Distributionally Robust Co-Optimization of Energy and Reserve Dispatch

Wei Wei, Feng Liu, *Member, IEEE*, and Shengwei Mei, *Fellow, IEEE*

***Abstract*—This paper proposes a two-stage distributionally robust optimization model for the joint energy and reserve dispatch (D-RERD for short) of bulk power systems with significant renewable energy penetration. Distinguished from the prevalent uncertainty set-based and worst-case scenario oriented robust optimization methodology, we assume that the output of volatile renewable generation follows some ambiguous distribution with known expectations and variances, the probability distribution function (pdf) is restricted in a functional uncertainty set. D-RERD aims at minimizing the total expected production cost in the worst renewable power distribution. In this way, D-RERD inherits the advantages from both stochastic optimization and robust optimization: statistical characteristic is taken into account in a data-driven manner without requiring the exact pdf of uncertain factors. We present a convex optimization-based algorithm to solve the D-RERD, which involves solving semidefinite programming (SDP), convex quadratic programming (CQP), and linear programming (LP). The performance of the proposed approach is compared with the emerging adaptive robust optimization (ARO)-based model on the IEEE 118-bus system. Their respective features are discussed in case studies.**

***Index Terms*—Energy and reserve dispatch, renewable generation, distributionally robust optimization, uncertainty.**

NOMENCLATURE

Main symbols and notations used in this paper are defined in this section for quick reference. Others are defined after their first appearances as required.

A. Indices

i	Index of generators.
j	Index of variable energy resource (VER) plants.
q	Index of loads.
s	Index of scenarios.

B. Parameters

a_i, b_i	Energy production cost coefficients of generator i.
C_j^V	Installed capacity of VER plant j, $C^V = \{C_j^V\}, \forall j$.
d_i^{g+}	Up-regulation cost coefficient of generator i.
d_i^{g-}	Down-regulation cost coefficient of generator i.
d_i^{r+}	Up-reserve cost coefficient of generator i.
d_i^{r-}	Down-reserve cost coefficient of generator i.
F_l	Power flow capacity of transmission line l.
P_i^l	Minimal output of generator i.
P_i^u	Maximal output of generator i.
p_q	Power demand of load q.
N_E	Number of extreme points.
N_G	Number of generators.
N_Q	Number of loads.
N_W	Number of VER plants.
Pr_s	Probability of scenario s.
R_i^+	Ramp-up limit of generator i.
R_i^-	Ramp-down limit of generator i.
Δt	Time duration of the current dispatch interval.
w_j^e	Generation capability forecast of VER plant j, vector $w^e = \{w_j^e\}, \forall j$.
w_j^g	Actual generation capability of VER plant j in the deterministic formulation, vector $w^g = \{w_j^g\}, \forall j$.
w_j^l	Minimal generation capability of VER plant j.
w_j^u	Maximal generation capability of VER plant j.
w_j^h+	Half of the forecast interval, $w_j^h = 0.5(w_j^u - w_j^l), \forall j$
π_{il}	Power transfer distribution factor (PTDF) from generator i to line l.
π_{jl}	PTDF from VER plant j to line l.
π_{ql}	PTDF from load q to line l.
Vr_j	Variance of the generation capability forecast error of VER plant j.
Σ	Covariance matrix of the VER generation capability forecast error.
Θ	Second-order moment matrix, $\Theta = \Sigma + w^e(w^e)^T$.
ρ_j	Generation curtailment cost of VER plant j.
ρ_q	Load-shedding cost of load q.

C. First-stage Variables

p_i^f	Scheduled output of generator i, vector $p^f = \{p_i^f\}, \forall i$.
r_i^+	Up-spinning reserve capacity offered by generator i, vector $r^+ = \{r_i^+\}, \forall i$.
r_i^-	Down-spinning reserve capacity offered by generator i, vector $r^- = \{r_i^-\}, \forall i$.
w_j^f	Scheduled output of VER plant j, $w^f = \{w_j^f\}, \forall j$.
x	Vector of the first-stage dispatch decision in the compact form, $x = \{p^f, r^+, r^-, w^f\}$.

Manuscript received February 26, 2015; revised July 21, 2015 and September 06, 2015; accepted October 07, 2015. Date of publication November 10, 2015; date of current version December 11, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 51577163, in part by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China under Grant 51321005, and in part by the State Grid Corporation of China under Grant SGSXDKY-DWKJ2015-001. Paper no. TSTE-00138-2015.

The authors are with the State Key Laboratory of Power Systems, Department of Electrical Engineering, Tsinghua University, 100084 Beijing, China (e-mail: wei-wei04@mails.tsinghua.edu.cn).

Digital Object Identifier 10.1109/TSTE.2015.2494010

D. Second-stage Variables

p_i^+	Up-regulation power of generator i in response to the actual VER generation, vector $p^+ = \{p_i^+\}, \forall i$.
p_i^-	Down-regulation power of generator i in response to the actual VER generation, vector $p^- = \{p_i^-\}, \forall i$.
p_q^l	Demand curtailment of load q in response to the actual VER generation, vector $p^l = \{p_q^l\}, \forall q$.
w_j^c	VER power spillage in plant j, $w^c = \{w_j^c\}, \forall j$
$\tilde{w}_j^g$	Random generation capability of VER plant j in the second stage of D-RERD, vector $\tilde{w}^g = \{\tilde{w}_j^g\}, \forall j$.
$\bar{w}_j^g$	Uncertain generation capability of VER plant j in the second stage of the ARO based energy and reserve dispatch (A-RERD for short), vector $\bar{w}^g = \{\bar{w}_j^g\}, \forall j$.
y	Vector of the second-stage dispatch decisions in the compact form, $y = \{p^+, p^-, p^l, w^c\}$.
H	Matrix variable in the dual of the second-stage problem of D-RERD.
h	Vector variable in the dual of the second-stage problem of D-RERD..
h_0	Scalar variable in the dual of the second-stage problem of D-RERD..
u	Dual variable of the operating constraints.

E. Sets

W	Uncertainty set used in A-RERD.
Ω	Functional uncertainty set used in D-RERD.
X	Feasible set of the first-stage decision x.
U	Feasible set of dual variable u.
B^V	Hypercube of VER generation capacity $B^V = \{w^g \mid 0 \le w^g \le C^V\}$.

F. Functions

$E_P(\cdot)$	Expectation over distribution P.
$\Pr(\cdot)$	Probability of an event.
$Q(x, w^g)$	Optimal value function of the second-stage problem under given x and w^g.
$\mathrm{tr}(\cdot)$	Trace of a matrix.
$\mathrm{vert}(\cdot)$	Extreme points of a polytope.
$Z_P(x)$	Optimal value function of the primal second-stage problem in D-RERD with given x.
$Z_D(x)$	Optimal value function of the dual second-stage problem in D-RERD with given x.

G. Abbreviations

ARO	Adjustable Robust Optimization.
BCP	Bi-convex Program.
BLP	Bilinear Program.
C&CG	Constraint-and-Column Generation.
CQP	Convex Quadratic Programming.
DCG	Delayed Constraint Generation.
DRO	Distributionally Robust Optimization.
ERD	Energy and Reserve Dispatch.
A-RERD	ARO based ERD.
D-RERD	DRO based ERD.
LMI	Linear Matrix Inequality.
LP	Linear Programming.
MILP	Mixed Integer Linear Programming.
PDF	Probability Distribution Function.
PSD	Positive Semidefinite
SDP	Semidefinite Programming
SO	Stochastic Optimization
VER	Variable Energy Resources.

I. INTRODUCTION

THE ADVANTAGES of renewable generations as clean and cheap energy resources have inspired the dramatic integration of wind and solar energy into power systems during the past decade [1]–[3]. However, the uncertain behavior of such variable energy resources (VERs) also increases the challenges of power system operation [4]–[7], such as the unit commitment and economic dispatch problem, where the generation should balance the demand on a moment-by-moment basis without violating the operating constraints of both generating units and transmission networks.

The variable nature of renewable generation is usually modeled through probability distribution functions (PDFs). Take wind power generation for example, according to a recent survey [8], the probability distribution of wind speed can be described by the Gamma distribution, Weibull distribution, Rayleigh distribution, beta distribution, square-root normal distribution, etc.. According to another study [9], the bimodal Weibull mixture distribution, 4-parameter Kappa distribution and 5-parameter Wakeby distribution are more suitable for offshore wind speeds. Provided the PDF of wind speed, the PDF of usable wind power can be derived from the equation of wind speed and wind turbine generation capability [10]. Despite these widely used PDFs, the Gaussian distribution is also used to approximate the wind power forecast error in short-term generation scheduling problems, [11], [12]. Provided with the PDFs of wind power, the unit commitment, economic dispatch and spinning reserve can be scheduled in the framework of stochastic optimization (SO), including the scenario based models [13]–[17] and the chance-constraints based models [18]–[21]. Most SO approaches minimize the expected cost in their objective functions.

In order to acquire a PDF that reflects the uncertain behavior of actual VER generation with high accuracy, sufficient historical data is required but not always available at hand. The adjustable robust optimization (ARO) proposed in [22] rises an appealing decision-making paradigm for power system generation scheduling problems without exploiting the PDFs of VER generation, and has resulted fruitful outcomes in recent studies, such as the robust unit commitment [23]–[26], robust economic dispatch [27], [28], robust optimal power flow [29], and joint energy and reserve dispatch [30]–[32]. Most of the aforementioned ARO based approaches minimize the cost in the worst-case scenario.

According to current researches, both SO and ARO provide powerful decision-making tools for power system optimization problems under uncertainties. If accurate PDFs are available, the SO method is usually preferred as it produces economically efficient solutions in the statistical point of view; otherwise, the ARO is a good alternative as it only requires the forecast data and can protect the system against a pre-defined uncertainty set, although it is likely to be more conservative than SO.

Inspired by recent advances in the distributionally robust optimization (DRO) [33], [34], this paper intends to develop a comprehensive method that combines the advantages of both SO and ARO, and addresses the energy and reserve dispatch problem with large-scale VER integrations in an economic and reliable way. The contribution of this paper is twofold:

1. We propose a two-stage distributionally robust optimization model for the energy and reserve dispatch (D-RERD for short). The first stage decides the set point of generators and VER plants, as well as the spinning reserve capacity preserved in each generator; In the second stage, corrective actions, including the re-dispatch of generators, renewable power spillage and load shedding, are deployed in real-time dispatch in response to the actual VER generation capability. In the proposed D-RERD, the PDF of uncertain VER generation belongs to a set of multivariate distributions with known expectations and variances. The expected cost of corrective actions under the worst-case renewable output distribution is minimized. Compared with the SO methods, D-RERD does not require the exact PDFs, which may not be available at hand. Compared with ARO methods, D-RERD accounts for the distribution property and makes full use of the variances information. We show D-RERD is equivalent to a semidefinite programming (SDP) and adopt the delayed constraint generation (DCG) algorithm outlined in [34] to solve it in a tractable manner. The master problem renders a relaxed SDP and is gradually tightened. The subproblem is a bi-convex program (BCP). We propose an alternating direction oracle for the non-convex subproblem by solving convex quadratic programming (CQP) and linear programming (LP) iteratively. In this regard, our method relies solely on convex optimization.
2. We compare the performances of the D-RERD with the conventional ARO based energy and reserve dispatch (A-RERD) on the IEEE 118-bus system, from the perspectives of computational efficiency, the total costs and the expected costs, and summarize their advantages and drawbacks, demonstrating that D-RERD is a promising approach for short-term generation scheduling problems under uncertainties. To the best of our knowledge, this is the first comparative study of ARO and DRO in power system applications.

The remaining parts of this paper are organized as follows. The mathematical model of D-RERD is presented and analyzed in Section II. The equivalent SDP formulation of D-RERD and its DCG algorithm is presented in Section III, following which the CQP-LP oracle for the sub-problem is described. The proposed method is applied to the IEEE 118-bus system and compared with A-RERD in Section IV. Conclusions are given in Section V. The mathematical formulation of A-RERD as well as its algorithm are briefly introduced in Appendix for quick reference.

II. Mathematical Formulation

The distinguishing feature of D-RERD stems from the model of uncertainty and the formulation of the second-stage problem. We will first introduce the deterministic formulation of the second-stage problem, then present its primal and dual form in D-RERD, finally give the whole model of D-RERD.

A. Deterministic Formulation of the Second-Stage Problem

Given the first stage decision $x = \{p^f, r^+, r^-, w^f\}$ and the actual VER generation capability w^g, the best corrective actions $y = \{p^+, p^-, p^l, w^c\}$ in real-time dispatch is the optimal solution of the following LP:

$$\min \sum_{i=1}^{N_G} (d_i^{g+} p_i^+ + d_i^{g-} p_i^-) + \sum_{j=1}^{N_W} \rho_j w_j^c + \sum_{q=1}^{N_Q} \rho_q p_q^l \tag{1.1}$$

$$\sum_{q=1}^{N_q} (p_q - p_q^l) = \sum_{j=1}^{N_W} (w_j^g - w_j^c) + \sum_{i=1}^{N_G} (p_i^f + p_i^+ - p_i^-) \tag{1.2}$$

$$\begin{aligned} -F_l \le \sum_{i=1}^{N_G} \pi_{il} (p_i^f + p_i^+ - p_i^-) + \sum_{j=1}^{N_W} \pi_{jl} (w_j^g - w_j^c) \\ - \sum_{q=1}^{N_q} \pi_{ql} (p_q - p_q^l) \le F_l, \forall l \end{aligned} \tag{1.3}$$

$$0 \le p_i^+ \le r_i^+, \forall i, \; 0 \le p_i^- \le r_i^-, \forall i \tag{1.4}$$

$$0 \le w_j^c \le w_j^g, \forall j, \; 0 \le p_q^l \le p_q, \forall q \tag{1.5}$$

where the objective (1.1) minimizes the cost of corrective actions in order to recover the operating constraints, including the re-dispatch cost of generators, the VER power spillage cost and the load shedding cost; constraint (1.2) is the power balancing condition; constraint (1.3) is the security limitation of transmission lines; constraints (1.4) stipulates the real-time adjustment of generators within its spinning reserve capacity preserved in the first stage; constraints (1.5) sets the bounds on the amount of VER power curtailment and load shedding, indicating that the VER power spillage and the load shedding could neither become negative nor exceed the maximal value. The compact form of LP (1.1)-(1.5) with given x and w^g is shown below

$$\begin{aligned} Q(x, w^g) = \min_y f^T y \\ s.t.\ By \le b - Ax - Cw^g \end{aligned} \tag{2}$$

where, matrices A, B, C and vector b correspond to the coefficients in constraints (1.2)-(1.5), vector f corresponds to the coefficients in objective (1.1). Function $Q(x, w^g)$ denotes the optimal value of LP (2) as a function of x and w^g.

B. Worst-Case Distribution Based Primal and Dual Form

In practice, the actual VER generation capability w^g is not known exactly in advance. In SO method, it is described by a random vector $\tilde{w}^g$. A set of scenarios $\{w^s\}, s = 1, 2, \cdots$ are sampled from the PDF of $\tilde{w}^g$, and the objective of the second stage is approximated by the weighted-sum function $\sum_s \Pr_s Q(x, w^s)$ (where $\Pr_s$ is the probability of scenario s).

This paper assumes the exact PDF of VER generation is unavailable due to the lack of enough historical data. In such circumstance, ARO method constructs an deterministic set W in variable $\bar{w}^g$ that covers most possible realizations of uncertainty, and uses the cost in the worst-case scenario $\max_{\bar{w}^g \in W} Q(x, \bar{w}^g)$ as the objective of the second stage.

Combining the idea of SO and ARO, we assume the VER generation in D-RERD follows some multivariate PDF, which is not known exactly and belongs to a functional uncertainty set Ω. We use the expected cost of corrective actions in the worst-case distribution as the objective of the second stage

$$\max_{P \in \Omega} \mathrm{E}_P[Q(x, \tilde{w}^g)] \tag{3.1}$$

where, E_P denotes the expectation operator over distribution P. The PDF P of $\tilde{w}^g$ belongs to the set Ω of all candidate PDFs which is characterized by the forecast vector w^e and covariance matrix Σ, and plays the similar role as the uncertainty set W in ARO

$$\Omega = \left\{ P \left| \begin{array}{l} \Pr\left[\tilde{w}^g \in \mathbb{R}^{N_W}\right] = 1 \\ \mathrm{E}_P\left[\tilde{w}^g\right] = w^e \\ \mathrm{E}_P\left[\tilde{w}^g (\tilde{w}^g)^T\right] = \Sigma + w^e (w^e)^T \end{array} \right. \right\} \tag{3.2}$$

Formulation (3) is different from SO because the PDF P is not known exactly except the forecast w^e and the covariance matrix Σ. It also differs from ARO because the functional uncertainty set Ω is defined for the PDF P rather than the possible realizations of the uncertain variable $\bar{w}^g$, in other words, the VER generation $\tilde{w}^g$ is still random variable described by PDF, thus the distribution property is explicitly taken into account. Moreover, the objective function in (3.1) is an expectation reflecting the statistical behavior of the second-stage cost, rather than that in ARO associates with only a single worst-case scenario. Because the PDF of VER generation is ambiguous, it is prudent to investigate the worst-case distribution $P^* \in \Omega$ that maximizes the expected cost in the second stage. It is clear that formulation (3) inherits the features of both ARO and SO, on the one hand, the requirement on the exact PDF is relaxed but the worst-case outcome is considered; on the other hand, distribution property of the uncertainty is also taken into account.

Compared with the ARO method, formulation (3) has some appealing features. It is known that in the uncertainty set W of ARO, every scenario is treated with the same importance, regardless of its distance from the forecast. In this regard, the conservativeness of ARO largely depends on the selection of uncertainty set. The trade-off is usually manually decided through the utilization of a budget constraint [35], which is a 1-norm inequality and controls the total deviation from the forecast by specifying a budget of uncertainty. On contrary, all information used in D-RERD can be derived from actual data, no trade-off should be made manually. In the functional uncertainty set Ω, because the variance is fixed, a scenario that deviates far away from the forecast would have a low probability. Moreover, it is usually important to tackle the risk caused by the "tail effect" in probability theory, which indicates the occurrence of actual VER generation that is far away from the forecast may induce heavy losses in spite of its low probability. Such phenomenon is naturally taken into account in D-RERD but difficult to be modeled in ARO, because probability information is ignored in uncertainty set W. If the dispatch strategy of ARO is eligible to protect the system against all rare events, the operational cost could be extremely high and unacceptable in practice.

Write problem (3) in an explicit form as

$$\begin{aligned} Z_P(x) = & \max_{f(\tilde{w}^g) \in \Omega} \int_{\mathbb{R}^{N_W}} Q(x, \tilde{w}^g) f(\tilde{w}^g) \mathrm{d}\tilde{w}^g \\ & s.t.\ f(\tilde{w}^g) \geq 0,\ \forall \tilde{w}^g \in \mathbb{R}^{N_W} \\ & \int_{\mathbb{R}^{N_W}} f(\tilde{w}^g) \mathrm{d}\tilde{w}^g = 1 : h_0 \\ & \int_{\mathbb{R}^{N_W}} \tilde{w}_j^g f(\tilde{w}^g) \mathrm{d}\tilde{w}^g = w_j^e : h_j, \\ & j = 1, 2, \cdots, N_W \\ & \int_{\mathbb{R}^{N_W}} \tilde{w}_j^g \tilde{w}_k^g f(\tilde{w}^g) \mathrm{d}\tilde{w}^g = \Sigma_{jk} + w_j^e w_k^e : H_{jk}, \\ & j, k = 1, 2, \cdots, N_W \end{aligned} \tag{4.1}$$

In problem (4.1), the decision variable is the PDF $f(\tilde{w}^g)$, which is a continuous function over $\mathbb{R}^{N_W}$, so the primal problem (4.1) is an infinite dimensional LP, if we regard the integral operator as a generalization of summation. Associating dual variables H_{jk}, h_j, and h_0 shown at the right side of each constraint, the dual form of problem (4.1) can be derived in the spirit of the duality theory of conic linear programming [34], [36] and shown below

$$\begin{aligned} Z_D(x) = & \min_{H, h, h_0} \mathrm{tr}(H^T \Theta) + h^T w^e + h_0 \\ & s.t.\ (\tilde{w}^g)^T H \tilde{w}^g + h^T \tilde{w}^g + h_0 \geq \\ & Q(x, \tilde{w}^g), \forall \tilde{w}^g \in \mathbb{R}^{N_W} \end{aligned} \tag{4.2}$$

where, the second-order moment matrix $\Theta = \Sigma + w^e (w^e)^T$, function tr() stands for the trace of a matrix. Problem (4.2) is a semi-infinite program with finite number of variables and infinitely many constraints. We leave the solution method to the next section. The relationship between the primal problem (4.1) and the dual problem (4.2) is stated below.

Proposition 1 [34], [37]: If the covariance matrix Σ is strict positive, strong duality holds, i.e. $Z_P(x) = Z_D(x) = Z(x)$.

If the renewable power uncertainty in different VER generation centers is independent, Σ will become a diagonal matrix with positive elements Vr_j being its non-zero elements. According to Proposition 1, strong duality holds for problem (4.1) and (4.2). In practice, the renewable generation in different VER centers may be correlated. It is well known in probability theory that the covariance matrix must be positive semidefinite. Nevertheless, we assume the covariance matrix Σ derived from actual data is always positive definite, thus strong duality holds true, regardless whether the randomness is independent or not.

C. Formulation of the D-RERD

The first stage of the D-RERD as well as A-RERD is the following convex quadratic problem that minimizes the total production cost subject to the operating constraints with respect to the VER output forecast w^e

$$\min \sum_{i=1}^{N_G} \left[a_i^2 (p_i^f)^2 + b_i p_i^f + d_i^{r+} r_i^+ + d_i^{r-} r_i^- \right] + \sum_{j=1}^{N_W} \rho_j (w_j^e - w_j^f) \tag{5.1}$$

$$s.t. \; P_i^l \le p_i^f - r_i^-, \, p_i^f + r_i^+ \le P_i^u, \forall i \tag{5.2}$$

$$\sum_{i=1}^{N_G} p_i^f + \sum_{j=1}^{N_W} w_i^f = \sum_{q=1}^{N_Q} p_q \tag{5.3}$$

$$-F_l \le \sum_{i=1}^{N_G} \pi_{il} p_i^f + \sum_{j=1}^{N_W} \pi_{jl} w_j^f - \sum_{q=1}^{N_q} \pi_{ql} p_q \le F_l, \forall l \tag{5.4}$$

$$0 \le r_i^+ \le R_i^+ \Delta t, \, 0 \le r_i^- \le R_i^- \Delta t, \forall i \tag{5.5}$$

$$0 \le w_j^f \le w_j^e, \forall j \tag{5.6}$$

where, objective (5.1) is the total production cost in the first stage, including the generation cost, the spinning reserve cost and VER spillage cost. Constraint (5.2) is the generation capacity limitation considering spinning reserve offer; constraints (5.3) and (5.4) are the power balancing condition and power flow restriction of transmission lines with respect to the VER generation forecast; constraint (5.5) enforces that the spinning reserve capacity offered by each generator cannot exceed its ramping limit in the considered dispatch interval; constraints (5.6) implies that the scheduled output of VER plant should neither become negative nor exceed its predicted generation capability. Load shedding is not allowed in the first stage. Because the quadratic terms in (5.1) are convex, CQP (5) can be directly modeled in commercial solvers, such as CPLEX. If a quadratic solver is not preferred, these quadratic terms can be linearized by the piecewise linear functions [38]. Without loss of generality, problem (5) can be arranged into a compact linear form as

$$\min_{x \in X} c^T x \tag{6}$$

where the constraint set X corresponds to constraints (5.2)-(5.6) which only involves the first-stage decision. Recall Proposition 1 and the dual form of the second-stage problem, the compact form of the D-RERD is given by

$$\min_{x \in X} c^T x + Z_D(x) \tag{7}$$

Some properties regarding the conservativeness and sensitivity of the D-RERD are analyzed as follows.

1. Regarding the conservativeness. On the one hand, the D-RERD is clearly more pessimistic than a corresponding SO model, because the specific probability distribution used in SO is a candidate of the functional uncertainty set Ω defined in (3.2), therefore, it yields a lower bound of the optimal value of problem (3.1). On the other hand, the D-RERD is expected to be less pessimistic than the corresponding ARO model in most cases, but generally not comparable in theory because the conservativeness of ARO also depends on the parameter of the uncertainty set which is manually supplied. More importantly, the tail effect in the worst-case distribution is also a source of conservativeness in D-RERD. It is worth mentioning that the actual output of wind farms belongs to a subset of $\mathbb{R}^{N_W}$. In view of this, we can impose a support set for the PDF in equation (3) to restrict the distribution range of $\tilde{w}_g$. If the support set in D-RERD equals the uncertainty set in A-RERD, the former must be less conservative than the latter.
2. Regarding the sensitivity. D-RERD incorporates a functional uncertainty set Ω, its optimal value provides an upper bound of the expected total cost as long as the PDF of VER generation belongs to the set Ω. In this regard, we can conclude that the optimal value of D-RERD would be less sensitive than the SO method to the perturbation of VER power distribution, although may be more conservative from an economical point of view.

III. Solution Approach

In this section, we develop convex optimization based algorithm to solve D-RERD (7) following the DCG framework in [34]. The master problem is reduced to a SDP, and the sub-problem relies on solving CQPs and LPs iteratively.

A. A SDP Reformulation and the DCG Algorithm

Two obstacles prevent D-RERD (7) from being solved directly:

1) The optimal value function $Q(x, \tilde{w}^g)$ in the constraint of problem (4.2) is not given in a closed form.
2) The dual problem (4.2) includes an infinite number of constraints in variables H, h and h_0.

To represent the optimal value function $Q(x, \tilde{w}^g)$, write out the dual of LP (2)

$$\max_{u \in U} u^T (b - Ax - C\tilde{w}^g)$$
$$U = \{u \,|\, B^T u = f, u \le 0\} \tag{8}$$

Because VER spillage and load shedding is allowed in RTD, LP (1) is always feasible and has a finite optimal value, so does its dual LP (2). Therefore, the optimal solution of LP (2) is bounded and can be found at one of the extreme points of set U, such that

$$\exists u^* \in \text{vert}(U) : \, Q(x, \tilde{w}^g) = (b - Ax - C\tilde{w}^g)^T u^*$$

where $\text{vert}(U) = \{u^1, u^2, \cdots, u^{N_E}\}$ stands for the vertices of polyhedron U, and $N_E = |\text{vert}(U)|$ is the cardinality of vert(U), In view of this equation, the constraint in (4.2) can be expressed as

$$(\tilde{w}^g)^T H \tilde{w}^g + h^T \tilde{w}^g + h_0 \ge (b - Ax - C\tilde{w}^g)^T u$$
$$\forall \tilde{w}^g \in \mathbb{R}^{N_W}, \forall u \in U \tag{9}$$

or in the discrete form

$$(\tilde{w}^g)^T H \tilde{w}^g + h^T \tilde{w}^g + h_0 \geq (b - Ax - C\tilde{w}^g)^T u^i \\ \forall \tilde{w}^g \in \mathbb{R}^{N_W}, i = 1, 2, \cdots, N_E \tag{10}$$

To eliminate variable $\tilde{w}^g$ and derive an explicit constraint in variables H, h, and h_0 from equation (10), notice that it can be written as positive quadratic functions in $\tilde{w}^g$ as

$$(\tilde{w}^g)^T H \tilde{w}^g + (h + C^T u^i)^T \tilde{w}^g + h_0 - (b - Ax) u^i \geq 0 \\ \forall \tilde{w}^g \in \mathbb{R}^{N_W}, i = 1, 2, \cdots, N_E$$

which has the following compact matrix form

$$\begin{bmatrix} \tilde{w}^g \\ 1 \end{bmatrix}^T \begin{bmatrix} H & 0.5(h + C^T u^i) \\ 0.5(h + C^T u^i)^T & h_0 - (b - Ax)^T u^i \end{bmatrix} \begin{bmatrix} \tilde{w}^g \\ 1 \end{bmatrix} \\ \geq 0, \forall \tilde{w}_g \in \mathbb{R}^{N_W}, i = 1, 2, \cdots, N_E$$

Therefore, we can conclude that constraint (10) is equivalent to the following positive semi-definite (PSD) constraints

$$\begin{bmatrix} H & 0.5(h + C^T u^i) \\ 0.5(h + C^T u^i)^T & h_0 - (b - Ax)^T u^i \end{bmatrix} \succeq 0 \\ i = 1, 2, \cdots, N_E \tag{11}$$

Finally, the D-RERD is equivalent to the following SDP

$$\min_{x,H,h,h_0} c^T x + \mathrm{tr}(H^T \Theta) + h^T w^e + h_0 \\ s.t. \begin{bmatrix} H & 0.5(h + C^T u^i) \\ 0.5(h + C^T u^i)^T & h_0 - (b - Ax)^T u^i \end{bmatrix} \succeq 0 \\ i = 1, 2, \cdots, N_E \\ x \in X \tag{12}$$

However, the number of vertices in set U (the cardinality of vert(U), or $|\text{vert}(U)|$ for short) can be extremely large. It is quite difficult to enumerate all of them. A more practical way is to use a subset of vert(U) and formulate a relaxed SDP, then check whether constraint (9) is fulfilled. If yes, the relaxation is exact and the optimal solution is found. If not, find a new vertex of U at which constraint (9) is violated, then add a cut to the relaxed problem so as to tighten the relaxation, till constraint (9) is satisfied. The DCG algorithm for the D-RERD is summarized below

Algorithm 1.

Step 1: Retrieve the VER generation forecast w^e and the covariance matrix Σ. Choose a small tolerance $\varepsilon > 0$, and an arbitrary initial vertex set $VE \subseteq$ vert(U).
Step 2: Solve the following master problem

$$\min_{x,H,h,h_0} c^T x + \mathrm{tr}(H^T \Theta) + h^T w^e + h_0 \\ s.t. \begin{bmatrix} H & 0.5(h + C^T u^i) \\ 0.5(h + C^T u^i)^T & h_0 - (b - Ax)^T u^i \end{bmatrix} \succeq 0 \\ \forall u^i \in VE, x \in X \tag{13.1}$$

Record the optimal value R^* and the optimal solution x^*.
Step 3: Solve the following sub-problem

$$\min w^T H w + h^T w + h_0 - (b - Ax - Cw)^T u \\ s.t. w \in \mathbb{R}^{N_W}, u \in U \tag{13.2}$$

The optimal value is r^*, the optimal solution is u^*.
Step 4: If $r^* \geq 0$, terminate, report the optimal value R^* and the optimal solution x^*; otherwise, $VE = VE \cup u^*$, go Step 2.

Algorithm 1 terminates in a finite number of iterations which is bounded by $|\text{vert}(U)|$. In fact, the actual iterations required by Algorithm 1 can be much less than $|\text{vert}(U)|$, because only a few vertices in vert(U) will contribute binding PSD constraints, moreover, the sub-problem (13.2) in step 3 always identifies the most critical element in $|\text{vert}(U)|$. It is worth mentioning that the sub-problem (13.2) is a nonlinear program due to the product term involving w and u. In spite of the fact that it can be solved by general nonlinear solvers, we will suggest a sequential convex optimization approach to solve it more efficiently.

B. CQP-LP Oracle for the Sub-Problem

This oracle is established on three observations:

1) The constraints on variable w and u are decoupled;
2) The optimal solution H of problem (13.1) is PSD. To see this, if matrix H has an eigenvalue with negative real part, constraint (11) will be violated because the matrix

$$\begin{bmatrix} H & 0.5(h + C^T u^i) \\ 0.5(h + C^T u^i)^T & h_0 - (b - Ax)^T u^i \end{bmatrix} \not\succeq 0$$

3) Given the dual variable H, h and h_0, sub-problem (13.2) is a BCP: fixing w, it is a LP; fixing u, it is a CQP because $H \succeq 0$ (see the previous point).

Above features motivate the following alternating CQP-LP algorithm

Algorithm 2.

Step 1: Choose a tolerance $\delta > 0$. Start with some w^*.
Step 2: Solve the following LP with w^*

$$\min_{u \in U} w^{*T} H w^* + h^T w^* + h_0 - (b - Ax - Cw^*)^T u$$

The optimal value is R_1, the optimal solution is u^*.
Step 3: Solve the following CQP with u^*

$$\min_{w} w^T H w + h^T w + h_0 - (b - Ax - Cw)^T u^*$$

The optimal value is R_2, the optimal solution is w^*.
Step 4: If $R_1 - R_2 < \delta$, terminate, report the optimal value R_2 and the optimal solution u^*; otherwise, go Step 2.

Because the optimal solution of a LP can always be found at one of the vertices of its feasible region, the optimal solution u^* reported by Algorithm 2 must be a vertex of U, thus can be used in Step 4 of Algorithm 1.

The convergence property of Algorithm 2 is addressed in [39]. Based on the three properties revealed at the beginning of this subsection, Algorithm 2 can provide a local optimal solution of BCP (13.2). However, if we use multiple properly chosen initial values of w^*, Algorithm 2 is able to provide a high quality solution which may be close to the global optimal one. We recommend two heuristics for choosing the initial values. If the number of VER plants N_W is small, one can use the 2^{N_W} extreme points of the hypercube $B^V = \{w^g \mid 0 \le w^g_j \le C^V_j, \forall j\}$ as the initial values, where C^V_j is the installed generation capacity of VER plant j. If N_W is large, one can use the following $N_W + 1$ points as the initial values

$$w^g = 0,\ w^g = C^V_j e^{N_W}_j, j = 1, 2, \cdots, N_W$$

where $e^{N_W}_j$ is the j-th column of an $N_W \times N_W$ identity matrix. Certainly, different initial values can be used in Algorithm 2 according to certain heuristics. The trick is that the vectors connecting the forecast and these initial values should cover most directions of the $\mathbb{R}^{N_W}$ space. We found Algorithm 2 is usually faster than directly solving (13.2) using a nonlinear solver, even trying multiple initial values, because LP and CQP are most tractable optimization problems. It is also helpful to restrict $w \in B^V$ when solving CQP in Step 3 if it is unbounded in the first a few steps.

Remark 1

By applying Algorithm 1 to solve D-RERD, we can obtain the first-stage decision x and the dual solution H, h, and h_0 of the second stage. We do not have direct knowledge on the worst-case PDF of VER generation. Nevertheless, the strategy which need to be deployed is only the first-stage decision x. The actual corrective action y should be determined from LP (2) after the uncertainty w^g is revealed. So the primal optimal solution of the second-stage problem is less important in application.

Remark 2

If we impose a support set for the distribution range of $\tilde{w}_g$ in Ω, the equivalent SDP will be different. For instance, if the support set is an ellipsoid, we have to use the S-Lemma [45] to claim positivity of a quadratic function over an ellipsoid; if the support set is a polytope, we can use the positivstellensatz condition [46] to claim positivity of a quadratic function over an semi-algebraic set. However, in both cases, the subproblem will not be biconvex, because H will no longer be a PSD matrix.

IV. Case Studies

In this section, the proposed D-RERD is compared with the A-RERD on the IEEE 118-bus system. The mathematical formulation and solution algorithm of A-RERD is provided in Appendix. Both methods do not require the exact PDF of VER generation, which is the premise of this paper. The data of the IEEE 118-bus system are provided online at: http://motor.ece.iit.edu/data/JEAS_IEEE118.doc. All numeric experiments are conducted on a laptop computer with Intel i5-3210M CPU and 4 GB memory. SDPs are solved by MOSEK, CQPs and LPs are solved by CPLEX 12.5.

A. Data

This test system possesses 54 generating units and 186 transmission lines. In the considered dispatch interval, the total demand is 5500MW. According to the setting in [16], the reserve cost coefficients d^{r+}_i/d^{r-}_i of each generator are assumed to be 10% of its cost coefficient b_i. The regulation cost coefficients d^{g+}_i/d^{g-}_i of each generator are assumed to be the per-unit production cost at the nominal operating point, i.e. $d^{g+}_i = d^{g-}_i = a_i P^u_i + b_i + c_i/P^u_i$. The ramping limits R^+_i/R^-_i of each generator is assumed to be 40% of its maximal output P^u_i. Nine wind farms are connected to the system at bus #70, #12, #17 (Area 1), #49, #59, #77, #80 (Area 2), #100, #92 (Area 3). Their predicted generation capability is 100 MW. The wind power curtailment cost is 5$/MWh. The load shedding cost is 500$/MWh. Other parameter without particular announcement is the same as those provided online.

B. Model of Uncertainty

In D-RERD, the wind generation follows some multivariate probability distribution with known expectations and variances, and the candidate PDFs are restricted in the functional uncertainty set Ω described in (3.2). We further assume the forecast error is independent so that the covariance matrix Σ is a diagonal matrix shown below

$$\Sigma = \begin{bmatrix} \mathrm{Vr}_1 & & \\ & \ddots & \\ & & \mathrm{Vr}_{N_W} \end{bmatrix}$$

The independent assumption usually holds for bulk power systems, where large-scale wind generation centers apart from each other at a distance more than several hundred kilometers. Nevertheless, in practical usage, the matrix Σ should be calculated from actual data, correlation is possible and not a limitation of our method.

In A-RERD, the uncertain wind generation is described by the uncertainty set (A2) in Appendix. In order to identify a proper uncertainty set that is neither too big nor too small, we assume the wind power forecast error follows Gaussian distribution, similar assumption has also been adopted in [11], [12] and [40], then the bound parameter w^l_j and w^u_j can be selected as [40]

$$w^l_j = w^e_j - z(\alpha)\mathrm{Vr}_j,\ w^u_j = w^e_j + z(\alpha)\mathrm{Vr}_j$$

where $z(\alpha)$ represents the α confidence level for the standard Gaussian distribution. In our tests, $\alpha = 99.9\%$ is adopted to guarantee the probability $\Pr[w^l \le w \le w^u]$ is high. Because there are 18 inequalities, $\Pr[w^l \le w \le w^u] = 0.999^{18} = 98.2\%$. According to [41], the root mean square error of the hourly-ahead forecast is about 10% of the predicted output. In our tests, $8\% \le \sqrt{\mathrm{Vr}}/w^e \le 14\%$ is investigated. The corresponding lower bound w^l and upper bound w^u in (A2) are shown in Table I.

The choice of the budget Γ depends on the number of random variables n involved in (A2). If n is large enough, the budget Γ can be selected as $\Gamma \sim O(\sqrt{n})$, which can be attributed to the central limit theorem in the probability theory [24]. According to the study in [24], Γ can vary from $0.5\sqrt{n}$ to $3\sqrt{n}$. In our tests, we increase Γ from 6 to 9 (equivalent to $2\sqrt{N_W} \le \Gamma \le 3\sqrt{N_W}$).

TABLE I
CONFIDENT INTERVALS OF A-RERD

$\sqrt{\mathrm{Vr}}/w^e$	w^l (MW)	w^u (MW)
8%	75.28	124.72
10%	69.10	130.90
12%	62.92	137.08
14%	56.74	143.26

TABLE II
COMPARISON OF COMPUTATIONAL TIMES

$\sqrt{\mathrm{Vr}}/w^e$	D-RERD (s)	A-RERD (s)			
		$\Gamma = 6$	$\Gamma = 7$	$\Gamma = 8$	$\Gamma = 9$
8%	12.9	7.53	8.51	9.23	12.7
10%	13.4	8.11	9.45	11.0	14.9
12%	12.8	8.56	9.16	10.5	13.5
14%	13.1	8.93	14.1	14.0	10.4

It should be pointed out that ARO method itself does not make any specific assumption on the PDF of wind generation. Here Gaussian distribution is used to select the bound parameter, rather than producing scenario and making decisions as what SO method does. There also exist other heuristic methods to select the bounds, which does not depend on the probability distribution, such as the fixed percentage method, where $w_j^l = (1-\beta)w_j^e$ and $w_j^u = (1+\beta)w_j^e$, β is a constant.

C. Results

We compare the performance of D-RERD and A-RERD from four aspects:

1) the computational efficiency.
2) the respective total cost and the second-stage cost.
3) the expected total cost when the forecast error follows a specific distribution.
4) the total cost of D-RERD in the worst-case scenario of A-RERD.

In view that the exact PDF of wind power may not be available to use due to the lack of enough historical data. Therefore, we will not implement a SO model in case study.

In Algorithm 2 of D-RERD, we use $N_W + 1$ initial values as those suggested at the end of Section III. In A-RERD, we use the MILP reformulation method in [31] to solve the nonconvex linear max-min sub-problems. The computation time is shown in Table II, suggesting that the computational efficiency of D-RERD and A-RERD are in general comparable, despite the fact that the latter can be solved faster when Γ is small. This result is inspiring because D-RERD is more close to SO models that accounts for the average performance over all possible scenarios. However, SO models usually take more time to solve than the ARO models which only accounts for the worst-case scenario. The reason is that the second-stage problem of D-RERD is simpler than those of the traditional SO models, in terms of problem scale and tractability.

The optimal values of D-RERD (total expected cost in the worst-case wind power distribution, which is a PDF) and A-RERD (total cost in the worst-case wind power realization, which is a fixed value) under different forecast accuracy are

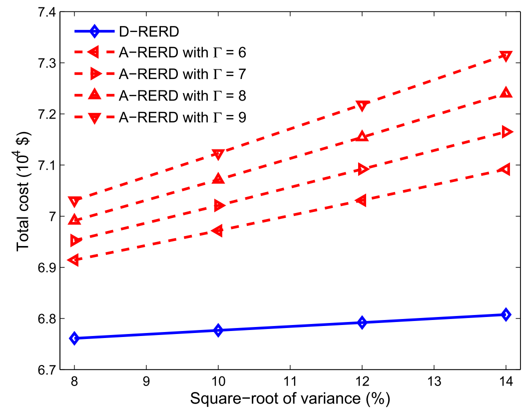

Fig. 1. Comparison of the total costs of D-RERD and A-RERD.

provided in Fig. 1, showing that the total cost of A-RERD are higher than that of D-RERD, due to the different criterions used in the second stage. Moreover, when the wind generation forecast becomes less accurate ($\sqrt{\mathrm{Vr}}/w^e$ becomes larger), the cost of D-RERD grows slower than that of A-RERD. This is also understandable, as D-RERD considers the expected cost in the second stage. Even in the worst-case distribution, most scenarios still appear to be near to the forecast due to the variance constraint in (4.1). If the majority of scenarios leave far away from the forecast, the moment constraint will be violated.

We compared the second-stage cost of D-RERD in the worst-case distribution with that in the Gaussian distribution. The former is the value of Z_D in (4.2), and the latter is obtained from Monte Carlo simulation, which proceeds as follows:

1) Solve D-RERD and fix the first-stage decision;
2) By assuming the forecast error follows Gaussian distribution with the provided mean and variance, generate 10000 samples of w^g. Solve the deterministic second-stage problem (2) for each scenario.
3) Calculate the mean of the second-stage cost.

We are aware that the wind generation may follow other distributions in practice, and if this is true, the results may change. Nevertheless, Gaussian distribution is not used for making decisions or computing the true expectation in practice, it is only used in Monte Carlo simulation for two purposes:

1) Demonstrating the gap between the expected cost in the worst-case distribution and a particular distribution, or how conservative the D-RERD may appear to be.
2) Evaluating the performances of different dispatch strategies offered by D-RERD and A-RERD.

The results are provided in Fig. 2, demonstrating the cost in the Gaussian distribution is about half of that in the worst-case distribution, which indicates the Gaussian distribution is not the worst-case on contrary. For comparison, the second-stage costs of A-RERD in the worst-case scenario under different budget Γ are shown in Fig. 3, which are much larger than that of D-RERD, because they are associated with the worst-case scenario.

Finally, we compare the expected total cost of D-RERD and A-RERD. We first acquire the expected second-stage cost of D-RERD and A-RERD by using the Monte Carlo simulation, and then add them with the corresponding first-stage cost. The

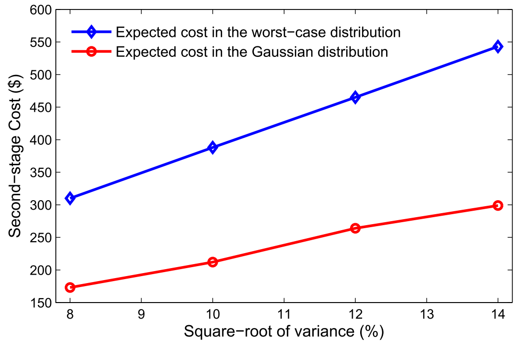

Fig. 2. Expected second stage cost of D-RERD in the worst-case distribution and Gaussian distribution.

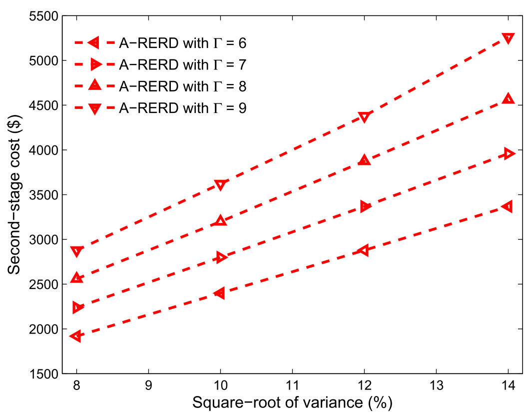

Fig. 3. Second stage cost of A-RERD under different budget Γ.

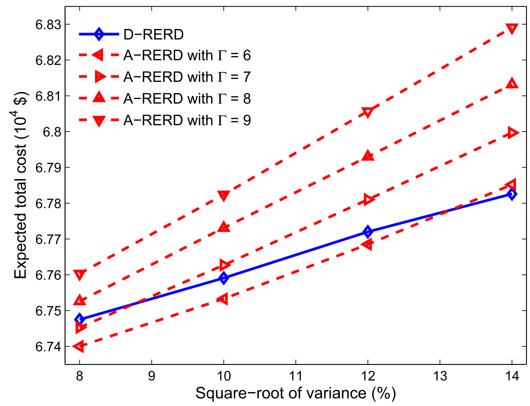

Fig. 4. Comparison of the expected costs of D-RERD and A-RERD.

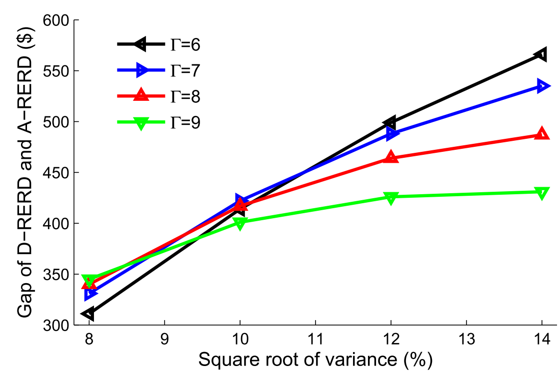

Fig. 5. Comparison of the costs of D-RERD and A-RERD in the worst-case scenario.

results are shown in Fig. 4, indicating that A-RERD is not necessarily always conservative than D-RERD in terms of the expected total cost. When Γ is small, A-RERD may have a lower expected total cost than D-RERD. With the forecast being less accurate, D-RERD is likely to have lower expected total cost. It is worth mentioning that this is partly because of the data set we used, where the reserve cost and regulation cost is in proportion to the marginal cost of production. If the simulation was setup in the opposite way such that units with high marginal cost, e.g., gas-fired units, have low reserve cost and regulation cost, and slower units with low marginal cost have high reserve cost and regulation cost, A-RERD will prefer to dispatch the former ones, since the second-stage cost will be smaller when reserves are fully deployed in the worst-case scenario. This would be suboptimal in terms of expectation, resulting in worse performance of A-RERD compared to D-RERD. Another reason is the Gaussian distribution we assumed while selecting the bound parameter of the uncertainty set. If other distribution is adopted, A-RERD may appear to be more pessimistic (due to slower decay rate of PDF and larger uncertainty set).

Compared with Fig. 1, the expected total cost of A-RERD is much less than the total cost in the worst-case scenario. In practice, the worst-case scenario is quite unlikely to happen, the realization of uncertainty is usually near to the forecast. In this regard, the A-RERD may not be that conservative than it was expected to be. The most important task in A-RERD is to derive the convincing parameters of the uncertainty set, especially the budget of uncertainty Γ. On contrary, there is no manually supplied parameter in D-RERD. Meanwhile, it performs well in a wide range of the forecast accuracy.

The gap between the optimal total costs of D-RERD and A-RERD in the worst-case scenario is shown in Fig. 5. The total cost of A-RERD is the optimal value of problem (A1) in Appendix, i.e., $C_T^A = c^T x_A^* + Q(x_A^*, w_A^*) = c^T x_A^* + f^T y_A^*$. The total cost of D-RERD C_T^D is procured in the following way: the first-stage decision x_D^* is provided by Algorithm 1, and the second-stage decision y_D^* is obtained from the optimal solution of LP (2) with the worst-case scenario w_A^* of A-RERD. In this way, the total cost of D-RERD is $C_T^D = c^T x_D^* + Q(x_D^*, w_A^*) = c^T x_D^* + f^T y_D^*$, which must be larger than C_T^A, because both stages of A-RERD are optimized for w_A^*. This assertion is certified by Fig. 5, from which we can see the gap $C_T^D - C_T^A$ is always positive. Fig. 5 also illustrates the following two facts: on the one hand, for a fixed budget Γ, with the forecast becoming less accurate, the gap grows larger, as the worst-case scenario w_A^* leaves farther from the forecast, and the results of D-RERD become less optimized for w_A^*; on the other hand, with the budget Γ growing, the gap is becoming less sensitive to the accuracy of forecast. These gaps demonstrate that deploying the strategy of D-RERD will not cause disastrous contingencies even under the worst-case scenarios of A-RERD, although may incur a higher but acceptable cost in these rare events.

Finally, the reserve capacity scheduled by D-RERD and A-RERD are given in Fig. 6. In this case, only up-reserve capacity is scheduled and down-reserve capacity is 0 because the wind

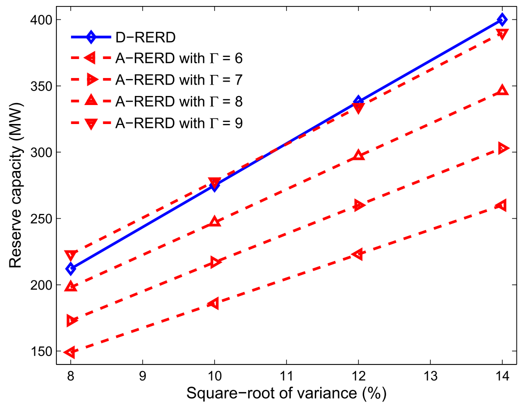

Fig. 6. Reserve capacity scheduled by D-RERD and A-RERD.

curtailment cost is lower than the down-regulation cost. As a result, excessive wind generation will be curtailed rather than deploying down-reserve capacity. The situation may change when both the wind power and reserve capacity participate in a joint energy and reserve market, where their respective cost is comparable. From Fig. 6, it seems D-RERD will dispatch more reserve capacity than A-RERD. The reason rests on the phenomenon called "tail effect", i.e. in the worst-case distribution, a larger share of wind power occurrence tends to realize as far away from the forecast as possible, subjecting to the constraints in (3.2). A-RERD is difficult to model such phenomenon because the uncertainty set W is a bounded polytope that does not involve any probability information. Despite the fact that more reserve capacity is scheduled, we have already known from Fig. 4 that the total expected cost of D-RERD is still lower than that of A-RERD in most cases. Moreover, the optimal value of D-RERD provides an upper bound of the expected total cost when the wind generation follows the PDFs in set Ω.

V. Conclusions

This paper proposes a distributionally robust optimization model for the energy and reserve dispatch problem, namely the D-RERD. Compared with the SO method which requires the exact PDF of wind generation, the data-driven D-RERD only needs the forecast and variance. In theory, D-RERD will be more pessimistic than the SO method. Meanwhile, D-RERD is expected to be less pessimistic than the ARO method in a wide range of forecast accuracy, but in general not comparable, because the conservativeness of the latter can vary when different uncertainty set is adopted. D-RERD is especially suitable to tackle the tail effect, which is difficult to be modeled in the conventional ARO method. The conservativeness of DRO can be further reduced by introducing a support set that restricts the distribution range of uncertainty, which is our ongoing research. Case studies on the IEEE 118-bus system demonstrate that D-RERD and A-RERD has similar computational efficiency. In view of these, D-RERD could become a promising alternative method for power system dispatch problems under uncertain renewable generations, and has the potential to fill the gap between the SO method and the ARO method.

Appendix

Recall the compact formulation of the first-stage and the second-stage problem in Section II, the mathematical model of A-RERD is a linear min-max-min problem shown below

$$\min_{x\in X} c^T x + \max_{\bar{w}^g \in W} \min_{y\in Y(x,\bar{w}^g)} f^T y \tag{A1}$$

where the set $Y(x,\bar{w}^g)$ is defined as

$$Y(x,\bar{w}^g) = \{y | By \le b - Ax - C\bar{w}^g\}$$

and W is a pre-defined uncertainty set as follows

$$W = \left\{ \bar{w}^g \left| \begin{array}{l} w_j^l \le \bar{w}_j^g \le w_j^u, \forall j \\ \sum_{j=1}^{N_W} \dfrac{|\bar{w}_j^g - w_j^e|}{w_j^h} \le \Gamma \end{array} \right. \right\} \tag{A2}$$

The first constraint enforces $\bar{w}_j^g$ within the interval $[w_j^l, w_j^u]$, the second constraint restricts the total deviation of the actual VER generation from the forecast, parameter Γ is called the budget of uncertainty. Cardinality constrained uncertainty set or general polytope uncertainty set can also be used. For more information about the above A-RERD, please refer to [30]–[32]. A-RERD (A1) is a standard ARO. Several decomposition algorithms are available to solve this particular tri-level optimization problem, such as the benders-type decomposition algorithm in [23] and [24], and the column-and-constraint generation (C&CG) algorithm in [42]. In this paper, we use the C&CG algorithm to solve A-RERD, which is briefly outlined below.

Algorithm 3.

Step 1: Set $LB = 0$, $UB = 1$, tolerance $\delta > 0$, $k = 1$ and $w^1 = w^e$.

Step 2: Solve the following master problem

$$\begin{aligned} LB = \min c^T x + \sigma \\ s.t.\, x \in X \\ Ax + By^r \le b - C\bar{w}^r, \forall r \le k \\ \sigma \ge f^T y^r,\ \forall r \le k \end{aligned}$$

Record the optimal value LB and the optimal decision x^*.

Step 3: Solve the following sub-problem with x^*

$$\max_{\bar{w}^g \in W} \min_{y \in Y(x^*,\bar{w}^g)} f^T y$$

$k = k + 1$, the optimal value is R_2, update $UB = c^T x^* + R_2$, the worst-case scenario under x^* is $\bar{w}^k$.

Step 4: If $UB - LB < \delta$, terminate, report the optimal solution x^* and the optimal value $0.5(LB + UB)$; otherwise, create new variable y^k and add the following constraints to the master problem, go to Step 2

$$\begin{aligned} Ax + By^k \le b - C\bar{w}^k \\ \sigma \ge f^T y^k \end{aligned}$$

Algorithm 3 terminates in $O(|\text{vert}(W)|)$ iterations [42] where $|\text{vert}(W)|$ denotes the number of extreme points of W, if it is a polytope or the cardinality of W if it is a discrete set with countable elements. The linear max-min problem in Step 3 is equivalent to the following bilinear program [43]

$$\begin{aligned} &\max_{u,\bar{w}^g} u^T(b - Ax - C\bar{w}^g) \\ &\quad s.t.\, \bar{w}^g \in W,\, u \in U \end{aligned} \tag{A3}$$

where u is the dual variable associating with the constraints in $Y(x, \bar{w}^g)$, set U is defined in (8). BLP (A3) can be solved by the alternating direction method in [27] and [44], or can be reformulated as a MILP such as the method in [23], [31], and [32], depending on the type of the uncertainty set W. In this paper, we adopt the method in [31], which is able to find the global optimal solution of BLP (A3) with W being a general polytope. In order to use the reformulation technique in [31], we use the following equivalent form of W

$$W = \left\{ \bar{w}^g \,\middle|\, \begin{array}{l} \exists v : 1^T v \le \Gamma,\, w_j^l \le \bar{w}_j^g \le w_j^u, \forall j \\ -v_j \le (\bar{w}_j^g - w_j^e)/w_j^h \le v_j, \forall j \end{array} \right\}$$

which is an explicit polyhedron with $2N_W$ variables and $4N_W + 1$ constraints. Therefore, the number of binary variables in the resulting MILP will be $4N_W + 1$ and independent from the power system model.

Acknowledgment

We would like to express our sincere gratitude and appreciation to the anonymous reviewers, who provided insightful comments and constructive suggestions on improving the quality of this paper.

References

[1] X. Yang, Y. Song, G. Wang, and W. Wang, “A comprehensive review on the development of sustainable energy strategy and implementation in China,” *IEEE Trans. Sustain. Energy*, vol. 1, no. 2, pp. 57–65, Jul. 2010.
[2] D. P. Manjure, Y. Mishra, S. Brahma, and D. Osborn, “Impact of wind power development on transmission planning at midwest ISO,” *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 845–852, Oct. 2012.
[3] N. Aparicio, I. MacGill, J. R. Abbad, and H. Beltran, “Comparison of wind energy support policy and electricity market design in Europe, the United States, and Australia,” *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 809–818, Oct. 2012.
[4] J. Kabouris and F. D. Kanellos, “Impacts of large-scale wind penetration on designing and operation of electric power systems,” *IEEE Trans. Sustain. Energy*, vol. 1, no. 2, pp. 107–114, Jul. 2010.
[5] Y. V. Makarov, P. V. Etingov, J. Ma, Z. Huang, and K. Subbarao, “Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures,” *IEEE Trans. Sustain. Energy*, vol. 2, no. 4, pp. 433–442, Oct. 2011.
[6] J. Kiviluoma *et al.*, “Short-term energy balancing with increasing levels of wind energy,” *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 769–776, Oct. 2012.
[7] L. Soder *et al.*, “Experience and challenges with short-term balancing in European systems with large share of wind power,” *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 853–861, Oct. 2012.
[8] J. A. Carta, P. Ramirez, and S. Velazquez, “A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands,” *Renew. Sustain. Energy Rev.*, vol. 13, no. 5, pp. 933–955, Jun. 2009.
[9] E. C. Morgan, M. Lackner, R. M. Vogel, and L. G. Baise, “Probability distributions for offshore wind speeds,” *Energy Convers. Manage.*, vol. 52, no. 1, pp. 15–26, Jan. 2011.
[10] D. Villanueva and A. Feijóo, “Wind power distributions: A review of their applications,” *Renew. Sustain. Energy Rev.*, vol. 14, no. 5, pp. 1490–1495, Jun. 2010.
[11] C. Lowery and M. O’Malley, “Impact of wind forecast error statistics upon unit commitment,” *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 760–768, Oct. 2012.
[12] R. Doherty and M. O’Malley, “A new approach to quantify reserve demand in systems with significant installed wind capacity,” *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 587–595, May 2005.
[13] J. M. Morales, A. J. Conejo, and J. Pérez-Ruiz, “Economic valuation of reserves in power systems with high penetration of wind power,” *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 900–910, May 2009.
[14] A. Papavasiliou, S. Oren, and R. O’Neill, “Reserve requirements for wind power integration: A scenario-based stochastic programming framework,” *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2197–2206, Nov. 2011.
[15] C. Sahin, M. Shahidehpour, and I. Erkmen, “Allocation of hourly reserve versus demand response for security-constrained scheduling of stochastic wind energy,” *IEEE Trans. Sustain. Energy*, vol. 4, no. 1, pp. 219–228, Jan. 2013.
[16] A. Ahmadi-Khatir, A. J. Conejo, and R. Cherkaoui, “Multi-area energy and reserve dispatch under wind uncertainty and equipment failures,” *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4373–4383, Nov. 2013.
[17] 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.
[18] 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.
[19] H. Wu, M. Shahidehpour, Z. Li, and W. Tian, “Chance-constrained day-ahead scheduling in stochastic power system operation,” *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1583–1591, Jul. 2014.
[20] H. Zhang and P. Li, “Chance constrained programming for optimal power flow under uncertainty,” *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2417–2424, Nov. 2011.
[21] X. Liu and W. Xu, “Economic load dispatch constrained by wind power availability: A here-and-now approach,” *IEEE Trans. Sustain. Energy*, vol. 1, no. 1, pp. 2–9, Apr. 2010.
[22] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski, “Adjustable robust solutions of uncertain linear programs,” *Math. Program.*, vol. 99, no. 2, pp. 351–376, Mar. 2004.
[23] 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.
[24] 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.
[25] Y. An and Z. Bo, “Exploring the modeling capacity of two-stage robust optimization: Variants of robust unit commitment model,” *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 109–122, Jan. 2015.
[26] A. Street, F. Oliveira, and J. M. Arroyo, “Contingency-constrained unit commitment with n–K security criterion: A robust optimization approach,” *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1581–1590, Aug. 2015.
[27] 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.
[28] Y. Zhang, N. Gatsis, and G. B. Giannakis, “Robust energy management for microgrids with high-penetration renewables,” *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 944–953, Oct. 2013.
[29] R. A. Jabr, “Adjustable robust OPF with renewable energy sources,” *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4742–4751, Nov. 2013.
[30] A. Street, A. Moreira, and J. M. Arroyo, “Energy and reserve scheduling under a joint generation and transmission security criterion: An adjustable robust optimization approach,” *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 3–14, Jan. 2014.
[31] M. Zugno and A. J. Conejo, “A robust optimization approach to energy and reserve dispatch in electricity markets,” *Eur. J. Oper. Res.*, vol. 247, no. 2, pp. 659–671, Dec. 2015.
[32] W. Wei, F. Liu, S. Mei, and Y. Hou, “Robust energy and reserve dispatch under variable renewable generation,” *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 369–380, Jan. 2015.
[33] E. Delage and Y. Ye, “Distributionally robust optimization under moment uncertainty with application to data-driven problems,” *Oper. Res.*, vol. 58, no. 3, pp. 595–612, May 2010.

[34] D. Bertsimas, X. Doan, K. Natarajan, and C. Teo, "Models for minimax stochastic linear optimization problems with risk aversion," *Math. Oper. Res.*, vol. 35, no. 3, pp. 580–602, Apr. 2010.

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

[36] A. Shapiro, "On duality theory of conic linear problems," in *Semi-infinite Programming*. Norwell, MA, USA: Kluwer, 2001, ch. 7, pp. 135–165.

[37] D. Bertsimas and I. Popescu, "Optimal inequalities in probability theory: A convex optimization approach," *SIAM J. Optim.*, vol. 15, no. 3, pp. 780–804, 2005.

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

[39] J. Korski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: a survey and extensions," *Math. Methods Oper. Res.*, vol. 66, no. 3, pp. 373–407, Dec. 2007.

[40] Y. Guan and J. Wang, "Uncertainty sets for robust unit commitment," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1439–1440, May 2014.

[41] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renew. Energy*, vol. 37, no. 1, pp. 1–8, Jan. 2012.

[42] B. Zeng and L. Zhao, "Solving two-stage robust optimization problems using a column-and-constraint generation method," *Oper. Res. Lett.*, vol. 41, no. 5, pp. 457–461, Sep. 2013.

[43] J. E. Falk, "A linear max-min problem," *Math. Program.*, vol. 5, no. 1, pp. 169–188, 1973.

[44] R. Jiang, J. Wang, M. Zhang, and Y. Guan, "Two-stage minimax regret robust unit commitment," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2271–2282, Aug. 2013.

[45] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge university press, 2004.

[46] M. Putinar, "Positive polynomials on compact semi-algebraic sets," *Indiana Univ. Math. J.*, vol. 42, no. 3, pp. 969–984, 1993.

Wei Wei received the B.Sc. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2008 and 2013, respectively. He was a Postdoctoral Research Fellow with Tsinghua University from 2013 to 2015. He was a Visiting Scholar at Cornell University, Ithaca, NY, USA, in 2014, and Harvard University, Cambridge, MA, USA, in 2015. He is currently a Research Assistant Professor with Tsinghua University. His research interests include applied optimization and energy economics.

Feng Liu (M'10) received the B.Sc. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1999 and 2004, respectively. He is currently an Associate Professor with Tsinghua University. His research interests include power system analysis and control, and renewable generation.

Shengwei Mei (F'15) received the B.Sc. degree in mathematics from Xinjiang University, Urumqi, China, the M.Sc. degree in operations research from Tsinghua University, Beijing, China, and the Ph.D. degree in automatic control from Chinese Academy of Sciences, Beijing, China, in 1984, 1989, and 1996, respectively. He is currently a Professor with Tsinghua University. His research interests include power system analysis and control, game theory and its application in power systems.