Two-Stage Robust Unit Commitment for Co-Optimized Electricity Markets: An Adaptive Data-Driven Approach for Scenario-Based Uncertainty Sets

Alexandre Velloso, *Student Member, IEEE,* Alexandre Street, *Senior Member, IEEE,* David Pozo, *Member, IEEE,* José M. Arroyo, *Senior Member, IEEE,* Noemi G. Cobos, *Student Member, IEEE*

Abstract—Two-stage robust unit commitment (RUC) models have been widely used for day-ahead energy and reserve scheduling under high renewable integration. The current state of the art relies on budget-constrained polyhedral uncertainty sets to control the conservativeness of the solutions. The associated lack of interpretability and parameter specification procedures, as well as the high computational burden exhibited by available exact solution techniques call for new approaches. In this work, we use an alternative scenario-based framework whereby uncertain renewable generation is characterized by a polyhedral uncertainty set relying on the direct specification of its vertexes. Moreover, we present a simple, yet efficient, adaptive data-driven procedure to dynamically update the uncertainty set vertexes with observed daily renewable-output profiles. Within this setting, the proposed data-driven RUC ensures protection against the convex hull of realistic scenarios empirically capturing the complex and time-varying intra-day spatial and temporal interdependences among renewable units. The resulting counterpart features advantageous properties from a computational perspective and can be effectively solved by the column-and-constraint generation algorithm until ϵ-global optimality. Out-of-sample experiments reveal that the proposed approach is capable of attaining efficient solutions in terms of cost and robustness while keeping the model tractable and scalable.

Index Terms—Adaptive data-driven approach, energy and reserve scheduling, renewable integration, robust optimization, scenario-based uncertainty set, unit commitment.

Nomenclature

This section lists the notation. Bold symbols are reserved to matrices (uppercase) and vectors (lowercase). Additional symbols with superscript "(k)" denote new variables corresponding to the k-th scenario selected by the solution method.

A. Sets

$\mathcal{F}$	Feasible set for the decision variables associated with thermal generators.
$\mathcal{H}$	Set of time periods.
$\mathcal{R}$	Set of renewable energy generators.
$\mathcal{S}$	Set of renewable energy generation scenarios.
$\mathcal{S}_j$	Subset of $\mathcal{S}$ including the scenarios selected until iteration j.
$\mathcal{U}$	Uncertainty set defined as the convex hull of $\mathcal{S}$.
$\mathcal{X}$	Set of first-stage generation-related variables.

B. Constants

Γ, Λ	Spatial and temporal budgets used in the budget-constrained uncertainty set formulation.
$\Delta_{ih}^+, \Delta_{ih}^-$	Upper and lower deviation bounds for renewable generator i and period h used in the budget-constrained uncertainty set formulation.
ϵ	Feasibility tolerance.
$\mathbf{A}$	Line-bus incidence matrix.
$\mathbf{B}$	Thermal generator-bus incidence matrix.
$\mathbf{c}^g$	Vector of fuel costs of thermal generators.
$\mathbf{c}^{dn}, \mathbf{c}^{up}$	Vectors of cost rates for down- and up-spinning reserves.
$\mathbf{d}_h$	Vector of nodal consumptions in period h.
$\mathbf{e}$	Vector of ones with appropriate dimension.
$\bar{\mathbf{f}}$	Vector of line capacities.
$\overline{\mathbf{G}}, \underline{\mathbf{G}}$	Diagonal matrices of maximum and minimum generation limits of thermal units.
$\overline{G}_i$	Capacity of thermal unit i.
K	Number of days for observed renewable energy generation data.
$\mathbf{P}$	Renewable generator-bus incidence matrix.
RD_i, RU_i	Ramp-down and ramp-up limits of thermal unit i within two consecutive periods.
$\mathbf{R}^{dn}, \mathbf{R}^{up}$	Diagonal matrices of downward and upward limits for corrective actions of thermal units within each period.
$\mathbf{S}$	Angle-to-flow matrix.
SD_i, SU_i	Shut-down and start-up ramp rates for thermal unit i.
$\mathbf{u}_{hk}$	Vector of renewable energy generation in period h for scenario k.
$\mathbf{U}_k$	Scenario k of $\mathcal{S}$.
$\hat{\mathbf{u}}_h$	Vector of expected day-ahead renewable energy generation levels in period h.
$\hat{u}_{ih}$	Expected day-ahead power output of renewable generator i in period h.

C. First-Level Decision Vectors

$\boldsymbol{\theta}_h$	Phase angles in period h.
$\mathbf{c}_h^{sd}, \mathbf{c}_h^{su}$	Shut-down and start-up costs in period h.
$\mathbf{f}_h$	Line power flows in period h.
$\mathbf{g}_h$	Nominal generation levels in period h.
$\mathbf{r}_h^{dn}, \mathbf{r}_h^{up}$	Down- and up-spinning reserves in period h.

$\mathbf{v}_h$ Binary on/off statuses in period h.

D. Second- and Third-Level Decision Variables

α_k Second-level decision variable representing the convex combination weight of scenario $\mathbf{U}_k$ used in the scenario-based uncertainty set $\mathcal{U}$.

$\boldsymbol{\theta}_h^{wc}$ Third-level decision vector representing the worst-case nodal phase angles in period h.

$\mathbf{f}_h^{wc}$ Third-level decision vector representing the worst-case line power flows in period h.

$\mathbf{g}_h^{wc}$ Third-level decision vector representing the worst-case generation redispatch in period h.

$\mathbf{s}_h^+, \mathbf{s}_h^-$ Third-level decision vectors representing the renewable energy spillage and load shedding in period h.

$\mathbf{U}$ Second-level decision matrix representing the generation levels for all renewable units across the day-ahead scheduling horizon.

$\mathbf{u}_h$ Second-level decision vector representing the h-th column of $\mathbf{U}$.

u_{ih} Second-level decision variable representing the i-th element of $\mathbf{u}_h$.

z_{ih}^+, z_{ih}^- Second-level decision variables, used in the budget-constrained uncertainty set formulation, representing upward and downward deviations from the expected value for renewable generator i in period h.

E. Dual Variables

$\boldsymbol{\beta}_h$ Vector of dual variables associated with the power balance equations in period h.

$\boldsymbol{\gamma}_h, \boldsymbol{\tau}_h$ Vectors of dual variables associated with generation redispatch limits in period h.

$\boldsymbol{\zeta}_h, \boldsymbol{\kappa}_h$ Vectors of dual variables associated with ramping-up and ramping-down constraints in period h.

$\boldsymbol{\sigma}_h, \boldsymbol{\pi}_h$ Vectors of dual variables associated with the power flow capacity constraints in period h.

$\boldsymbol{\varsigma}_h, \boldsymbol{\xi}_h$ Vectors of dual variables associated with load shedding and renewable energy spillage limits in period h.

$\boldsymbol{\omega}_h$ Vector of dual variables associated with Kirchhoff's second law constraints in period h.

F. Functions

$\Phi(\cdot)$ Worst-case system power imbalance.

$\mathbf{a}(\cdot), \mathbf{b}(\cdot)$ Vector functions used to enforce ramping limits.

I. INTRODUCTION

NON-DISPATCHABLE renewable energy generation (REG) has undergone a sharp increase in the last decades and is already one of the major components in some electricity markets. High integration of these intermittent and variable energy sources brings additional challenges to short-term power system operation that are well known and have been widely discussed [1]–[12]. Briefly, REG variability, especially from wind power units, is driven by complex time-varying spatial and temporal dynamics [13]. In order to benefit from REG resources, a mix of conventional generation and expensive operational actions are both needed to constantly deploy (up and down) reserves in a fast and reliable way. For this reason, the uncertainty inherent to REG should be precisely accounted for in the scheduling and dispatch models used to determine appropriate levels of energy and reserves, such as those adopted in currently implemented co-optimized electricity markets [6], [11], [12], [14]–[18].

Due to the appealing tradeoff between tractability and accuracy, two-stage adaptive robust optimization has been used to deal with uncertainty in day-ahead generation scheduling [2]–[12]. The interested reader is referred to [19] for a detailed literature review. In such robust unit commitment (RUC) models, a trilevel optimization problem is built to characterize the min(decision)-max(uncertainty)-min(decision) structure. Within such a scheme, the first-level problem determines, before the observation of the uncertain parameters, the day-ahead commitment for each generator. In the second-level problem, the worst-case scenario of uncertainties is selected within a given polyhedral uncertainty set as a function of the first-level decisions. Finally, in the third level, the best operational reaction (redispatch) is obtained for the second-level scenario within the first-level scheduled resources.

A key aspect in RUC is the way that uncertainties are represented. The success of an RUC model mainly depends on the selection of an uncertainty set that is capable of capturing the main patterns present in the uncertain parameters while keeping model tractability. Valuable and thorough discussions on the subject can be found in [19] and [20].

Most previously reported two-stage RUC models [2]–[12] rely on the budget-constrained polyhedral uncertainty set presented in [21]. The specification of a budget-constrained uncertainty set is made through linear inequality constraints defining the boundaries of a polyhedron. Such boundaries are set up by componentwise box-like limits and linear (budget) constraints limiting the number of components deviating from their nominal scenario. In this framework, each vertex of the polyhedron representing the uncertainty set is indirectly determined by the intersection of constraints, which may hinder the physical interpretation of the scenarios.

Variants of conventional budget-constrained uncertainty sets have been proposed in [22] and applied in [5]–[10] to better model the variability of renewable power generation. In [5], an interesting approach based on linear models was proposed to improve the dynamics of REG scenarios. The use of parametric linear models was also described in [6]. For those cases, the polyhedral uncertainty sets were defined based on affine constraints representing linear models for renewable injections. In addition, budget constraints were applied to the residuals of the models to control the conservativeness of solutions. In [7], An and Zeng introduced variants of robust models based on linear expected-value operators (averages) applied to multiple worst-case operational costs. In such a work, each worst case resulted from a different budget-constrained uncertainty set centered on an exogenously generated scenario. Multiple spatial and temporal budget constraints were presented in [8] to increase the modeling capability. In [9], a flexible uncertainty set was characterized in terms of a user-defined parameter to

capture the risk of misestimating the box-like limits for REG levels. In [10], a flexible uncertainty set was adjusted over time to provide a tradeoff between economics and robustness of the generation schedule. Notwithstanding, the approaches presented in [5]–[10] rely on budget-constrained uncertainty sets. Therefore, the modeling choices for describing REG variability, which features complex, nonlinear, and time-varying dynamics [13], are restricted to linear models due to tractability issues.

From a computational perspective, another drawback associated with the use of budget-constrained uncertainty sets is the combinatorial growth of the number of vertexes of the polyhedron with respect to the number of uncertain parameters. The resulting RUC models based on budget-constrained uncertainty sets are challenging instances of trilevel programming that are generally addressed through decomposition-based methods. The state-of-the-art techniques such as the column-and-constraint generation algorithm (CCGA) [23] and Benders decomposition [3] involve the iterative solution of a master problem and a subproblem, also known as oracle subproblem. In the related literature, the oracle subproblem represents the worst-case-scenario search procedure corresponding to the two lowermost optimization levels of the trilevel counterpart. Hence, the oracle is an instance of bilevel programming, which, in the presence of budget-constrained uncertainty sets, is NP-hard [3]. Solution techniques available to tackle the oracle subproblem can be categorized in two groups. On the one hand, heuristic yet efficient methods, such as variants of the outer-approximation algorithm [24], were applied in [2]–[5] and [8]. On the other hand, in [6], [7], [9]–[12], well-known linearization procedures, relying on the binary representation of the vertexes of the polyhedron characterizing the uncertainty set, were used to obtain exact yet computationally expensive single-level equivalents based on mixed-integer linear programming (MILP). Thus, existing solution methodologies for RUC models with budget-constrained uncertainty sets either efficiently provide a solution without being able to acknowledge optimality or rely on exact MILP-based NP-hard models that are challenging to solve in practice.

Motivated by the above issues of existing works [2]–[12] and the wide availability of REG data, the objective of this paper is to propose an alternative to the use of budget-constrained uncertainty sets in RUC. In this work, we consider the RUC for co-optimized electricity markets, i.e., the centralized robust joint scheduling of energy and reserves targeting total cost minimization. Here, as suggested in [25], the uncertainty characterization is directly connected to data. To that end, we propose modeling the REG uncertainty in day-ahead RUC by an alternative scenario-based polyhedral uncertainty set that is built through a novel data-driven approach. Based on the general scenario-based uncertainty set description provided in [26], we define a new polyhedral uncertainty set as the convex hull of a set of exogenously generated multivariate points, or scenarios, capturing relevant information regarding the uncertainty process over a given time window. Thus, differently from [7], each vertex of the polyhedron representing the uncertainty set is defined as one of these exogenous scenarios. In the proposed data-driven framework, scenarios represent observed daily renewable-generation profiles, i.e., matrices whose dimension is given by the number of renewable units and the number of time periods of the scheduling horizon, typically 24 hours. Hereinafter, the proposed data-driven scenario-based uncertainty set is referred to as DDUS.

Two recent examples of successful application of the scenario-based uncertainty sets first proposed in [26] can be found in [27] and [28]. Within a finance context, the vertexes of the uncertainty set were generated directly from most recent observed data in [27]. Using a general mathematical setting, in [28], sampled points were endogenously selected to belong to the uncertainty set through embedded statistical hypothesis tests. In the context of RUC, however, this alternative framework has not been explored yet despite its relevant benefits.

From a modeling perspective, the use of the proposed uncertainty characterization for RUC is advantageous in several aspects as compared with previous models relying on budget-constrained uncertainty sets [2]–[12]. First, the true underlying uncertainty process drives the construction of the polyhedron representing the uncertainty set, which is made up of vertexes with high physical interpretability. As a consequence, the resulting RUC features relevant information about the complex and time-varying temporal and spatial dependences found in REG within the scheduling horizon. Moreover, the novel data-driven procedure devised to build polyhedral uncertainty sets through their vertexes is an entirely exogenous adaptive and nonparametric process. Therefore, the proposed approach paves the way for the use of a wide range of existing scenario-generation methods as alternatives to the proposed data-driven procedure. For instance, any nonlinear model or data-processing scheme useful for defining or preprocessing the scenarios, such as clustering, data categorization, or filtering processes based on weather-related and real-time dispatch information, can be used to generate vertexes for the uncertainty set.

The incorporation of the proposed DDUS in RUC is also beneficial from a methodological perspective. Similar to existing models [2]–[12], the resulting data-driven formulation, denoted by DDRUC, is suitable for the state-of-the-art CCGA. In addition, as a salient feature, the proposed scenario-based robust framework is characterized by a relevant property: one of the multivariate points within the given time window of observed data is the worst-case vertex provided by the optimal solution of the oracle subproblem. This property results in an oracle that is solvable in polynomial time [28], unlike the oracle subproblems described in [2]–[12]. It is also worth mentioning that, according to our empirical findings, which are consistent with those reported in [27], the use of DDUS typically requires a narrow time window to attain high-quality solutions. This aspect is particularly relevant for the practical adoption by system operators to schedule generation in electricity markets.

The contributions of this paper are threefold:

1) To raise awareness of the modeling capability and computational advantages of the scenario-based polyhedral uncertainty sets proposed in [26] to address REG uncertainty in RUC problems. Within this general framework, we propose defining scenarios as matrices representing

the hourly generation profiles of all renewable units within a day. Hence, for the first time in the RUC literature, REG variability is described by a polyhedral uncertainty set relying on the convex hull of a polynomial set of multivariate points.

2) To propose a nonparametric data-driven procedure that defines the vertexes of the resulting polyhedral uncertainty set directly from observed data, thereby embedding the true complex and time-varying interdependences among renewable units.

3) To present a novel data-driven two-stage robust unit commitment model that is scalable, easy to specify, and suitable for the exact CCGA due to the resulting computationally inexpensive and polynomial-time-solvable oracle subproblem.

The rest of this paper is organized as follows. In Section II, the alternative uncertainty set is described. The proposed RUC model is formulated in Section III. The solution methodology is presented in Section IV. An evaluation procedure for assessing the devised model and numerical experience are reported in Section V. Finally, this paper is concluded in Section VI.

II. Uncertainty Set Characterization

In generation scheduling under high penetration of renewable-based generation, uncertain data comprise the day-ahead power output for each renewable generator $i \in \mathcal{R}$ and each time period $h \in \mathcal{H}$. Thus, we represent uncertainty by a renewable generator-by-time matrix, $\mathbf{U} \in \mathbb{R}^{|\mathcal{R}| \times |\mathcal{H}|}$, where $\mathbf{U} = [\mathbf{u}_1, ..., \mathbf{u}_h, ..., \mathbf{u}_{|\mathcal{H}|}]$ and $\mathbf{u}_h \in \mathbb{R}^{|\mathcal{R}|}$ is the uncertainty vector whose components u_{ih} correspond to the available generation of each renewable generator i in period h. Note that generation levels of different REG units may present spatial, temporal, and cross-lagged dependences.

In robust optimization, uncertainty is modeled through uncertainty sets. Before presenting the proposed DDUS, a general formulation for conventional budget-constrained uncertainty sets is provided based on the above matrix characterization.

A. Conventional Budget-Constrained Uncertainty Sets

In robust generation scheduling, uncertainty sets typically rely on fluctuation intervals representing the support of the uncertainties [2]–[4], [11], [12], and spatial and/or temporal budget constraints modeling the conservativeness or risk aversion of the decision maker [2]–[4], [8], [11], [12]. Thus, conventional budget-constrained uncertainty sets can be defined as the set of matrices $\mathbf{U}$ such that:

$$u_{ih} = \hat{u}_{ih} + \Delta_{ih}^+ z_{ih}^+ - \Delta_{ih}^- z_{ih}^- \quad \forall i \in \mathcal{R}, \ \forall h \in \mathcal{H} \quad (1)$$

$$0 \leq z_{ih}^+, z_{ih}^- \leq 1 \quad \forall i \in \mathcal{R}, \ \forall h \in \mathcal{H} \quad (2)$$

$$\sum_{i \in \mathcal{R}} z_{ih}^+ + z_{ih}^- \leq \Gamma|\mathcal{R}| \quad \forall h \in \mathcal{H} \quad (3)$$

$$\sum_{h \in \mathcal{H}} z_{ih}^+ + z_{ih}^- \leq \Lambda \quad \forall i \in \mathcal{R}. \quad (4)$$

As per (1) and (2), decision variables z_{ih}^+ and z_{ih}^- determine the available generation for renewable generator i in period h, represented by u_{ih}. The lower and upper box-like limits,

$\hat{u}_{ih} - \Delta_{ih}^-$ and $\hat{u}_{ih} + \Delta_{ih}^+$, are, respectively, the minimum and maximum possible values for u_{ih}. The first budget constraint (3) limits, for each period, the number of renewable generating units that may deviate from their expected value by parameter Γ, which can vary from 0 to 100%. As in [21], this constraint means that up to $\lfloor \Gamma|\mathcal{R}| \rfloor$ generators[1] can simultaneously change their production from their expected values $\hat{u}_{ih}$ to their upper or lower bounds in the same period. The second budget constraint (4) limits the number of periods in which each renewable generator $i \in \mathcal{R}$ can deviate from its expected production. For the sake of simplicity, parameter Λ represents an integer number of periods ranging between 0 and $|\mathcal{H}|$.

B. Scenario-Based Uncertainty Sets Driven by Data

In order to build a flexible, easy to specify, and practical uncertainty set that allows scalability for the available exact solution techniques for RUC, we propose a scenario-based polyhedral uncertainty set [26]–[28] to model REG uncertainty. To that end, the uncertainty set is defined as the convex hull of a set of multivariate points representing scenarios, i.e., REG profiles. Among the various exogenous approaches that could be used to generate the scenarios, we propose the use of a data-driven scheme. Historical daily profiles are therefore directly used as scenarios, thereby embedding relevant information about the true underlying uncertainty process in each vertex of the uncertainty set. Mathematically, the proposed DDUS, $\mathcal{U}$, is cast as follows:

$$\mathcal{U} = \left\{ \mathbf{U} \in \mathbb{R}^{|\mathcal{R}| \times |\mathcal{H}|} \middle| \mathbf{U} = \sum_{k=1}^{K} \alpha_k \mathbf{U}_k, \sum_{k=1}^{K} \alpha_k = 1, \alpha_k \geq 0 \right\} (5)$$

where $\{\mathbf{U}_k\}_{k=1,...,K}$ is the set of daily REG profiles corresponding to K previous days, which is hereinafter referred to as $\mathcal{S}$. Note that $\mathcal{U}$ is the convex hull of $\mathcal{S}$, thereby representing the smallest convex set that contains every scenario in $\mathcal{S}$. Moreover, by definition, the vertexes of the polyhedral uncertainty set $\mathcal{U}$ are in $\mathcal{S}$ [29].

The capability of DDUS to capture temporal and spatial dependences is illustrated in Figs. 1 and 2, where two-dimensional projections of the multidimensional DDUS are shown for the data of the second case study examined in Section V, for the period between 10/31/2012 and 12/15/2012. Fig. 1 shows a high-dependence pattern, exhibited in bus 5 for 12:00 p.m. and 1:00 p.m., and a low-dependence pattern, observed in the same bus for 12:00 p.m. and 6:00 p.m. Likewise, different spatial dependences can be visualized in Fig. 2 for 9:00 a.m.

An interesting interpretation for the use of observed generation profiles as vertexes of the uncertainty set is that we are implicitly performing an endogenous stress test for each feasible solution of the RUC problem whereby relevant multidimensional dependences found in the true hourly REG within a day are empirically accounted for. Furthermore, by using the proposed DDUS within an adaptive decision-making scheme, the evolution of the REG dependences across time can

---

[1]In cases where $\Gamma|\mathcal{R}|$ is not an integer, an additional renewable generator is allowed to deviate to make up for the residual fraction, $\Gamma|\mathcal{R}| - \lfloor \Gamma|\mathcal{R}| \rfloor$.

be dynamically updated. Hence, strong assumptions about the nature of the true model behind data are not required.
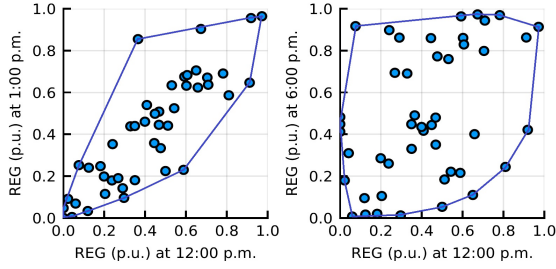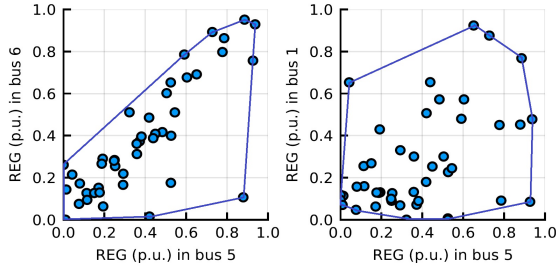

Fig. 1. Examples of temporal dependences.

Fig. 2. Examples of spatial dependences.

Moreover, note that the proposed framework can be adapted to consider sophisticated REG forecasts. As an example, any exogenous day-ahead forecasting procedure could be used to detrend/deseasonalize historical data and generate forecast errors. The errors computed for the K previous days could thus be added to an updated day-ahead forecast to generate data-driven adjusted scenarios to build $\mathcal{S}$. With this simple prepro-cessing scheme, the shaping capability of the proposed data-driven approach could be straightforwardly combined with state-of-the-art forecasting methods, which will be addressed in future works.

It is also worth mentioning that, if, instead of using real data, our data-driven model were built based on a Monte Carlo simulation of a time-series model, as done under a stochastic programming framework, our model can be viewed as a sampled version of a chance-constrained model. For instance, according to Calafiore [30], if the number of scenarios K (generated through Monte Carlo sampling methods) in DDUS is greater than $|\mathcal{R}||\mathcal{H}|/\varepsilon - 1$, where $\varepsilon \in (0,1)$ is a user-defined parameter that can be made arbitrarily small, the expected probability of violations (shedding load or curtailing REG) remains below ε. As per our numerical experiments, the CCGA, in practice, converges with a much smaller subset of the K sampled scenarios. Hence, in the case where our model is used to reproduce sample-based chance-constrained problems as done in [30], our robust approach would be a computationally efficient solution methodology.

III. TWO-STAGE ROBUST UNIT COMMITMENT MODEL

For expository purposes, we use a standard two-stage ro-bust optimization framework with a single uncertainty set [19], which is a particular instance of the general notion of stochastic robust optimization presented in [7] and [31]. This model is simpler to describe and analyze yet bringing out the main features of the proposed data-driven approach. Based on previous works on robust energy and reserve scheduling [6], [11], [12] and industry practice [14], [32], a base-case dispatch for energy is considered while ensuring ϵ-feasibility for all realizations within the uncertainty set. The mathematical formulation for the proposed DDRUC problem is as follows:

$$\min_{\mathcal{X}, [\boldsymbol{\theta}_h, \mathbf{f}_h]_{h \in \mathcal{H}}} \sum_{h \in \mathcal{H}} \left\{ \begin{matrix} \mathbf{e}'(\mathbf{c}_h^{sd} + \mathbf{c}_h^{su}) + (\mathbf{c}^g)'\mathbf{g}_h \\ + (\mathbf{c}^{up})'\mathbf{r}_h^{up} + (\mathbf{c}^{dn})'\mathbf{r}_h^{dn} \end{matrix} \right\} \quad (6)$$

subject to:

$$\mathbf{Af}_h + \mathbf{Bg}_h + \mathbf{P}\hat{\mathbf{u}}_h = \mathbf{d}_h \qquad\qquad \forall h \in \mathcal{H} \quad (7)$$

$$-\overline{\mathbf{f}} \leq \mathbf{f}_h \leq \overline{\mathbf{f}} \qquad\qquad\qquad \forall h \in \mathcal{H} \quad (8)$$

$$\mathbf{f}_h = \mathbf{S}\boldsymbol{\theta}_h \qquad\qquad\qquad \forall h \in \mathcal{H} \quad (9)$$

$$\underline{\mathbf{G}}\mathbf{v}_h + \mathbf{r}_h^{dn} \leq \mathbf{g}_h \leq \overline{\mathbf{G}}\mathbf{v}_h - \mathbf{r}_h^{up} \qquad \forall h \in \mathcal{H} \quad (10)$$

$$\mathbf{0} \leq \mathbf{r}_h^{up} \leq \mathbf{R}^{up}\,\mathbf{v}_h \qquad\qquad \forall h \in \mathcal{H} \quad (11)$$

$$\mathbf{0} \leq \mathbf{r}_h^{dn} \leq \mathbf{R}^{dn}\,\mathbf{v}_h \qquad\qquad \forall h \in \mathcal{H} \quad (12)$$

$$-\mathbf{b}(\mathbf{v}_h, \mathbf{v}_{h-1}) \leq \mathbf{g}_h - \mathbf{g}_{h-1} \leq \mathbf{a}(\mathbf{v}_h, \mathbf{v}_{h-1}) \;\forall h \in \mathcal{H} \quad (13)$$

$$\{\mathbf{c}_h^{sd}, \mathbf{c}_h^{su}, \mathbf{v}_h\}_{h \in \mathcal{H}} \in \mathcal{F} \qquad\qquad (14)$$

$$\Phi(\mathcal{X}) \leq \epsilon \qquad\qquad (15)$$

$$\Phi(\mathcal{X}) = \max_{\mathbf{U} \in \mathcal{U}} \left\{ \min_{\substack{[\boldsymbol{\theta}_h^{wc}, \mathbf{f}_h^{wc}, \\ \mathbf{g}_h^{wc}, \mathbf{s}_h^+, \\ \mathbf{s}_h^-]_{h \in \mathcal{H}}}} \sum_{h \in \mathcal{H}} \mathbf{e}'(\mathbf{s}_h^+ + \mathbf{s}_h^-) \right. \qquad (16)$$

subject to:

$$\mathbf{Af}_h^{wc} + \mathbf{Bg}_h^{wc} + \mathbf{Pu}_h =$$
$$\mathbf{d}_h + \mathbf{s}_h^+ - \mathbf{s}_h^- \;\; : (\boldsymbol{\beta}_h) \quad \forall h \in \mathcal{H} \; (17)$$

$$-\overline{\mathbf{f}} \leq \mathbf{f}_h^{wc} \leq \overline{\mathbf{f}} \qquad : (\boldsymbol{\sigma}_h, \boldsymbol{\pi}_h) \; \forall h \in \mathcal{H} \; (18)$$

$$\mathbf{f}_h^{wc} = \mathbf{S}\boldsymbol{\theta}_h^{wc} \qquad\quad : (\boldsymbol{\omega}_h) \qquad \forall h \in \mathcal{H} \; (19)$$

$$\mathbf{g}_h^{wc} - \mathbf{g}_{h-1}^{wc} \leq \mathbf{a}(\mathbf{v}_h, \mathbf{v}_{h-1}) \quad : (\boldsymbol{\zeta}_h) \qquad \forall h \in \mathcal{H} \; (20)$$

$$\mathbf{g}_{h-1}^{wc} - \mathbf{g}_h^{wc} \leq \mathbf{b}(\mathbf{v}_h, \mathbf{v}_{h-1}) \quad : (\boldsymbol{\kappa}_h) \qquad \forall h \in \mathcal{H} \; (21)$$

$$\mathbf{g}_h - \mathbf{r}_h^{dn} \leq \mathbf{g}_h^{wc} \leq \mathbf{g}_h + \mathbf{r}_h^{up} : (\boldsymbol{\gamma}_h, \boldsymbol{\tau}_h) \; \forall h \in \mathcal{H} \; (22)$$

$$\mathbf{0} \leq \mathbf{s}_h^- \leq \mathbf{d}_h \qquad\qquad : (\boldsymbol{\varsigma}_h) \qquad \forall h \in \mathcal{H} \; (23)$$

$$\left. \mathbf{0} \leq \mathbf{s}_h^+ \leq \mathbf{Pu}_h \qquad\qquad : (\boldsymbol{\xi}_h) \qquad \forall h \in \mathcal{H} \right\} (24)$$

where $\mathcal{X}$ is the set of first-level variables related to generation, i.e., $\mathcal{X} = [\mathbf{c}_h^{sd}, \mathbf{c}_h^{su}, \mathbf{g}_h, \mathbf{r}_h^{dn}, \mathbf{r}_h^{up}, \mathbf{v}_h]_{h \in \mathcal{H}}$, and third-level dual variables are shown in parentheses.

The two-stage DDRUC (6)–(24) is formulated as a (min-max-min) trilevel optimization problem. The first optimization level (6)–(15) determines the on/off statuses of generating units, as well as the energy and reserve scheduling. The ob-jective function (6) comprises shut-down costs, start-up costs, production costs, and up- and down-reserve costs. Constraint (7) represents nodal power balance under a dc power flow model. Expression (8) represents transmission line power flow limits, while Kirchhoff's second law is accounted for through (9). The limits for generation levels and up- and down-spinning reserves are imposed in expressions (10)–(12). Inter-period ramping limits are modeled by (13), where the components of the auxiliary vector functions $\mathbf{a}(\cdot)$ and $\mathbf{b}(\cdot)$ are [11]: $a_i(v_{ih}, v_{ih-1}) = RU_i v_{ih-1} + SU_i(v_{ih} - v_{ih-1}) + \overline{G}_i(1 - v_{ih})$ and $b_i(v_{ih}, v_{ih-1}) = RD_i v_{ih} + SD_i(v_{ih-1} - v_{ih}) + \overline{G}_i(1 - v_{ih-1})$. Following [33], equation (14) represents constraints

related to shut-down costs, start-up costs, and minimum up and down times. Equation (15) ensures redispatch capability within a feasibility tolerance, ϵ, under the set $\mathcal{X}$ for all plausible realizations within the uncertainty set.

The second-level problem comprises the outer maximization problem in (16). The goal of the second-level problem is to find the weights $\{\alpha_k\}_{k=1,\ldots,K}$ in (5) corresponding to the worst-case uncertainty realization $\mathbf{U} \in \mathcal{U}$ for a given value of $\mathcal{X}$. The measure of worst case is given by the minimum power imbalance function, which receives as inputs the values of the first- and second-level variables, $\mathcal{X}$ and $\mathbf{U}$, respectively. The third-level problem, i.e., the inner minimization problem (16)–(24), plays the role of the minimum power imbalance function. The objective of this problem is to find a redispatch solution that minimizes the total sum of the mismatch variables artificially introduced in the power balance constraint (17). The mismatch variables, $\mathbf{s}_h^-$ and $\mathbf{s}_h^+$, can be interpreted as load shedding and REG spillage, respectively. Constraints (17)–(21) are analogous to first-level expressions (7)–(9) and (13). The maximum operational deviation from nominal scheduled generation is controlled by constraint (22), whereby generation redispatch levels $\mathbf{g}_h^{wc}$ are limited by up- and down-reserves.

Within this approach, by ensuring ϵ-feasibility for a given set of observed data points, the proposed DDRUC provides solutions that are optimized and tight for a given stress test setup. Therefore, the resulting reserve procurement can be interpreted as the least-cost schedule that allows reserve deliverability for realistic system-stress conditions. The proposed approach differs from current industry practice, where stress tests are applied *ex post* as offline validation procedures.

IV. SOLUTION METHODOLOGY

This work leverages from the fact that the two lowermost levels (16)–(24) correspond to a maximization, within a polyhedral uncertainty set $\mathcal{U}$, of a convex function given by the output of the inner minimization in (16) as a function of $\mathbf{U}$. Therefore, from standard results of convex analysis, one of the vertexes belongs to the optimal solution set [29]. In other words, at the optimal solution, all α_k are equal to 0 except that corresponding to the worst-case vertex, which is equal to 1, thereby being binary valued. Hence, we can replace the continuous polyhedral uncertainty set $\mathcal{U}$ in the outer maximization of (16) with the discrete set of scenarios $\mathcal{S}$. By doing so, problem (6)–(24) can be cast as a single-level MILP-based equivalent wherein expressions (16)–(24) are replaced with one set of redispatch constraints (17)–(24) for each one of the K scenarios in $\mathcal{S}$. Unfortunately, addressing such a full-scenario-based equivalent by the branch-and-cut algorithm may lead to intractability.

Alternatively, problem (6)–(24), and its single-level equivalent, likewise, are suitable for the CCGA [23], which ensures finite convergence to global optimality by iteratively solving a master problem and an oracle subproblem. At each iteration, the master problem, which is a relaxed version of the full-scenario-dependent equivalent problem, finds a trial solution that is evaluated in terms of operational feasibility by the oracle procedure. The oracle returns to the master problem violated constraints that will change the master problem solution at the next iteration. The algorithm terminates when the

oracle does not find any violated constraint, thereby certifying the incumbent trial solution as globally optimal.

A. Master Problem

The master problem is a relaxed version of problem (6)–(24). Thus, at each iteration j of the CCGA, the master problem is an instance of MILP as follows:

$$\min_{\substack{\mathcal{X},[\boldsymbol{\theta}_h,\mathbf{f}_h]_{h\in\mathcal{H}} \\ [\boldsymbol{\theta}_h^{(k)},\mathbf{f}_h^{(k)},\mathbf{g}_h^{(k)},\mathbf{s}_h^{+(k)}, \\ \mathbf{s}_h^{-(k)}]_{h\in\mathcal{H},k=1,\ldots,|\mathcal{S}_j|}}} \sum_{h\in\mathcal{H}} \Big\{ \mathbf{e}'(\mathbf{c}_h^{sd} + \mathbf{c}_h^{su}) + (\mathbf{c}^g)'\mathbf{g}_h$$
$$+ (\mathbf{c}^{up})'\mathbf{r}_h^{up} + (\mathbf{c}^{dn})'\mathbf{r}_h^{dn} \Big\} \quad (25)$$

subject to:

Constraints (7)–(14) $\quad$ (26)

Redispatch constraints for scenario k, $k = 1,\ldots,|\mathcal{S}_j|$ $\quad$ (27)

$$\sum_{h\in\mathcal{H}} \mathbf{e}'(\mathbf{s}_h^{+(k)} + \mathbf{s}_h^{-(k)}) \leq \epsilon, \quad k = 1,\ldots,|\mathcal{S}_j| \quad (28)$$

where set $\mathcal{S}_j$ comprises the scenarios of $\mathcal{S}$ selected by the oracle subproblem until iteration j of the CCGA.

Expressions (25)–(26) are identical to (6)–(14). As per (27), a set of redispatch constraints is iteratively added. Such redispatch constraints correspond to (17)–(24) where variables $\boldsymbol{\theta}_h^{wc}$, $\mathbf{f}_h^{wc}$, $\mathbf{g}_h^{wc}$, $\mathbf{s}_h^+$, and $\mathbf{s}_h^-$ are respectively replaced with new variables $\boldsymbol{\theta}_h^{(k)}$, $\mathbf{f}_h^{(k)}$, $\mathbf{g}_h^{(k)}$, $\mathbf{s}_h^{+(k)}$, and $\mathbf{s}_h^{-(k)}$, whereas $\mathbf{u}_h$ is replaced with the corresponding scenario of $\mathcal{S}_j$ selected at previous iterations by the oracle subproblem. Finally, in (28), the system power imbalance under every scenario in $\mathcal{S}_j$ is bounded by the threshold ϵ.

B. Oracle Subproblem

The oracle subproblem corresponds to the two lowermost optimization levels (16)–(24) for given upper-level decisions provided by the preceding master problem. This bilevel program can be solved in two different ways:

1) By the enumeration of all feasible values for variables α_k. This solution approach, hereinafter referred to as the *inspection-based oracle*, consists in the serial or parallel evaluation of the imbalance through the solution of the third-level problem for all K scenarios in $\mathcal{S}$. The inspection-based oracle is thus equivalent to searching through all vectors $\{\alpha_k\}_{k=1,\ldots,K}$ that are candidates for optimality, i.e., those with one entry equal to 1 and all other entries equal to 0. Note that each optimization problem is a linear program, which runs in polynomial time with interior point methods. As the value of K is not related to the instance size, the inspection-based oracle can be easily implemented as a computational routine that receives a pair $(\mathcal{X}^*, \mathcal{S})$ and returns the worst-case scenario $\mathbf{U}^*$ in $\mathcal{S}$ and the associated power imbalance $\Phi(\mathcal{X}^*)$ in polynomial time [28].

2) By the application of branch-and-cut algorithms (readily available in off-the-shelf MILP solvers) to a single-level MILP-based equivalent subproblem. Based on a discrete representation of $\mathcal{S}$, this solution approach consists in casting the original bilevel subproblem as a single-level MILP equivalent, denoted by *MILP-based oracle* subproblem. Such an equivalent requires modeling variables

α_k as binary, using the dual of the lower level of the subproblem, and applying well-known integer algebra results [34] to recast the resulting bilinear terms as linear expressions. The MILP-based oracle subproblem is thus a linearized version (relying on standard disjunctive constraints) of this mixed-integer nonlinear problem:

$$\Phi(\mathcal{X}) = \max_{\substack{[\alpha_k]_{k\in\mathcal{K}}, \\ [\boldsymbol{\beta}_h, \boldsymbol{\gamma}_h, \boldsymbol{\zeta}_h, \\ \boldsymbol{\kappa}_h, \boldsymbol{\xi}_h, \\ \boldsymbol{\pi}_h, \boldsymbol{\sigma}_h, \\ \boldsymbol{\varsigma}_h, \boldsymbol{\tau}_h, \\ \boldsymbol{\omega}_h, \mathbf{u}_h]_{h\in\mathcal{H}}}} \sum_{h\in\mathcal{H}} \left\{ \boldsymbol{\beta}'_h (\mathbf{d}_h - \mathbf{P}\mathbf{u}_h) \right. $$
$$-\boldsymbol{\pi}'_h \overline{\mathbf{f}} - \boldsymbol{\sigma}'_h \overline{\mathbf{f}}$$
$$+\boldsymbol{\gamma}'_h (\mathbf{g}_h - \mathbf{r}_h^{dn})$$
$$-\boldsymbol{\tau}'_h (\mathbf{g}_h + \mathbf{r}_h^{up})$$
$$-\boldsymbol{\zeta}'_h \mathbf{a}(\mathbf{v}_h, \mathbf{v}_{h-1})$$
$$+ \boldsymbol{\kappa}'_h \mathbf{b}(\mathbf{v}_h, \mathbf{v}_{h-1})$$
$$\left. -\boldsymbol{\varsigma}'_h \mathbf{d}_h - \boldsymbol{\xi}'_h \mathbf{P}\mathbf{u}_h \right\} \quad (29)$$

subject to:

$$\mathbf{u}_h = \sum_{k\in\mathcal{S}} \alpha_k \mathbf{u}_{hk} \qquad \forall h \in \mathcal{H} \quad (30)$$

$$\sum_{k\in\mathcal{S}} \alpha_k = 1 \qquad\qquad (31)$$

$$\alpha_k \in \{0,1\} \qquad\qquad \forall k \in \mathcal{S} \quad (32)$$

$$\boldsymbol{\beta}_h - \boldsymbol{\varsigma}_h \le \mathbf{e} \qquad\qquad \forall h \in \mathcal{H} \quad (33)$$

$$-\boldsymbol{\beta}_h - \boldsymbol{\xi}_h \le \mathbf{e} \qquad\qquad \forall h \in \mathcal{H} \quad (34)$$

$$-\mathbf{A}'\boldsymbol{\beta}_h + \boldsymbol{\pi}_h - \boldsymbol{\sigma}_h - \boldsymbol{\omega}_h = \mathbf{0} \qquad \forall h \in \mathcal{H} \quad (35)$$

$$-\mathbf{S}'\boldsymbol{\omega}_h = \mathbf{0} \qquad\qquad \forall h \in \mathcal{H} \quad (36)$$

$$-\boldsymbol{\gamma}_h + \boldsymbol{\tau}_h - \mathbf{B}'\boldsymbol{\beta}_h + \boldsymbol{\zeta}_h$$
$$-\boldsymbol{\zeta}_{h+1} - \boldsymbol{\kappa}_h + \boldsymbol{\kappa}_{h+1} \ge \mathbf{0} \qquad \forall h \in \mathcal{H} \quad (37)$$

$$\boldsymbol{\pi}_h, \boldsymbol{\sigma}_h, \boldsymbol{\tau}_h, \boldsymbol{\gamma}_h, \boldsymbol{\zeta}_h, \boldsymbol{\kappa}_h, \boldsymbol{\varsigma}_h, \boldsymbol{\xi}_h \ge \mathbf{0} \quad \forall h \in \mathcal{H}. \quad (38)$$

In (29)–(38), the objective function (29) and constraints (33)–(38) represent, respectively, the objective function and constraints of the dual formulation for the third-level problem presented in (16)–(24). The proposed DDUS is equivalently modeled by constraints (30)–(32), where $\mathbf{u}_{hk}$ stands for the h-th column of the k-th data point, $\mathbf{U}_k$. According to (5) and the aforementioned results of convex analysis, $\mathcal{U}$ is replaced with $\mathcal{S}$ in (30) and (31), whereas variables α_k are characterized as binary in (32). Note that problem (29)–(38) is a mixed-integer nonlinear optimization problem with bilinear terms in the objective function (29). Such nonlinearities involve products of $\mathbf{u}_h$ and third-level dual variables. As is customary in the literature [6], [11], [12], an equivalent MILP formulation for (29)–(38), i.e., the MILP-based oracle, is achieved by applying standard integer algebra results [34] that allow recasting the nonlinear products in (29) as linear expressions.

C. Algorithm

For given $\mathcal{S}$ and ϵ, the proposed CCGA works as follows:

Algorithm CCGA($\mathcal{S}$, ϵ)

1: Initialization: $j \leftarrow 0$ and $\mathcal{S}_j \leftarrow \emptyset$.
2: Solve the master problem (25)–(28) to obtain $\mathcal{X}^*$.
3: Solve the subproblem for $(\mathcal{X}^*, \mathcal{S})$ to obtain $\Phi(\mathcal{X}^*)$ and the worst-case scenario $\mathbf{U}^*$.
4: **if** $\Phi(\mathcal{X}^*) \le \epsilon$ **then:** STOP
5: **else:** $j \leftarrow j + 1$, $\mathcal{S}_j \leftarrow \mathcal{S}_{j-1} \cup \{\mathbf{U}^*\}$, and go to step 2
6: **end if**.

The algorithm iteratively adds violated constraints into the master problem and terminates when infeasibility is within ϵ.

In order to reduce the search space, vertex-identification algorithms [35] could be implemented *ex ante* to determine the subset of $\mathcal{S}$ comprising the vertexes of $\mathcal{U}$. However, in a high-dimensional case, it is very unlikely to find a point in $\mathcal{S}$ that is not a vertex of $\mathcal{U}$. Hence, the use of vertex-identification procedures, in practice, may deteriorate the performance of the CCGA as only an insignificant number of points of $\mathcal{S}$, most likely none, would be excluded for being in the interior of the DDUS.

V. NUMERICAL RESULTS

This section reports results from an illustrative 4-bus system and the IEEE 118-bus test system over a 24-hour time span. For the sake of reproducibility, data for both test systems can be downloaded from [36]. For expository purposes, wind-related uncertainty is considered. The source of data for wind power generation is the Global Energy Competition (GEFCom) [37], [38].

As explained in the next section, the proposed DDRUC has been assessed with two benchmark models based on two-stage robust optimization and two-stage stochastic programming. Such an assessment has been conducted through backtesting on historical data observed over a given set of days.

The benchmark RUC model, hereinafter referred to as BRUC, relies on the budget-constrained uncertainty set formulated in Section II-A. DDRUC has been implemented for different time windows, i.e., for different values of K, whereas different combinations of budgets Γ and Λ were considered for BRUC. For quick reference, the instances of both models are denoted by DDRUC(K) and BRUC(Γ,Λ), respectively. For a given day d, the uncertainty set for DDRUC(K) is directly defined using the REG profiles observed in the K previous days, i.e., considering $\mathcal{S}$ as $\{\mathbf{U}_k\}_{k=d-K,\dots,d-1}$.

For the sake of a fair comparison, BRUC is also adaptively adjusted across the rolling-horizon study, similarly to DDRUC, in order to prevent over-conservative solutions. This is done through a moving window of previous observed days within which the lower and upper box-like limits used in the budget-constrained uncertainty sets are defined according to the corresponding hourly minimum and maximum production limits for each renewable unit. Our tests indicated that the length of such a moving window is not as relevant for BRUC as the selection of K is for DDRUC. Hence, based on trial and error, the budget-constrained uncertainty sets for BRUC(Γ,Λ) were built using the previous 35 days.

The instances of DDRUC were addressed by the CCGA with both the inspection- and the MILP-based oracle sub-problems. In contrast, BRUC was solved by the CCGA with a modified version of the MILP-based oracle subproblem, as done in [11]. To that end, expressions (30) and (31) in problem (29)–(38), representing the uncertainty set constraints, were replaced with (1)–(4), whereas the linearization scheme mentioned in Section IV-B was applied to the resulting bilinear terms.

Additionally, in order to provide a comparison between the proposed data-driven robust approach and stochastic programming, a benchmark stochastic unit commitment model,

hereinafter referred to as BSUC, was analyzed. BSUC is a mixed-integer linear program structurally similar to the DDRUC master problem (25)–(28). Three differences characterize BSUC, namely 1) an expected imbalance cost term is added to the objective function (25), 2) the set of scenarios in (27) comprises uncertainty realizations obtained from a scenario-generation procedure using probabilistic information, and 3) imbalance requirements are eliminated by dropping constraint (28). The value of the imbalance cost was set to $500/MWh, which corresponds approximately to 25 times the average value (or 13 times the highest value) of the fuel costs, $\mathbf{c}^g$, for the thermal generators.

The scenario-generation procedure for BSUC was based on a multivariate lognormal distribution inferred from the previous 100 days. First, the historical data were normalized and a logarithmic transformation was applied. Then, a multivariate normal distribution was estimated and 500 independent and equally distributed scenarios were generated based on the expected value vector and the covariance matrix. Finally, the inverse transformations were applied to rescale the scenarios. As for the solution methodology, BSUC was tackled by the branch-and-cut algorithm, which was initialized with the previous day's solution for binary scheduling variables and reserve contributions.

Simulations were run using Gurobi 7.0.2 under JuMP (Julia 0.5) on a Xeon E5-2680 processor at 2.5 GHz and 128 GB of RAM.

A. Evaluation Methodology

The performances of DDRUC(K), BRUC(Γ,Λ), and BSUC have been compared in terms of their tradeoff between cost and robustness. The evaluation methodology consists in conducting a rolling-horizon out-of-sample backtest over a set of days for which realistic REG data are available.

For each day d within the backtest horizon, the instances under assessment are solved using available historical REG data for past days, i.e., excluding day d. The solution of such instances of DDRUC, BRUC, and BSUC yields the corresponding values of the total cost for that day. Subsequently, solution robustness is quantified by computing the infeasibility of the generation dispatch associated with the resulting generation schedules and the actual REG scenario observed for day d. To that end, based on industry practice, a single-period version of the inner minimization problem in (16)–(24) is solved for each hour of this day. Hence, based on the resulting hourly levels of load shedding and REG spillage, out-of-sample statistics for robustness are devised.

B. Illustrative 4-Bus System

First, we consider an illustrative test system consisting of 4 buses, 4 transmission lines, 14 thermal generators, and 2 wind farms [36]. Wind generation data were rescaled from buses 2 and 5 from [37]. The optimality gap of Gurobi was set at 0% and the CCGA was implemented until no imbalance was found for the instances of DDRUC and BRUC. Hence, such simulations were run to optimality for $\epsilon = 0$. As for BSUC, a time limit of 1,500 seconds was considered. The evaluation backtest was conducted for 336 days of realistic

TABLE I
4-BUS SYSTEM –RESULTS FROM DDRUC, BRUC, AND BSUC

Instance	Avg. time (s)	Avg. iter. no.	Avg. cost ($)	LOLP (%)	PWS (%)
DDRUC(42)	20.0 (34.7)	13.4	62122	1.56	2.67
DDRUC(35)	17.7 (28.8)	12.8	61894	1.85	3.53
DDRUC(28)	15.2 (23.2)	12.0	61592	2.33	4.15
DDRUC(20)	11.9 (17.0)	10.9	61121	3.91	5.43
BRUC(90, 24)	157.7	6.2	66256	4.13	1.00
BRUC(90, 1)	688.7	71.5	63401	5.36	1.09
BRUC(70, 24)	115.9	6.5	62349	12.44	2.86
BRUC(50, 24)	13.0	6.2	60571	21.60	5.25
BSUC	1494.0	–	61967	9.78	2.96

wind power generation data. Table I and Fig. 3 summarize the results from the backtest for several instances of DDRUC, BRUC, and BSUC.

Columns 2–4 of Table I respectively provide the average computing times, iteration numbers, and costs over the 336 runs of the models being examined. For the instances of DDRUC, the first figure in column 2 corresponds to the method using the inspection-based oracle subproblem, whereas the second figure in parentheses is associated with the approach relying on the solution of the MILP-based equivalent for the oracle subproblem. Column 5 presents a reliability index referred to as loss of load probability (LOLP), which is defined as the fraction of hours with load shedding exceeding 0.1% of the system load. Analogously, column 6 reports the probability of wind spillage (PWS), defined as the fraction of hours in which wind spillage exceeds 0.1% of the available wind power generation. Thus, the values of LOLP and PWS listed in Table I represent valuable statistical information on the robustness of the solutions provided by each model.

Table I shows the impact of K on the computational effort required by DDRUC. It should be emphasized that the number of iterations of the CCGA grows at a less than linear rate with K whereas computing times behave similarly. As a relevant result, the impact of K on computing times is lower when the subproblem is solved by the vertex inspection, which outperforms the MILP solver for all instances.

As compared with BRUC, DDRUC is in general more effective from a computational perspective. For the instances with no temporal budget constraints ($\Lambda = 24$), the average numbers of iterations for BRUC are approximately half of those required to solve the instances of DDRUC. However, such reduced numbers of iterations do not imply shorter computing times because the MILP-based oracle for BRUC is computationally expensive. Moreover, for $\Lambda = 1$, which is far less conservative, BRUC requires significantly more iterations to converge on average, thereby increasing the computational burden. An exception is the instance BRUC(50, 24), which is a particular case with no temporal budget constraints and a spatial budget corresponding to precisely a single renewable generator deviating from the expected production. Interestingly, this was the case for which BRUC achieved the worst solution robustness in terms of LOLP, which is 5.5 times higher than the worst level attained by DDRUC. Regarding BSUC, its required computational effort
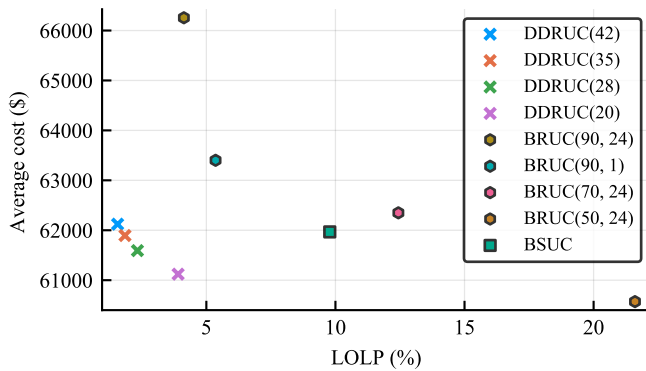
Fig. 3.   Average cost vs. LOLP for the 4-bus system.

TABLE II
4-BUS SYSTEM – AVERAGE COST BREAKDOWN (10^3 \$)

Instance	Total	Production	Reserve	Start-up and shut-down
DDRUC(35)	61.9	55.7	6.1	0.1
DDRUC(20)	61.1	55.5	5.5	0.1
BRUC(90, 24)	66.3	57.2	9.0	0.1
BRUC(90, 1)	63.4	56.2	7.1	0.1
BSUC	62.0	55.7	6.2	0.1

is considerably greater than those of DDRUC and BRUC. The 1,500-s time limit was reached almost every day of the rolling-window study.

As can also be observed in Table I, as K grows from 20 to 42, both LOLP and PWS decrease. Note also that, in comparison with BRUC and BSUC, all solutions provided by DDRUC gave rise to acceptably small LOLP values ranging between 1.56% and 3.91%, while keeping the average cost at reasonable levels within a narrow 1.6% band. The superiority of DDRUC over BRUC and BSUC is thus evidenced in terms of the tradeoff between cost and robustness. Such a tradeoff is further illustrated in Fig. 3, where robustness is expressed in terms of the out-of-sample reliability index LOLP. Note that the solutions to DDRUC dominate in Pareto sense (lower cost and lower LOLP) almost all solutions provided by BRUC. The only non-dominated BRUC instance is the aforementioned BRUC(50, 24), which, albeit incurring the lowest average cost, leads to an unreasonably high LOLP. Analogously, despite featuring a reasonable average cost, the solution provided by BSUC gave rise to a very high value of LOLP, namely 9.78%, which was substantially outperformed by all instances of DDRUC. Hence, the BSUC solution was dominated by the instances DDRUC(20), DDRUC(28), and DDRUC(35).

A breakdown of the average costs provided in column 4 of Table I is presented in Table II for representative instances of DDRUC, BRUC, and BSUC models. Thus, average production costs, average reserve costs, and average start-up and shut-down costs are reported. It can be observed that the largest cost component is the average production cost, representing between 86% and 91% of the average total cost for all instances. However, the average reserve cost, which amounts to between 9% and 14% of the average total cost, is the most relevant in terms of cost difference for any pair of instances. Overall, reserve costs for DDRUC and BSUC are
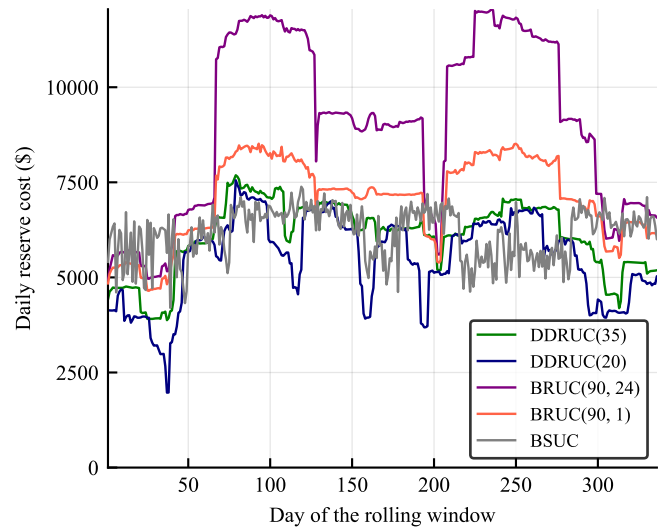

Fig. 4.   Daily reserve costs for the 4-bus system.

considerably lower than those for BRUC. For further details, daily reserve costs are shown in Fig. 4. As compared to BRUC, DDRUC and BSUC were able to address the next-day REG profile with lower levels of reserves. Moreover, daily reserve costs for DDRUC are less variable than those for BSUC due to the inherent stochastic nature of the scenario-generation procedure.

For illustration purposes, Fig. 5 depicts two two-dimensional projections of the scenarios in $\mathcal{S}$ for the 43rd day of the rolling window. The 8 scenarios colored in red correspond to the scenarios selected by the CCGA for DDRUC(35). Not surprisingly, some of the selected scenarios present very high or very low REG profiles. Such scenarios are more likely to yield higher imbalance costs (from either wind spillage or load shedding) at intermediate iterations of the CCGA. Salient features of the selected scenarios are also characterized by the scatter plot in Fig. 6, which shows the total 24-hour REG and the largest absolute variation of renewable generation between two consecutive hours of each scenario. Note that the two scenarios with the largest absolute REG variation between two consecutive periods were also selected. These scenarios also represent a challenge to the system as they stress its ramping capability and require higher reserve levels.

C. 118-Bus System

The second case study is a modified version of the IEEE 118-bus test system with 10 wind farms. The optimality gap of Gurobi was set at 0.05% for the master problem, the subproblem was solved to optimality through the vertex inspection, and ϵ was equal to 0.1% of the total system demand.

For this case study, we first assessed the performance of DDRUC(35) with a second instance, denoted by DDRUC(35+10), whereby intra-day dynamics of wind power generation for the same period in the previous year were considered. Thus, the scenarios for DDRUC(35+10) were associated with the previous 35 days and a ten-day period from the year before, beginning five days prior to the date corresponding to the day under analysis in the backtest scheme. For
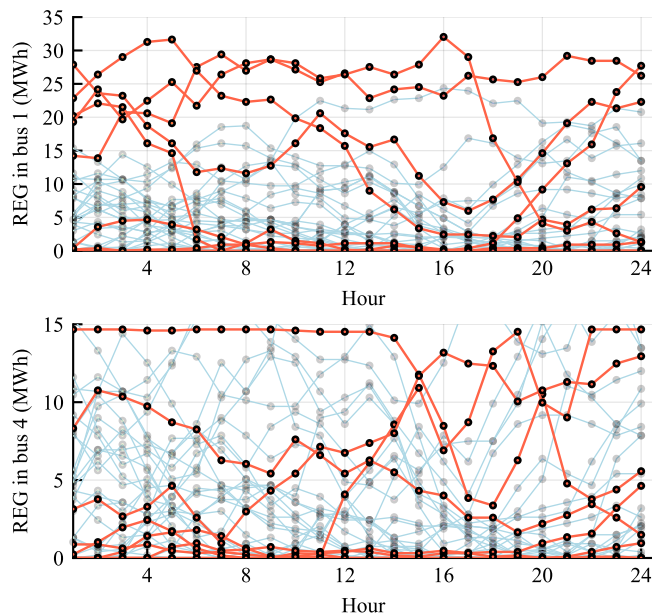
Fig. 5. Scenarios in $\mathcal{S}$ for the 43rd day of the rolling window for the 4-bus system.
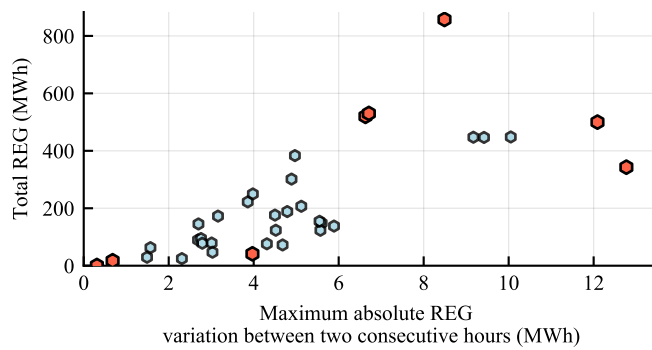

Fig. 6. Scatter plot for the scenarios selected by DDRUC(35) in the 43rd day of the rolling window for the 4-bus system.

both instances, the backtest was performed for 214 days under realistic wind power generation data [38]. The results listed in Table III indicate that the proposed DDRUC can provide a robust and cost-effective generation schedule. Moreover, it is interesting to note that the cost increase due to the consideration of 10 additional vertexes from the previous year is negligible, while the values of LOLP and PWS respectively drop from 3.2% and 5.8% to 2.7% and 4.6%.

It is also relevant to highlight the low computational burden exhibited by both DDRUC instances. The significance of such a result was validated by performing the 214-day backtest with BRUC, which was the only benchmark model used for this case study due to the poor computational behavior featured by BSUC for the illustrative example. Unfortunately, for BRUC, the CCGA failed to converge in practical computing times.

DDRUC has been further assessed with BRUC through a reduced backtest relying on a randomly selected sequence of 10 days. Simulations were run for a feasibility tolerance ϵ of 0.1% of the system demand and a practical 1-hour time limit. For the instances of BRUC examined, the time limit was reached during the solution of the master problem. For such instances, the CCGA was allowed to run until the

TABLE III
118-BUS SYSTEM – RESULTS FROM THE 214-DAY BACKTEST

	DDRUC(35)	DDRUC(35+10)
Average time (s)	774	792
Maximum time (s)	1752	2220
Average no. of iter.	8.5	8.4
Average cost (10^4 \$)	110.4	110.5
LOLP (%)	3.2	2.7
PWS (%)	5.8	4.6

TABLE IV
118-BUS SYSTEM – RESULTS FROM THE 10-DAY BACKTEST

Instance	Avg. time (s)	Avg. iter. no.	Avg. cost (10^4 \$)	LOLP (%)	PWS (%)
DDRUC(42)	494	5.0	110.2	1.67	0.83
DDRUC(35+10)	696	5.7	110.2	1.67	0.83
DDRUC(35)	378	5.6	110.1	1.67	1.25
DDRUC(28)	804	6.6	109.6	1.25	5.00
BRUC(70, 24)	3847	3.0	109.6	8.33	1.67
BRUC(50, 24)	3741	3.0	108.3	19.17	5.83

corresponding iteration was completed. Table IV summarizes the results from the 10-day backtest. The reported results suggest that the proposed DDRUC is a scalable alternative for RUC.

VI. CONCLUSION

This paper has described a new, comprehensive, and parsimonious method to characterize REG uncertainty in day-ahead RUC. The proposed method relies on an alternative scenario-based polyhedral uncertainty set that is built through a novel data-driven approach. Unlike conventional budget-constrained uncertainty sets, the proposed polyhedral uncertainty set is characterized directly from data through the convex hull of a set of previously observed REG profiles. Thus, relevant empirical information regarding the existing complex and time-varying dynamics of REG sources is embedded in the vertexes of the polyhedron. In addition, a salient feature of the proposed uncertainty set is its capability to consider any exogenous model (or scheme) for adjusting, including or excluding (filtering or clustering) scenarios based on existing methods used in industry. Moreover, the resulting robust counterpart for generation scheduling can be efficiently solved by the CCGA until ϵ-global optimality. This relevant practical aspect stems from the reduced complexity of the oracle subproblem, which is solvable in polynomial time. For the case studies analyzed in the paper, a relatively small number of past renewable generation profiles were required by the proposed data-driven approach to outperform two benchmarks, namely an RUC model based on a conventional budget-constrained uncertainty set and a stochastic unit commitment model. Numerical results indicate that the proposed method might be a practical, scalable, easy to specify, and cost-efficient alternative tool for managing wind power variability in the unit commitment problem.

Future research will consider 1) exogenous day-ahead forecasting methods for nominal scenarios, and 2) data-processing

schemes based on forecasting errors to generate the vertexes for the uncertainty set. In addition, alternative budget-constrained uncertainty sets will be analyzed. Another relevant avenue of research is the extension of the proposed data-driven approach within a stochastic robust optimization framework.

REFERENCES

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

[2] 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.

[3] D. Bertsimas, E. Litvinov, X. A. 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.

[4] R. Jiang, M. Zhang, G. Li, and Y. Guan, "Two-stage network constrained robust unit commitment problem," *Eur. J. Oper. Res.*, vol. 234, no. 3, pp. 751–762, May 2014.

[5] Á. 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.

[6] A. Moreira, A. Street, and J. M. Arroyo, "Energy and reserve scheduling under correlated nodal demand uncertainty: An adjustable robust optimization approach," *Int. J. Electr. Power Energy Syst.*, vol. 72, pp. 91–98, Nov. 2015.

[7] Y. An and B. Zeng, "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.

[8] C. Dai, L. Wu, and H. Wu, "A multi-band uncertainty set based robust SCUC with spatial and temporal budget constraints," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4988–5000, Nov. 2016.

[9] C. Wang, F. Liu, J. Wang, F. Qiu, W. Wei, S. Mei, and S. Lei, "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.

[10] C. Shao, X. Wang, M. Shahidehpour, X. Wang, and B. Wang, "Security-constrained unit commitment with flexible uncertainty set for variable wind power," *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 1237–1246, Jul. 2017.

[11] N. G. Cobos, J. M. Arroyo, and A. Street, "Least-cost reserve offer deliverability in day-ahead generation scheduling under wind uncertainty and generation and network outages," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3430–3442, Jul. 2018.

[12] N. G. Cobos, J. M. Arroyo, N. Alguacil, and A. Street, "Network-constrained unit commitment under significant wind penetration: A multistage robust approach with non-fixed recourse," *Appl. Energy*, vol. 232, pp. 489–503, Dec. 2018.

[13] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, *The State-of-the-Art in Short-Term Prediction of Wind Power: A Literature Overview, 2nd ed.* DTU, Denmark, 2011.

[14] J. M. Arroyo and F. D. Galiana, "Energy and reserve pricing in security and network-constrained electricity markets," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 634–643, May 2005.

[15] J. Wang, M. Shahidehpour, and Z. Li, "Contingency-constrained reserve requirements in joint energy and ancillary services auction," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1457–1468, Aug. 2009.

[16] Y. Chen, P. Gribik, and J. Gardner, "Incorporating post zonal reserve deployment transmission constraints into energy and ancillary service co-optimization," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 537–549, Mar. 2014.

[17] Federal Energy Regulatory Commission. (2014, Dec.). Operator-initiated commitments in RTO and ISO markets. Price formation in organized wholesale electricity markets. Docket no. AD14-14-000. Washington, DC, USA.

[18] Regulatory Authority for Energy, "The Greek grid and exchange code," 2019. [Online]. Available: http://www.rae.gr/old/en/codes/main.htm

[19] X. A. Sun, "Robust optimization in electric power systems," *in Advances and Trends in Optimization with Engineering Applications*, T. Terlaky, M. F. Anjos, and S. Ahmed, Eds. Philadelphia, PA, USA: SIAM, 2017, pp. 357–365.

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

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

[22] C. Bandi and D. Bertsimas, "Tractable stochastic analysis in high dimensions via robust optimization," *Math. Program.*, vol. 134, no. 1, pp. 23–70, Aug. 2012.

[23] 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.

[24] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms, 3rd ed.* Hoboken, NJ, USA: John Wiley & Sons, 2006.

[25] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM Rev.*, vol. 53, no. 3, pp. 464–501, 2011.

[26] D. Bertsimas and D. B. Brown, "Constructing uncertainty sets for robust linear optimization," *Oper. Res.*, vol. 57, no. 6, pp. 1483–1495, Nov.–Dec. 2009.

[27] B. Fernandes, A. Street, D. Valladão, and C. Fernandes, "An adaptive robust portfolio optimization model with loss constraints based on data-driven polyhedral uncertainty sets," *Eur. J. Oper. Res.*, vol. 255, no. 3, pp. 961–970, Dec. 2016.

[28] D. Bertsimas, V. Gupta, and N. Kallus, "Data-driven robust optimization," *Math. Program.*, vol. 167, no. 2, pp. 235–292, Feb. 2018.

[29] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, vol. 6. Belmont, MA, USA: Athena Scientific, 1997.

[30] G. C. Calafiore, "On the expected probability of constraint violation in sampled convex programs," *J. Optim. Theory Appl.*, vol. 143, no. 2, pp. 405–412, Nov. 2009.

[31] C. Liu, C. Lee, H. Chen, and S. Mehrotra, "Stochastic robust mathematical programming model for power system optimization," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 821–822, Jan. 2016.

[32] F. D. Galiana, F. Bouffard, J. M. Arroyo, and J. F. Restrepo, "Scheduling and pricing of coupled energy and primary, secondary, and tertiary reserves," *Proc. IEEE*, vol. 93, no. 11, pp. 1970–1983, Nov. 2005.

[33] M. Carrión and J. M. Arroyo, "A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1371–1378, Aug. 2006.

[34] C. A. Floudas, *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications.* New York, NY, USA: Oxford Univ. Press, 1995.

[35] J. Riko, "Dynamic planar convex hull," Ph.D. dissertation, *University of Aarhus*, Aarhus, Denmark, 2002.

[36] A. Velloso, A. Street, D. Pozo, J. M. Arroyo, and N. G. Cobos, "Two-stage robust unit commitment for co-optimized electricity markets: An adaptive data-driven approach for scenario-based uncertainty sets – Case studies data," 2019. [Online]. Available: https://drive.google.com/open?id=1ngPlDyAyr-BXY9WvPsI1av1Q-fg4APGR

[37] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *Int. J. Forecast.*, vol. 30, no. 2, pp. 357–363, Apr.–Jun. 2014.

[38] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *Int. J. Forecast.*, vol. 32, no. 3, pp. 896–913, Jul.–Sep. 2016.