A Distributionally Robust AC Network-Constrained Unit Commitment

Shahab Dehghan, *Senior Member, IEEE*, Petros Aristidou, *Senior Member, IEEE*,
Nima Amjady, *Senior Member, IEEE*, and Antonio J. Conejo, *Fellow, IEEE*

***Abstract*—This paper presents a distributionally robust network-constrained unit commitment (DR-NCUC) model considering AC network modeling and uncertainties of demands and renewable productions. The proposed model characterizes uncertain parameters using a data-driven ambiguity set constructed by training samples. The non-convex AC power flow equations are approximated by convex quadratic and McCormick relaxations. Since the proposed min-max-min DR-NCUC problem cannot be solved directly by available solvers, a new decomposition algorithm with proof of convergence is reported in this paper. The master problem of this algorithm is solved using both primal and dual cuts, while the max-min sub-problem is solved using the primal-dual hybrid gradient method, obviating the need for using duality theory. Also, an active set strategy is proposed to enhance the tractability of the decomposition algorithm by ignoring the subset of inactive constraints. The proposed model is applied to a 6-bus test system and the IEEE 118-bus test system under different conditions. These case studies illustrate the performance of the proposed DR-NCUC model to characterize uncertainties and the superiority of the proposed decomposition algorithm over other decomposition approaches using either primal or dual cuts.**

***Index Terms*—Convexification, Decomposition, Distributionally Robust Optimization, Unit Commitment, Uncertainty.**

NOMENCLATURE

A. Functions

$\mathbb{E}(\bullet)$	Expectation operator over •.
OC^{UC}	Commitment costs [$].
OC^{ED}	Dispatch costs, being OC^{ED^c}/OC^{ED^e} the optimal dispatch costs with convex/exact power flow equations [$].
$\boldsymbol{Q}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{\Lambda})$	Dispatch costs in compact form [$].
$\mathcal{s}(\widetilde{\mathbb{P}},\mathbb{P}_{\mathrm{e}})$	Distance between $\widetilde{\mathbb{P}}$ and $\mathbb{P}_{\mathrm{e}}$.
$\|\bullet\|_1/\|\bullet\|_2$	L_1/L_2 norm of vector •.

B. Indices

d	Index of demands.
h/h'	Index of hours.
i	Index of iterations of the outer loop of the decomposition algorithm.
j	Index of iterations of the inner loop (i.e., the PDHG method) of the decomposition algorithm.
n/m	Index of nodes.
(n,m)	Sending and receiving nodes of a transmission line.
s	Index of training samples, being s^* the index of the *worst-case* one.
u	Index of units.

C. Matrices and Vectors

$\boldsymbol{A}$	Vector of commitment costs.
$\widetilde{\mathbb{P}}$	Vector of unknown probability distributions, being $\tilde{p}_s$ its s^{th} entry.
$\mathbb{P}_{\mathrm{e}}$	Vector of empirical probability distributions, being p_{e_s} its s^{th} entry.
$\boldsymbol{X}$	Vector of binary commitment variables.
$\boldsymbol{X^i}$	Vector of binary commitment variables fixed at iteration i.
$\boldsymbol{Y}$	Vector of continuous dispatch variables.
$\boldsymbol{Y}_s$	Vector of continuous dispatch variables for training sample s.
$\boldsymbol{Y}_{si}$	Vector of continuous dispatch variables for training sample s at iteration i.
$\boldsymbol{\Upsilon}$	Vector of dual variables corresponding to the constraint $\boldsymbol{X}=\boldsymbol{X^i}$ in the SP at iteration i.
$\boldsymbol{\Upsilon^i}$	Vector of dual variables fixed at iteration i.
$\boldsymbol{\Lambda}$	Vector of the uncertain parameters.
$\widehat{\boldsymbol{\Lambda}}$	Vector of forecast values of the uncertain parameters.
$\boldsymbol{\Lambda}^s$	Training sample s of the uncertain parameters.
$\Psi_\bullet$	Vectors of decision variables in • (i.e., either MP1, MP2, MP3, or SP), being Ψ_{SP_s} the vector of decision variables in the SP for training sample s.

D. Parameters

B_{nn}	Shunt susceptance at node n [mho].
B_{nm}	Susceptance of the transmission line connecting nodes n and m [mho].
C_u^0, C_u^1, C_u^2	Coefficients of the cost function of thermal/RES unit u.
C^{UD}	Penalty cost of unsupplied active load [$/MWh].
C_u^{SD}	Shutdown cost of thermal unit u [$].
C_u^{SU}	Start-up cost of thermal unit u [$].
DT_u/UT_u	Minimum down-time/up-time limits of thermal unit u [h].
G_{nm}	Conductance of the transmission line connecting nodes n and m [mho].
$\hat{l}_{nm}$	Forecasted current rating of the transmission line connecting nodes n and m [A].
N^{HO}	Number of historical observations of the uncertain parameters, being N^{HO_s} the number of historical observations inside subinterval s of the sample space.
N^s	Number of training samples.
$\tilde{p}^{si}$	Worst-case realization for the probability of training sample s fixed at iteration i.
q^{si}	Dispatch-related cost of training sample s at iteration i [$].
$\hat{P}_{dh}$	Forecasted load for demand d at hour h [MW].
$\hat{P}_{uh}$	Forecasted (available) power for RES (solar or wind) unit u at hour t [MW].
PF_{dh}	Power factor of the load of demand d at hour h.
R_{nm}/X_{nm}	Resistance/reactance of the transmission line connecting nodes n and m [ohm].
RD_u/RU_u	Maximum ramp-down/ramp-up limits of thermal unit u [MW/h].
α/β	Scaling parameters in the maximization/minimization problem of the PDHG method.

This work was supported in part by Engineering and Physical Sciences Research Council (EPSRC) in the U.K. under Grant EP/R030243/1.

S. Dehghan is with the School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK (s.dehghan@leeds.ac.uk).
P. Aristidou is with the Department of Electrical Engineering, Computer Engineering, and Informatics, Cyprus University of Technology, Cyprus (petros.aristidou@cut.ac.cy).
N. Amjady is with the Department of Electrical Engineering, Semnan University, Semnan, Iran (amjady@semnan.ac.ir).
A. J. Conejo is with the Department of Integrated Systems Engineering and the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, USA (conejo.1@osu.edu).

$\epsilon_{in}/\epsilon_{out}$ Convergence tolerance of the inner/outer loop.
ρ Confidence level.

E. Sets

Ω^{D} Set of all demands, being Ω^{D_n} the set of demands connected to node n.
Ω^{DC} Set of dual cuts.
Ω^{H} Set of all scheduling hours.
Ω^{L} Set of all transmission lines, being Ω^{L_n} the set of lines connected to node n.
$\Omega^{\mathcal{M}_+}$ Set of all probability measures.
Ω^{MPASS} Set of reduced constraints pertaining to transmission lines in the master problem.
Ω^{MPVASS} Set of violated constraints pertaining to transmission lines in the master problem.
Ω^{N} Set of all nodes.
Ω^{PC} Set of primal cuts.
Ω^{R} Set of all RES units, being Ω^{R_n} the set of RES units connected to node n.
$\Omega^{\boldsymbol{\mathcal{S}}}$ Data-driven ambiguity set.
$\Omega^{\boldsymbol{\mathcal{S}}_1}$ Data-driven ambiguity set based on the L_1 norm.
Ω^{SPASS} Set of reduced constraints pertaining to transmission lines in the sub-problem.
Ω^{SPVASS} Set of violated constraints pertaining to transmission lines in the sub-problem.
Ω^{T} Set of all thermal units, being Ω^{T_n} the set of thermal units connected to node n.
$\Omega^{\boldsymbol{\mathcal{X}}}$ Feasible space of all commitment variables.
$\Omega^{\boldsymbol{\mathcal{Y}}(\boldsymbol{X},\boldsymbol{\Lambda})}$ Feasible space of all dispatch variables.
$\Omega^{\boldsymbol{\mathcal{Y}}_{s^*}(\boldsymbol{X},\boldsymbol{\Lambda}^{s^*})}$ Feasible space of the dispatch variables of the *worst-case* training sample s^*.
$\Omega^{\boldsymbol{\Lambda}}$ Sample space for the uncertain parameters.

F. Variables

p_{dh}^{UD} Unsupplied active power of demand d at hour h [MW].
p_{nmh} $/q_{nmh}$ Active/reactive power flow through the transmission line connecting nodes n and m at hour h [MW].
$\tilde{p}_s$ Unknown probability of training sample s.
p_{uh}/q_{uh} Active/reactive power production of unit u at hour h [MW].
q_s Dispatch-related cost of training sample s [$].
v_{nh} Voltage magnitude of node n at hour h [V].
x_{uh} Binary variable indicating the commitment status of thermal unit u at hour h (i.e., $x_{uh} = 0/1$: on/off).
x_{uh}^{SD}/x_{uh}^{SU} Binary variable indicating whether or not thermal unit u starts up or shuts down at the beginning of hour h (i.e., $x_{uh}^{SD} = 1/0$ shuts down/otherwise and $x_{uh}^{SU} = 1/0$ starts up/otherwise).
θ_{nh} Voltage angle of node n at hour h [rad].
θ_{nmh} Voltage angle difference between nodes n and m [rad].
$\gamma_{\bullet}$ Auxiliary variables in $\bullet$ (i.e., either MP1, MP2, or MP3), being χ, cc_{nmh}^{c}, cs_{nmh}^{c}, l_{nmh}^{c}, v_{nh}^{c}, vv_{nmh}^{c}, vc_{nmh}^{c}, vs_{nmh}^{c} other auxiliary variables.

Hereafter, $\overline{\bullet}/\underline{\bullet}$ are the maximum/minimum limit of quantity $\bullet$. Other parameters/variables are conveniently defined within the text.

I. Introduction

A. Motivation and Background

In recent years, the increasing integration of intermittent renewable energy resources (RES) into power systems has imposed additional uncertainties to the network-constrained unit commitment (NCUC) problem. Previously, the NCUC problem has been solved by stochastic optimization (SO) [1-3] and robust optimization (RO) [4, 5]. In general, SO-based NCUC models (S-NCUC) obtain a commitment schedule that is optimal on average for all scenarios characterizing uncertainty sources, while RO-based NCUC models (R-NCUC) obtain a commitment schedule that is optimal for the worst-case realization of the uncertain parameters. Although it is vital to ensure that AC power flow equations are satisfied in RES-integrated power systems, most of the previous SO and RO-based NCUC models have utilized DC power flow simplifications, ignoring voltage magnitudes and reactive power. To enhance accuracy, a two-stage SO-based AC NCUC model has been presented in [6]. This model being a mixed-integer nonlinear and non-convex stochastic problem is solved using a two-level decomposition approach with dual cuts. Furthermore, a two-stage RO-based AC NCUC model has been presented in [7]. This model being as well a mixed-integer nonlinear and non-convex robust problem is solved using a three-level decomposition approach with both primal and dual cuts as well as convex relaxations of non-convex terms in the AC power flow equations.

Generally, the conservatism of the solution in SO-based models is lower than that in RO-based ones, while the required historical data in RO-based models is typically less demanding than that in SO-based ones [6-8]. To combine the benefits of both SO-based and RO-based NCUC models and to resolve their limitations, a number of distributionally robust NCUC (DR-NCUC) models [9, 10] have been introduced in the literature to obtain a commitment schedule which is optimal for the worst-case realization of probability distributions of uncertain parameters. However, to the best of the authors' knowledge, these models use DC power flow equations, and thus there is no DR-NCUC model with AC power flow equations reported in the literature. Therefore, this paper presents a DR-NCUC model with AC power flow equations rather than DC ones aiming at enhancing the accuracy of the optimal solution under different uncertainties.

B. Contributions

The main characteristics of this research work can be summarized as follows:

- Introducing a tri-level min-max-min DR-NCUC model with AC power flow equations modeling demands and RES power productions as uncertain data. Although distributionally robust optimization (DRO) has been presented in the literature to solve the unit commitment problem in recent years, previous works have either excluded power flow equations or only included DC power flow equations rather than AC ones. For instance, [11] excludes network constraints, and [9] only includes DC power flow equations rather than AC ones. Therefore, a DR-NCUC model with AC power flow equations has been introduced in this paper to enhance the accuracy and practicability of previous DR-NCUC models. Quadratic and McCormick relaxations are utilized to recast the non-convex terms in the AC power flow equations as convex ones. To the best of the authors' knowledge, there is no distributionally robust unit commitment model with convex AC power flow equations reported in the literature.

• Representing the uncertainties of demands and RES power productions using a training-based data-driven ambiguity set. The accuracy of the proposed DR-NCUC model can be controlled by adjusting the number of training samples while constructing the ambiguity set. In this paper, the term "training samples" refers to historical observations of uncertain parameters [12]. There are different ambiguity sets reported in the literature, including discrepancy-based ambiguity sets, moment-based ambiguity sets, shape-preserving ambiguity sets, and Kernel-based ambiguity sets [13]. However, to the best of the authors' knowledge, there is no training-based data-driven ambiguity set reported for the AC NCUC problem.

• Proposing a decomposition approach to solve the tri-level min-max-min DR-NCUC model using primal/dual cuts and a primal-dual hybrid gradient (PDHG) method. It is worth noting that the proposed decomposition approach with the PDHG method efficiently solves the proposed DR-NCUC model without resorting to duality theory, Karush-Kuhn-Tucker (KKT) conditions, and Big-M transformation. Also, the convergence of the proposed decomposition approach is proved.

• Reducing the computational burden of the proposed DR-NCUC model by (i) incorporating the worst-case training sample (instead of all training samples) as primal cuts into the master problem, and (ii) proposing an effective active set strategy (ASS) to ignore inactive constraints.

C. Paper Organization

The rest of this paper is organized as follows. In Section II, a deterministic NCUC (D-NCUC) model with convex approximations of the AC power flow equations is presented. In Section III, the DR-NCUC model is formulated as a tri-level optimization problem, and then, an effective decomposition algorithm is presented to solve it. In Section IV, the proposed model and decomposition algorithm are illustrated on a 6-bus test system and the IEEE 118-bus test system. Finally, Section V concludes the paper.

II. Deterministic NCUC Model

In this section, a D-NCUC model is presented by extending the NCUC formulation in [14]:

$$\min OC^{UC} + OC^{ED} \tag{1a}$$

s.t.

$$OC^{UC} = \sum_{\forall h\in\Omega^H}\sum_{\forall u\in\Omega^T}(C_u^{SU}\cdot x_{uh}^{SU} + C_u^{SD}\cdot x_{uh}^{SD} + C_u^0\cdot x_{uh}) \tag{1b}$$

$$OC^{ED} = \sum_{\forall h\in\Omega^H}\sum_{\forall u\in\{\Omega^R,\Omega^T\}}(C_u^1\cdot p_{uh} + C_u^2\cdot p_{uh}^2) + \sum_{\forall h\in\Omega^H}\sum_{\forall d\in\Omega^D}(C^{UD}\cdot p_{dh}^{UD}) \tag{1c}$$

We note that for notational brevity the term "unit" and the symbol "u" refer to all RES and thermal units, while a different set is defined for each of them (i.e., Ω^R and Ω^T). The objective function (1a) is the total operation cost including OC^{UC} and OC^{ED}. Constraint (1b) represents the commitment-related costs, including start-up, shutdown, and fixed costs of thermal units; while constraint (1c) represents the dispatch-related costs, including variable costs of RES and thermal units and penalty costs of unsupplied loads. In our study, only variable costs are considered for RES units in (1c) [7].

Constraints on commitment variables:

$$x_{u(h-1)} - x_{uh} + x_{uh}^{SU} \geq 0 \quad \forall u\in\Omega^T, \forall h\in\Omega^H \tag{1d}$$

$$x_{uh} - x_{u(h-1)} + x_{uh}^{SD} \geq 0 \quad \forall u\in\Omega^T, \forall h\in\Omega^H \tag{1e}$$

$$x_{uh} - x_{u(h-1)} \leq x_{uh'} \quad \forall u\in\Omega^T, \forall h'\in[h+1, \min\{UT_u - 1{,}24\}], \forall h\in\Omega^H \tag{1f}$$

$$x_{u(h-1)} - x_{uh} \leq 1 - x_{uh'} \quad \forall u\in\Omega^T, \forall h'\in[h+1, \min\{DT_u - 1{,}24\}], \forall h\in\Omega^H \tag{1g}$$

Constraints (1d) and (1e) are logical relations of commitment-related variables, indicating start-up, shutdown, and on/off status of thermal units in all hours of the scheduling horizon. Constraints (1f) and (1g) enforce minimum up-time and down-time requirements, respectively, for all thermal units.

Constraints on thermal units:

$$\underline{P_u}\cdot x_{uh} \leq p_{uh} \leq \overline{P}_u\cdot x_{uh} \quad \forall u\in\Omega^T, \forall h\in\Omega^H \tag{1h}$$

$$\underline{Q_u}\cdot x_{uh} \leq q_{uh} \leq \overline{Q}_u\cdot x_{uh} \quad \forall u\in\Omega^T, \forall h\in\Omega^H \tag{1i}$$

$$-RD_u \leq p_{uh} - p_{u(h-1)} \leq RU_u \quad \forall u\in\Omega^T, \forall h\in\Omega^H \tag{1j}$$

Constraints (1h) and (1i) limit the active and reactive power productions of each thermal unit, respectively. Also, constraint (1j) enforces ramp-up and ramp-down limits for all thermal units.

Constraints on RES units:

$$0 \leq p_{uh} \leq \hat{P}_{uh} \quad \forall u\in\Omega^R, \forall h\in\Omega^H \tag{1k}$$

Constraint (1k) limits the active power production of renewable units considering potential spillage.

Nodal balance constraints:

$$\sum_{\forall u\in\{\Omega^{R_n},\Omega^{T_n}\}} p_{uh} = \sum_{\forall d\in\Omega^{D_n}}\left(\hat{P}_{dh} - p_{dh}^{UD}\right) + \sum_{\forall(n,m)\in\Omega^{L_n}} p_{nmh} + \sum_{\forall(m,n)\in\Omega^{L_n}} p_{mnh} \quad \forall n\in\Omega^N, \forall h\in\Omega^H \tag{1l}$$

$$\sum_{\forall u\in\Omega^{T_n}} q_{uh} = \sum_{\forall d\in\Omega^{D_n}}\left(\tan(\cos^{-1} PF_{dh})\cdot\left(\hat{P}_{dh} - p_{dh}^{UD}\right)\right) + \sum_{\forall(n,m)\in\Omega^{L_n}} q_{nmh} + \sum_{\forall(m,n)\in\Omega^{L_n}} q_{mnh} \quad \forall n\in\Omega^N, \forall h\in\Omega^H \tag{1m}$$

Constraints (1l) and (1m) represent the nodal active and reactive power balances, respectively.

Constraints on transmission lines:

$$p_{nmh} = G_{nm}\cdot v_{nh}^c - G_{nm}\cdot vc_{nmh}^c - B_{nm}\cdot vs_{nmh}^c \quad \forall(n,m)\in\Omega^L, \forall h\in\Omega^H \tag{1n}$$

$$q_{nmh} = -\left(\tfrac{B_{nn}}{2} + B_{nm}\right)\cdot v_{nh}^c + B_{nm}\cdot vc_{nmh}^c - G_{nm}\cdot vs_{nmh}^c \quad \forall(n,m)\in\Omega^L, \forall h\in\Omega^H \tag{1o}$$

$$R_{nm}\cdot l_{nmh}^c = p_{nmh} + p_{mnh} \quad \forall(n,m)\in\Omega^L, \forall h\in\Omega^H \tag{1p}$$

$$X_{nm}\cdot l_{nmh}^c = q_{nmh} + q_{mnh} \quad \forall(n,m)\in\Omega^L, \forall h\in\Omega^H \tag{1q}$$

$$l_{nmh}^c = (G_{nm}^2 + B_{nm}^2)\cdot(v_{nh}^c + v_{mh}^c - 2\cdot vc_{nmh}^c) \quad \forall(n,m)\in\Omega^L, \forall h\in\Omega^H \tag{1r}$$

$$l_{nmh}^c \leq \hat{l}_{nm}^2 \quad \forall(n,m)\in\Omega^L, \forall h\in\Omega^H \tag{1s}$$

Constraints (1n) and (1o) evaluate the active and reactive line flows, respectively. In this paper, a quadratic relaxation method based on McCormick envelopes is used to obtain a convex approximation for the exact nonlinear and non-convex AC line flow equations [15]. Accordingly, auxiliary variables v_{nh}^c, vc_{nmh}^c, and vs_{nmh}^c in (1n) and (1o) represent convex transformations of nonlinear/non-convex terms v_{nh}^2, $v_{nh}\cdot v_{mh}\cdot\cos(\theta_{nh}-\theta_{mh})$, and $v_{nh}\cdot v_{mh}\cdot\sin(\theta_{nh}-\theta_{mh})$ in the exact AC line flow equations, respectively. Constraints (1p) and (1q) represent the active and reactive power loss, respectively, of each transmission line. Additionally, constraint (1r) calcu-

lates the squared current of each transmission line, while constraint (1s) enforces the current rating of each transmission line.

Constraints on voltage magnitudes and angles:

$$\underline{V}_n \le v_{nh} \le \overline{V}_n \quad \forall n \in \Omega^N, \forall h \in \Omega^H \tag{1t}$$

$$\underline{V}_n^2 \le v_{nh}^c \le \overline{V}_n^2 \quad \forall n \in \Omega^N, \forall h \in \Omega^H \tag{1u}$$

$$\underline{\theta}_{nm} \le \overbrace{\theta_{nh} - \theta_{mh}}^{\theta_{nmh}} \le \overline{\theta}_{nm} \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1v}$$

Constraint (1t) limits nodal voltage magnitudes while constraint (1u) limits their squared values. Constraint (1v) bounds the difference of voltage angles between the nodes of all lines.

Constraints on auxiliary variables for the convex relaxations:

$$v_{nh}^c \ge v_{nh}^2 \quad \forall n \in \Omega^N, \forall h \in \Omega^H \tag{1w}$$

$$v_{nh}^c \le \left(\underline{V}_n + \overline{V}_n\right) \cdot v_{nh} - \underline{V}_n \cdot \overline{V}_n \quad \forall n \in \Omega^N, \forall h \in \Omega^H \tag{1x}$$

$$cc_{nmh}^c \le 1 - \left(\frac{1-\cos(\overline{\theta}_{nm})}{\overline{\theta}_{nm}^2}\right) \cdot \theta_{nmh}^2 \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1y}$$

$$cc_{nmh}^c \ge \cos\left(\overline{\theta}_{nm}\right) \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1z}$$

$$cs_{nmh}^c \le \cos\left(\frac{\overline{\theta}_{nm}}{2}\right) \cdot \left(\theta_{nmh} - \frac{\overline{\theta}_{nm}}{2}\right) + \sin\left(\frac{\overline{\theta}_{nm}}{2}\right) \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1za}$$

$$cs_{nmh}^c \ge \cos\left(\frac{\overline{\theta}_{nm}}{2}\right) \cdot \left(\theta_{nmh} + \frac{\overline{\theta}_{nm}}{2}\right) - \sin\left(\frac{\overline{\theta}_{nm}}{2}\right) \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zb}$$

$$vv_{nmh}^c \ge \underline{V}_n \cdot v_{mh} + \underline{V}_m \cdot v_{nh} - \underline{V}_n \cdot \underline{V}_m \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zc}$$

$$vv_{nmh}^c \ge \overline{V}_n \cdot v_{mh} + \overline{V}_m \cdot v_{nh} - \overline{V}_n \cdot \overline{V}_m \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zd}$$

$$vv_{nmh}^c \le \underline{V}_n \cdot v_{mh} + \overline{V}_m \cdot v_{nh} - \underline{V}_n \cdot \overline{V}_m \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1ze}$$

$$vv_{nmh}^c \le \overline{V}_n \cdot v_{mh} + \underline{V}_m \cdot v_{nh} - \overline{V}_n \cdot \underline{V}_m \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zf}$$

$$vc_{nmh}^c \ge \underline{VV}_{nmh}^c \cdot cc_{nmh}^c + \underline{CC}_{nmh}^c \cdot vv_{nmh}^c - \underline{VV}_{nmh}^c \cdot \underline{CC}_{nmh}^c \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zg}$$

$$vc_{nmh}^c \ge \overline{VV}_{nmh}^c \cdot cc_{nmh}^c + \overline{CC}_{nmh}^c \cdot vv_{nmh}^c - \overline{VV}_{nmh}^c \cdot \overline{CC}_{nmh}^c \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zh}$$

$$vc_{nmh}^c \le \underline{VV}_{nmh}^c \cdot cc_{nmh}^c + \overline{CC}_{nmh}^c \cdot vv_{nmh}^c - \underline{VV}_{nmh}^c \cdot \overline{CC}_{nmh}^c \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zi}$$

$$vc_{nmh}^c \le \overline{VV}_{nmh}^c \cdot cc_{nmh}^c + \underline{CC}_{nmh}^c \cdot vv_{nmh}^c - \overline{VV}_{nmh}^c \cdot \underline{CC}_{nmh}^c \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zj}$$

$$vs_{nmh}^c \ge \underline{VV}_{nmh}^c \cdot cs_{nmh}^c + \underline{CS}_{nmh}^c \cdot vv_{nmh}^c - \underline{VV}_{nmh}^c \cdot \underline{CS}_{nmh}^c \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zk}$$

$$vs_{nmh}^c \ge \overline{VV}_{nmh}^c \cdot cs_{nmh}^c + \overline{CS}_{nmh}^c \cdot vv_{nmh}^c - \overline{VV}_{nmh}^c \cdot \overline{CS}_{nmh}^c \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zl}$$

$$vs_{nmh}^c \le \underline{VV}_{nmh}^c \cdot cs_{nmh}^c + \overline{CS}_{nmh}^c \cdot vv_{nmh}^c - \underline{VV}_{nmh}^c \cdot \overline{CS}_{nmh}^c \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zm}$$

$$vs_{nmh}^c \le \overline{VV}_{nmh}^c \cdot cs_{nmh}^c + \underline{CS}_{nmh}^c \cdot vv_{nmh}^c - \overline{VV}_{nmh}^c \cdot \underline{CS}_{nmh}^c \quad \forall (n,m) \in \Omega^L, \forall h \in \Omega^H \tag{1zn}$$

Auxiliary variables v_{nh}^c, cc_{nmh}^c, and cs_{nmh}^c as well as convex quadratic constraints (1w)-(1x), (1y)-(1z), and (1za)-(1zb) are used to approximate nonlinear/non-convex terms v_{nh}^2, $\cos(\theta_{nh} - \theta_{mh})$, and $\sin(\theta_{nh} - \theta_{mh})$, respectively. Moreover, auxiliary variables vv_{nmh}^c, vc_{nmh}^c, and vs_{nmh}^c as well as convex linear constraints (1zc)-(1zf), (1zg)-(1zj), and (1zk)-(1zn) represent McCormick envelopes for product terms $v_{nh} \cdot v_{mh}$, $v_{nh} \cdot v_{mh} \cdot \cos(\theta_n - \theta_m)$, and $v_{nh} \cdot v_{mh} \cdot \sin(\theta_n - \theta_m)$, respectively, where nonlinear and non-convex terms $\cos(\theta_n - \theta_m)$ and $\sin(\theta_n - \theta_m)$ are replaced by their auxiliary variables cc_{nmh}^c and cs_{nmh}^c as defined in (1y)-(1z) and (1za)-(1zb), respectively. That is, $v_{nh} \cdot v_{mh} \cdot \cos(\theta_{nh} - \theta_{mh}) \cong v_{nh} \cdot v_{mh} \cdot cc_{nmh}^c \cong vv_{nmh}^c \cdot cc_{nmh}^c \cong vc_{nmh}^c$ and $v_{nh} \cdot v_{mh} \cdot \sin(\theta_{nh} - \theta_{mh}) \cong v_{nh} \cdot v_{mh} \cdot cs_{nmh}^c \cong vv_{nmh}^c \cdot cs_{nmh}^c \cong vs_{nmh}^c$.

The D-NCUC model (1a)-(1zn) is a convex mixed-integer quadratically-constrained quadratic-programming (MIQCQP) problem, which can be solved by available commercial optimization tools. For notational brevity, the proposed D-NCUC model is rewritten in a compact matrix form as given below:

$$\min_{\boldsymbol{X} \in \Omega^{\boldsymbol{X}}, \boldsymbol{Y} \in \Omega^{\boldsymbol{Y}(\boldsymbol{X},\boldsymbol{\Lambda})}} \boldsymbol{A}^\top \cdot \boldsymbol{X} + \boldsymbol{Q}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\Lambda}) \tag{2}$$

where $\boldsymbol{\Lambda}$ is fixed on $\widehat{\boldsymbol{\Lambda}}$ (i.e., $\boldsymbol{\Lambda} = \widehat{\boldsymbol{\Lambda}}$) in the D-NCUC model.

III. The Proposed Model and Solution Approach

In this section, the proposed distributionally robust model and decomposition approaches are presented.

A. Distributionally Robust NCUC Model

The proposed min-max-min DR-NCUC can be compactly written as:

$$\min_{\boldsymbol{X} \in \Omega^{\boldsymbol{X}}} \left(\boldsymbol{A}^\top \cdot \boldsymbol{X} + \max_{\widetilde{\mathbb{P}} \in \Omega^{\boldsymbol{S}}} \left(\min_{\boldsymbol{Y} \in \Omega^{\boldsymbol{Y}(\boldsymbol{X},\boldsymbol{\Lambda})}} \mathbb{E}_{\widetilde{\mathbb{P}}} \{ \boldsymbol{Q}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\Lambda}) \} \right) \right) \tag{3}$$

Model (3) is solved to identify the optimal solution with minimum expected costs under the worst-case realization of the unknown probability distributions $\widetilde{\mathbb{P}}$ belonging to the data-driven ambiguity set $\Omega^{\boldsymbol{S}}$ as given below:

$$\Omega^{\boldsymbol{S}} = \left\{ \widetilde{\mathbb{P}} \in \Omega^{\mathcal{M}_+} : \mathcal{s}\left(\widetilde{\mathbb{P}}, \mathbb{P}_{\mathrm{e}}\right) \le \rho \right\} \tag{4}$$

where $\Omega^{\mathcal{M}_+}$ represents the set of all probability measures on the probability space $(\Omega^{\boldsymbol{\Lambda}}, \mathcal{F}^{\Omega})$ being $\Omega^{\boldsymbol{\Lambda}}$ the sample space of $\boldsymbol{\Lambda}$ and $\mathcal{F}^{\Omega}$ the σ-algebra of $\Omega^{\boldsymbol{\Lambda}}$ [16]. Also, $\mathcal{s}\left(\widetilde{\mathbb{P}}, \mathbb{P}_{\mathrm{e}}\right)$ represents the distance between the unknown probability distribution $\widetilde{\mathbb{P}}$ and the empirical probability distribution $\mathbb{P}_{\mathrm{e}}$ (i.e., the training probability distribution constructed using historical data). Additionally, ρ represents the confidence level of the data-driven ambiguity set as a function of the sample size of historical data [16, 17]. In this paper, without loss of generality, the L_1 norm is used to calculate the distance between $\widetilde{\mathbb{P}}$ and $\mathbb{P}_{\mathrm{e}}$. Consequently, a data-driven ambiguity set based on the L_1 norm is given by:

$$\Omega^{\boldsymbol{S}_1} = \left\{ \widetilde{\mathbb{P}} \in \mathbb{R}_+^{N^S} : \mathcal{s}\left(\widetilde{\mathbb{P}}, \mathbb{P}_{\mathrm{e}}\right) = \left\| \widetilde{\mathbb{P}} - \mathbb{P}_{\mathrm{e}} \right\|_1 \triangleq \sum_{s=1}^{N^S} \left| \tilde{p}_s - p_{\mathrm{e}_s} \right| \le \rho \ , \ \sum_{s=1}^{N^S} \tilde{p}_s = 1 \right\} \tag{5}$$

where the variable $\tilde{p}_s$ and the parameter p_{e_s} represents the s^{th} entries of $\widetilde{\mathbb{P}}$ and $\mathbb{P}_{\mathrm{e}}$, respectively, and N^S represents the number of training samples to construct the data-driven ambiguity set in (5). To construct such a set, initially, a bounded interval is considered for each uncertain parameter based on its historical data. Then, a set of training samples (i.e., realizations of uncertain parameters with their corresponding probabilities) are generated for each uncertain parameter within its bounded interval. In summary, given N^{HO} historical observations of the uncertain parameters, the sample space can be constructed by finding variation intervals of the uncertain parameters (i.e.,

lower and upper bounds) by means of their historical observations. Then, the sample space (i.e., the variation intervals of the uncertain parameters) can be divided into N^s equidistant subintervals. For $s = 1, \dots, N^s$, the center of subinterval s represents the training sample s (i.e., $\boldsymbol{\Lambda}^s$) while its corresponding training probability (i.e., p_{e_s}) is equal to the number of historical observations laying within subinterval s (i.e., N^{HO_s}) divided by the total number of historical observations (i.e., N^{HO}). In other words, $p_{\mathrm{e}_s} = \frac{N^{HO_s}}{N^{HO}}$ where $\sum_{s=1}^{N^s} N^{HO_s} = N^{HO}$, and consequently, $\sum_{s=1}^{N^s} p_{\mathrm{e}_s} = 1$.

B. Decomposition Approaches

Problem (3) cannot be directly solved by commercial solvers. Hence, three decomposition algorithms are proposed:

- **Algorithm 1 (A1):** The first decomposition algorithm incorporates primal cutting planes [18] in the master problem (MP1) and uses the PDHG method [19] in the sub-problem (SP).
- **Algorithm 2 (A2):** The second decomposition algorithm incorporates dual cutting planes in the master problem (MP2) and uses the PDHG method in the sub-problem.
- **Algorithm 3 (A3):** The third decomposition algorithm is similar to the second decomposition algorithm except for its master problem (MP3). The master problem in the third decomposition algorithm incorporates dual cutting planes, as in the second decomposition algorithm, and all dispatch-related constraints for the *worst-case* training sample out of all training samples used for constructing the data-driven ambiguity set.

It is worthwhile to note that the SP, as a max-min optimization problem, is identical in all A1-A3. MP1 includes all commitment-related variables and incorporates primal cutting planes of all previous iterations. This consists of all dispatch-related variables and constraints for all training samples with their corresponding worst-case probabilities obtained from solving the SP at each iteration. MP2 only includes commitment-related variables and incorporates dual cutting planes of all previous iterations. This relies on the dual variables of the complicating variables in the SP (i.e., fixed commitment-related variables) obtained from solving the SP at each iteration. MP3 includes commitment-related variables and incorporates dual cutting planes of all previous iterations, similarly to A2, as well as the dispatch-related variables and constraints only for the worst-case training sample obtained from solving the SP at each iteration. In summary, the number of both variables and constraints in MP1 increases with iterations, while the number of variables in MP2 does not change with iterations, and only one constraint is added to MP2 at each iteration. The number of variables and the number of constraints in MP3 only increase in the first iteration to incorporate the dispatch-related variables and constraints of the worst-case training sample. However, in the next iterations, the number of variables of MP3 does not change with iterations, and only one constraint is added to MP3 at each iteration similar to MP2. Accordingly, after a specific number of iterations, the computational complexity of MP2/MP3 is significantly less than that of MP1 as M2/MP3 includes a lower number of variables and constraints. Additionally, and more importantly, MP3 provides a more accurate lower bound at each iteration in comparison with MP2 as MP3 includes the dispatch-related variables and constraints of the worst-case training sample.

In the following, the SP (i.e., the inner loop) as well as MP1, MP2, and MP3 (i.e., the outer loop) pertaining to algorithms A1, A2, and A3, respectively, are presented in detail. Hereafter, superscript/subscript "s" is used for parameters/variables.

1) Commitment-Related Master Problem (Outer Loop):

Master Problem in Algorithm 1: MP1 in A1 at iteration i finds the optimal values of the commitment variables subject to the commitment-related constraints in (6b) and *the primal cuts* in (6c) as given below:

$$\min_{\Psi_{\mathrm{MP1}}} \overbrace{\boldsymbol{A}^{\intercal} \cdot \boldsymbol{X}}^{\Xi_{\mathrm{MP}}} + \gamma_{\mathrm{MP1}} \tag{6a}$$

s.t.

$$\boldsymbol{X} \in \Omega^{\boldsymbol{X}} \tag{6b}$$

$$\gamma_{\mathrm{MP1}} \geq \sum_{s=1}^{N^s} \tilde{p}^{si} \cdot \boldsymbol{Q}(\boldsymbol{X}, \boldsymbol{Y}_{si}, \boldsymbol{\Lambda}^s) \quad s = 1, \dots, N^s, \forall i \in \Omega^{PC} \;\; ; \;\; \boldsymbol{Y}_{si} \in \Omega^{\boldsymbol{y}(\boldsymbol{X}, \boldsymbol{\Lambda}^s)} \quad s = 1, \dots, N^s, \forall i \in \Omega^{PC} \tag{6c}$$

where $\Psi^{\mathrm{MP1}} = \{\boldsymbol{X} \;\; , \;\; \boldsymbol{Y}_{si} \;\; s = 1, \dots, N^s, \forall i \in \Omega^{PC} \;\; , \;\; \gamma_{\mathrm{MP1}}\}$.

Master Problem in Algorithm 2: MP2 in A2 at iteration i finds the optimal values of the commitment variables subject to the commitment-related constraints in (7b) and *the dual cuts* in (7c) as given below:

$$\min_{\Psi_{\mathrm{MP2}}} \overbrace{\boldsymbol{A}^{\intercal} \cdot \boldsymbol{X}}^{\Xi_{\mathrm{MP}}} + \gamma_{\mathrm{MP2}} \tag{7a}$$

s.t.

$$\boldsymbol{X} \in \Omega^{\boldsymbol{X}} \tag{7b}$$

$$\gamma_{\mathrm{MP2}} \geq \sum_{s=1}^{N^s} \tilde{p}^{si} \cdot q^{si} + (\boldsymbol{X} - \boldsymbol{X}^i)^{\intercal} \cdot \boldsymbol{\Upsilon}^{\boldsymbol{i}} \quad \forall i \in \Omega^{DC} \tag{7c}$$

where $\Psi^{\mathrm{MP2}} = \{\boldsymbol{X} \;\; , \;\; \gamma_{\mathrm{MP2}}\}$.

Master Problem in Algorithm 3: MP3 in A3 at iteration i finds the optimal values of the commitment variables subject to the commitment-related constraints in (8b), dispatch-related constraints for *the worst-case training sample* in (8c), and *the dual cuts* in (8d), as given below:

$$\min_{\Psi_{\mathrm{MP3}}} \overbrace{\boldsymbol{A}^{\intercal} \cdot \boldsymbol{X}}^{\Xi_{\mathrm{MP}}} + \gamma_{\mathrm{MP3}} \tag{8a}$$

s.t.

$$\boldsymbol{X} \in \Omega^{\boldsymbol{X}} \tag{8b}$$

$$\gamma_{\mathrm{MP3}} \leq \boldsymbol{Q}\left(\boldsymbol{X}, \boldsymbol{Y}_{s^*}, \boldsymbol{\Lambda}^{s^*}\right) \;\; ; \;\; \boldsymbol{Y}_{s^*} \in \Omega^{\boldsymbol{y}_{s^*}\left(\boldsymbol{X}, \boldsymbol{\Lambda}^{s^*}\right)} \tag{8c}$$

$$\gamma_{\mathrm{MP3}} \geq \sum_{s=1}^{N^s} \tilde{p}^{si} \cdot q^{si} + (\boldsymbol{X} - \boldsymbol{X}^i)^{\intercal} \cdot \boldsymbol{\Upsilon}^{\boldsymbol{i}} \quad \forall i \in \Omega^{DC} \tag{8d}$$

where $\Psi^{\mathrm{MP3}} = \{\boldsymbol{X} \;\; , \;\; \boldsymbol{Y}_{s^*} \;\; , \;\; \gamma_{\mathrm{MP3}}\}$.

It is worthwhile to note that the dispatch-related cost of the worst-case training sample is higher than the worst-case expected cost of all training samples. Hence, $\gamma \leq \boldsymbol{Q}\left(\boldsymbol{X}, \boldsymbol{Y}_{s^*}, \boldsymbol{\Lambda}^{s^*}\right)$ is considered in (8c) with the purpose of connecting dispatch-related constraints and their corresponding cost to the objective function of the master problem and thus enhancing the computational performance of this master problem.

After solving the master problem at iteration i (i.e., either MP1, MP2, or MP3), the commitment decisions are fixed for the SP at iteration i (i.e., $\boldsymbol{X}^i$). Additionally, the lower bound of the decomposition algorithm (either A1, A2, or A3) is updated using the optimal solution of the master problem at iteration i of the outer loop (i.e., $LB = \Xi_{\mathrm{MP}} + \gamma_{\boldsymbol{\cdot}}$), as indicated in the

flowchart of Fig. 1.

2) Dispatch-Related Sub-Problem (Outer Loop): Given the values of the commitment variables obtained from solving the master problem at iteration i of the outer loop (i.e., $\boldsymbol{X}^i$ in either MP1, MP2, or MP3), the sub-problem can be written as follows:

$$\max_{\widetilde{\mathbb{P}}\in\Omega^{\boldsymbol{S}}}\left(\min_{\forall\boldsymbol{\chi}=\boldsymbol{X}^i\ ,\ \boldsymbol{Y}\in\Omega^{\boldsymbol{\mathcal{Y}}(\boldsymbol{\chi},\boldsymbol{\Lambda})}}\mathbb{E}_{\widetilde{\mathbb{P}}}\{\boldsymbol{Q}(\boldsymbol{\chi},\boldsymbol{Y},\boldsymbol{\Lambda})\}\right) \tag{9}$$

where $\boldsymbol{\chi}$ (i.e., the vector of *auxiliary continuous variables*) is equal to $\boldsymbol{X}^i$. Without loss of generality, a discrete sample space is considered in this paper [16, 17]. Accordingly, the max-min sub-problem (9) can be rewritten as follows:

$$\max_{\forall\tilde{p}_s\in\Omega^{\boldsymbol{S}}}\left(\min_{\forall\boldsymbol{\chi}=\boldsymbol{X}^i\ ,\ \forall\boldsymbol{Y}_s\in\Omega^{\boldsymbol{\mathcal{Y}}_s(\boldsymbol{\chi},\boldsymbol{\Lambda}^s)}}\sum_{s=1}^{N^S}\tilde{p}_s\cdot\boldsymbol{Q}(\boldsymbol{\chi},\boldsymbol{Y}_s,\boldsymbol{\Lambda}^s)\right)=$$
$$\max_{\forall\tilde{p}_s\in\Omega^{\boldsymbol{S}}}\left(\min_{\Psi_{\mathrm{SP}}}\overbrace{\sum_{s=1}^{N^S}\tilde{p}_s\cdot q_s}^{\Xi_{\mathrm{SP}}}\right) \tag{10}$$

where $\Psi_{\mathrm{SP}}=\{q_s\mid q_s\geq\boldsymbol{Q}(\boldsymbol{\chi},\boldsymbol{Y}_s,\boldsymbol{\Lambda}^s)\ \ s=1,\dots,N^S\ \ ,\ \ \boldsymbol{Y}_s\in\Omega^{\boldsymbol{\mathcal{Y}}_s(\boldsymbol{\chi},\boldsymbol{\Lambda}^s)}\ \ s=1,\dots,N^S\ \ ,\ \ \boldsymbol{\chi}=\boldsymbol{X}^i\ :\ \boldsymbol{\Upsilon}\}$ and $\boldsymbol{\Upsilon}$ represents the vector of dual variables corresponding to the constraint $\boldsymbol{\chi}=\boldsymbol{X}^i$. In previous research works in the power system area, the max-min sub-problem (10) has been solved by using (i) duality theory or KKT conditions to recast the max-min optimization problem into a single maximization problem, and (ii) a Big-M transformation to recast the bilinear terms into linear ones that generally results in an NP-hard MIQCQP optimization sub-problem [7, 9, 20]. Moreover, duality theory [11] and duality theory with Big-M transformation [21] have been used to solve DRO problems. On the contrary, the PDHG method is used in this work to solve the max-min sub-problem, which does not require duality theory, KKT conditions, and Big-M transformations. Given X^i obtained from solving the master problem at iteration i of the outer loop, the PDHG algorithm can be summarized as follows:

PDHG Algorithm (Inner Loop):

Step 1) Generate N^S training samples for all uncertain parameters using historical data within the sample space $\Omega^{\boldsymbol{\Lambda}}$.

Step 2) Set $j=0$, and parameters $\alpha>0$ and $\beta>0$ where $\alpha\cdot\beta<1$. Additionally, initialize parameters $\tilde{p}^{s(j=0)}$, using historical data [22], and $q^{s(j=0)}$, by independently solving the minimization problem $\min_{\Psi_{\mathrm{SP}_s}} q_s$ for $s=1,\dots,N^S$, where $\Psi_{\mathrm{SP}_s}=\{q_s\mid q_s\geq\boldsymbol{Q}(\boldsymbol{\chi},\boldsymbol{Y}_s,\boldsymbol{\Lambda}^s)\ \ ,\ \ \boldsymbol{Y}_s\in\Omega^{\boldsymbol{\mathcal{Y}}_s(\boldsymbol{\chi},\boldsymbol{\Lambda}^s)}\ \ ,\ \ \boldsymbol{\chi}=\boldsymbol{X}^i\}$.

Step 3) Set $j=j+1$.

Step 4) Calculate:

$$\tilde{p}^{sj}=\underset{\forall\tilde{p}_s\in\Omega^{\boldsymbol{S}}}{\mathrm{argmax}}\left\{\sum_{s=1}^{N^S}(\tilde{p}_s\cdot q^{s(j-1)})-\frac{1}{2\alpha}\cdot\sqrt[2]{\sum_{s=1}^{N^S}(\tilde{p}_s-\tilde{p}^{s(j-1)})^2}\right\} \tag{11}$$

Step 5) Set $\check{p}^{sj}=2\tilde{p}^{sj}-\tilde{p}^{s(j-1)}$.

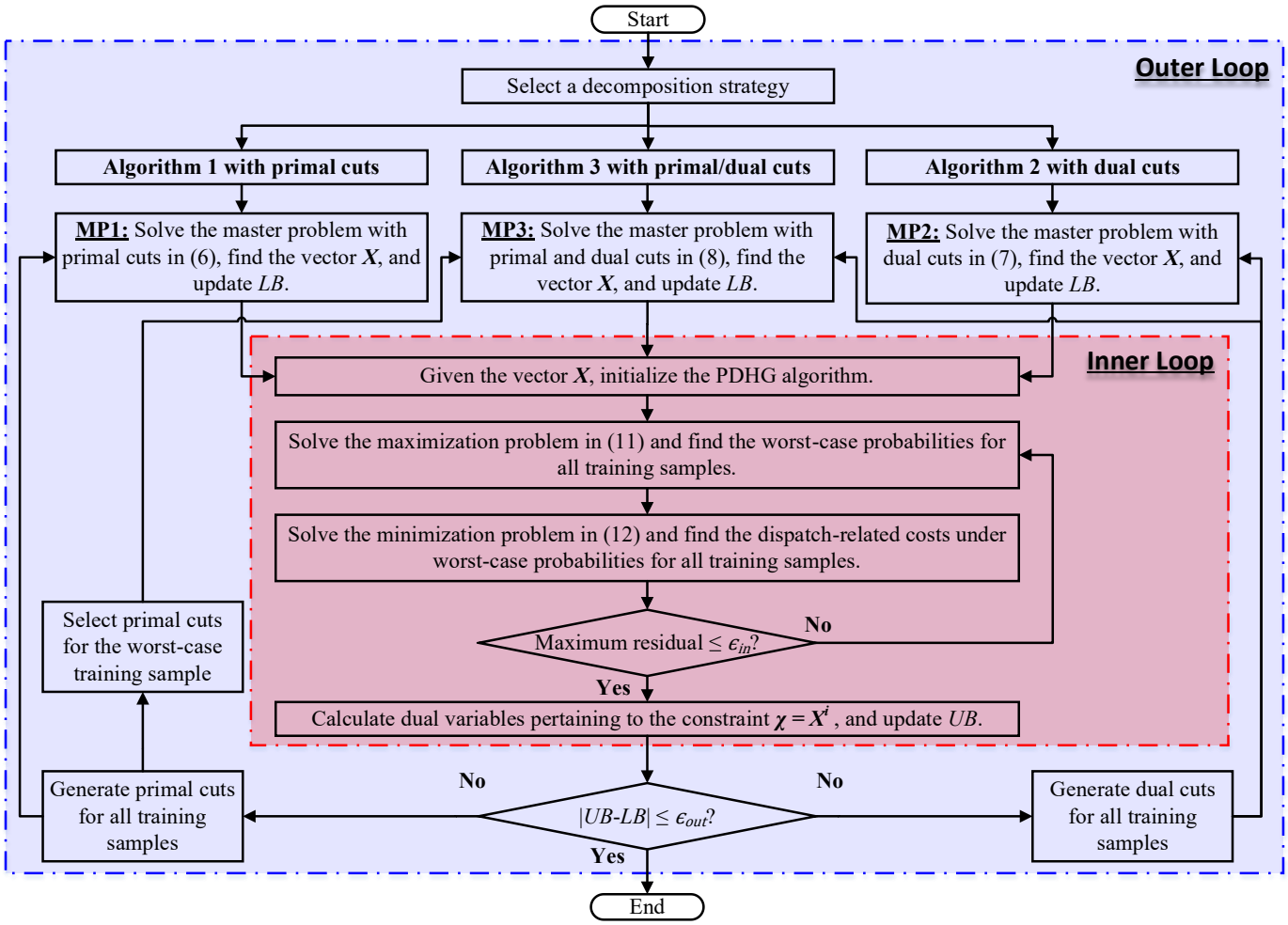

Fig. 1. Flowchart of the proposed decomposition approach.

Step 6) Calculate:

$$q^{sj} = \underset{\forall q_s \geq \boldsymbol{Q}(\boldsymbol{\chi},\boldsymbol{Y}_s,\boldsymbol{\Lambda}^s)\ ,\ \forall \boldsymbol{\chi}=\boldsymbol{X}^i\ ,\ \forall \boldsymbol{Y}_s \in \Omega^{\boldsymbol{Y}_s(\boldsymbol{\chi},\boldsymbol{\Lambda}^s)}}{\operatorname{argmin}} \left\{ \sum_{s=1}^{N^s} (\check{p}^{sj} \cdot q_s) + \frac{1}{2\beta} \cdot \sqrt[2]{\sum_{s=1}^{N^s} (q_s - q^{s(j-1)})^2} \right\} \quad (12)$$

Step 7) If the maximum residual is less than the convergence tolerance of the inner loop, i.e., $\max(\left\| q^{sj} - q^{s(j-1)} \right\|_2, \left\| \tilde{p}^{sj} - \tilde{p}^{s(j-1)} \right\|_2) \leq \epsilon_{in}$, report $\tilde{p}^{sj}$ and q^{sj} for $s = 1, \dots, N^s$ and $\boldsymbol{\Upsilon}$, i.e., the vector of dual variables corresponding to the constraint $\boldsymbol{\chi} = \boldsymbol{X}^i$ obtained from solving the minimization problem $\min_{\forall q_s \geq \boldsymbol{Q}(\boldsymbol{\chi},\boldsymbol{Y}_s,\boldsymbol{\Lambda}^s)\ ,\ \forall \boldsymbol{\chi}=\boldsymbol{X}^i\ ,\ \forall \boldsymbol{Y}_s \in \Omega^{\boldsymbol{Y}_s(\boldsymbol{\chi},\boldsymbol{\Lambda}^s)}} \left(\sum_{s=1}^{N^s} \tilde{p}^{sj} \cdot q_s \right)$, and terminate the PDHG algorithm; otherwise, return to Step 3.

It is worth noting that each of the optimization problems in Step 4 and Step 6 of the PDHG algorithm with the norm-2 regularization terms can be recast into a convex/concave optimization problem with quadratic constrains and polynomial complexity [23].

After satisfying the convergence criterion for the inner loop (i.e., that of the PDHG algorithm) of the decomposition approach, the optimal value of the sub-problem under the worst-case probability distributions at iteration i of the outer loop is obtained in terms of $\Xi_{\text{SP}} = \sum_{s=1}^{N^s} (\tilde{p}^{sj} \cdot q^{sj})$ where j indicates the last iteration of the inner loop, $\tilde{p}^{si} = \tilde{p}^{sj}$, and $q^{si} = q^{sj}$ for $s = 1, \dots, N^s$. Additionally, the upper bound of the decomposition algorithm is updated by means of the optimal solution of the master problem (either MP1, MP2, or MP3) and the SP at iteration i of the outer loop (i.e., $UB = \min[UB, \Xi_{\text{MP}} + \Xi_{\text{SP}}]$), as indicated in the flowchart of Fig. 1. If $|UB - LB| \leq \epsilon_{out}$, the optimal solution has been attained, and the decomposition algorithm is terminated. Otherwise, a cut is sent to the master problem as given below:

Cuts in Master Problem 1: A primal cut is sent to MP1 using $\tilde{p}^{si} = \tilde{p}^{sj}$ for $s = 1, \dots, N^s$ as indicated in (6c).

Cuts in Master Problem 2: A dual cut is sent to MP2 using $\boldsymbol{X}^i$ and $\boldsymbol{\Upsilon}^i$ as well as $\tilde{p}^{si} = \tilde{p}^{sj}$ and $q^{si} = q^{sj}$ for $s = 1, \dots, N^s$ as indicated in (7c).

Cuts in Master Problem 3: All dispatch-related constraints for the worst-case training sample and a dual cut are sent to MP3 using $\boldsymbol{X}^i$ and $\boldsymbol{\Upsilon}^i$ as well as $\tilde{p}^{si} = \tilde{p}^{sj}$ and $q^{si} = q^{sj}$ for $s = 1, \dots, N^s$ as indicated in (8c) and (8d).

The convergence of the three decomposition approaches A1, A2, and A3 is proved in the Appendix.

IV. Case Studies

In this section, the proposed DR-NCUC model is tested on the Wood and Wollenberg 6-bus test system [24] as well as on the IEEE 118-bus test system [25] under different conditions. All studies are carried out on a server with 100 Intel Xeon cores and 120 GB of RAM in GAMS software using Mosek solver where $\epsilon_{in} = \epsilon_{out} = 10^{-2}$ and $\alpha = 10^2$ and $\beta = 10^{-3}$. The convergence of the PDHG algorithm can be proved provided that $\alpha \cdot \beta < 1$ [19]. In this paper, the adopted values for α and β offer the most effective performance for the PDHG algorithm.

A. Simulation Results on the 6-Bus Test System

In this section, the proposed DR-NCUC model is tested on a modified 6-bus test system [24] and its optimal solution is compared with the optimal solutions of the deterministic (i.e., D-NCUC), and robust (i.e., R-NCUC) models. The test system includes three demands at nodes 4, 5, and 6, and two 25-MW RES units (i.e., wind farms) connected to nodes 1 and 2. The daily profiles of demands and RES power productions are constructed using the normalized daily profile of demands and RES power productions in the UK on 14th February 2018 [26]. Also, the uncertainties of demands and RES power productions are characterized using training scenarios within the sample space $\Omega^{\boldsymbol{\Lambda}} = \{\tilde{p}_{dh} \in [0.8 \cdot \hat{P}_{dh}, 1.2 \cdot \hat{P}_{dh}] \quad \forall d \in \Omega^D, \forall h \in \Omega^H \ ; \ \tilde{p}_{uh} \in [0.7 \cdot \hat{P}_{uh}, 1.3 \cdot \hat{P}_{uh}] \quad \forall u \in \Omega^R, \forall h \in \Omega^H\}$.

1) Analyzing Total In-Sample and Out-of-Sample Costs for the Different Models: The total in-sample costs of the D-NCUC, DR-NCUC, and R-NCUC models are presented in Table I, where five samples are considered for the DR-NCUC model, and the confidence level is set to 0.05. According to Table I, the total in-sample cost of the DR-NCUC model is smaller than that of the R-NCUC model and larger than that of the D-NCUC model. To evaluate the performance of different NCUC models under various realizations of uncertain parameters with different distributions, an out-of-sample analysis is carried out where 5000 samples are generated for each distribution of each uncertain parameter using a Latin hypercube sampling method [27]. Regarding the out-of-sample analysis: first) each NCUC model is solved (i.e., D-NCUC, DR-NCUC, and R-NCUC), second) a multi-period optimal power flow (OPF) is solved for each sample by fixing all commitment variables on their optimal values obtained from solving the NCUC model, and third) the expected total cost is calculated over all 5000 samples. In this study, six distributions with different parameters are considered, including:

BH: *High-variance* Beta distribution with shape factors $a = b = 1$.

BM: *Mid-variance* Beta distribution with shape factors $a = b = 2$.

BL: *Low-variance* Beta distribution with shape factors $a = b = 3$.

NH: *High-variance* Normal distribution with mean values $\mu_{dh} = \hat{P}_{dh}$ and $\mu_{uh} = \hat{P}_{uh}$ and standard deviations $\sigma_{dh} = 0.07\hat{P}_{dh}$ and $\sigma_{uh} = 0.09\hat{P}_{uh}$.

NM: *Mid-variance* Normal distribution with mean values $\mu_{dh} = \hat{P}_{dh}$ and $\mu_{uh} = \hat{P}_{uh}$ and standard deviations $\sigma_{dh} = 0.05\hat{P}_{dh}$ and $\sigma_{uh} = 0.07\hat{P}_{uh}$.

NL: *Low-variance* Normal distribution with mean values $\mu_{dh} = \hat{P}_{dh}$ and $\mu_{uh} = \hat{P}_{uh}$ and standard deviations $\sigma_{dh} = 0.03\hat{P}_{dh}$ and $\sigma_{uh} = 0.05\hat{P}_{uh}$.

It is worthwhile to mention that the [0,1] interval for each uncertain parameter in Beta distributions is mapped into its corresponding interval in the support space (i.e., $\Omega^{\boldsymbol{\Lambda}} = \{\tilde{p}_{dh} \in [0.8 \cdot \hat{P}_{dh}, 1.2 \cdot \hat{P}_{dh}] \quad \forall d \in \Omega^D, \forall h \in \Omega^H \ ; \ \tilde{p}_{uh} \in [0.7 \cdot \hat{P}_{uh}, 1.3 \cdot \hat{P}_{uh}] \quad \forall u \in \Omega^R, \forall h \in \Omega^H\}$). The out-of-sample results of the NCUC models with these distributions

are presented in Table II. For all distributions, the proposed DR-NCUC has lower total out-of-sample cost than R-NCUC and D-NCUC. Also, R-NCUC has lower total out-of-sample cost than D-NCUC in all distributions of Table II. In addition, Table II shows that by increasing the variance of Beta and Normal distributions (which typically occurs when the uncertain parameters become more volatile), the total out-of-sample costs of all NCUC models increase. However, increments in DR-NCUC and R-NCUC outcomes are much smaller than that in D-NCUC because DR-NCUC and R-NCUC characterize deviations from the forecasted values of uncertain parameters and provide robustness against different realizations of uncertain parameters, unlike D-NCUC. Accordingly, the DR-NCUC/D-NCUC has the best/worst out-of-sample performance for all out-of-sample cases in Table II.

TABLE I
TOTAL IN-SAMPLE COSTS [$] FOR DIFFERENT NCUC MODELS (6-BUS)

Model	D-NCUC	DR-NCUC	R-NCUC
Total Costs [$]	48715	51890	62974

TABLE II
TOTAL OUT-OF-SAMPLE COSTS [$] FOR DIFFERENT NCUC MODELS (6-BUS)

Model	BH	BM	BL	NH	NM	NL
D-NCUC	92541	75316	67771	64943	56798	50964
DR-NCUC	**50361**	**49829**	**49730**	**49840**	**49694**	**49685**
R-NCUC	50584	50556	50545	50570	50533	50527

TABLE III
TOTAL IN-SAMPLE COSTS VS. DIFFERENT CONFIDENCE LEVELS

Confidence Level	Total Costs [$]
0.00	50467
0.02	51036
0.04	51605
0.06	52175
0.08	52744
0.10	53313

TABLE IV
COMPUTATIONAL COMPLEXITIES OF A1-A3 ON THE 6-BUS TEST SYSTEM

Decomposition Algorithm	Number of Iterations		Execution Time [s]
	Outer Loop	Inner Loop	
A1	3	6	174
A2	43	93	298
A3	18	41	93

2) Evaluating the Impact of the Confidence Level on the Total Cost: In Table III, the impact of the confidence level on the total in-sample cost is evaluated by increasing the confidence level value from 0.00 to 0.10 in the step of 0.02. According to Table III, the total in-sample cost increases by increasing the confidence level from 0.00 to 0.10. Increasing the confidence level results in higher robustness against different realizations of the uncertain parameters. This higher robustness comes at a cost, as shown in Table III. Hence, the conservatism of the optimal solution and its associated cost can be controlled by increasing/decreasing the confidence level in the proposed DR-NCUC model.

3) Analyzing the Computational Complexity of the Decomposition Algorithms: To analyze the computational complexity of the proposed decomposition algorithms, the DR-NCUC model is solved using algorithms A1, A2, and A3, presented in Section III, and their computational performances are illustrated in Table IV. The solutions obtained by decomposition algorithms A1-A3 have negligible differences, while their computational performances for the same convergence tolerance are significantly different. According to Table IV, A1 converges in 174s with 3 iterations in the outer loop and 6 iterations in the inner loop, A2 converges in 298s with 43 iterations in the outer loop and 93 iterations in the inner loop, and A3 converges in 93s with 18 iterations in the outer loop and 41 iterations in the inner loop. Therefore, A3 has the lowest computation time as compared to A1 and A2. Also, A1 that has a lower number of iterations than A3 requires a higher computation time than A3. The reason is that A1 incorporates all dispatch-related constraints to the master problem at each iteration, while A3 only adds the dispatch-related constraints of the worst-case training sample to the master problem at the first iteration.

B. Simulation Results on the IEEE 118-Bus Test System

In this section, the tractability of the proposed DR-NCUC approach is evaluated on a modified IEEE 118-bus test system [25]. This test system includes 54 units, 186 lines, and six 100-MW RES units (i.e., wind farms) connected to nodes 10, 30, 50, 70, 90, and 110. The daily profiles of demands and RES power productions as well as the sample space Ω^{Λ} of their uncertainties are constructed as described in the section pertaining to the 6-bus test system.

1) Analyzing Total In-Sample and Out-of-Sample Costs for the Different Models: Similar to the previous study on the 6-bus test system, in-sample scenarios are used within the NCUC models to obtain optimal commitment decisions, and consequently, find their total in-sample costs. In general, deterministic approaches assume that no uncertain parameter of the optimization problem deviates from its forecast value. On the contrary, non-deterministic approaches assume that the input parameters of the optimization problem may favorably or unfavorably deviate from their forecast values. The main characteristics of the in-sample scenarios of each NCUC model can be summarized as follows:

- The D-NCUC model considers only one in-sample scenario corresponding to the forecast value of each uncertain parameter.
- The S-NCUC model considers a specific number of in-sample scenarios with known probability distributions corresponding to favorable/unfavorable deviations of the uncertain parameters from their forecast values.
- The R-NCUC model considers one in-sample scenario corresponding to the worst-case deviations of the uncertain parameters from their forecast values.
- The DR-NCUC model considers a specific number of in-sample scenarios with unknown but bounded probability distributions corresponding to favorable/unfavorable deviations of the uncertain parameter from their forecast values.

Different numbers of in-sample scenarios can be considered within the S-NCUC and DR-NCUC models. Here, without loss of generality, five in-sample scenarios are used in the S-NCUC and DR-NCUC models to characterize the uncertainties. In the S-NCUC model, it is assumed that uncertain parameters follow uniform probability distributions, and thus, identical probabilities are considered for all in-sample scenari-

os in this model. Also, the confidence level is set to 0.05 in the DR-NCUC model.

The total in-sample costs for the D-NCUC, S-NCUC, R-NCUC, and DR-NCUC models are provided in Table V. Since each of the D-NCUC, S-NCUC, R-NCUC, and DR-NCUC models has a different uncertainty modeling approach, and consequently, considers different in-sample scenarios, their performances may not be truly evaluated using in-sample scenarios. Especially, the D-NCUC model obtains the lowest in-sample cost in Table V as this model ignores all uncertainties of the problem and considers unrealistically idealized operation conditions. On the contrary, every out-of-sample scenario is unseen for the NCUC models as out-of-sample scenarios are outside the design framework of each NCUC model. Thus, the out-of-sample analysis using a sufficiently large number of different unseen out-of-sample scenarios can truly evaluate the performance of each NCUC model through their out-of-sample costs. In this paper, an identical set of 5000 different scenarios is considered in the out-of-sample analysis to appropriately capture the uncertainty spectrum and fairly calculate the out-of-sample cost for each NCUC model. Similar to the 6-bus test system, six distributions with different parameters (i.e., BH, BM, BL, NH, NM, NL) are considered to perform the out-of-sample analysis.

The total out-of-sample costs for the D-NCUC, S-NCUC, R-NCUC, and DR-NCUC models are provided in Table VI, wherein the best out-of-sample cost in each case (distribution) is indicated in bold. In summary:

- The D-NCUC model results in the highest out-of-sample cost as compared to the S-NCUC, R-NCUC, and DR-NCUC models for all six distributions BH, BM, BL, NH, NM, and NL since D-NCUC ignores all uncertainties.
- The out-of-sample cost in the S-NCUC model is less than the R-NCUC model for cases with low volatility in the uncertain parameters (i.e., distributions BL, NM, and NL). On the contrary, the out-of-sample cost in the R-NCUC model with high conservativeness is less than the S-NCUC model for cases with high volatility in the uncertain parameters (i.e., distributions BH, BM, and NH).
- The DR-NCUC model offers a lower out-of-sample cost as compared to the D-NCUC, S-NCUC, and R-NCUC models in all examined cases for the six distributions BH, BM, BL, NH, NM, and NL.

TABLE V
TOTAL IN-SAMPLE COSTS [M$] FOR DIFFERENT NCUC MODELS (118-BUS)

Model	D-NCUC	S-NCUC	DR-NCUC	R-NCUC
Total Costs [M$]	0.894032	0.897897	0.935734	1.175673

TABLE VI
TOTAL OUT-OF-SAMPLE COSTS [M$] FOR DIFFERENT NCUC MODELS (118-BUS)

Model	BH	BM	BL	NH	NM	NL
D-NCUC	1.072240	0.987709	0.947071	0.998280	0.968628	0.932189
S-NCUC	0.944359	0.921399	0.910328	0.924219	0.916084	0.906027
DR-NCUC	**0.904567**	**0.903026**	**0.902234**	**0.902981**	**0.902346**	**0.901515**
R-NCUC	0.921311	0.919843	0.919088	0.919792	0.919185	0.918390

2) Analyzing the Computational Complexity of the Decomposition Algorithms Using the Active Set Strategy: To enhance the performance of the proposed solution approach, an effective ASS is introduced to ignore inactive constraints in the master problem and sub-problem as follows:

Active Set Strategy

Step 1) Solve a multi-period OPF for each training sample to identify the worst-case training sample among all training samples. Binary decision variables in the NCUC are relaxed into continuous ones. Report active constraints pertaining to transmission lines under the worst-case training sample to MP3 (i.e., Ω^{MPASS}) and the SP (i.e., Ω^{SPASS}).
Step 2) Solve MP3 for the reduced set of reported constraints pertaining to transmission lines (i.e., Ω^{MPASS}).
Step 3) Check that all constraints of all transmission lines under the worst-case training sample are satisfied. If so, update LB as indicated in A3, and go to Step 4. Otherwise, add all violated constraints (i.e., Ω^{MPVASS}) to the set of active constraints in MP3 (i.e., $\Omega^{MPASS} \cup \Omega^{MPVASS} \rightarrow \Omega^{MPASS}$) and go to Step 2.
Step 4) Solve the SP for the reduced set of constraints belonging to transmission lines (i.e., Ω^{SPASS}) as indicated in A3.
Step 5) Check that all constraints of all transmission lines under all training samples are satisfied. If so, update UB, as indicated in A3, and go to Step 6. Otherwise, add the violated constraints (i.e., Ω^{SPVASS}) to the set of active constraints in the SP (i.e., $\Omega^{SPASS} \cup \Omega^{SPVASS} \rightarrow \Omega^{SPASS}$) and go to Step 4.
Step 6) Check the convergence criterion in A3. If so, terminate and report the optimal solution. Otherwise, generate appropriate cuts, as indicated in A3, and go to Step 2.

The main difference between the proposed ASS and previous ones reported in the literature is that the proposed strategy considers transmission line constraints for the worst-case training sample, instead of all training samples, in the master problem of the decomposition algorithm A3.

The results of the proposed DR-NCUC model, solved by A1-A3 with/without the ASS method, on the modified IEEE 118-bus test system are presented in Table VII. The ASS method can be used for A1 and A2 similarly. Although ASS increases the number of iterations in A1-A3, it results in lower computation time in all algorithms A1-A3. Although A1 has the lowest number of iterations in comparison with A2 and A3 with/without ASS, it has the highest computation time with/without ASS in Table VII, while its computation time was lower/higher than the computation time of A2/A3 in Table IV. The reason is that in the modified IEEE 118-bus test system, MP1 in A1, considering all training samples, is a computationally cumbersome MIQCQP optimization problem at each iteration of the decomposition algorithm. In Table VII, A3 with/without ASS has the lowest computation time among A1-A3. Additionally, and more importantly, the computation time of A3 with ASS is 12m, 34s, which is a reasonable computation time for solving a DR-NCUC problem with a 24-hour planning horizon. Both Tables IV and VII show that the number of iterations of A3 is lower than that of A2 in all cases, which numerically confirms the convergence theorem presented in the Appendix.

Finally, it is worth noting that the maximum convexification error between the exact non-convex AC power flow equations and the convexified ones is less than 1% in both test systems. To calculate the convexification error, first, the optimal unit commitment schedule, obtained from solving the pro-

posed distributionally robust AC NCUC problem, is fixed. Afterward, with such fixed unit commitment schedule, the optimal dispatch-related costs under convex and exact power flow equations, denoted by OC^{ED^c} and OC^{ED^e}, are obtained by solving quadratically-constrained quadratic-programming (QCQP) and nonlinear programming (NLP) multi-period OPF problems, respectively. The convexification error is calculated by $\left|\frac{OC^{ED^e}-OC^{ED^c}}{OC^{ED^e}}\right|$ [15]. It is worthwhile to mention that the convexification error under different loading conditions can be effectively reduced by using different approaches in the literature, such as bound tightening techniques [28, 29] and infeasibility cut-based approaches [30].

TABLE VII
RESULTS OF DR-NCUC MODEL SOLVED BY DIFFERENT DECOMPOSITION ALGORITHMS WITH/WITHOUT ASS ON THE IEEE 118-BUS TEST SYSTEM

Decomposition Algorithm	Number of Iterations		Execution Time [s]
	Outer Loop	Inner Loop	
A1 without ASS	3	7	18950
A1 with ASS	6	14	14560
A2 without ASS	71	138	2548
A2 with ASS	84	171	1034
A3 without ASS	17	39	1952
A3 with ASS	23	52	754

V. CONCLUSIONS

This paper presents a min-max-min DR-NCUC model to obtain the optimal unit commitment solution under the worst-case probability distributions of uncertain demands and RES productions. A training-based data-driven ambiguity set is introduced to characterize unknown probability distributions, and a confidence level is used to control the conservatism of the optimal solution and the deviations of unknown probability distributions from probability distributions of training samples. The non-convex AC power flow equations are approximated by quadratic and McCormick relaxations. Since the proposed min-max-min DR-NCUC problem cannot be solved by optimization solvers directly, a new decomposition algorithm is introduced in this paper. The proposed algorithm benefits from both primal and dual cuts in addition to the PDHG method. Also, it uses an effective ASS to enhance computational efficiency. The main conclusions of two case studies can be summarized as follows: i) the out-of-sample results illustrate the superiority of the proposed DR-NCUC model over both D-NCUC, S-NCUC, and R-NCUC models; ii) A3 with primal cuts for the worst-case training sample, in addition to dual cuts, in MP3 and the PDHG method in the SP outperforms both A1 with primal cuts for all training samples and A2 with dual cuts; iii) the proposed ASS method enhances the computational performance of all algorithms A1-A3; and iv) the quadratic and McCormick relaxations offer a convex optimization problem that can be tractably solved with a reasonable convexification error of 1%.

VI. APPENDIX

Theorem: *Let $N_{|\Omega^s|}$ be the number of extreme points in the data-driven ambiguity set (i.e., $\Omega^{\boldsymbol{S}}$), $N_{|\Omega^T|}$ be the number of commitment variables, and the innermost convex minimization problem in (3) has "complete recourse". Then, (i) A1-A3 converge in a finite number of iterations, and (ii) A1 and A2 converge in $O\left(N_{|\Omega^s|}\right)$ and $O\left(N_{|\Omega^s|}\cdot 2^{N_{|\Omega^T|}}\right)$ iterations, respectively, and A3 converges in $O(A3) \le O\left(N_{|\Omega^s|}\cdot 2^{N_{|\Omega^T|}}\right)$ iterations.*

Proof: The max-min optimization problem (10) is bi-level over $\Omega^{\boldsymbol{S}}$ and $\Omega^{\boldsymbol{y}_s(\boldsymbol{\chi},\boldsymbol{\Lambda}^s)}$, two disjoint convex sets. The ambiguity set $\Omega^{\boldsymbol{S}}$ of the max problem is polyhedral and convex with a finite number of extreme points not dependent on commitment variables and uncertain parameters. The optimal solution of this problem is one of the extreme points of the ambiguity set. The feasible space $\Omega^{\boldsymbol{y}_s(\boldsymbol{\chi},\boldsymbol{\Lambda}^s)}$ of the min problem pertaining to operation variables $\boldsymbol{y}_s$ for each training sample s depends on commitment variables and uncertain parameters fixed to particular values for each training sample. Given specific values for commitment variables, the min problem is convex, and its optimal solution can be found with a finite number of iterations.

A. Proof of Convergence for Algorithm 1

Since the ambiguity set $\Omega^{\boldsymbol{S}}$ is a polyhedral convex set with a finite number of extreme points, after a finite number of iterations, the extreme points for the training sample s of the ambiguity set would be repeated at two successive iterations i and $(i-1)$ (i.e., $\tilde{p}^{si}=\tilde{p}^{s(i-1)}$), where the optimal solution of the SP at iteration i is $\Xi^i_{\mathrm{SP}}=\sum_{s=1}^{N^S}\tilde{p}^{si}\cdot q^{si}$. In A1, primal cutting planes, generated from solving the SP at each iteration, are added to MP1, as given by (6c), for all training samples. Since $\tilde{p}^{si}=\tilde{p}^{s(i-1)}$ for all training samples, identical primal cutting planes are added to (6c) in MP1 at successive iterations i and $(i-1)$, and thus, $\gamma^i_{\mathrm{MP1}}=\gamma^{(i-1)}_{\mathrm{MP1}}$. Therefore, the same solution is obtained from solving (6) at two successive iterations i and $(i-1)$, i.e., $\boldsymbol{X}^i=\boldsymbol{X}^{(i-1)}$, and consequently, $LB^i=\Xi^i_{\mathrm{MP}}+\gamma^i_{\mathrm{MP1}}=\boldsymbol{A}^{\intercal}\cdot\boldsymbol{X}^i+\gamma^i_{\mathrm{MP1}}=\boldsymbol{A}^{\intercal}\cdot\boldsymbol{X}^{(i-1)}+\gamma^{(i-1)}_{\mathrm{MP1}}=\Xi^{(i-1)}_{\mathrm{MP}}+\gamma^{(i-1)}_{\mathrm{MP1}}=LB^{(i-1)}$. Since 1) the optimal solution of the inner minimization for each training sample in the SP is a function of commitment variables and uncertain parameters, 2) $\boldsymbol{X}^i=\boldsymbol{X}^{(i-1)}$, $q^{si}=q^{s(i-1)}$, and 3) uncertain parameters are fixed to particular values in successive iterations i and $(i-1)$ for all training samples, consequently, $UB^i=\Xi^i_{\mathrm{MP}}+\Xi^i_{\mathrm{SP}}=\boldsymbol{A}^{\intercal}\cdot\boldsymbol{X}^i+\sum_{s=1}^{N^S}\tilde{p}^{si}\cdot q^{si}=\boldsymbol{A}^{\intercal}\cdot\boldsymbol{X}^{(i-1)}+\sum_{s=1}^{N^S}\tilde{p}^{s(i-1)}\cdot q^{s(i-1)}=\Xi^{(i-1)}_{\mathrm{MP}}+\Xi^{(i-1)}_{\mathrm{SP}}=UB^{(i-1)}$, therefore, upper and lower bounds remain unchanged at successive iterations i and $(i-1)$. In addition, based on the proposed decomposition approach in Fig. 1, $LB^i=\boldsymbol{A}^{\intercal}\cdot\boldsymbol{X}^i+\gamma^i_{\mathrm{MP1}}\ge\boldsymbol{A}^{\intercal}\cdot\boldsymbol{X}^{(i-1)}+\sum_{s=1}^{N^S}\tilde{p}^{s(i-1)}\cdot q^{s(i-1)}=UB^{(i-1)}$. From $LB^i\ge UB^{(i-1)}$ and $UB^i=UB^{(i-1)}$, we obtain $LB^i\ge UB^i$. At the same time, in general, we have $LB^i\le UB^i,\forall i$. These two inequalities yield $LB^i=UB^i$. Consequently, based on the stopping condition specified in the flowchart of Fig. 1, A1 converges at iteration i for any specified convergence tolerance (ϵ_{out}) value. Since the number of iterations of A1 is a function of the number of extreme points of the ambiguity set (i.e., $N_{|\Omega^s|}$), as described above, A1 converges in at most $O\left(N_{|\Omega^s|}\right)$ iterations.

B. Proof of Convergence for Algorithm 2

In A2, after a finite number of iterations, we obtain $\tilde{p}^{si} = \tilde{p}^{s(i-1)}$, since there is a finite number of extreme points in polyhedral convex ambiguity set Ω^{s} (as described in the proof of convergence of A1), and $\boldsymbol{X}^{i} = \boldsymbol{X}^{(i-1)}$, since there is a finite number of unit commitment configurations. In A2, dual cutting planes, generated from solving the SP at each iteration, are added to MP2, as given in (7c), for all training samples. Also, the optimal solution of the SP at iteration i is: $\Xi^{i}_{\mathrm{SP}} = \sum_{s=1}^{N^{S}} \tilde{p}^{si} \cdot q^{si}$. Since 1) the SP is a max-min optimization problem over two disjoint convex sets, 2) the feasible space of its inner minimization problem is a function of commitment variables and uncertain parameters, 3) $\boldsymbol{X}^{i} = \boldsymbol{X}^{(i-1)}$, and 4) uncertain parameters are fixed to particular values in successive iterations i and $(i-1)$ for all training samples, then, the optimal solution of the inner minimization problem and the dual variables of the complicating commitment variables in the SP are also repeated (i.e., $q^{si} = q^{s(i-1)}$ and $\boldsymbol{\Upsilon}^{i} = \boldsymbol{\Upsilon}^{(i-1)}$). This result and $\tilde{p}^{si} = \tilde{p}^{s(i-1)}$ for all training samples yield that identical dual cutting planes are added as (7c) to MP2 at successive iterations i and $(i-1)$, and thus, $\gamma^{i}_{\mathrm{MP2}} = \gamma^{(i-1)}_{\mathrm{MP2}}$. Considering this result and $\boldsymbol{X}^{i} = \boldsymbol{X}^{(i-1)}$, we can write: $LB^{i} = \Xi^{i}_{\mathrm{MP}} + \gamma^{i}_{\mathrm{MP2}} = \boldsymbol{A}^{\intercal} \cdot \boldsymbol{X}^{i} + \gamma^{i}_{\mathrm{MP2}} = \boldsymbol{A}^{\intercal} \cdot \boldsymbol{X}^{(i-1)} + \gamma^{(i-1)}_{\mathrm{MP2}} = \Xi^{(i-1)}_{\mathrm{MP}} + \gamma^{(i-1)}_{\mathrm{MP2}} = LB^{(i-1)}$. Moreover, considering $\boldsymbol{X}^{i} = \boldsymbol{X}^{(i-1)}$, $\tilde{p}^{si} = \tilde{p}^{s(i-1)}$, and $q^{si} = q^{s(i-1)}$, we can write: $UB^{i} = \Xi^{i}_{\mathrm{MP}} + \Xi^{i}_{\mathrm{SP}} = \boldsymbol{A}^{\intercal} \cdot \boldsymbol{X}^{i} + \sum_{s=1}^{N^{S}} \tilde{p}^{si} \cdot q^{si} = \boldsymbol{A}^{\intercal} \cdot \boldsymbol{X}^{(i-1)} + \sum_{s=1}^{N^{S}} \tilde{p}^{s(i-1)} \cdot q^{s(i-1)} = \Xi^{(i-1)}_{\mathrm{MP}} + \Xi^{(i-1)}_{\mathrm{SP}} = UB^{(i-1)}$. In addition, based on the proposed decomposition approach in Fig. 1, similarly to A1, $LB^{i} = \boldsymbol{A}^{\intercal} \cdot \boldsymbol{X}^{i} + \gamma^{i}_{\mathrm{MP2}} \geq \boldsymbol{A}^{\intercal} \cdot \boldsymbol{X}^{(i-1)} + \sum_{s=1}^{N^{S}} \tilde{p}^{s(i-1)} \cdot q^{s(i-1)} = UB^{(i-1)}$. Considering $LB^{i} \geq UB^{(i-1)}$ and $UB^{i} = UB^{(i-1)}$, we obtain $LB^{i} \geq UB^{i}$. At the same time, in general, we have $LB^{i} \leq UB^{i}, \forall i$. These two inequalities yield $LB^{i} = UB^{i}$. Consequently, based on the stopping condition specified in the flowchart of Fig. 1, A2 converges at iteration i for any specified convergence tolerance (ϵ_{out}) value. Since the number of iterations is a function of the number of extreme points of the ambiguity set (i.e., $N_{|\Omega^{S}|}$) and the number of unit commitment combinations (i.e., $2^{N_{|\Omega^{T}|}}$), A2 converges in at most $O\left(N_{|\Omega^{S}|} \cdot 2^{N_{|\Omega^{T}|}}\right)$ iterations.

C. Proof of Convergence for Algorithm 3

The convergence proof for A3 can be presented similarly to A2 as both MP2 in A2 and MP3 in A3 utilize dual cutting planes in (7c) and (8d), respectively. However, A3 shrinks the feasible space of MP3 as compared to MP2 and bounds the worst-case expected dispatch-related costs using the dispatch related cost of the worst-case training sample, rather than all training samples, in constraint (8c). Accordingly, it converges in at most $O(A3) \leq O(A2) = O\left(N_{|\Omega^{S}|} \cdot 2^{N_{|\Omega^{T}|}}\right)$ iterations.

It is noteworthy to mention that the big O notation is used here to describe the convergence rate of each decomposition algorithm by means of an *a priori upper bound*. Therefore, the number of iterations is not necessarily equal to such a priori upper bound. In other words, each decomposition algorithm (i.e., A1, A2, and A3) can find ϵ-optimal solutions in a finite number of iterations, notably much less than those of the a priori upper bound, as verified by the different case studies in Section IV.

REFERENCES

[1] L. Wu, M. Shahidehpour, and T. Li, "Stochastic security-constrained unit commitment," *IEEE Transactions on Power Systems,* vol. 22, no. 2, pp. 800-811, 2007.

[2] J. M. Morales, A. J. Conejo, and J. Pérez-Ruiz, "Economic valuation of reserves in power systems with high penetration of wind power," *IEEE Transactions on Power Systems,* vol. 24, no. 2, pp. 900-910, 2009.

[3] A. Papavasiliou, and S. S. Oren, "Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network," *Operations Research Letters,* vol. 61, no. 3, pp. 578-592, 2013.

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

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

[6] A. Nasri, S. J. Kazempour, A. J. Conejo, and M. Ghandhari, "Network-constrained AC unit commitment under uncertainty: A Benders' decomposition approach," *IEEE Transactions on Power Systems,* vol. 31, no. 1, pp. 412-422, 2015.

[7] N. Amjady, S. Dehghan, A. Attarha, and A. Conejo, "Adaptive robust network-constrained AC unit commitment," *IEEE Transactions on Power Systems,* vol. 32, no. 1, pp. 672-683, 2016.

[8] Q. P. Zheng, J. Wang, and A. L. Liu, "Stochastic optimization for unit commitment—A review," *IEEE Transactions on Power Systems,* vol. 30, no. 4, pp. 1913-1924, 2014.

[9] C. Zhao, and R. Jiang, "Distributionally robust contingency-constrained unit commitment," *IEEE Transactions on Power Systems,* vol. 33, no. 1, pp. 94-102, 2017.

[10] C. Duan, L. Jiang, W. Fang, and J. Liu, "Data-driven affinely adjustable distributionally robust unit commitment," *IEEE Transactions on Power Systems,* vol. 33, no. 2, pp. 1385-1398, 2017.

[11] A. Gourtani, H. Xu, D. Pozo, and T.-D. Nguyen, "Robust unit commitment with *n*-1 security criteria," *Mathematical Methods of Operations Research,* vol. 83, no. 3, pp. 373-408, 2016.

[12] A. Esteban-Pérez, and J. M. Morales, "Partition-based distributionally robust optimization via optimal transport with order cone constraints," *arXiv preprint arXiv:.01769*, 2019.

[13] H. Rahimian, and S. Mehrotra, "Distributionally robust optimization: A review," *arXiv preprint arXiv:.05659*, 2019.

[14] M. F. Anjos, and A. Conejo, "Unit commitment in electric energy systems," *Foundations Trends® in Electric Energy Systems,* vol. 1, no. 4, pp. 220-310, 2017.

[15] H. Hijazi, C. Coffrin, and P. Van Hentenryck, "Convex quadratic relaxations for mixed-integer nonlinear programs in power systems," *Mathematical Programming Computation,* vol. 9, no. 3, pp. 321-367, 2017.

[16] C. Zhao, and Y. Guan, "Data-driven risk-averse stochastic optimization with Wasserstein metric," *Operations Research Letters,* vol. 46, no. 2, pp. 262-267, 2018.

[17] Y. Liu, X. Yuan, S. Zeng, and J. Zhang, "Primal–dual hybrid gradient method for distributionally robust optimization problems," *Operations Research Letters,* vol. 45, no. 6, pp. 625-630, 2017/11/01/, 2017.

[18] B. Zeng, and L. Zhao, "Solving two-stage robust optimization problems using a column-and-constraint generation method," *Operations Research Letters,* vol. 41, no. 5, pp. 457-461, 2013/09/01/, 2013.

[19] B. He, and X. Yuan, "Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective," *SIAM Journal on Imaging Sciences,* vol. 5, no. 1, pp. 119-149, 2012.

[20] C. Ruiz, and A. J. Conejo, "Robust transmission expansion planning," *European Journal of Operational Research,* vol. 242, no. 2, pp. 390-401, 2015.

[21] A. Velloso, D. Pozo, and A. Street, "Distributionally Robust Transmission Expansion Planning: a Multi-scale Uncertainty Approach," *IEEE Transactions on Power Systems,* vol. 35, no. 5, pp. 3353-3365, 2020.

[22] C. Zhao, and Y. Guan, "Data-driven stochastic unit commitment for integrating wind generation," *IEEE Transactions on Power Systems,* vol. 31, no. 4, pp. 2587-2596, 2015.
[23] F. Alizadeh, and D. Goldfarb, "Second-order cone programming," *Mathematical Programming,* vol. 95, no. 1, pp. 3-51, 2003.
[24] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power generation, operation, and control*: John Wiley & Sons, 2013.
[25] "IEEE 118-Bus Test System," July, 2020; http://motor.ece.iit.edu/data/.
[26] "Open Power System Data," July 2020; https://open-power-system-data.org/data-sources.
[27] S. Dehghan, A. Kazemi, and N. Amjady, "Multi-objective robust transmission expansion planning using information-gap decision theory and augmented ε-constraint method," *IET Generation, Transmission, & Distribution,* vol. 8, no. 5, pp. 828-840, 2014.
[28] C. Coffrin, H. L. Hijazi, and P. Van Hentenryck, "Strengthening convex relaxations with bound tightening for power network optimization," *International Conference on Principles and Practice of Constraint Programming*, pp. 39-57, 2015.
[29] M. R. Ebrahimi, and N. Amjady, "Contingency-constrained operation optimization of microgrid with wind and solar generations: A decision-driven stochastic adaptive-robust approach," *IET Renewable Power Generation,* vol. 15, no. 2, pp. 326-341, Feb. 2021.
[30] M. Esmaili, M. Ghamsari-Yazdel, N. Amjady, and C. Chung, "Convex model for controlled islanding in transmission expansion planning to improve frequency stability," *IEEE Transactions on Power Systems,* vol. 36, no. 1, pp. 58-67, Jan. 2021.