Data-Driven Affinely Adjustable Distributionally Robust Unit Commitment

Chao Duan, *Student Member, IEEE*, Lin Jiang, *Member, IEEE*, Wanliang Fang, and Jun Liu, *Member, IEEE*

***Abstract*—This paper proposes a data-driven affinely adjustable distributionally robust method for unit commitment considering uncertain load and renewable generation forecasting errors. The proposed formulation minimizes expected total operation costs, including the costs of generation, reserve, wind curtailment, and load shedding, while guaranteeing the system security. Without any presumption about the probability distribution of the uncertainties, the proposed method constructs an ambiguity set of distributions using historical data and immunizes the operation strategies against the worst case distribution in the ambiguity set. The more historical data is available, the smaller the ambiguity set is and the less conservative the solution is. The formulation is finally cast into a mixed integer linear programming whose scale remains unchanged as the amount of historical data increases. Numerical results and Monte Carlo simulations on the 118- and 1888-bus systems demonstrate the favorable features of the proposed method.**

***Index Terms*—Ambiguity, chance constraints, distributionally robust optimization, uncertainty, unit commitment.**

NOMENCLATURE

$\mathcal{B}, \mathcal{L}, \mathcal{T}$	Set of all buses, lines, and time periods.
$\mathcal{G}_b$	Set of all generators at bus b.
SU_i^b/SD_i^b	Start-up/shut-down costs of unit i at bus b.
$F_i^b(\cdot)$	Cost function of unit i at bus b.
$R_i^{b,up}/R_i^{b,dn}$	Upward/downward reserve availability price of unit i at bus b.
$Q_i^{b,up}/Q_i^{b,dn}$	Upward/downward reserve procurement price of unit i at bus b.
C_{ls}/C_{wc}	Penalty price of load shedding/wind curtailment.
MU_i^b/MD_i^b	Minimum up/down time of unit i at bus b.
RU_i^b/RD_i^b	Ramp up/down rate limit of unit i at bus b.
$\overline{RU}_i^b/\overline{RD}_i^b$	Start-up ramp-up/shut-down ramp-down rate limit of unit i at bus b.
L_i^b/U_i^b	Output power lower/upper bound of unit i at bus b.
C_l	Capacity of transmission line l.
K_l^b	Load shift factor from bus b to line l.
w_{it}^b	Binary decision variable: "1" if unit i at bus b is on in time t; "0" otherwise.
u_{it}^b	Binary decision variable: "1" if unit i at bus b is started up in time t; "0" otherwise.
v_{it}^b	Binary decision variable: "1" if unit i at bus b is shut down in time t; "0" otherwise.
x_{it}^b	AGC setting point of unit i at bus b in time t.
a_{it}^b	AGC participation factor of unit i at bus b in time t.
$r_{it}^{b,up}$	Upward reserve of unit i at bus b in time t.
$r_{it}^{b,dn}$	Downward reserve of unit i at bus b in time t.
$\tilde{p}_{it}^b$	Uncertain actual power output of unit i at bus b in time t.
$\tilde{l}_t^b$	Uncertain composite load at bus b in time t.
$\hat{l}_t^b$	Forecasted composite load at bus b in time t.
$\mathbb{P}$	A probability measure/distribution.
$\mathbb{E}_{\mathbb{P}}$	Expectation respect to probability measure $\mathbb{P}$.
$\mathcal{P}_0(\mathcal{S})$	Set of all probability measures with support $\mathcal{S}$.
$I_{\mathcal{S}}(\cdot)$	Indicator function of set $\mathcal{S}$, i.e., $I_{\mathcal{S}}(x) = 1$ when $x \in \mathcal{S}$ and 0 otherwise.
$(x)^+$	$\max\{x, 0\}$.

Manuscript received August 22, 2016; revised January 10, 2017, April 3, 2017, and June 29, 2017; accepted August 15, 2017. Date of publication August 18, 2017; date of current version February 16, 2018. This work was supported in part by National Key Research and Development Program of China under Grant 2016YFB0901903, in part by National Natural Science Foundation of China under Grant 51428702 and in part by Engineering and Physical Sciences Research Council under Grant EP/L014351/1. Paper no. TPWRS-01279-2016. *(Corresponding author: Wanliang Fang.)*

C. Duan is with the Department of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China, and also with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, U.K. (e-mail: duanchao@stu.xjtu.edu.cn).

L. Jiang is with Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, U.K. (e-mail: ljiang@liverpool.ac.uk).

W. Fang and J. Liu are with the Department of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: eewlfang@mail.xjtu.edu.cn; eeliujun@mail.xjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2017.2741506

I. INTRODUCTION

UNIT Commitment (UC) aims at reducing costs and improving reliability by optimal scheduling and dispatching generation units. Integration of intermittent renewable energy and market-driven operation have brought uncertainties to both generation and demand sides. It is, therefore, necessary to incorporate uncertainties into UC. Vast literatures are devoted to the stochastic optimization for UC [1]. Among all the methods proposed so far, the stochastic programming (SP) and the robust optimization (RO) attract the most attention.

SP approach assumes operational uncertainties follow predefined probability distributions which can be learned from historical data. The objective is to minimize the expectation of generation costs. Since continuous distributions are usually numerically intractable, they are replaced by discrete scenarios. Hence the objective becomes minimizing the weighted-average generation costs over parallel scenarios for two-stage problems

[2]–[4] or a scenario tree for multi-stage problems [5]–[7]. The solution quality of SP relies on the representativeness of selected scenarios. Monte Carlo simulation based method is usually employed to generate scenarios [3], [7]. A huge number of scenarios are often required to comprehensively represent the underlying stochastic nature [8], which results in prohibitively high computational burden [1], [8]. Therefore, SP is often equipped with scenario reduction methods to control computational complexity [7], [8]. However, the dilemma between quality and complexity of scenarios always exists.

In contrast to SP, RO does not require any probabilistic information of the uncertainties. Instead, randomness is represented by a deterministic uncertainty set containing worst-case scenarios. The objective is minimizing the worst-case costs regarding all possible realizations of the uncertainty set. The two-stage adaptive optimization framework has been the subject of many contributions [9]–[11]. One difficulty appears in this framework is that the robust counterpart of the second-stage problem is bilinear and non-convex. This bilinear problem can be either solved by the outer approximation method [9] (only local optimality is guaranteed) or rewritten into a mixed integer linear program (MILP) using the big-M method [10], [11] (under some assumptions on the uncertainty set [11]) and solved by generic MILP solvers. Although RO provides security against the worst-case scenario, it may also yield over-conservative solutions resulting from the sheer ignorance of underlying probabilistic information. Paper [12] proposes a UC formulation combining SP and RO to melt the bright sides of both approaches.

Note that the SP assumes the underlying probability distribution of uncertainties to be precisely known, whereas RO ignores the probabilistic information. In practice, the probability distribution truly exists but must be estimated from historical data and is, therefore, itself uncertain. To better modeling and tackling uncertainties, in several very recent studies, distributionally robust optimization (DRO) has been introduced to power system optimization problems including unit commitment [13], [14], energy reverse dispatch [15], [16], reverse scheduling [17], [18] and DC optimal power flow [19], [20]. DRO assumes that the true probability distribution of uncertain parameters lies in an ambiguity set (of probability distributions) and immunizes the operation strategies against all distributions in the ambiguity set. Different ways to construct the ambiguity set leads to different DRO approaches with the different degree of conservatism and computational efficiency. Paper [13] adopted a scenario-based approach where the random variable representing wind generation is assumed to have finite support. Statistical inference technique is employed to construct ambiguity sets for the discrete distributions. This method is data-driven and more data leads to the less conservative solution. Other papers assume that the random variables have continuous distributions. In [14], the support of a one-dimensional random variable is partitioned into several segments, and the ambiguity set imposes an upper bound for the expectation in each segment. As the number of segments increases, the probability distribution can be characterized with more details. The ambiguity sets employed in [15]–[18] and [20] are the sets of all probability distributions with given mean and covariance. Paper [16] further assumes unimodality to reduce conservatism, and reference [20] further considers the uncertainties of mean and covariance. Different from all others, paper [19], [21] assumes the distribution type is known as a priori while the mean and covariance are subject to uncertainties. To clear the jungle of ambiguity sets, we raise four criteria to judge the quality of ambiguity sets.

1) *Tractability:* DRO problems are reformulated as deterministic optimization problems to be solved by numerical methods. The numerical tractability of the corresponding deterministic problems decides the solvability and practicability of the DRO approach. Therefore, the designed ambiguity set must allow a tractable and efficient reformulation of the DRO problems. For example, the DRO problems in [13] and [14] are reformulated as (mixed integer) linear programmings (LP) whereas those in [15]–[18] and [20] are reformulated as semidefinite programmings (SDP). LP is much more tractable than SDP.
2) *Statistical Foundation:* Is there a sound statistical foundation for determining the parameters of the ambiguity set? For instance, the ambiguity set in [13] is based on the statistical inference which guarantees that the ambiguity set contains the true distribution with given confidence level. In contrast, no theory is provided in [14] to guide the parameter selection.
3) *Scalability:* How does the computational burden change when encoding more detailed probabilistic information in the ambiguity set? For example, the numbers of decision variables and constraints of the deterministic reformulations grow linearly with the number of bins in [13] and the number of segments in [14].
4) *Data-exploiting Ability:* Can the ambiguity set become smaller when more historical data is available? Intuitively, the more data we have, the more accurate we can deduce about the underlying true distribution, which leads to smaller ambiguity set. DRO with smaller ambiguity set results in the less conservative solution. Such data-exploiting ability is evident in the ambiguity set in [13].

This paper discusses a data-driven affinely adjustable distributionally robust unit commitment (AA-DRUC). The contribution is twofold. Firstly, we propose a novel ambiguity set based on a non-parametric confidence band of the cumulative distribution function (CDF) of the random variable. Secondly, assuming the generation units respond affinely to the total forecasting error of renewable generation, we present a UC problem formulation that minimizes the expected operation costs corresponding to the worst-case distribution in the proposed ambiguity set while explicitly considers spinning reserve, wind curtailment, and load shedding. Distributionally robust chance constraints are employed to guarantee reserve and transmission adequacy. The proposed method possesses the following features:

1) The problem is finally formulated as a mixed integer linear programming (MILP) for which off-the-shelf solvers are available. (tractability)
2) All parameters in the ambiguity set are automatically obtained through non-parametric inference. (statistical foundation)

3) The method is data-driven in the sense that no prior knowledge about the probability distribution of the uncertainties is needed and the historical data is directly incorporated in the solution process. The more historical data is available, the less conservative the solution is. (data-exploiting ability)
4) The scale of the MILP remains unchanged as the amount of historical data and the number of uncertain renewable sources increase. (scalability)

The rest of this paper is organized as follows. Section II introduces the ambiguity set constructed from confidence band of CDF. Section III presents the problem formulation and solution approach of the AA-DRUC. Case studies are reported in Section IV. Finally, we draw the conclusion and make discussion in Section V.

II. Ambiguity Set for Univariate Distribution

Consider a one-dimensional random variable ξ whose probability distribution is unknown whereas sample set S is available. The ambiguity set $\mathcal{P}$ is a set of probability distributions consistent with observed sample set S. $\mathcal{P}$ should have the following desired properties: 1) $\mathcal{P}$ contains the underlying true probability distribution; 2) to reduce conservatism, $\mathcal{P}$ can be made as small as possible by incorporating more observed data; 3) the structure of $\mathcal{P}$ allows the reformulation of distributionally robust optimization problems into tractable deterministic problems. In this section, we provide one approach to construct the ambiguity set $\mathcal{P}$ with above properties based on confidence bands for cumulative distribution function (CDF) from non-parametric statistics.

Let $F(x) = \mathbb{P}^*\{\xi \le x\}$ be the CDF of true distribution $\mathbb{P}^*$. The $1-\alpha$ confidence bands for $F(x)$ is a pair of sample-dependent functions $\underline{P}(x)$ and $\overline{P}(x)$ for which $\underline{P}(x) \le F(x) \le \overline{P}(x)$, $\forall x \in \mathbb{R}$ with probability $1-\alpha$ over the choice of sample set S. Deriving finite sample confidence bands of CDF is a basic problem in non-parametric statistics [22]. The most widely used method is based on Kolmogorov-Smirnov (KS) statistic [22], but the bands obtained by KS test are well-known to be unfavorably wide in the tails of the distribution. Recently, the Dirichlet method provides even sensitivity in different parts of the distribution and the closed-form approximation formula given in [23] make this method commendably easy to use. We summarize the method from [23] in the following Lemma.

Lemma 1: ([23]) Let $S = \{\hat{\xi}^{(1)}, \hat{\xi}^{(2)}, \ldots, \hat{\xi}^{(n)}\}$ be the ascendingly ordered sample set of random variable ξ generated independently according to true distribution $\mathbb{P}^*$ with continuous CDF $F(x)$. $B^{\alpha}_{k,n}$ denotes the α-quantile of the $\beta(k, n+1-k)$ distribution. For given n and α, define $\underline{p}_k = B^{\tilde{\alpha}/2}_{k,n}$ and $\overline{p}_k = B^{1-\tilde{\alpha}/2}_{k,n}$ where

$$\tilde{\alpha} = \exp\left(-c_1(\alpha) - c_2(\alpha)\sqrt{\ln[\ln(n)]} - c_3(\alpha)[\ln(n)]^{c_4(\alpha)}\right) \tag{1}$$

with $c_1(\alpha) = -2.75 - 1.04\ln(\alpha)$, $c_2(\alpha) = 4.76 - 1.20\alpha$, $c_3(\alpha) = 1.15 - 2.39\alpha$, and $c_4(\alpha) = -3.96 + 1.72\alpha^{0.171}$. Add $\hat{\xi}^{(0)} = -\infty$ and $\hat{\xi}^{(n+1)} = \infty$ to the ascending sequence of the sample set S, and define $\underline{p}_0 = 0$ and $\overline{p}_{n+1} = 1$. Then

$$\underline{P}(x) = \max\{\underline{p}_k : \hat{\xi}^{(k)} \le x\} \tag{2}$$

$$\overline{P}(x) = \min\{\overline{p}_k : \hat{\xi}^{(k)} \le x\} \tag{3}$$

are the $1-\alpha$ confidence bands for $F(x)$.

The underlying principle for Lemma 1 is that the spacings $F(\hat{\xi}^{(1)})$, $F(\hat{\xi}^{(2)}) - F(\hat{\xi}^{(1)}), \ldots$, $F(\hat{\xi}^{(n)}) - F(\hat{\xi}^{(n-1)})$ are random variable having n-variate Dirichlet distribution $D(1, \ldots, 1; 1)$ if S is independently sampled from an identical continuous distribution. Equivalently, the random variables $F(\hat{\xi}^{(1)})$, $F(\hat{\xi}^{(2)})$, $\ldots$, $F(\hat{\xi}^{(n)})$ follow the ordered n-variate Dirichlet distribution $D^*(1, \ldots, 1; 1)$. Then marginal distributions are $F(\hat{\xi}^{(k)}) \sim \beta(k, n+1-k)$, for $k = 1, \ldots, n$. Finally, the ordered Dirichlet distribution $D^*(1, \cdots, 1; 1)$ determines the point-wise $1-\tilde{\alpha}(\alpha, n)$ coverage of $F(\hat{\xi}^{(k)})$ that yields the overall $1-\alpha$ coverage of the CDF. Goldman and Kaplan gives the close-form approximation formula (1) for the mapping $\tilde{\alpha}(\alpha, n)$ through extensive simulations [23].

Therefore, given sample set S and significance level $1-\alpha$, the confidence bands of CDF can be readily computed using Lemma 1. α is generally set to be equal to or smaller than 0.1 in engineering practice, and it is set to be 0.05 in this paper. Note that $\underline{P}(x)$ and $\overline{P}(x)$ have the following properties: 1) they are stair-step functions that take values $\underline{p}_k$ and $\overline{p}_k$ at $\hat{\xi}^{(k)}$, respectively; 2) the empirical CDF $\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{\hat{\xi}^{(i)} \le x\}}$ is lower and upper bounded by $\underline{P}(x)$ and $\overline{P}(x)$, i.e. $\underline{P}(x) \le \hat{F}(x) \le \overline{P}(x)$; 3) as the size of the sample set $n \to \infty$, $\sup|\underline{P}(x) - \overline{P}(x)| \to 0$. In other words, $\underline{P}(x)$ and $\overline{P}(x)$ represent the reliable information that can be extracted from finite samples and the information becomes more and more accurate as the size of the sample set grows. Fig. 1 illustrates the evolution of the obtained CDF confidence bands as the size of the sample set increases.

Note that confidence bands for CDF do not contain the support information of the random variable. We further introduce Devroye-Wise method [24] to estimate the support $[\underline{\xi}, \overline{\xi}]$. Define $\delta = \max_{2 \le i \le n} |\hat{\xi}^{(i+1)} - \hat{\xi}^{(i)}|$. It is suggested in [25] to choose $[\underline{\xi}, \overline{\xi}] = [\hat{\xi}^{(1)} - \delta/2, \hat{\xi}^{(n)} + \delta/2]$ which is proved to converge to the true support in probability.

Based on the confidence bands for CDF and the estimated support, the ambiguity set $\mathcal{P}$ employed in this paper takes the form

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0([\underline{\xi}, \overline{\xi}]) \,\middle|\, \mathbb{P}[\xi \le \hat{\xi}^{(k)}] \in [\underline{p}_k, \overline{p}_k],\ k = 1, \ldots, n \right\} \tag{4}$$

where $\mathcal{P}_0([\underline{\xi}, \overline{\xi}])$ denotes the set of all probability measures whose supports are the interval $[\underline{\xi}, \overline{\xi}]$. The proposed structure of ambiguity $\mathcal{P}$ is designed to encode the information from confidence bands for CDF and does not assume any prior knowledge about the distribution type. Due to the convergence property of the confidence bands shown in Fig. 1, the ambiguity set $\mathcal{P}$ is made smaller and smaller by incorporating more and more historical data. Moreover, the structure defined in (4) allows very efficient reformulation of distributionally robust optimization

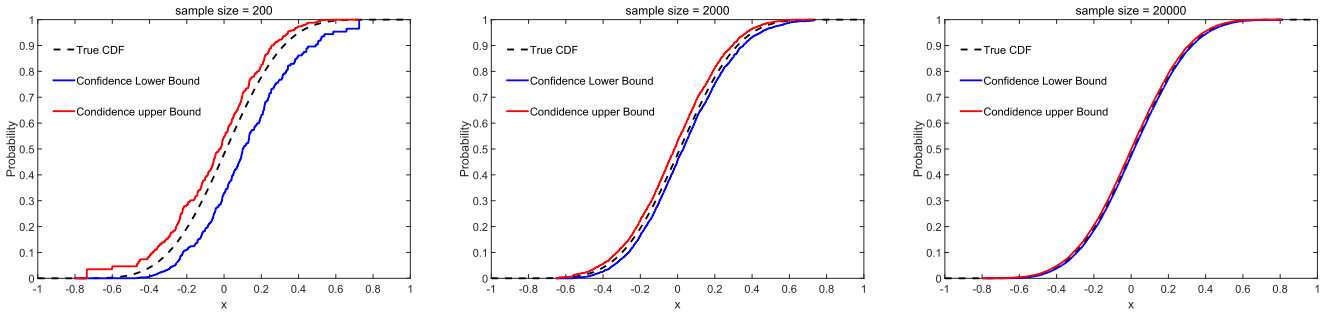

Fig. 1. Confidence bands of CDF constructed from 200, 2000 and 20000 samples of the random variable.

problems, which will be analyzed in Section III. Compared with uncertainty set used in RO, ambiguity set is a set of probability distributions (measures) while uncertainty set is a set of possible realizations of the random variable. Analogous to RO which considers the worst-cast realization in the uncertainty set, the proposed DRO considers the worst-case distribution in the ambiguity set.

III. Affinely Adjustable Distributionally Robust Unit Commitment

A. Feasible Set

The constraints for the binary commitment variables are as follows: $\forall i \in \mathcal{G}_b, \forall b \in \mathcal{B}, \forall t \in \mathcal{T}$

$$-w^b_{i(t-1)} + w^b_{it} - w^b_{ik} \le 0, \quad \forall k \in [t+1, \min\{t + MU^b_i - 1, T\}] \tag{5}$$

$$w^b_{i(t-1)} - w^b_{it} + w^b_{ik} \le 1, \quad \forall k \in [t+1, \min\{t + MD^b_i - 1, T\}] \tag{6}$$

$$-w^b_{i(t-1)} + w^b_{it} - u^b_{it} \le 0 \tag{7}$$

$$w^b_{i(t-1)} - w^b_{it} - v^b_{it} \le 0 \tag{8}$$

$$w^b_{it}, u^b_{it}, v^b_{it} \in \{0, 1\} \tag{9}$$

where constraints (5) and (6) represent the minimum up-time and minimum down-time restrictions, and constraints (7) and (8) indicate the relationship between on/off status and start up/down operations.

As for the continuous dispatch problem, we adopt the affinely adjustable approach [21], [26]–[29], i.e. the actual power outputs of generators respond affinely to the total forecasting error of composite loads:

$$\tilde{p}^b_{it} = x^b_{it} + a^b_{it} \sum_{b \in \mathcal{B}} (\tilde{l}^b_t - \hat{l}^b_t) \tag{10}$$

where x^b_{it} is the generator setting point and a^b_{it} is the generator participation factor in response to the total difference between uncertain composite load $\tilde{l}^b_t$ and its forecast value $\hat{l}^b_t$. This affine policy has its own limitations. First, the affine policy is only a conservative approximation to the optimal recourse action. Second, only system-level probabilistic information is exploited. However, it possesses some indispensable advantages as follows. 1) Tractability: the robust counterparts of affinely adjustable approach is usually tractable convex problems whereas those of fully adjustable approach are non-convex problems. 2) Scalability: due to such aggregated treatment with uncertainties, we only need to consider a one-dimensional rather than high-dimensional random variable in each time period. Therefore, the computation burden does not increase with the number of uncertain sources; moreover, thanks to the designed ambiguity set and the related reformulation technique discussed later, the computation burden remains unchanged when using more historical data. 3) Practicability: due to the system operators ability to aggregate uncertainty across all renewable sources, the system-wide forecast is much more accurate. System-wide wind and load prediction is usually used for day-ahead generation scheduling in practice. In addition, affinely adjustable approach is directly compatible with automatic generation control (AGC) systems where generators respond to area control error (ACE) according to participation factors [30].

To reduce the dimensions of random variables, define

$$\tilde{s}_t = \sum_{b \in \mathcal{B}} (\tilde{l}^b_t - \hat{l}^b_t) \tag{11}$$

$$\tilde{h}^l_t = \sum_{b \in \mathcal{B}} K^b_l (\tilde{l}^b_t - \hat{l}^b_t) \tag{12}$$

Let $\mathbb{P}^l_t$ denote the probability distribution for 2-dimensional random variable $(\tilde{s}_t, \tilde{h}^l_t)$ with marginal distributions $\mathbb{P}^s_t$ and $\mathbb{P}^h_t$. Using the historical data of $\tilde{s}_t$ and $\tilde{h}^l_t$, we can construct the ambiguity sets as defined in (4) for $\mathbb{P}^s_t$ and $\mathbb{P}^h_t$, denoted as $\mathcal{P}^s_t$ and $\mathcal{P}^h_t$, respectively. Then the ambiguity set of joint distribution $\mathbb{P}^l_t$ is defined as $\mathcal{P}^l_t = \{\mathbb{P}^l_t \in \mathcal{P}_0(\mathcal{R}^2) | \mathbb{P}^s_t \in \mathcal{P}^s_t, \mathbb{P}^h_t \in \mathcal{P}^h_t\}$.

The feasible set of the economic dispatch is described as follows: $\forall i \in \mathcal{G}_b, \forall b \in \mathcal{B}, \forall l \in \mathcal{L}, \forall t \in \mathcal{T}$

$$\sum_{b \in \mathcal{B}} \sum_{i \in \mathcal{G}_b} x^b_{it} - \sum_{b \in \mathcal{B}} \tilde{l}^b_t = 0 \tag{13}$$

$$\sum_{b \in \mathcal{B}} \sum_{i \in \mathcal{G}_b} a^b_{it} = 1 \tag{14}$$

$$0 \le a^b_{it} \le w^b_{it} \tag{15}$$

$$L^b_i w^b_{it} + r^{b,dn}_{it} \le x^b_{it} \le U^b_i w^b_{it} - r^{b,up}_{it} \tag{16}$$

$$(x^b_{it}+r^{b,up}_{it})-\left(x^b_{i(t-1)}-r^{b,dn}_{i(t-1)}\right)\le\left(2-w^b_{i(t-1)}-w^b_{it}\right)\overline{RU}^b_i+\left(1+w^b_{i(t-1)}-w^b_{it}\right)RU^b_i \tag{17}$$

$$\left(x^b_{i(t-1)}+r^{b,up}_{i(t-1)}\right)-\left(x^b_{it}-r^{b,dn}_{it}\right)\le\left(2-w^b_{i(t-1)}-w^b_{it}\right)\overline{RD}^b_i+\left(1-w^b_{i(t-1)}+w^b_{it}\right)RD^b_i \tag{18}$$

$$\inf_{\mathbb{P}^s_t\in\mathcal{P}^s_t}\mathbb{P}^s_t\left[-r^{b,dn}_{it}\le a^b_{it}\tilde{s}_t\le r^{b,up}_{it},\forall i\in\mathcal{G}_b,b\in\mathcal{B}\right]\ge 1-\beta \tag{19}$$

$$\inf_{\mathbb{P}^l_t\in\mathcal{P}^l_t}\mathbb{P}^l_t\left[-C_l\le\sum_{b\in\mathcal{B}}K^b_l\left(\sum_{i\in\mathcal{G}_b}(x^b_{it}+a^b_{it}\tilde{s}_t)-\hat{l}^b_t\right)-\tilde{h}^l_t\le C_l\right]\ge 1-\beta-\gamma,\ \forall l\in\mathcal{L} \tag{20}$$

Equality constraints (13) and (14) together ensure the total generation-consumption balance at every time period in the presence of forecasting errors of composite loads. Constraint (15) enforces limits on generator participation factors. Constraints (16)–(18) together ensure upward and downward reserve of each generator are actually procurable considering generator capacity and ramp rate limits. Distributionally robust chance constraints (DRCC) (19) and (20) guarantee the adequacy of the generator reserve and transmission line capacity with high probability. DRCC ensures the system reliability for all probability distributions in the ambiguity sets, which provide robustness for the system operation.

The DRCC (19) and (20) admit deterministic safe approximation in light of the definition of the ambiguity set (4), which is revealed in the following two lemmas.

Lemma 2: Let $\underline{S}_t(x)$ and $\overline{S}_t(x)$ be the confidence bands for the CDF of random variable $\tilde{s}_t$ defined in (2) and (3). If $-r^{b,dn}_{it}\le a^b_{it}\tilde{s}_t\le r^{b,up}_{it},\forall i\in\mathcal{G}_b,b\in\mathcal{B},\forall\tilde{s}_t\in[\underline{s}'_t,\overline{s}_t]$ where $\underline{s}'_t=\overline{S}_t^{-1}(\beta_1)$ and $\overline{s}'_t=\underline{S}_t^{-1}(1-\beta_2)$ with $\beta_1+\beta_2=\beta$, then DRCC (19) is satisfied.

Proof: Just need to notice that $\forall\mathbb{P}^s_t\in\mathcal{P}^s_t$,

$$\begin{aligned}&\mathbb{P}^s_t[\tilde{s}_t\notin[\underline{s}'_t,\overline{s}'_t]]\\&=\mathbb{P}^s_t[\tilde{s}_t<\underline{s}'_t]+\mathbb{P}^s_t[\tilde{s}_t>\overline{s}'_t]\\&\le\overline{S}_t(\underline{s}'_t)+1-\underline{S}_t(\overline{s}'_t)=\beta.\end{aligned} \tag{21}$$

■

Therefore, DRCC (19) can be safely replaced by the deterministic robust counterpart: $\forall i\in\mathcal{G}_b,b\in\mathcal{B}$

$$-r^{b,dn}_{it}\le a^b_{it}\underline{s}'_t \tag{22a}$$

$$a^b_{it}\overline{s}'_t\le r^{b,up}_{it}. \tag{22b}$$

Lemma 3: Let $\underline{S}_t(x)$ and $\overline{S}_t(x)$ be the confidence bands for the CDF of random variable $\tilde{s}_t$, and $\underline{H}^l_t(x)$ and $\overline{H}^l_t(x)$ be the confidence bands for the CDF of random variable $\tilde{h}^l_t$. If $-C_l\le\sum_{b\in\mathcal{B}}K^b_l(\sum_{i\in\mathcal{G}_b}(x^b_{it}+a^b_{it}\tilde{s}_t)-\hat{l}^b_t)-\tilde{h}^l_t\le C_l$, $\forall\tilde{s}_t\in[\underline{s}'_t,\overline{s}'_t]$, $\tilde{h}^l_t\in[\underline{h}'_{tl},\overline{h}'_{tl}]$, $\forall l\in\mathcal{L}$ where $\underline{s}'_t=\overline{S}_t^{-1}(\beta_1)$, $\overline{s}'_t=\underline{S}_t^{-1}(1-\beta_2)$ with $\beta_1+\beta_2=\beta$, and $\underline{h}'_{tl}=(\overline{H}^l_t)^{-1}(\gamma/2)$ and $\overline{h}'_{tl}=(\underline{H}^l_t)^{-1}(1-\gamma/2)$, then DRCC (20) is satisfied.

Proof: Just need to notice that $\forall\mathbb{P}^l_t\in\mathcal{P}^l_t$

$$\begin{aligned}&\mathbb{P}_t\left[\tilde{s}_t\notin[\underline{s}'_t,\overline{s}'_t]\text{ or }\tilde{h}^l_t\notin[\underline{h}'_{tl},\overline{h}'_{tl}]\right]\\&\le\mathbb{P}_t[\tilde{s}_t<\underline{s}'_t]+\mathbb{P}_t[\tilde{s}_t>\overline{s}'_t]+\mathbb{P}_t[\tilde{h}^l_t<\underline{h}'_{tl}]+\mathbb{P}_t[\tilde{h}^l_t>\overline{h}'_{tl}]\\&=\mathbb{P}^s_t[\tilde{s}_t<\underline{s}'_t]+\mathbb{P}^s_t[\tilde{s}_t>\overline{s}'_t]+\mathbb{P}^l_t[\tilde{h}^l_t<\underline{h}'_{tl}]+\mathbb{P}^l_t[\tilde{h}^l_t>\overline{h}'_{tl}]\\&\le\overline{S}_t(\underline{s}'_t)+1-\underline{S}_t(\overline{s}'_t)+\overline{H}^l_t(\underline{h}'_{tl})+1-\underline{H}^l_t(\overline{h}'_{tl})=\beta+\gamma\end{aligned} \tag{23}$$

■

Similarly, DRCC (20) is replaced by: $\forall l\in\mathcal{L}$

$$\sum_{b\in\mathcal{B}}K^b_l\left(\sum_{i\in\mathcal{G}_b}(x^b_{it}+a^b_{it}\underline{s}'_t)-\hat{l}^b_t\right)-\underline{h}'_{tl}\le C_l \tag{24a}$$

$$-C_l\le\sum_{b\in\mathcal{B}}K^b_l\left(\sum_{i\in\mathcal{G}_b}(x^b_{it}+a^b_{it}\underline{s}'_t)-\hat{l}^b_t\right)-\overline{h}'_{tl} \tag{24b}$$

$$\sum_{b\in\mathcal{B}}K^b_l\left(\sum_{i\in\mathcal{G}_b}(x^b_{it}+a^b_{it}\overline{s}'_t)-\hat{l}^b_t\right)-\underline{h}'_{tl}\le C_l \tag{24c}$$

$$-C_l\le\sum_{b\in\mathcal{B}}K^b_l\left(\sum_{i\in\mathcal{G}_b}(x^b_{it}+a^b_{it}\overline{s}'_t)-\hat{l}^b_t\right)-\overline{h}'_{tl} \tag{24d}$$

Remark 1: In Lemmas 2 and 3, any positive values of β_1 and β_2 with $\beta_1+\beta_2=\beta$ will make (22) and (24) safe approximations to (19) and (20). Here β_1 and β_2 have clear engineering meanings. β_1 is the tolerable probability of wind curtailment whereas β_2 is the tolerable probability of load shedding. We leave them as tuning parameters of the method and let the users choose suitable values according to the system reliability standards.

Remark 2: Note that the ambiguity set for $\mathbb{P}^l_t$ is formed by directly combining the information from its marginal distributions. In other words, only marginal distributional information is encoded in $\mathcal{P}^l_t$ and the dependency information of the two marginals are not exploited. This inevitably brings additional conservatism to the chance constraints for the transmission line flow. However, this treatment avoids the consideration of high-dimensional statistics, which contributes to the highly scalable algorithm.

B. Objective Function

By Lemmas 2 and 3, the system can safely respond to the total forecasting error $\tilde{s}_t$ in the interval $[\underline{s}'_t,\overline{s}'_t]$. To ensure the system security, the system operator resorts to load shedding when $\tilde{s}_t$ exceeds $\overline{s}'_t$ and initiates wind curtailment when $\tilde{s}_t$ goes below $\underline{s}'_t$, shown in Fig. 2.

Therefore, the total operation costs including those of unit start-up/shut-down, generation, reserve availability, reserve procurement, load shedding and wind curtailment, are written

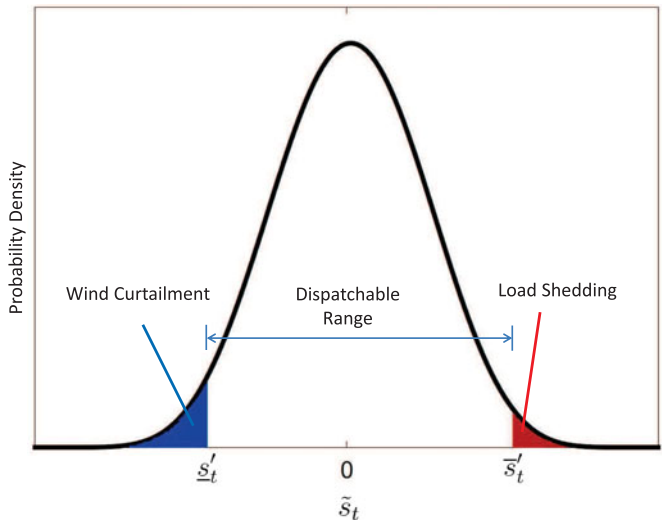

Fig. 2. Distribution of forecasting error with illustration of wind curtailment and load shedding.

as

$$\begin{aligned} &F(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x},\boldsymbol{a},\boldsymbol{r}) \\ &=\sum_{t\in\mathcal{T}}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{G}_b}(SU_i^b u_{it}^b + SD_i^b v_{it}^b + F_i^b(x_{it}^b) + R_i^{b,up} r_{it}^{b,up} \\ &\quad + R_i^{b,dn} r_{it}^{b,dn}) + \sum_{t\in\mathcal{T}} \max_{\mathbb{P}_t^s\in\mathcal{P}_t^s} \mathbb{E}_{\mathbb{P}_t^s}[Q_t(\boldsymbol{a}_t,\tilde{s}_t) + C_t(\tilde{s}_t)] \end{aligned} \tag{25}$$

where $\boldsymbol{a}_t = (a_{it}^b)_{i\in\mathcal{G}_b, b\in\mathcal{B}}$, $\boldsymbol{a} = (\boldsymbol{a}_t)_{t\in\mathcal{T}}$ and similar definitions apply for $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{x}$ and $\boldsymbol{r}$; moreover,

$$\begin{aligned} Q_t(\boldsymbol{a}_t,\tilde{s}_t) = \sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{G}_b}\Big(&Q_i^{b,up}(a_{it}^b \min\{\tilde{s}_t, \overline{s}'_t\})^+ \\ &+ Q_i^{b,dn}(-a_{it}^b \max\{\tilde{s}_t, \underline{s}'_t\})^+\Big) \end{aligned} \tag{26}$$

$$C_t(\tilde{s}_t) = C_{ls}(\tilde{s}_t - \overline{s}'_t)^+ + C_{wc}(-\tilde{s}_t + \underline{s}'_t)^+. \tag{27}$$

Reserve procurement costs is represented by Q_t in which the first and second terms are related to upward and downward reserve procurement, respectively. Also, the first term of C_t denotes load shedding costs and the second represents wind curtailment costs.

Remark 3: Note that an explicit formulation for the amount of upward reserve procurement is $(\min\{a_{it}^b\tilde{s}_t, r_{it}^{b,up}\})^+$. Nevertheless, we can show $(\min\{a_{it}^b\tilde{s}_t, r_{it}^{b,up}\})^+ = (a_{it}^b \min\{\tilde{s}_t, \overline{s}'_t\})^+$ at the optimal solution of the problem. Suppose $(a_{it}^{b*}, r_{it}^{b,up*})$ is at the optimal solution with $r_{it}^{b,up*} > a_{it}^{b*}\overline{s}'_t$. Observe that $r_{it}^{b,up\#} = a_{it}^{b*}\overline{s}'_t$ is also a feasible solution for the upward reserve (just check constraint (16)–(18) and (22)). In addition, $(\min\{a_{it}^{b*}\tilde{s}_t, r_{it}^{b,up\#}\})^+ = (\min\{a_{it}^{b*}\tilde{s}_t, a_{it}^{b*}\overline{s}'_t\})^+ \le (\min\{a_{it}^{b*}\tilde{s}_t, r_{it}^{b,up*}\})^+$ and $r_{it}^{b,up\#} < r_{it}^{b,up*}$, so the costs for $(a_{it}^{b*}, r_{it}^{b,up\#})$ is strictly less than that of $(a_{it}^{b*}, r_{it}^{b,up*})$, which contradicts the optimality of $(a_{it}^{b*}, r_{it}^{b,up*})$. Therefore, we have $r_{it}^{b,up} = a_{it}^b\overline{s}'_t$ at the optimal solution. Then we have $(\min\{a_{it}^b\tilde{s}_t, r_{it}^{b,up}\})^+ = (\min\{a_{it}^b\tilde{s}_t, a_{it}^b\overline{s}'_t\})^+ = (a_{it}^b \min\{\tilde{s}_t, \overline{s}'_t\})^+$ where the second equality follows from the non-negativity of a_{it}^b. It is the same case for the downward reserve.

Remark 4: Note that an explicit formulation for the amount of load shedding should be computed for each bus and then summed up, i.e. $\sum_{b\in\mathcal{B}}(\sum_{i\in\mathcal{G}_b} a_{it}^b\tilde{s}_t - \sum_{i\in\mathcal{G}_b} r_{it}^{b,up})^+$. Following Remark 2, we have $\sum_{b\in\mathcal{B}}(\sum_{i\in\mathcal{G}_b} a_{it}^b\tilde{s}_t - \sum_{i\in\mathcal{G}_b} r_{it}^{b,up})^+ = \sum_{b\in\mathcal{B}}(\sum_{i\in\mathcal{G}_b} a_{it}^b\tilde{s}_t - \sum_{i\in\mathcal{G}_b} a_{it}^b\overline{s}'_t)^+ = (\tilde{s}_t - \overline{s}'_t)^+$ where the second equality comes from non-negativity of a_{it}^b and (14). It is also the same case for wind curtailment.

C. Evaluation of Worst-Case Costs

In the objective function of the proposed formulation, we need to evaluate the worst-case expectation of the piece-wise linear function of the random variable taking the general form

$$\begin{aligned} &\max_{\mathbb{P}_t^s\in\mathcal{P}_t^s} \mathbb{E}_{\mathbb{P}_t^s}[Q_t(\boldsymbol{a}_t,\tilde{s}_t) + C_t(\tilde{s}_t)] \\ &\quad = \max_{\mathbb{P}_t^s\in\mathcal{P}_t^s} \mathbb{E}_{\mathbb{P}_t^s}[\max\{f_t^{up}(\tilde{s}_t), f_t^{dn}(\tilde{s}_t)\}] \end{aligned} \tag{28}$$

where

$$f_t^{up}(\tilde{s}_t) = g_t^{up}\tilde{s}_t + (C_{ls} - g_t^{up})(\tilde{s}_t - \overline{s}'_t)^+ \tag{29}$$

$$f_t^{dn}(\tilde{s}_t) = -g_t^{dn}\tilde{s}_t + (C_{wc} - g_t^{dn})(-\tilde{s}_t + \underline{s}'_t)^+ \tag{30}$$

with $g_t^{up} = \sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{G}_b} Q_i^{b,up} a_{it}^b$ and $g_t^{dn} = \sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{G}_b} Q_i^{b,dn} a_{it}^b$. The derivation of equality (28) can be found in Appendix B.

It turns out that the worst-case expectation (28) can be evaluated by solving a linear programming (LP). The main results are stated in Theorem 1 inspired by [31].

Lemma 4: Let $\hat{s}_t^{(1)}, \hat{s}_t^{(2)}, \cdots, \hat{s}_t^{(n)}$ be the ascendingly ordered samples of random variable $\tilde{s}_t$. Without loss of generality, assume $\hat{s}_t^{(k)} \le 0, \forall k \le m$ and $\hat{s}_t^{(k)} > 0, \forall k > m$. $[\underline{s}_t, \overline{s}_t]$ is the estimated support of random variable $\tilde{s}_t$ by Devroye-Wise method. The ambiguity set $\mathcal{P}_t^s$ is constructed as in (4), i.e.

$$\mathcal{P}_t^s = \left\{\mathbb{P}\in\mathcal{P}_0([\underline{s}_t, \overline{s}_t]) \middle| \mathbb{P}\left[\tilde{s}_t \le \hat{s}_t^{(k)}\right] \in \left[\underline{p}_t^k, \overline{p}_t^k\right], k = 1, \cdots, n\right\} \tag{31}$$

For notational convenience, let $\hat{s}_t^{(0)} = \underline{s}_t$, $\hat{s}_t^{(n+1)} = \overline{s}_t$ and $\underline{p}_t^{n+1} = \overline{p}_t^{n+1} = 1$. The worst-case expectation (28) is equal to the optimum of the following LP:

$$\begin{aligned} \min_{\substack{\underline{\lambda}_t^k, \overline{\lambda}_t^k \\ k=m-1,\cdots,m+2}} &\sum_{k=m-1}^{m+2}\left(\overline{\lambda}_t^k\overline{p}_t^k - \underline{\lambda}_t^k\underline{p}_t^k\right) + g_t^{dn}s_t^{dn} + g_t^{up}s_t^{up} \\ &+ C_{wc}m_t^{dn} + C_{ls}m_t^{up} \end{aligned}$$

$$\text{s.t.}\begin{cases}\underline{\lambda}_t^k \geq 0,\ \overline{\lambda}_t^k \geq 0, k = m-1, \cdots, m+2 \\ \sum_{i=k}^{m+2}\left(\overline{\lambda}_t^i - \underline{\lambda}_t^i\right) + f_t^{up}\left(\hat{s}_t^{(m+3)}\right) \geq f_t^{dn}\left(\hat{s}_t^{(k-1)}\right), \\ \qquad\qquad k = m-1, m, m+1 \\ \sum_{i=k}^{m+2}\left(\overline{\lambda}_t^i - \underline{\lambda}_t^i\right) + f_t^{up}\left(\hat{s}_t^{(m+3)}\right) \geq f_t^{up}\left(\hat{s}_t^{(k)}\right), \\ \qquad\qquad k = m+1, m+2\end{cases} \tag{32}$$

where s_t^{dn}, s_t^{up}, m_t^{dn} and m_t^{up} are defined in (33)–(36), respectively.

$$\begin{aligned} s_t^{dn} = \sum_{k=1}^{m-2}\Big[\hat{s}_t^{(k)} + \Big(-\hat{s}_t^{(k)} + \underline{s}_t'\Big)^+ - \hat{s}_t^{(k-1)} \\ - \Big(-\hat{s}_t^{(k-1)} + \underline{s}_t'\Big)^+\Big]\overline{p}_t^k \end{aligned} \tag{33}$$

$$\begin{aligned} s_t^{up} = \sum_{k=m+3}^{n}\Big[\hat{s}_t^{(k)} - \Big(\hat{s}_t^{(k)} - \overline{s}_t'\Big)^+ - \hat{s}_t^{(k+1)} + \Big(\hat{s}_t^{(k+1)} \\ - \overline{s}_t'\Big)^+\Big]\underline{p}_t^k + \Big[\hat{s}_t^{(n+1)} - \Big(\hat{s}_t^{(n+1)} - \overline{s}_t'\Big)^+\Big]\overline{p}_t^{n+1}. \end{aligned} \tag{34}$$

$$m_t^{dn} = \sum_{k=1}^{m-2}\Big[\Big(-\hat{s}_t^{(k-1)} + \underline{s}_t'\Big)^+ - (-\hat{s}_t^{(k)} + \underline{s}_t')^+\Big]\overline{p}_t^k \tag{35}$$

$$\begin{aligned} m_t^{up} = \sum_{k=m+3}^{n}\Big[\Big(\hat{s}_t^{(k)} - \overline{s}_t'\Big)^+ - (\hat{s}_t^{(k+1)} - \overline{s}_t')^+\Big]\underline{p}_t^k \\ + (\hat{s}_t^{(n+1)} - \overline{s}_t')^+\overline{p}_t^{n+1}. \end{aligned} \tag{36}$$

Proof: See Appendix A. ■

Note that the LP (32) is a small-scale problem with only 4 decision variables and the problem scale is irrelevant to the number of historical data. In other words, incorporating more historical data does not bring higher computational burden.

Therefore, evaluating the worst-case expectation in the objective function (25) with LP (32) and replacing the DRCC (19) and (20) with deterministic linear constraints (22) and (24) yield a MILP. To sum up, for given reliability level $1-\beta$ and historical data of forecasting errors, the proposed AA-DRUC takes the form

$$\begin{aligned} O(\beta_1, \beta_2, C_{ls}, C_{wc}) = \min\ & F(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{x}, \boldsymbol{a}, \boldsymbol{r}) \\ \text{s.t.}\ & (5) \sim (9), (13) \sim (18), (22)(24). \end{aligned} \tag{37}$$

D. Application Modes

In the above problem formulation, we have included both the reliability indices β_1/β_2 and the emergency control prices C_{ls}/C_{wc}. Usually, either of the two is presented as the input to the optimization model in practice. In the conventional vertically integrated systems, electric utilities operate as monopolies and enforce a reliability standard across the network to limit the occurrence of load shedding and wind curtailment. Under such circumstances, reliability indices β_1/β_2 are given as prior knowledge, and the costs of load shedding and wind curtailment are irrelevant to the decision process. We call this application mode the **Mode-I**. To obtain optimal operational strategy in this mode, we only need to solve the problem $O(\beta_1, \beta_2, 0, 0)$. By comparison, the operation of the modern restructured power systems is price-driven, and the load shedding and wind curtailment are considered ancillary services provided by the consumers and wind farms. Hence the access to these services come at a cost. We name this application mode the **Mode-II**. Note that high reliability results in low emergency control costs but high reverse costs, and vice versa. Therefore, there is a pair of optimal reliability indices β_1^*/β_2^* which minimize the total operation costs, and the solution to the corresponding MILP (37) should be considered the optimal operational strategy. To find out the optimal reliability indices β_1^*/β_2^*, we need to solve the higher-level optimization problem $\min_{\beta_1,\beta_2} O(\beta_1, \beta_2, C_{ls}, C_{wc})$. Since $O(\beta_1, \beta_2, C_{ls}, C_{wc})$ can be conveniently evaluated by solving MILP (37), the optimization over β_1, β_2 can be done by any derivative-free search method, e.g. the Nelder-Mead simplex Method [32].

E. Elimination of Redundant Line Capacity Constraints

In practical operation of power systems, the line capacity constraints (24) are only active at very few transmission lines during some periods. Hence most of the constraints in (24) are redundant for the optimization problem (37). Identifying and eliminating those redundant constraints before solving the MILP could significantly improve computational efficiency. The fast identification method proposed in [33] can be extended to our problem formulation with minor modification.

Consider the following problems:

$$\begin{aligned} &\Lambda_{t,\max}^l\left(\tilde{s}_t, \tilde{h}_t^l\right)\left(\Lambda_{t,\min}^l\left(\tilde{s}_t, \tilde{h}_t^l\right)\right) \\ &= \arg\max_{\boldsymbol{x}_t, \boldsymbol{\alpha}_t}(\min)\sum_{b\in\mathcal{B}} K_l^b\left(\sum_{i\in\mathcal{G}_b}(x_{it}^b + a_{it}^b\tilde{s}_t) - \hat{l}_t^b\right) - \tilde{h}_t^l \end{aligned} \tag{38}$$

subject to

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{G}_b}\left(x_{it}^b + a_{it}^b\tilde{s}_t\right) = \sum_{b\in\mathcal{B}}\hat{l}_t^b + \tilde{s}_t \tag{39a}$$

$$x_{it}^b + a_{it}^b\underline{s}_t' \geq 0 \tag{39b}$$

$$x_{it}^b + a_{it}^b\overline{s}_t' \leq U_i^b. \tag{39c}$$

where (39a) is obtained by multiplying (14) by $\tilde{s}_t$ and adding to (13); (39b) and (39c) are deduced from (17), (22) by relaxing w_{it}^b to 0 or 1 when necessary. Therefore, the feasible sets of the above optimization problems are relaxations of the feasible set of the original MILP model. Minimization (maximization) w.r.t. the feasible set defined by (39) yields a lower (upper) bound of the minimum (maximum) w.r.t the feasible set of the original MILP model. The objective function (38) is just the possible line flow at each line in each time period. Similar to the analysis in [33], we have the following lemma.

Lemma 5: For any $l \in \mathcal{L}$ and $t \in \mathcal{T}$, we have

1) If $\Lambda^l_{t,\max}(\underline{s}_t, \underline{h}^l_t) \le C_l$, constraint (24a) is inactive;
2) If $\Lambda^l_{t,\max}(\overline{s}_t, \underline{h}^l_t) \le C_l$, constraint (24c) is inactive;
3) If $\Lambda^l_{t,\min}(\underline{s}_t, \overline{h}^l_t) \ge -C_l$, constraint (24b) is inactive;
4) If $\Lambda^l_{t,\min}(\overline{s}_t, \overline{h}^l_t) \ge -C_l$, constraint (24d) is inactive.

Define $p^b_t = \sum_{i\in\mathcal{G}_b} p^b_{it}$ where $p^b_{it} = x^b_{it} + a^b_{it}\tilde{s}_t$ and $U^b = \sum_{i\in\mathcal{G}_b} U^b_i$, then we have

$$\begin{aligned}&\Lambda^l_{t,\max}\left(\tilde{s}_t, \tilde{h}^l_t\right)\left(\Lambda^l_{t,\min}\left(\tilde{s}_t, \tilde{h}^l_t\right)\right)\\ &\quad= \arg\max_{\boldsymbol{p}_t}(\min)\sum_{b\in\mathcal{B}} K^b_l p^b_t - \sum_{b\in\mathcal{B}} K^b_l \hat{l}^b_t - \tilde{h}^l_t\\ &\quad\text{s.t.}\begin{cases}\sum_{b\in\mathcal{B}} p^b_t = \sum_{b\in\mathcal{B}} \hat{l}^b_t + \tilde{s}_t\\ 0 \le p^b_t \le U^b\end{cases}\end{aligned} \tag{40}$$

LP (40) has a analytical solution according to the analysis in [33]. Let $b_1, b_2, \cdots, b_{|\mathcal{B}|}$ be a permutation of $1, 2, \ldots, |\mathcal{B}|$ such that $\{K^{b_1}_l, K^{b_2}_l, \ldots, K^{b_{|\mathcal{B}|}}_l\}$ are in descending (ascending) order, and there exists an integer $1 \le m \le |\mathcal{B}|$ such that $\sum_{k=1}^{m-1} U^{b_k} \le \sum_{b\in\mathcal{B}} \hat{l}^b_t + \tilde{s}_t \le \sum_{k=1}^{m} U^{b_k}$. Then

$$\begin{aligned}&\Lambda^l_{t,\max}\left(\tilde{s}_t, \tilde{h}^l_t\right)\left(\Lambda^l_{t,\min}\left(\tilde{s}_t, \tilde{h}^l_t\right)\right)\\ &= \sum_{k=1}^{m-1}(K^{b_k}_l - K^{b_m}_l)U^b + K^{b_m}_l\left(\sum_{b\in\mathcal{B}} \hat{l}^b_t + \tilde{s}_t\right) - \sum_{b\in\mathcal{B}} K^b_l \hat{l}^b_t - \tilde{h}^l_t.\end{aligned} \tag{41}$$

Based on the analytical expression (41), Lemma 5 gives a computationally cheap way to identify most of the inactive line capacity constraints in (24).

IV. Case Studies

This section presents numerical results on the IEEE 118-bus system and the 1888-bus French very high voltage system. The 118-bus system has 54 units and 186 lines. In the base case, 10 wind farms with each capacity of 80 MW are installed across the 118-bus system. The 1888-bus system has 290 units and 2531 lines. In the base case, we install 20 wind farms with each capacity of 500 MW over the network. The network data of the test systems is extracted from MATPOWER 5.1 and the unit data is from http://motor.ece.iit.edu/data. The hourly load profile is obtained from [27] and hourly forecasting wind power curves are from NREL WIND Toolkit. In addition, the reserve prices $R^{b,up}_i/R^{b,dn}_i$ and the reserve procurement prices $Q^{b,up}_i/Q^{b,dn}_i$ are assumed to be 10% and 110% of the coefficients of the linear terms of quadratic generator cost functions. Wind curtailment and load shedding costs are 100 \$/MWh and 500 \$/MWh, respectively. The proposed method is programmed in MATLAB with Gurobi [34] as the MILP solver running on a Win 8 PC with a 3.0 GHz CPU and 24 GB RAM. The MIP gap is set to be 10e-4.

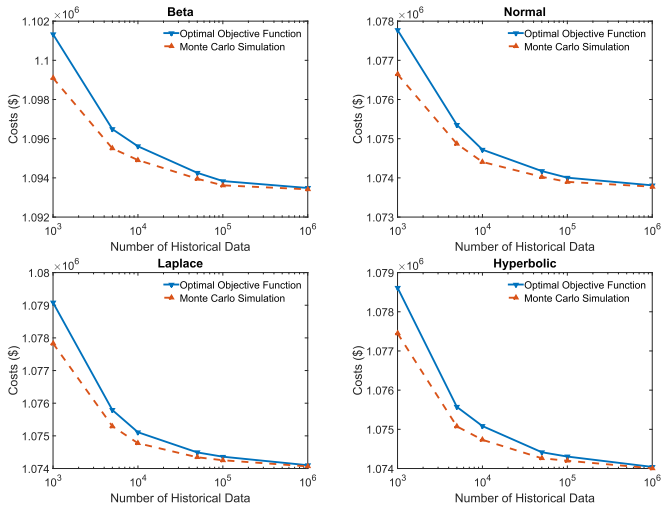

Fig. 3. Evolution of objective function and simulated costs as the increase of available historical data in application mode-I on 118-bus system. Four types of probability distributions, i.e., beta, normal, Laplace and hyperbolic, are used to simulate the underlying true distribution.

A. Distributional Robustness and Data-Exploiting Feature of the Proposed Method

To test the distributional robustness and data-exploiting feature of the proposed method, we generate wind power forecasting errors from four different types of probability distributions, including beta, normal, laplace and hyperbolic [35]. The mean and standard deviation of the forecasting errors are based on the typical day-ahead forecasting errors in U.S. reported in [36], i.e. $\mu = 0.0117$ p.u., $\sigma = 0.1187$ p.u. Historical sample sets of different sizes are used in the proposed method to reveal the relation between sample size and solution quality. After solving each problem, Monte Carlo simulation (MCS) with another 10^6 samples generated from the corresponding distribution is employed to test the practical and *out-of-sample* performance of the proposed AA-DRUC.

Figs. 3 and 4 illustrate the effects of incorporating more historical data on the optimal objective function and the operation costs from MCS on both test systems in application mode-I with $\beta_1 = 0.03, \beta_2 = 0.01$ and $\gamma = 0$. As shown in both figures, no matter what distribution the forecasting errors follow, the optimal values of AA-DRUC objective function always act as the upper bounds of practical operation costs from MCS. This is due to the fact that the proposed method considers the worst-case distribution in the ambiguity set constructed from historical data, but the underlying true distribution usually differs from the worst one. However, by incorporating more historical data, the ambiguity set shrinks and the worst-cast distribution in the ambiguity set approaches the true distribution. Therefore, as shown in Figs. 3 and 4, the values of the optimal objective function and the gap between optimal objective function and operation costs from MCS decrease as the amount of available historical data increases. This reveals the value of data, i.e. the more data we use, the less conservative thus the more economical the operation strategy is. Figs. 5–8 further compare the probability of load shedding and wind curtailment from MCS under different

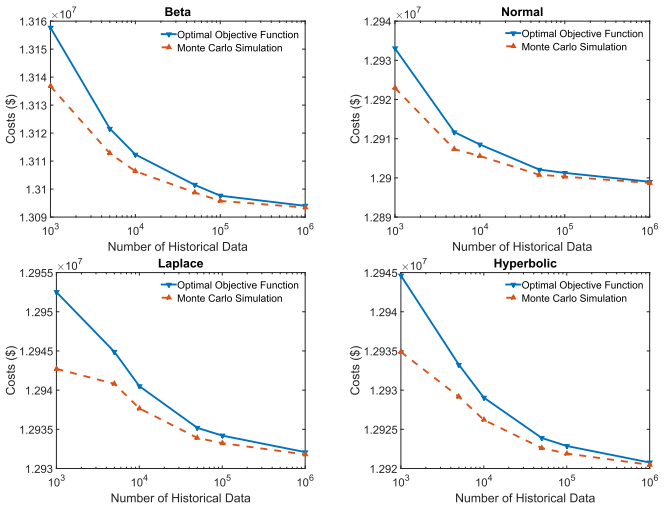

Fig. 4. Evolution of objective function and simulated costs as the increase of available historical data in application mode-I on 1888-bus system. Four types of probability distributions, i.e., beta, normal, Laplace and hyperbolic, are used to simulate the underlying true distribution.

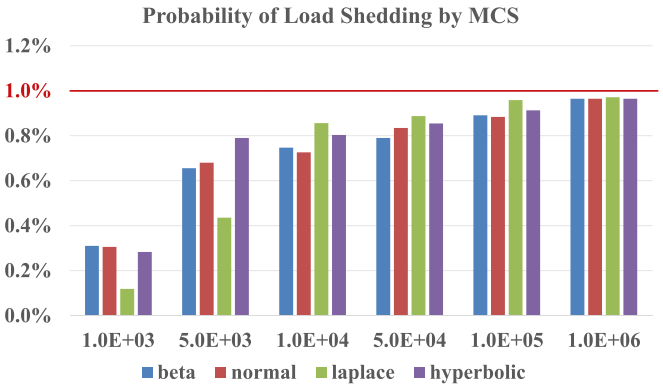

Fig. 5. Probability of load shedding by MCS on 118-bus system in application mode-I under different number of historical data and different underlying true distributions of uncertainty. The reliability requirement is $< 1.0\%$.

types of distributions and different amount of available historical data. It is evident from Figs. 5–8 that the prespecified reliability requirements are always satisfied by the proposed method whatever distribution the uncertain forecasting error obeys, which is an explicit demonstration of the distributional robustness of the proposed method. As more and more historical data is available, the practical reliability level of the operation strategy presses on towards the pre-specified reliability indices, which reduces the conservatism of the DRCC. Moreover, Figs. 9 and 10 show the evolution of objective function and simulated costs as the increase of available historical data in the application mode-II on both test systems. As β_1, β_2 increases, the dispatchable range illustrated in Fig. 2 shrinks and the allocated system reserve decreases reserve costs decreases, whereas the probability and expected costs of wind curtailment and load shedding increase. As a result, there exists a pair of optimal values of β_1, β_2 which can be located by the Nelder-Mead simplex method implemented in MATLAB command `fminsearch`. The optimal reliability indices are labeled near the corresponding optimal values of the objective function on Figs. 9 and 10. The results in the

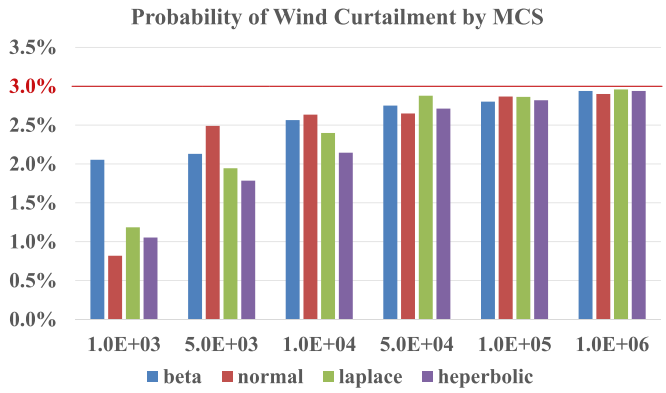

Fig. 6. Probability of wind curtailment by MCS on 118-bus system in application mode-I under different number of historical data and different underlying true distributions of uncertainty. The reliability requirement is $< 3.0\%$.

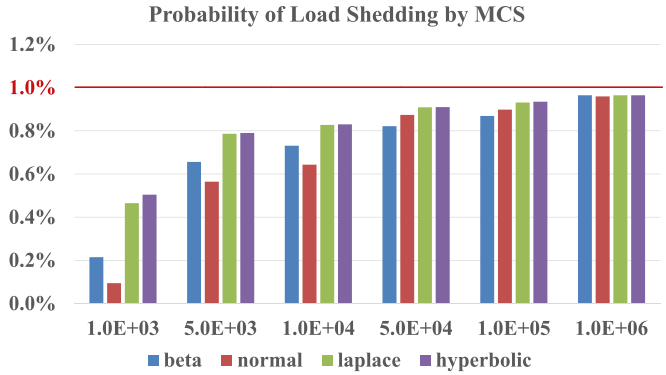

Fig. 7. Probability of load shedding by MCS on 1888-bus system in application mode-I under different number of historical data and different underlying true distributions of uncertainty. The reliability requirement is $< 1.0\%$.

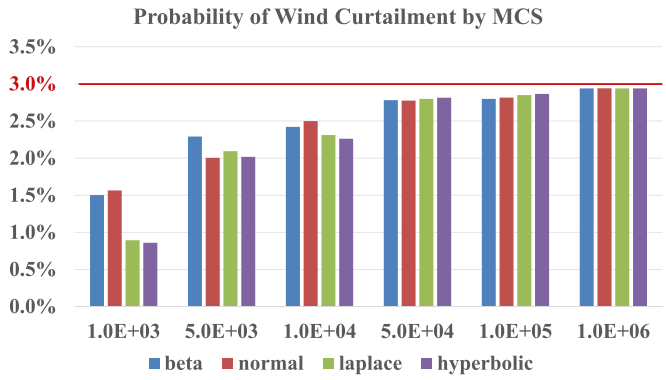

Fig. 8. Probability of wind curtailment by MCS on 1888-bus system in application mode-I under different number of historical data and different underlying true distributions of uncertainty. The reliability requirement is $< 3.0\%$.

application mode-II resemble those of application mode-I: the gap between the optimal objective function and the costs from MCS vanishes as more historical data is available.

B. Computational Efficiency and Scalability

Table I presents the solver time in seconds of the proposed method as increasing amount of historical data is available. It confirms the observation we made in Section III-C that the solution time is irrelevant to the number of historical data points

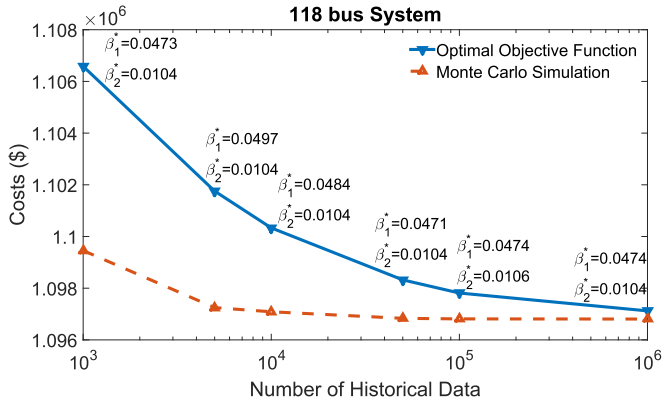

Fig. 9. Evolution of objective function and simulated costs as the increase of available historical data in application mode-II on 118-bus system. β_1^* and β_2^* are the optimal reliability indices to achieve the corresponding objective value.

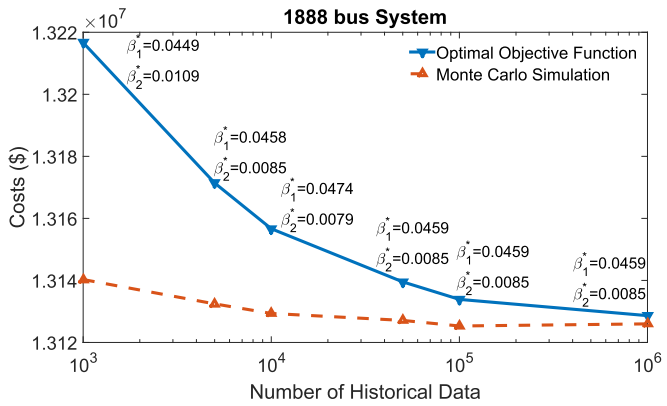

Fig. 10. Evolution of objective function and simulated costs as the increase of available historical data in application mode-II on 1888-bus system. β_1^* and β_2^* are the optimal reliability indices to achieve the corresponding objective value.

TABLE I
SOLVER TIME (SEC.) V.S. NUMBER OF HISTORICAL DATA

	1.E + 03	5.E + 03	1.E + 04	5.E + 04	1.E + 05	1.E + 06
case118	2.7	4.1	3.7	3.6	3.9	2.8
case1888	42.7	43.6	44.0	41.8	43.5	43.3

TABLE II
SOLVER TIME (SEC.) V.S. NUMBER OF WIND FARMS

wind farms No.	5	10	15	20	25	30
case118	3.1	4.5	3.5	2.8	3.6	2.6
case1888	36.9	41.8	39.5	45.0	46.4	49.6

employed. Table II, on the other hand, compares the solver time of the proposed method when different numbers of wind farms are installed across the systems. Due to the simple affine policy of reserve procurement employed in the problem formulation, the solver time is also irrelevant to the number of wind farms. Compared with the solver time of the distributionally robust model for UC in [14], the AA-DRUC proposed in this paper can be solved orders-of-magnitude faster. The high computational

TABLE III
SUMMARY OF THE RESERVE ALLOCATION OF 118-BUS SYSTEM DURING 4:00∼ 5:00 A.M

unit no.	participation factor	upward reserve (MW)	procurement price ($/MW)	total reserve
4	0.319	86.153	14.176	270.441 MW
27	0.048	13.105	9.173	
28	0.342	92.528	9.173	
36	0.078	21.220	14.176	
40	0.212	57.435	14.176	

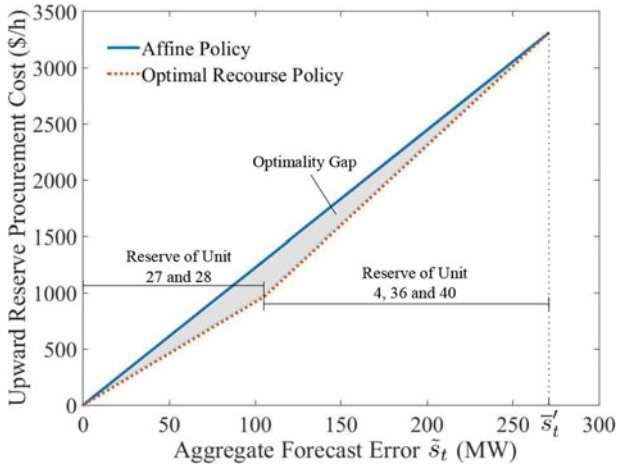

Fig. 11. Illustration of optimality gap between affine reserve procurement policy and optimal recourse reserve procurement policy on 118-bus system at 5 a.m.

efficiency of the proposed method is largely attributed to the redundant constraint identification method discussed in Section III-E which helps to eliminate about $88\% \sim 95\%$ line capacity constraints on both test systems.

C. Assessing the Conservatism of Affine Policy

Although the affine policy has brought high tractability and scalability for the proposed method, it inevitably introduces some conservatism. We first illustrate how the conservatism is introduced in the affine reserve procurement process. Table III summarizes the reserve allocation by the proposed method on the 118-bus system during 4:00∼ 5:00 a.m. Five units are delegated as spinning reserve units with total reserve 270.44 MW. These units are of two different reserve procurement prices. When wind power forecasting error occurs, to minimize the reserve procurement costs, the cheaper units should be completely utilized before resorting to any more expensive ones. Hence the reserve procurement costs corresponding to the optimal recourse policy is represented by the red dotted line in Fig. 11. However, the affine policy in the proposed model initiates all the reserve units simultaneously with different participation factors. The reserve procurement costs corresponding to the affine policy is shown with the blue line in Fig. 11. Clearly, there is a gap between the two lines due to the discrepancy of reserve prices. When beta distribution is used to simulate the forecasting error, the expected costs in 24 h are summarized in Table IV. The affine

TABLE IV
COMPARION BETWEEN AFFINE RESERVE PROCUREMENT AND FULLY ADAPTIVE RESERVE PROCUREMENT

Reserve Procurement Cost by Affine Policy	18756.14	Total operation Cost	1128619.02
Reserve Procurement Cost by Optimal Recourse Policy	17882.79	Percentage Gap w.r.t. Reserve Procurement Cost	4.88%
Gap	873.35	Percentage Gap w.r.t. Total Cost	0.08%

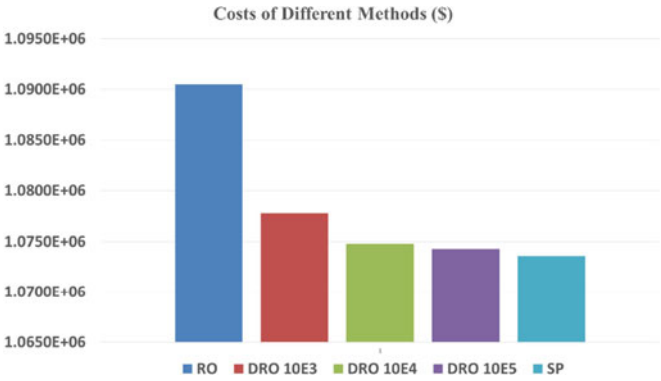

Fig. 12. Comparison of operation costs by MCS on 118-bus system using different methods, including RO, SP and the proposed DRO with 10^3, 10^4 and 10^5 historical data points.

policy has increased the reserve procurement costs by \$873.35 which accounts for 4.88% of the reserve procurement costs and 0.08% of the total operation costs. The above numerical study confirms the observation made in [26] and [28] that the simple affine policy (10) introduces minor conservatism.

D. Comparing Different Methods to Deal With Uncertainties

To compare the proposed DRO with the other two methods to deal with uncertainties, we also implement the RO and SP in our problem formulation in application mode-I. The simple affine policy (10) is used in three methods to provide a clear baseline for the comparison. The uncertainty set employed in RO is the estimated support of the random variable. The SP assumes the random variables follow the normal distribution with mean and variance being the sample mean and sample variance of the historical data. Laplace distribution is taken as the underlying true distribution for the random variables. After solving each problem, MCS is employed to test the practical performance, including operation costs and reliability guarantee, of the corresponding strategies. Fig. 12 compares the operation costs of different methods, including RO, SP and the proposed DRO with 10^3, 10^4 and 10^5 historical data points. The operation costs of the RO are the highest whereas those of the SP are the lowest. The DROs with different amount of data are intermediates between RO and SP. The costs of DRO go close to those of SP as more historical data is available. The probability of load shedding and wind curtailment for different methods are further compared in Fig. 13. RO has the far-more-than-required level of reliability due to the ignorance of probabilistic information. The

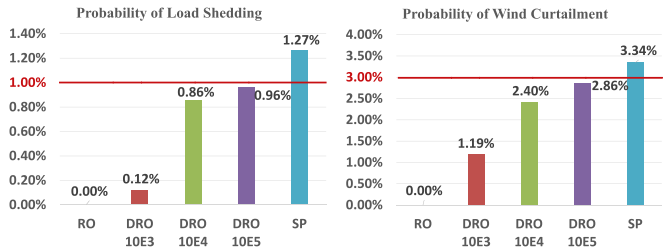

Fig. 13. Comparison of probability of load shedding and wind curtailment by MCS on 118-bus system using different methods, including RO, SP and the proposed DRO with 10^3, 10^4 and 10^5 historical data points.

TABLE V
COMPARISON AMONG DIFFERENT METHODS TO DEAL WITH UNCERTAINTIES

	the proposed DRO	RO	SP
exploit probabilistic information	yes	no	yes
presumption on distribution	no	no	yes
reliability guarantee when true distribution is unknown	yes	yes	no
data-exploiting: the more data, the less conservative	yes	no	no

SP, however, does not guarantee the required reliability level because the underlying true distribution differs from the normality assumption made in the method. The proposed DRO approach always provide safe reliability guarantee and the practical reliability level approaches the pre-specified reliability indices as more data is available. We summarize the features of the three methods in Table V.

V. CONCLUSION AND DISCUSSION

This paper applies the idea of distributionally robust optimization to the UC under uncertainty. Based on the non-parametric inference theory, an ambiguity set that contains the true probability distribution of uncertainties is constructed from observed historical data. The proposed model considers the worst-case distribution in the ambiguity set thereby achieves operational robustness. Moreover, the proposed ambiguity set shrinks to the true distribution as the amount of historical data increases. Therefore, the conservatism of the solution can be reduced by incorporating more data. In addition, the scale of the optimization problem remains unchanged when using more data.

Recently, dynamic uncertainty set [29], [37] and data-driven uncertainty set [38] have been proposed to reduce the conservatism of RO. Compared with the dynamic uncertainty set method which explicitly models the correlation and dynamics of uncertainties, the proposed approach to deal with uncertainty is still static in nature. But the solution of the proposed method is much more simple and direct. The treatment of the distributionally robust chance constraints in this paper is basically the same as the data-driven uncertainty set with probabilistic guarantee [38], but different hypothesis tests are employed. More importantly, the essential difference between the ambiguity-set-based DRO and the uncertainty-set-based RO is that the former minimizes expected costs w.r.t the worst-case distribution whereas

the latter minimizes the worst-case costs. Therefore, DRO exploits much more probabilistic information than RO. This paper and reference [14] share some similarities as both works employ the idea of DRO and affine policy. The ambiguity sets used in both papers allow reformulation of DRO problems into tractable MILPs. Unlike that in [14], the ambiguity set proposed in this paper has sound statistical foundation and data-exploiting ability. Meanwhile, the proposed approach is more scalable than that in [14] since the scale of MILP in this paper remains unchanged as the amount of historical data and the number of uncertain renewable sources increase.

However, the proposed method also has limitations. The ambiguity set for $\mathbb{P}_t^l$ only incoporates the information from its marginal distributions, which brings additional conservatism to the chance constraints for line flow. One future direction to remedy this issue is to consider a Wasserstein-metric-based ambiguity set which can effectively deal with multivariate distributions [39]. In addition, the simple affine policy is merely a conservative approximation to the optimal recourse action. Therefore, another direction is to apply the proposed ambiguity set and reformulation technique to evaluate the operation risk in the risk-constrained UC [40]. A fully adaptive data-driven UC method can then be obtained by leveraging the results from this paper and reference [40].

APPENDIX A
PROOF OF LEMMA 4

Proof Step 1: We first show the worst-case expectation can be evaluated by solving a LP.

The evaluation of worst-case expectation (28) is a infinite dimensional linear optimization problem of the form

$$\max_{\mathbb{P}\in\mathcal{P}_0} \int_{[\hat{s}_{(0)},\hat{s}_{(n+1)}]} \max\{f_t^{up}(\tilde{s}_t), f_t^{dn}(\tilde{s}_t)\}\mathbb{P}(d\tilde{s}_t) \tag{42a}$$

$$\underline{p}_k \le \int_{[\hat{s}_{(0)},\hat{s}_{(n+1)}]} I_{[\hat{s}_{(0)},\hat{s}_{(k)}]}(\tilde{s}_t)\mathbb{P}(d\tilde{s}_t) \le \overline{p}_k, \tag{42b}$$

$$\forall k = 1, \cdots, n+1.$$

By using the conic duality theory, assigning dual variables $(\underline{\lambda}_k, \overline{\lambda}_k)$ to constraints (42b), the dual of problem (42) is given by

$$\inf_{\underline{\lambda},\overline{\lambda}} \sum_{k=1}^{n+1}(\overline{\lambda}_k\overline{p}_k - \underline{\lambda}_k\underline{p}_k) \tag{43a}$$

$$\text{s.t.} \quad \underline{\lambda}_k \ge 0, \overline{\lambda}_k \ge 0,\ 1 \le k \le n+1 \tag{43b}$$

$$\sum_{k=1}^{n+1}(\overline{\lambda}_k - \underline{\lambda}_k)I_{[\hat{s}_{(0)},\hat{s}_{(k)}]}(\xi) - \max\{f_t^{up}(\tilde{s}_t), f_t^{dn}(\tilde{s}_t)\} \ge 0,$$

$$\forall \xi \in [\hat{s}_{(0)}, \hat{s}_{(n+1)}] \tag{43c}$$

which is a finite dimensional optimization problem with semi-infinite constraint (43c). Strong duality, i.e. the optimal values of (42) and (43) coincide, is guaranteed by [41, Th. 1]. Thus, we focus on the dual problem in the sequel. The semi-infinite constraint (43c) is equivalent to

$$\begin{cases} \inf\limits_{\xi\in[\hat{s}_{(0)},\hat{s}_{(n+1)}]} \left\{\sum_{k=1}^{n+1}(\overline{\lambda}_k - \underline{\lambda}_k)I_{[\hat{s}_{(0)},\hat{s}_{(k)}]}(\xi) - f_t^{up}(\tilde{s}_t)\right\} \ge 0 \\ \inf\limits_{\xi\in[\hat{s}_{(0)},\hat{s}_{(n+1)}]} \left\{\sum_{k=1}^{n+1}(\overline{\lambda}_k - \underline{\lambda}_k)I_{[\hat{s}_{(0)},\hat{s}_{(k)}]}(\xi) - f_t^{dn}(\tilde{s}_t)\right\} \ge 0 \end{cases} \tag{44}$$

Observe that $[\hat{s}_{(0)}, \hat{s}_{(n+1)}]$ can be partitioned into $n+1$ mutually disjoint sets $[\hat{s}_{(k-1)}, \hat{s}_{(k)}]$, $k = 1, \ldots, n+1$. Therefore constraint (44) splits into $n+1$ parts, i.e. $\forall k = 1, \ldots, n+1$,

$$\begin{cases} \inf\limits_{\xi\in[\hat{s}_{(k-1)},\hat{s}_{(k)}]} \left\{\sum_{i=k}^{n+1}(\overline{\lambda}_i - \underline{\lambda}_i) - f_t^{up}(\tilde{s}_t)\right\} \ge 0 \\ \inf\limits_{\xi\in[\hat{s}_{(k-1)},\hat{s}_{(k)}]} \left\{\sum_{i=k}^{n+1}(\overline{\lambda}_i - \underline{\lambda}_i) - f_t^{dn}(\tilde{s}_t)\right\} \ge 0 \end{cases} \tag{45}$$

Since $f_t^{up}(\tilde{s}_t)$ is monotone increasing and $f_t^{dn}(\tilde{s}_t)$ is monotone decreasing, (45) is equivalent to

$$\begin{cases} \sum_{i=k}^{n+1}(\overline{\lambda}_i - \underline{\lambda}_i) - f_t^{up}(\hat{s}_{(k)}) \ge 0 \\ \sum_{i=k}^{n+1}(\overline{\lambda}_i - \underline{\lambda}_i) - f_t^{dn}(\hat{s}_{(k-1)}) \ge 0 \end{cases} \tag{46}$$

By replacing semi-infinite constraint (43c) with constraint (46) and noticing $f_t^{up}(\tilde{s}_t)$ and $f_t^{dn}(\tilde{s}_t)$ intersect at $\tilde{s}_t = 0$, problem (43) is equivalent to the following LP:

$$\min_{\underline{\lambda}_t,\overline{\lambda}_t} \sum_{k=1}^{n+1}\left(\overline{\lambda}_t^k\overline{p}_t^k - \underline{\lambda}_t^k\underline{p}_t^k\right) \tag{47}$$

$$\text{s.t.} \begin{cases} \underline{\lambda}_t^k \ge 0,\ \overline{\lambda}_t^k \ge 0, \forall 1 \le k \le n+1 \\ \sum_{i=k}^{n+1}\left(\overline{\lambda}_t^i - \underline{\lambda}_t^i\right) - f_t^{dn}(\hat{s}_{(k-1)}) \ge 0, \forall 1 \le k \le m+1 \\ \sum_{i=k}^{n+1}\left(\overline{\lambda}_t^i - \underline{\lambda}_t^i\right) - f_t^{up}(\hat{s}_{(k)}) \ge 0, \forall m+1 \le k \le n+1 \end{cases} \tag{48}$$

Step 2: We then show the above LP can be significantly simplified. Let

$$x_t^k = \overline{\lambda}_t^k - \underline{\lambda}_t^k,\ 1 \le k \le n+1 \tag{49}$$

and

$$h_t^k = \begin{cases} f_t^{dn}(\hat{s}_{(k-1)}), & k \le m \\ \max\{f_t^{dn}(\hat{s}_{(k-1)}), f_t^{up}(\hat{s}_{(k)})\}, & k = m+1 \\ f_t^{up}(\hat{s}_{(k)}), & k \ge m+2. \end{cases} \tag{50}$$

LP (47) and (48) is equivalent to

$$\begin{aligned} \min_{\boldsymbol{x}_t,\underline{\boldsymbol{\lambda}}_t} & \sum_{k=1}^{n+1}\underline{\lambda}_t^k(\overline{p}_t^k - \underline{p}_t^k) + x_t^k\overline{p}_t^k \\ \text{s.t. } & x_t^k + \underline{\lambda}_t^k \ge 0, \underline{\lambda}_t^k \ge 0\ k = 1, \ldots, n+1 \\ & \sum_{i=k}^{n+1} x_t^i \ge h_t^k. \end{aligned} \tag{51}$$

Since $\overline{p}_t^k - \underline{p}_t^k \geq 0, \forall\ 1 \leq k \leq n+1$, the optimal value for $\underline{\lambda}_t^k$ is always $\max\{-x_t^k, 0\}$, i.e. the LP (51) is simplified to

$$\begin{aligned} \min_{\boldsymbol{x}_t} \ & \sum_{k=1}^{n+1} \max\{x_t^k \overline{p}_t^k, x_t^k \underline{p}_t^k\} \\ \text{s.t.} \ & \sum_{i=k}^{n+1} x_t^i \geq h_t^k,\ k = 1, \ldots, n+1. \end{aligned} \tag{52}$$

By Further defining $z_t^k = \sum_{i=k}^{n+1} x_t^i,\ \ \forall 1 \leq k \leq n+1$ and $z_t^{n+2} = 0$, problem (52) can be written as

$$\begin{aligned} \min_{\boldsymbol{z}_t} f_t(\boldsymbol{z}_t) = & \sum_{k=1}^{n+1} \max\{(z_t^k - z_t^{k+1})\overline{p}_t^k, (z_t^k - z_t^{k+1})\underline{p}_t^k\} \\ \text{s.t.} \ & z_t^k \geq h_t^k,\ k = 1, \ldots, n+1. \end{aligned} \tag{53}$$

Note that problem (53) is convex and $\forall k = 1, \ldots, m-1, m+3, \cdots, n+1$,

$$\begin{aligned} & \left.\frac{\partial f_t}{\partial z_t^k}\right|_{(h_t^1, \ldots, h_t^{m-1}, z_t^m, z_t^{m+1}, z_t^{m+2}, h_t^{m+3}, \ldots, h_t^{n+1})} \\ & \quad = \begin{cases} \overline{p}_t^k - \overline{p}_t^{k-1},\ 1 \leq k \leq m-1 \\ \underline{p}_t^k - \underline{p}_t^{k-1},\ m+3 \leq k \leq n \\ \overline{p}_t^{n+1} - \underline{p}_t^n,\ k = n+1 \end{cases} > 0 \end{aligned} \tag{54}$$

i.e. going along any feasible direction at point $(h_t^1, \ldots, h_t^{m-1}, z_t^m, z_t^{m+1}, z_t^{m+2}, h_t^{m+3}, \ldots, h_t^{n+1})$ lead to the increase of the objective function (more strictly, it is the optimality condition based on tangent cone of the feasible set, see [42, Th. 3.1]). Therefore, z_t^k takes optimal value at $h_t^k, \forall k = 1, \ldots, m-1, m+3, \ldots, n+1$. Equivalently, $x_t^k = h_t^k - h_t^{k+1},\ \ \forall k = 1, \ldots, m-2, m+3, \ldots, n$ and $x_t^{n+1} = h_t^{n+1}$ at the optimal solution of LP (51). It is obvious that $x_t^k > 0,\ 1 \leq k \leq m-2$ and $x_t^k < 0,\ m+3 \leq k \leq n+1$, so $\underline{\lambda}_t^k = 0,\ 1 \leq k \leq m-2$ and $\underline{\lambda}_t^k = -x_t^k,\ m+3 \leq k \leq n+1$. Substituting above results into (51) and noticing (49), (50), (29) and (30) yield the simplified LP (32), which completes the proof. ■

APPENDIX B
DERIVATION OF (28)

To show (28), we only need to be aware of the following equalities:

$$(\min\{\tilde{s}_t, \overline{s}_t'\})^+ = \left(\tilde{s}_t - (\tilde{s}_t - \overline{s}_t')^+\right) \cdot I_{[0,\infty)}(\tilde{s}_t) \tag{55}$$

$$(-\max\{\tilde{s}_t, \underline{s}_t'\})^+ = \left(-\tilde{s}_t - (-\tilde{s}_t + \underline{s}_t')^+\right) \cdot I_{(-\infty,0]}(\tilde{s}_t) \tag{56}$$

which can be verified by a classified calculation and noting that $\overline{s}_t' > 0$ and $\underline{s}_t' < 0$. Then what follows are just steps of direct calculation:

$$\begin{aligned} & Q_t(\boldsymbol{a}_t, \tilde{s}_t) + C_t(\tilde{s}_t) \\ & = \sum_{b \in \mathcal{B}} \sum_{i \in \mathcal{G}_b} \left(Q_i^{b,up}(a_{it}^b \min\{\tilde{s}_t, \overline{s}_t'\})^+ + Q_i^{b,dn}(-a_{it}^b \max\{\tilde{s}_t, \underline{s}_t'\})^+ \right) \\ & \quad + C_{ls}(\tilde{s}_t - \overline{s}_t')^+ + C_{wc}(-\tilde{s}_t + \underline{s}_t')^+ \\ & = g_t^{up}(\min\{\tilde{s}_t, \overline{s}_t'\})^+ + g_t^{dn}(-\max\{\tilde{s}_t, \underline{s}_t'\})^+ \\ & \quad + C_{ls}(\tilde{s}_t - \overline{s}_t')^+ + C_{wc}(-\tilde{s}_t + \underline{s}_t')^+ \\ & = g_t^{up}\left(\tilde{s}_t - (\tilde{s}_t - \overline{s}_t')^+\right) \cdot I_{[0,\infty)}(\tilde{s}_t) + C_{ls}(\tilde{s}_t - \overline{s}_t')^+ \\ & \quad + g_t^{dn}\left(-\tilde{s}_t - (-\tilde{s}_t + \underline{s}_t')^+\right) \cdot I_{(-\infty,0]}(\tilde{s}_t) + C_{wc}(-\tilde{s}_t + \underline{s}_t')^+ \\ & = \left(g_t^{up}\tilde{s}_t + (C_{ls} - g_t^{up})(\tilde{s}_t - \overline{s}_t')^+\right) \cdot I_{[0,\infty)}(\tilde{s}_t) \\ & \quad + \left(-g_t^{dn}\tilde{s}_t + (C_{wc} - g_t^{dn})(-\tilde{s}_t + \underline{s}_t')^+\right) \cdot I_{(-\infty,0]}(\tilde{s}_t) \\ & = f_t^{up}(\tilde{s}_t) \cdot I_{[0,\infty)}(\tilde{s}_t) + f_t^{dn}(\tilde{s}_t) \cdot I_{(-\infty,0]}(\tilde{s}_t) \\ & = \max\{f_t^{up}(\tilde{s}_t), f_t^{dn}(\tilde{s}_t)\}. \end{aligned}$$

REFERENCES

[1] Q. Zheng, J. Wang, and A. Liu, "Stochastic optimization for unit commitment: A review," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1913–1924, Jul. 2015.

[2] Q. Zheng, J. Wang, P. 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.

[3] Q. Wang, Y. Guan, and J. Wang, "A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 206–215, Feb. 2012.

[4] M. Ban, J. Yu, M. Shahidehpour, and Y. Yao, "Integration of power-to-hydrogen in day-ahead security-constrained unit commitment with high wind penetration," *J. Mod. Power Syst. Clean Energy*, vol. 5, no. 3, pp. 337–349, 2017.

[5] S. Takriti, J. Birge, and E. Long, "A stochastic model for the unit commitment problem," *IEEE Trans. Power Syst.*, vol. 11, no. 3, pp. 1497–1508, Aug. 1996.

[6] P. Carpentier, G. Gohen, J.-C. Culioli, and A. Renaud, "Stochastic optimization of unit commitment: A new decomposition framework," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 1067–1073, May 1996.

[7] L. Wu, M. Shahidehpour, and T. Li, "Stochastic security-constrained unit commitment," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 800–811, May 2007.

[8] V. Pappala, I. Erlich, K. Rohrig, and J. Dobschinski, "A stochastic model for the optimal operation of a wind-thermal power system," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 940–950, May 2009.

[9] D. Bertsimas, E. Litvinov, X. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 52–63, Feb. 2013.

[10] R. Jiang, J. Wang, and Y. Guan, "Robust unit commitment with wind power and pumped storage hydro," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 800–810, May 2012.

[11] C. Lee, C. Liu, S. Mehrotra, and M. Shahidehpour, "Modeling transmission line constraints in two-stage robust unit commitment problem," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1221–1231, May 2014.

[12] C. Zhao and Y. Guan, "Unified stochastic and robust unit commitment," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3353–3361, Aug. 2013.

[13] C. Zhao and Y. Guan, "Data-driven stochastic unit commitment for integrating wind generation," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2587–2596, Jul. 2016.

[14] P. Xiong, P. Jirutitijaroen, and C. Singh, "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.

[15] W. Wei, F. Liu, and S. Mei, "Distributionally robust co-optimization of energy and reserve dispatch," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 289–300, Jan. 2016.

[16] W. Wei, J. Wang, and S. Mei, "Dispatchability maximization for co-optimized energy and reserve dispatch with explicit reliability guarantee," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3276–3288, Jul. 2016.
[17] Q. Bian, H. Xin, Z. Wang, D. Gan, and K. P. Wong, "Distributionally robust solution to the reserve scheduling problem with partial information of wind power," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2822–2823, Sep. 2015.
[18] Z. Wang, Q. Bian, H. Xin, and D. Gan, "A distributionally robust co-ordinated reserve scheduling model considering CVaR-based wind power reserve requirements," *IEEE Trans. Sustain. Energy*, vol. 7, no. 2, pp. 625–636, Apr. 2016.
[19] M. Lubin, Y. Dvorkin, and S. Backhaus, "A robust approach to chance constrained optimal power flow with renewable generation," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3840–3849, Sep. 2016.
[20] Y. Zhang, S. Shen, and J. Mathieu, "Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1378–1388, Mar. 2016.
[21] D. Bienstock, M. Chertkov, and S. Harnett, "Chance-constrained optimal power flow: Risk-aware network control under uncertainty," *SIAM Rev.*, vol. 56, no. 3, pp. 461–495, 2014.
[22] L. Wasserman, *All of Nonparametric Statistics*. New York, NY, USA: Springer, 2006.
[23] M. Goldman and D. M. Kaplan, "Evenly sensitive KS-type inference on distributions," Tech. Rep. 13–19, Working paper, 2015. [Online]. Available at: http://faculty.missouri.edu/kaplandm
[24] L. Devroye and G. L. Wise, "Detection of abnormal behavior via nonparametric estimation of the support," *SIAM J. Appl. Math.*, vol. 38, no. 3, pp. 480–488, 1980.
[25] I. Ba *et al.*, "Set estimation and nonparametric detection," *Can. J. Statist.*, vol. 28, no. 4, pp. 765–782, 2000.
[26] R. A. Jabr, "Adjustable robust OPF with renewable energy sources," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4742–4751, Nov. 2013.
[27] R. Jabr, S. Karaki, and J. Korbane, "Robust multi-period opf with storage and renewables," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2790–2799, Sep. 2015.
[28] A. Lorca, X. A. Sun, E. Litvinov, and T. Zheng, "Multistage adaptive robust optimization for the unit commitment problem," *Oper. Res.*, vol. 64, no. 1, pp. 32–51, 2016.
[29] A. Lorca and X. Sun, "Multistage robust unit commitment with dynamic uncertainty sets and energy storage," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 1678–1688, May 2017.
[30] H. Bevrani, *Robust Power System Frequency Control*, vol. 85. New York, NY, USA: Springer, 2009.
[31] W. Wiesemann, D. Kuhn, and M. Sim, "Distributionally robust convex optimization," *Oper. Res.*, vol. 62, no. 6, pp. 1358–1376, 2014.
[32] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," *SIAM J. Optim.*, vol. 9, no. 1, pp. 112–147, 1997.
[33] Q. Zhai, X. Guan, J. Cheng, and H. Wu, "Fast identification of inactive security constraints in SCUC problems," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1946–1954, Nov. 2010.
[34] *Gurobi Optimizer Reference Manual*, Gurobi Optimization, Inc., Houston, TX, USA, 2015.
[35] W. Hu, Y. Min, Y. Zhou, and Q. Lu, "Wind power forecasting errors modelling approach considering temporal and spatial dependence," *J. Mod. Power Syst. Clean Energy*, vol. 5, no. 3, pp. 489–498, 2017.
[36] B.-M. Hodge *et al.*, "Wind power forecasting error distributions: An international comparison," Nat. Renewable Energy Lab., Golden, CO, USA, Tech. Rep. NREL/CP-5500–56130, 2012.
[37] A. Lorca and X. A. Sun, "Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1702–1713, Jul. 2015.
[38] D. Bertsimas, V. Gupta, and N. Kallus, "Data-driven robust optimization," *Math. Program.*, Springer, pp. 1–58, 2013.
[39] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu, "Distributionally robust chance-constrained voltage-concerned DC-OPF with Wasserstein metric," arXiv:1706.05538, 2017.
[40] C. Wang *et al.*, "Robust risk-constrained unit commitment with large-scale wind generation: An adjustable uncertainty set approach," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 723–733, Jan. 2017.
[41] K. Isii, "On sharpness of Tchebycheff-type inequalities," *Ann. Inst. Statist. Math.*, vol. 14, no. 1, pp. 185–197, 1962.
[42] A. Dhara and J. Dutta, *Optimality Conditions in Convex Optimization: A Finite-dimensional View*. Boca Raton, FL, USA: CRC Press, 2011.

Chao Duan (S'14) was born in Chongqing, China, in 1989. He received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2012. He is currently working toward the Ph.D. degree at Xi'an Jiaotong University, Xi'an, China, and the University of Liverpool, Liverpool, U.K. His research interests include stochastic optimization, stability analysis, and robust control of power systems.

Lin Jiang (M'00) received the B.Sc. and M.Sc. degrees from Huazhong University of Science and Technology, China, and the Ph.D. degree from the University of Liverpool, U.K., in 1992, 1996, and 2001, respectively, all in electrical engineering. From 2001 to 2003, he was a Postdoctoral Research Assistant at the University of Liverpool. From 2003 to 2005, he was a Postdoctoral Research Associate in the Department of Automatic Control and Systems Engineering, University of Sheffield. From 2005 to 2007, he was a Senior Lecturer at the University of Glamorgan and moved to the University of Liverpool in 2007. He is a currently a Reader in the University of Liverpool, Liverpool, U.K. His current research interests include control and analysis of power system, smart grid, and renewable energy.

Wanliang Fang was born in Henan, China, in 1958. He received the B.S. and M.S. degrees from Xi'an Jiaotong University, Xi'an, China, and the Ph.D. degree from Hong Kong Polytechnic University, Hong Kong, in 1982, 1988, and 1999, respectively, all in electrical engineering. He is currently a Professor of electrical engineering at Xi'an Jiaotong University. His research interests include power system stability analysis and control, FACTS and HVDC.

Jun Liu (S'09–M'10) received the B.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2012, respectively, all in electrical engineering. He is currently an Associate Professor in the Department of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China. From September 2008 to August 2010, he was a Visiting Scholar in Texas A&M University, College Station, TX, USA. His research interests include renewable energy integration, power system operation and control, power system stability, HVDC, and FACTS.