t]^\wedge}) = \hat{X}_{[t]^\wedge}^t \hat{\boldsymbol{\xi}}_{[t]^\wedge}, \forall t \in [T] \tag{39}$$

$$\boldsymbol{y}_t(\boldsymbol{\xi}_{[t]^\wedge}) = \bar{Y}_{[t]^\wedge}^t \bar{\boldsymbol{\xi}}_{[t]^\wedge}, \forall t \in [T] \tag{40}$$

Compared with classical decision rules that only model continuous variables [26], (39) and (40) further consider binary variables while maintaining linearity. Furthermore, the decision rule (40) grants continuous variables to follow commonly piecewise linear functions, which is essential in mixed integer problems. In next subsection, we address the linear inequality constraints of $(23)^2$.

5) Reformulation of inequality constraints

Firstly, decision rules (39) and (40) are substituted into $(23)^2$, then $\mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi}_{S^t})\right)$ is introduced, and constraints $(23)^2$ can be reformulated as:

$$\sum_{\tau=1}^t \left[\boldsymbol{B}_\tau^t \bar{Y}_{[\tau]^\wedge}^\tau \bar{\boldsymbol{\xi}}_{[\tau]^\wedge} + \boldsymbol{A}_\tau^t \hat{X}_{[\tau]^\wedge}^\tau \hat{\boldsymbol{\xi}}_{[\tau]^\wedge}\right] \leq \boldsymbol{D}^t \boldsymbol{\xi}_{[t]^\wedge}, \forall t \in [T] \tag{41}$$

where $S^t = \{(i,\tau) | i \in [R], \tau \in [t]\} \cup \{(0,0)\}$, $\forall \tilde{\boldsymbol{\xi}}_{[t]^\wedge} \in \mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi}_{S^t})\right)$, all of these conditions are met below and will not be described further.

Let $\vec{\tilde{\boldsymbol{\xi}}}_{[t]^\wedge} = [\boldsymbol{\xi}_{[t]^\wedge}; \bar{\boldsymbol{\xi}}_{[\tau]^\wedge}; \hat{\boldsymbol{\xi}}_{[\tau]^\wedge}]$, we rewrite $\mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi}_{S^t})\right)$ to be:

$$\mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi}_{S^t})\right) = \left\{ \vec{\tilde{\boldsymbol{\xi}}}_{[t]^\wedge} \middle| \begin{matrix} \boldsymbol{W}_{[t]^\wedge}^1 \vec{\tilde{\boldsymbol{\xi}}}_{[t]^\wedge} + \boldsymbol{U}_{[t]^\wedge}^1 \vec{\boldsymbol{\lambda}}_{[t]^\wedge} \geq \boldsymbol{h}_{[t]^\wedge}^1 \\ \boldsymbol{W}_{[t]^\wedge}^2 \vec{\tilde{\boldsymbol{\xi}}}_{[t]^\wedge} + \boldsymbol{U}_{[t]^\wedge}^2 \vec{\boldsymbol{\lambda}}_{[t]^\wedge} = \boldsymbol{h}_{[t]^\wedge}^2 \end{matrix} \right\} \tag{42}$$

(41) can be rewritten as:

$$Z_{[t]^\wedge}^t \vec{\tilde{\boldsymbol{\xi}}}_{[t]^\wedge} \geq 0, \forall t \in [T] \tag{43}$$

where

$$Z_{[t]^\wedge}^t = [-\boldsymbol{D}^t \quad \mathfrak{B}^t \quad \mathfrak{A}^t]$$

$$\mathfrak{B}^t = \left[\boldsymbol{B}_{[t]}^t \bar{Y}_0^{[t]} \quad \boldsymbol{B}_{[1:t]}^t \bar{Y}_1^{[1:t]} \quad \boldsymbol{B}_{[2:t]}^t \bar{Y}_2^{[2:t]} \quad \dots \quad \boldsymbol{B}_{[t:t]}^t \bar{Y}_t^{[t:t]}\right]$$

$$\mathfrak{A}^t = \left[\boldsymbol{A}_{[t]}^t \hat{X}_0^{[t]} \quad \boldsymbol{A}_{[1:t]}^t \hat{X}_1^{[1:t]} \quad \boldsymbol{A}_{[2:t]}^t \hat{X}_2^{[2:t]} \quad \dots \quad \boldsymbol{A}_{[t:t]}^t \hat{X}_t^{[t:t]}\right]$$

By applying standard robust counterpart reformulation techniques, (43) can be reconstructed as the following set of constraints:

$$\begin{cases} \Lambda_{[t]^\wedge}^k \boldsymbol{W}_{[t]^\wedge}^k = Z_{[t]^\wedge}^k \\ \left(\boldsymbol{U}_{[t]^\wedge}^k\right)^{\mathrm{T}} \Lambda_{[t]^\wedge}^k = 0 \\ \left(\boldsymbol{h}_{[t]^\wedge}^k\right)^{\mathrm{T}} \Lambda_{[t]^\wedge}^k \geq 0 \\ \Lambda_{[t]^\wedge}^1 \geq 0 \end{cases} \forall t \in [T], k = 1,2 \tag{44}$$

6) Reformulation of equality constraints

Similarly, constraints $(23)^3$ can be re-expressed as follows:

$$\sum_{\tau=1}^t \left[\boldsymbol{\mathcal{B}}_\tau^t \bar{Y}_{[\tau]^\wedge}^\tau \bar{\boldsymbol{\xi}}_{[\tau]^\wedge} + \boldsymbol{\mathcal{A}}_\tau^t \hat{X}_{[\tau]^\wedge}^\tau \hat{\boldsymbol{\xi}}_{[\tau]^\wedge}\right] = \boldsymbol{d}^t, \forall t \in [T] \tag{45}$$

(45) can be rewritten as

$$\mathcal{Z}_{[t]^\wedge}^t \vec{\tilde{\boldsymbol{\xi}}}_{[t]^\wedge} = \boldsymbol{d}^t, \forall t \in [T] \tag{46}$$

where $\mathcal{Z}_{[t]^\wedge}^t$ and $Z_{[t]^\wedge}^t$ are arranged in the same way.

$$\mathcal{Z}_{[t]^\wedge}^{t,j} \vec{\tilde{\boldsymbol{\xi}}}_{[t]^\wedge} = \boldsymbol{d}^{t,j}, \forall t \in [T] \tag{47}$$

$\mathcal{Z}_{[t]^\wedge}^{t,j}$, and $\boldsymbol{d}^{t,j}$ are the jth rows of $\mathcal{Z}_{[t]^\wedge}^t$ and $\boldsymbol{d}^t$ respectively.

For any $\tilde{\mathcal{V}}_{i,t}$, it is not very hard to verify that all the points in this set are linearly independent. The dimension of $\tilde{\mathcal{V}}_{i,t}$ is $2r_{i,t}$ and $|\tilde{\mathcal{V}}_{i,t}| = 2r_{i,t}$. Taking the separability of $\mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi}_{S^t})\right)$, we can construct $\mathrm{r} = \sum_{(i,t) \in S^t} 2r_{i,t}$ linearly independent vectors $(\vec{\tilde{\boldsymbol{\xi}}}_{[t]^\wedge}^k, k \in [\mathrm{r}])$ which belong to $\mathrm{conv}\left(\mathrm{cl}(\tilde{\Xi}_{S^t})\right)$.

So (47) imply that

$$\mathcal{Z}_0^{t,j} \boldsymbol{e} + \mathcal{Z}_{[t]}^{t,j} \vec{\tilde{\boldsymbol{\xi}}}_{[t]}^k = \boldsymbol{d}^{t,j}, \forall k \in [\mathrm{r}], \forall t \in [T] \tag{48}$$

Finally, (46) is equivalent to

$$\mathcal{Z}_0^t \boldsymbol{e} = \boldsymbol{d}^t$$

$$\mathcal{Z}_{[t]}^t = \boldsymbol{0}$$

7) Reformulation of DR objective function

Let $\ell\left(\vec{\tilde{\boldsymbol{\xi}}}\right)=\sum_{t=1}^{T}\left[\boldsymbol{c}_t^{\mathrm{T}}\boldsymbol{x}_t\left(\boldsymbol{\xi}_{[t]^{\wedge}}\right)+\boldsymbol{d}_t^{\mathrm{T}}\boldsymbol{y}_t\left(\boldsymbol{\xi}_{[t]^{\wedge}}\right)\right]=c^{\mathrm{T}}\vec{\tilde{\boldsymbol{\xi}}}$, (49)

where c^{T} and $Z_{[t]^{\wedge}}^{t}$ are arranged in the same way.

The objective function is ambitious to be reformulated because it is an optimization problem to guarantee worst-case expectation on ambiguity set. Fortunately, [16] developed the strong duality theory gives the reformulation of the worst-case expectation, i.e., the inner of (23) can be rewritten as:

$$\min_{\lambda_j,\beta\geq 0}\sum_{j=1}^{M}\lambda_j+\varepsilon\beta \tag{50}$$

$$\text{s.t.}\left\{\ell\left(\vec{\tilde{\boldsymbol{\xi}}}\right)-M\lambda_j-\beta d\left(\vec{\tilde{\boldsymbol{\xi}}},\vec{\tilde{\boldsymbol{\xi}}}^{j}\right)\leq 0,,\forall j=1\ldots,M\right.$$

Let $\tilde{\lambda}_j=M\lambda_j$, the constraint in (50) can be written:

$$\sup_{\vec{\tilde{\boldsymbol{\xi}}}\in\mathrm{conv}\left(\mathrm{cl}\left(\Xi_{s^t}\right)\right)}\left\{\ell\left(\vec{\tilde{\boldsymbol{\xi}}}\right)-\beta d\left(\vec{\tilde{\boldsymbol{\xi}}},\vec{\tilde{\boldsymbol{\xi}}}^{j}\right)\right\}\leq\tilde{\lambda}_j,\forall j=1\ldots,M \tag{51}$$

Lemma 1: $\omega d\left(\boldsymbol{\lambda},\boldsymbol{\lambda}^{j}\right)=\sup_{\|\boldsymbol{p}\|_{*}\leq\omega}\langle p,\boldsymbol{\lambda}-\boldsymbol{\lambda}^{j}\rangle$. (52)

Then, by using (52) and mixed decision rules (39) and (40) into (51), and interchanging the outer maximization over $\vec{\tilde{\boldsymbol{\xi}}}$ and the infimum over $\tilde{\boldsymbol{p}}^{j}$, which is allowed by the minimax theorem, we obtain the following reformulation of (51):

$$\inf_{\|\tilde{\boldsymbol{p}}^{j}\|_{*}\leq\beta}\left\{\langle\tilde{\boldsymbol{p}}^{j},\vec{\tilde{\boldsymbol{\xi}}}^{j}\rangle+S_{\mathrm{conv}\left(\mathrm{cl}\left(\Xi_{s^t}\right)\right)}\left(c^{\mathrm{T}}-\tilde{\boldsymbol{p}}^{j}\right)\right\} \tag{53}$$

Furthermore, the second part of (53) can be equivalented to:

$$\inf_{\boldsymbol{y}_j}\left\{\langle\boldsymbol{h},\boldsymbol{y}_j\rangle:[\boldsymbol{W}^{\mathrm{T}};\boldsymbol{U}^{\mathrm{T}}]\boldsymbol{y}_j=\left[c^{\mathrm{T}}-\tilde{\boldsymbol{p}}^{j};0\right],\boldsymbol{y}_j\leq 0\right\} \tag{54}$$

According to (54), we have

$$\tilde{\boldsymbol{p}}^{j}=c^{\mathrm{T}}-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}_j \tag{55}$$

According to (51), (53)-(55), (50) is equivalent to

$$\min_{\lambda_j,\beta\geq 0,\boldsymbol{y}_j\geq 0}\sum_{j=1}^{M}\lambda_j+\varepsilon\beta \tag{56}$$

$$\text{s.t.}\begin{cases}\langle c^{\mathrm{T}}-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}_j,\vec{\tilde{\boldsymbol{\xi}}}^{j}\rangle+\langle\boldsymbol{h},\boldsymbol{y}_j\rangle\leq\tilde{\lambda}_j\\ \left\|c^{\mathrm{T}}-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}_j\right\|_{*}\leq\beta & \forall j=1\ldots,M\\ \boldsymbol{U}^{\mathrm{T}}\boldsymbol{y}_j=0\end{cases}$$

C. Resulted DR-MDR Model

$$\min_{\lambda_j,\beta\geq 0,\boldsymbol{y}_j\geq 0,\Lambda_{[t]^{\wedge}}\geq 0}\sum_{j=1}^{M}\lambda_j+\varepsilon\beta \tag{57}$$

$$\text{s.t.}\begin{cases}\langle c^{\mathrm{T}}-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}_j,\vec{\tilde{\boldsymbol{\xi}}}^{j}\rangle+\langle\boldsymbol{h},\boldsymbol{y}_j\rangle\leq\tilde{\lambda}_j\\ \left\|c^{\mathrm{T}}-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}_j\right\|_{*}\leq\beta\\ \boldsymbol{U}^{\mathrm{T}}\boldsymbol{y}_j=0\\ \Lambda_{[t]^{\wedge}}\boldsymbol{W}_{[t]^{\wedge}}=Z_{[t]^{\wedge}}\\ \boldsymbol{U}_{[t]^{\wedge}}^{\mathrm{T}}\Lambda_{[t]^{\wedge}}=0\\ \boldsymbol{h}_{[t]^{\wedge}}^{\mathrm{T}}\Lambda_{[t]^{\wedge}}\geq 0\\ \mathcal{Z}_0^{t}\boldsymbol{e}=\boldsymbol{d}^{t}\\ \mathcal{Z}_{[t]}^{t}=\boldsymbol{0}\end{cases}\qquad\forall t\in[T],\forall j=[M]$$

IV. Numerical Results and Analysis

A. Simulation Setup

This paper selects IEEE test systems for simulations to authenticate the advantages and superiority of our model. The simulations are conducted on a laptop with Intel Core i5-4200M CPU@2.60GHz and 8GB RAM running Windows 10 operating system. Commercial solver MOSEK and CPLEX from Yalmip are used to solve the model. For the sake of brevity, data of systems are not listed here [31]. The codes and cases are available for freely download from GitHub [32]. We took one year's historical data as samples, and a sample set divided by monthly data, i.e., every 30 groups of daily historical wind data we took as a sample set. Specifically, Fig. 3. Shows the wind power in a sample set (30-samples). The X-axis represents scheduling periods of 24 hours, Z-axis represents the samples in the current group, and Y-axis represents the specific wind power of the corresponding samples in the corresponding period. It is not difficult to see from Fig. 3 that the more severe the curve fluctuation is, the more unstable the wind power is, and the more difficult the model is to adapt to such a complex scenario. For example, the wind power of Sample 30 is more unstable than Sample 28. In this chapter, we will use multiple sample sets to conduct simulation experiments to verify that the proposed model outperforms the traditional model in various scenarios.

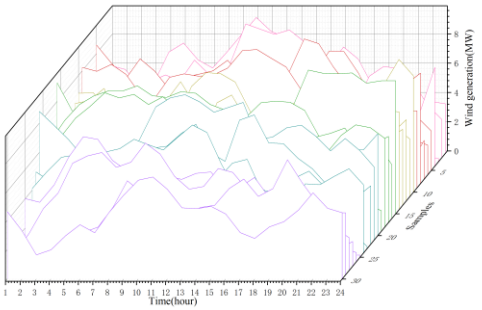

Fig. 3. Waterfall diagram of wind historical data

B. Comparison with typical model

For the sake of displaying the results clearly, the control experiment was carried out based on a 6-bus 12-periods test system in this subsection. Fig. 4 shows the total cost curve of our fully adaptive multistage distributionally robust model based on Wasserstein-metric and mixed decision rules (DR-MDR, (23)) and the typical multistage distributionally robust model (DR-TPMS, (21) and (22) refer to [12], [17], [18]) in different ambiguity set (the larger the ε is, the larger the ambiguity set is) and different sample sets. Different colors marked different sample sets (SPLs), and the cost curve of DR-MDR and DR-TPMS were marked by lines and dotted lines, respectively. Since the scheduling scheme obtained is applicable to any wind power in the ambiguity set, after obtaining the coefficients of integral $\hat{X}_{[t]^{\wedge}}^{t}$ and continuous $\hat{Y}_{[t]^{\wedge}}^{t}$ decision rules, wind power $\boldsymbol{\xi}_{[t]^{\wedge}}$ is put into (39), (40) (for example, unit state $u_{[G],t}\left(\boldsymbol{\xi}_{[t]^{\wedge}}\right)=\hat{X}_{[t]^{\wedge}}^{[G],t}\hat{\boldsymbol{\xi}}_{[t]^{\wedge}}$) to obtain all decision variables (the 18th samples in the sample set is selected in this subsection), different decision variables can be obtained for different wind power, then using (7) to calculate the generation cost. DR-TPMS is consistent, except that there is no integer decision coefficient. Unit output and wind generation comparison between DR-MDR and DR-TPMS in sample 3 with fixed $\varepsilon=2.25$ are shown in Fig. 5 and Fig. 6, respectively. (Actually, ε should not be fixed in practical application, in this part, we set it just for better comparison). And Fig. 7 shows the three dimensions histogram of load loss comparison of DR-MDR and DR-TPMS, X-axis represents periods, Y-axis represents the current model and node, and Z-axis represents the load loss.

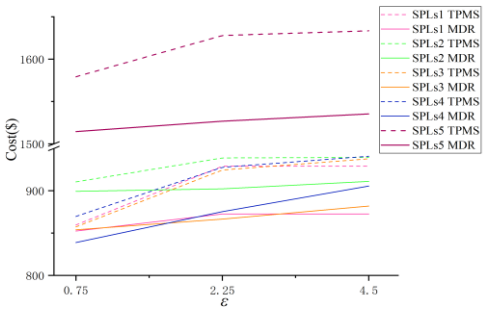

Fig. 4. System total cost comparison chart

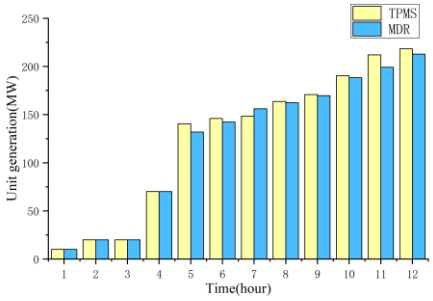

Fig. 5. Unit output comparison diagram

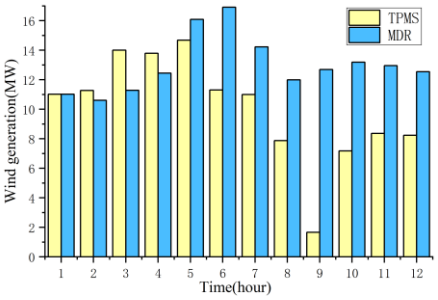

Fig. 6. Wind generation comparison diagram

Firstly, according to Fig. 4, the total cost of DR-MDR is less than the counterpart of DR-TPMS in any sample set. Such a conclusion is the same as the 1-1 system sample and the theoretical derivation of (22) and (23). Because of comparing to DR-TPMS, DR-MDR further processes quick-start unit decision variables and expands the feasible region, which is more likely to find a better solution, i.e., a scheduling scheme with lower generation costs. Meanwhile, when the wind fluctuation is greater (sample 5), the difference in generation cost between DR-MDR and DR-TPMS also increases significantly, which means that DR-MDR has a more obvious advantage than DR-TPMS in the case of unstable wind in the actual environment. In terms of ambiguity, the generation costs of DR-MDR and DR-TPMS increase with the increase of ε. Then DR-TPMS fixes the on/off state of units, therefore, DR-MDR makes more reasonable use of wind power to reduce the output of thermal units, which is consistent with the results in Fig. 5 and Fig. 6. DR-MDR's wind generation is more than DR-TPMS, while thermal generation is less than DR-TPMS in most periods. Take period 6 as an example, as Fig. 3 shows the wind power fluctuates greatly at this time (i.e., the variance is large, and the waterfall drop of each sample is large), which results in DR-MDR using more wind power generation than DR-TPMS. Correspondingly, the total thermal generation of DR-MDR also decreases in this period. However, in such a small-period test system, the state of DR-MDR and DR-TPMS do not have differences because the test system has a small number of periods. In order to meet the load requirements, the unit state scheduling is basically on as far as possible, and the state scheduling scheme with a larger number of periods may be different. It will better reflect the advantages of DR-MDR, which will be verified by a 24-period test system in the following subsection.

Finally, it is manifest from Fig. 6 and Fig. 7 that DR-MDR also reduces the load loss of each node in most periods. Unit 3 is a quick-start unit and as thermal units are not fully started in the first four periods, there is no significant difference between DR-MDR and DR-TPMS. However, with the change of time, there is more scheduling space after the thermal units start, and the difference in the corresponding load loss of the two models is more prominent. At the same time, the reduction of load loss not only reduces the total power generation cost of the DR-MDR scheme but also meets the load requirements better than DR-TPMS while controlling the power generation cost from the perspective of the user load.

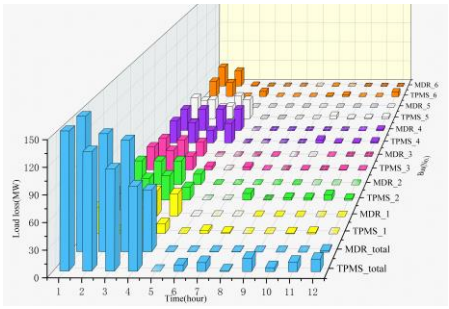

Fig. 7. Three dimensions histogram of load loss comparison

C. *Results on other IEEE test systems*

Previous experiments on the 1-1 and 6-12 systems verified that the proposed model could better adapt to wind power than the typical multistage model. In this subsection, we will conduct simulations on a larger test system to verify that the advantages of DR-MDR are more significant based on a larger system and more extended period.

Theoretically, the larger the system scale, the longer the scheduling period, the larger the feasible region of DR-MDR, and the more likely it is to find a better solution due to the application of MDR. However, as period increase the number of integer variables coefficient $\hat{X}^t_{[t]^\wedge}$ and continuous variables coefficient $\hat{Y}^t_{[t]^\wedge}$ will boost. In this case, the high computational cost is not conducive to simulation. Therefore, we will choose 12 hours as the scheduling period on the 30-bus system. Fig. 8 shows the total cost comparison between DR-MDR and DR-TPMS on 6-12, 6-24, and 30-12 test systems. At the same time, in order to verify that DR-MDR can generate different on/off unit state schemes compared with DR-TPMS. We selected a 6-24 system for simulation and put multiple wind power data and coefficient $\hat{X}^t_{[t]^\wedge}$ into (39) to get multiple unit state scheduling schemes $[u_{[G],1}, u_{[G],2} \dots u_{[G],t}]^1$, $[u_{[G],1}, \dots u_{[G],t}]^2, \dots, [u_{[G],1}, \dots u_{[G],t}]^n$. The results are shown in Fig. 9. The abscissa represents periods, the ordinate represents the on/off state of the unit, and 1 represents on, and 0 represents off.

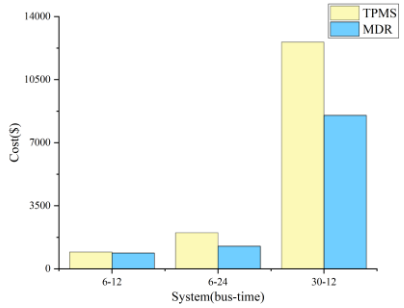

Fig. 8. Large-scale system total cost comparison diagram

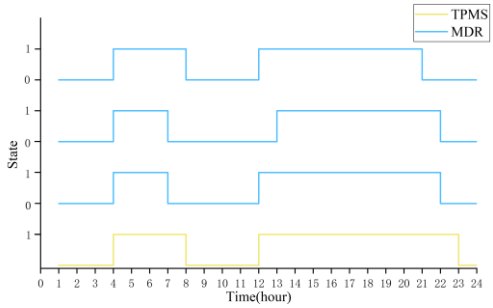

Fig. 9. Unit 1 statement diagram

It can be seen from Fig. 9 that since the on/off state of unit was determined before the scheduling of DR-TPMS, the line chart of the on/off state of DR-TPMS does not change with different wind power data. Therefore, there is only one yellow curve representing the on/off state of the DR-TPMS in Fig. 9. However, the on/off state scheduling scheme obtained by DR-MDR is an affine function of wind power, thus, when different wind power situations occur in the actual scenario, the scheduling scheme may also be different. As shown in Fig. 9, there are three different blue curves representing DR-MDR. It is not difficult to see that the scheduling scheme with different on/off state scheduling for different wind power can make better use of wind power. Then, it can be found from Fig. 8 that when the system is small (6-12), the total system costs of DR-MDR and DR-TPMS are similar. However, as the system scale or the number of periods increases, the amount of wind power that can be used for optimization increases dramatically, and the difference in total costs between DR-MDR and DR-TPMS increases. The results from the two case studies are consistent with our theoretical derivation given in Section III.

V. Conclusion

This paper proposed a fully adaptive distributionally robust multistage uncertain generation scheduling framework. Our model provides a unifying framework for several existing models, it can obtain various typical models by adjusting decision rules and relevant periods. Therefore, our model can better adopt wind power and curtail fuel's consumption, which lower greenhouse gas emissions and has a significant improvement for the environment. Then in order to address the intractability of the model, we used optimization theory and improved MDR to reformulate the original model to MILP. Experiments are carried out to show the superiority and availability of the proposed model. Although the proposed model significantly affected optimizing solutions under uncertainty, there is still room for further improvement: In this paper, we do not consider reducing the number of observed periods, which greatly increases the computational burden of the model and the periods far apart are meaningless to the scheduling scheme in actual scheduling.

References

[1] J Zhu. "Optimization of power system operation," Hoboken:Wiley, 2015.
[2] P F. Bouffard, F. D. Galiana, "Stochastic Security for Operations Planning With Significant Wind Power Generation," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 306-316, May 2008.
[3] Y. Zhang, H. H. -C. Iu, et al, "Cooperative Dispatch of BESS and Wind Power Generation Considering Carbon Emission Limitation in Australia," *IEEE Trans. Ind. Inform.*, vol. 11, no. 6, pp. 1313-1323, Dec. 2015.
[4] R. Singh, R. C. Bansal, "Optimization of an Autonomous Hybrid Renewable Energy System Using Reformed Electric System Cascade Analysis," *IEEE Trans. Ind. Inform.*, vol. 15, no. 1, pp. 399-409, Jan. 2019.
[5] H. Zhao, J. Zhao, et al "Cooperative Wind Farm Control With Deep Reinforcement Learning and Knowledge-Assisted Learning," *IEEE Trans. Ind. Inform.*, vol. 16, no. 11, pp. 6912-6921, Nov. 2020.
[6] Van Ackooij, W, et al. "Large-scale unit commitment under uncertainty: an updated literature survey," *Annals of Operations Research*, vol. 271, no. 1, pp.11-85, 2018.
[7] J. An, F. Yin, M. Wu, J. She and X. Chen, "Multisource Wind Speed Fusion Method for Short-Term Wind Power Prediction," *IEEE Trans. Ind. Inform.*, vol. 17, no. 9, pp. 5927-5937, Sept. 2021.
[8] HY Su et al, "Enhanced Wind Generation Forecast Using Robust Ensemble Learning," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 912-915, Jan. 2021.
[9] C. Peng, P. Xie, L. Pan and R. Yu, "Flexible Robust Optimization Dispatch for Hybrid Wind/Photovoltaic/Hydro/Thermal Power System," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 751-762, March 2016.
[10] C. Duan, L. Jiang, W. Fang, J. Liu and S. Liu, "Data-Driven Distributionally Robust Energy-Reserve-Storage Dispatch," *IEEE Trans. Ind. Inform.*, vol. 14, no. 7, pp. 2826-2836, July 2018.
[11] R. Zhu, H. Wei and X. Bai, "Wasserstein Metric Based Distributionally Robust Approximate Framework for Unit Commitment," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2991-3001, July 2019.
[12] Xiong, P., Jirutitijaroen, P., Singh, C. "A distributionally robust optimization model for unit commitment considering uncertain wind power generation," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 39-49, Jan. 2017.
[13] L. Yang, Y. Yang, G. Chen and Z. Y. Dong, "Distributionally Robust Framework and Its Approximations Based on Vector and Region Split for Self-scheduling of Generation Companies," *IEEE Trans. Ind. Inform.*, vol. 18, no. 8, pp. 5231-5241, Aug. 2022.
[14] Y. Chen, Q. Guo, et al, "A Distributionally Robust Optimization Model for Unit Commitment Based on Kullback–Leibler Divergence," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5147-5160, Sept. 2018.
[15] C. Duan, W. Fang, L. Jiang, L. Yao, and L. Jun, "Distributionally robust chance-constrained approximate AC-OPF with Wasserstein metric," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4924-4936, Sep. 2018.
[16] C. Zhao and Y. Guan, "Data-driven risk-averse stochastic optimization with Wasserstein metric," *Oper. Res. Lett.*, vol. 46, pp. 262-267, 2018.
[17] Á. Lorca, X A Sun, et al, "Multistage Adaptive Robust Optimization for the Unit Commitment Problem," *Oper. Res.*, vol. 64, no. 1, pp. 32-51, 2016.
[18] Y. Zhou, M. Shahidehpour, et al, "Multistage Robust Look-Ahead Unit Commitment with Probabilistic Forecasting in Multi-Carrier Energy Systems," *IEEE Trans. Sustain. Energy*, vol. 12, no. 1, pp. 70-82, Jan. 2021.
[19] Feng W, Feng Y, Zhang Q, "Multistage distributionally robust optimization for integrated production and maintenance scheduling," *AIChE Journal*, vol. 67, no. 9, May. 2021.
[20] S. Buhan and I. Çadırcı, "Multistage Wind-Electric Power Forecast by Using a Combination of Advanced Statistical Methods," *IEEE Trans. Ind. Inform.*, vol. 11, no. 5, pp. 1231-1242, Oct. 2015.
[21] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations," *Oper. Res.*, vol. 58, no. 4, pp. 902 – 917, 2010.
[22] Garstka S J, Wets J B. "On decision rules in stochastic programming," *Math. Program*, vol. 7, no. 1, pp.117-143, 1974.
[23] Ben-Tal, A., Goryashko, et al, A,"Adjustable robust solutions of uncertain linear programs," *Math. Program.*, vol. 99, no. 2, pp. 351–376, 2004.
[24] Georghiou A, Wiesemann W, Kuhn D, "Generalized decision rule approximations for stochastic programming via liftings," *Math. Program.*, vol.152, pp.301–338, 2015.
[25] Delage, E. and Ye, Y. "Distributionally robust optimization under moment

uncertainty with application to data-driven problems," *Oper. Res.*, vol. 58, no. 3, pp. 595-612, 2010.

[26] Chao Shang, Fengqi You. Distributionally robust optimization for planning and scheduling under uncertainty. *Computers and Chemical Engineering*, vol. 110, pp. 53-68, 2018.

[27] Meysam C, Jianqiang C, et al. "Computationally efficient Approximations for Distributionally Robust Optimization," *Opt online*, 2020.

[28] C. Zhao, J. Wang, and J. Watson, "Multi-stage robust unit commitment considering wind and demand response uncertainties," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2708-2717, Aug. 2013.

[29] Dyer M., Stougie L. "Computational complexity of stochastic programming problems," *Math. Program.*, vol. 106, no. 3, pp. 423–432, 2006.

[30] Bertsimas D, et al, "Binary decision rules for multistage adaptive mixed integer optimization," *Math. Program.*, vol. 167, no. 3, pp. 1-39, 2017.

[31] All IEEE test system (6 bus-118 bus). [Online]. Available: https://icseg.iti.illinois.edu/power-cases/.

[32] All datas and codes in this paper. [Online]. Available: https://github.com/linfengYang/Fully-Adaptive-Distributionally-Robust-Multi-stage-Framework-for-Uncertain-Unit-Commitment-Based-on/.