Data-Driven Variable Polyhedral Uncertainty Set for Renewable Planning

Wenda Chen, Hongxing Ye, *Senior Member, IEEE*, Shuyi Zhang

***Abstract*—With the high penetration of renewables, the uncertainty of renewable generation becomes an important factor in power system operation and planning. In the planning model, renewable installed capacity is a decision variable, which makes the uncertainty set variable. This letter presents a novel tractable approach with a variable polyhedral uncertainty set (VPUS) to capture the temporal-spatial correlations of renewable generation. A modified surrogate affine policy (SAP) is introduced to solve the problem. Numerical tests are conducted on the planning problem of the simplified transmission system. The results indicate that the proposed approach can reduce investment and operational costs for flexibility.**

***Index Terms*—Variable polyhedral uncertainty set, renewable planning, surrogate affine policy, correlations**

I. Introduction

RENEWABLE energy such as photovoltaic (PV) has increased significantly in recent decades. However, the uncertainty of renewable generation poses a challenge to the operation and planning of power systems. Robust optimization (RO) and Distributionally RO have been applied in power system operations. When renewable installed capacity is treated as a variable, the uncertainty set becomes variable and makes the traditional affine policy (AP) computationally intractable. The work [1] proposes a surrogate affine policy (SAP) to address the variable box uncertainty set (VBUS), and authors in [2] further extend it to multistage problems.

The uncertainty set is a vital part of the RO. The early RO literature considers box uncertainty sets (BUS) [1], [3], where temporal–spatial correlations are poorly captured or neglected. Recent studies indicate that correlation-aware polyhedral uncertainty sets (PUS) can reduce conservativeness and outperform BUS, enabling higher utilization of energy storage (ES) and lower renewable curtailment [4], [5]. However, capturing uncertainty correlations with PUS typically leads to computational intractability in the existing literature. Bridging this gap via a scalable approach is appealing. We propose a data-driven variable polyhedral uncertainty set (VPUS) that captures temporal and spatial correlations in renewable generation, and we develop a VPUS-based surrogate-affine policy (SAP) to address the renewable planning problem.

II. Data-Driven Variable Uncertainty Set

To capture the temporal-spatial correlations, we propose a novel VPUS that can capture such correlations and a data-driven method to construct it.

W. Chen, H. Ye, and S. Zhang are with STED Inst., Xi'an Jiaotong University. This work is support by the National Science and Technology Major Project 2025ZD0805000. Corresponding Author: Hongxing Ye, Email: yehxing@xjtu.edu.cn

A. Mathematical Formulation

Let $\mathcal{N}^{\mathrm{r}}$ and $\mathcal{T}=\{0,1,...,T\}$ denote the sets of renewable bases and time periods, respectively, and $N^{\mathrm{r}}=|\mathcal{N}^{\mathrm{r}}|$. Let $\boldsymbol{P}^{\mathrm{f}}$, $\boldsymbol{\varepsilon}$, and $\hat{\boldsymbol{P}}$ denote the forecast power, the uncertainty, and the materialized power for the renewable bases, respectively. Then, we have

$$\hat{\boldsymbol{P}}=\boldsymbol{P}^{\mathrm{f}}+\boldsymbol{\varepsilon}, \tag{1}$$

where $\hat{\boldsymbol{P}},\boldsymbol{P}^{\mathrm{f}},\boldsymbol{\varepsilon}\in\mathbb{R}^{D}$, and $D=N^{\mathrm{r}}\cdot(T+1)$. Let C_n denote the installed capacity of the nth renewable base. Let $\boldsymbol{I}_{(T+1)}\in\mathbb{R}^{(T+1)\times(T+1)}$ and $\otimes$ be the identity matrix and the kronecker product, respectively. By introducing

$$\boldsymbol{H}=\boldsymbol{I}_{(T+1)}\otimes\begin{bmatrix}C_1 & & \\ & \ddots & \\ & & C_{N^{\mathrm{r}}}\end{bmatrix},$$

we present the normalized uncertainty as $\boldsymbol{H}^{-1}\boldsymbol{\varepsilon}$, where $\boldsymbol{H}\in\mathbb{R}^{D\times D}$. If $C_n=0$, it can be removed from $\boldsymbol{H}$, making $\boldsymbol{H}$ invertible. A VPUS is then formulated as a function of $\boldsymbol{H}$

$$\mathcal{U}(\boldsymbol{H})=\{\boldsymbol{\varepsilon}\in\mathbb{R}^{D}|-\boldsymbol{u}^{\mathrm{low}}\le\boldsymbol{A}\boldsymbol{H}^{-1}\boldsymbol{\varepsilon}\le\boldsymbol{u}^{\mathrm{up}}\}, \tag{2}$$

where $\boldsymbol{u}^{\mathrm{low}},\boldsymbol{u}^{\mathrm{up}}$ and $\boldsymbol{A}$ are parameters attained from normalized uncertainty of historical data. Once $\boldsymbol{H}$ is determined, $\mathcal{U}(\boldsymbol{H})$ becomes a constant polytope. The proposed VPUS $\mathcal{U}(\boldsymbol{H})$ has two advantages: (i) it captures temporal–spatial correlations and is therefore less conservative; (ii) it varies with the renewable installed capacities $\boldsymbol{H}$ and can be incorporated directly into planning models where $\boldsymbol{H}$ is a decision variable.

B. Data-Driven Polytope Generation

Now, we present a data-driven method to attain $\boldsymbol{A}$, $\boldsymbol{u}^{\mathrm{low}}$ and $\boldsymbol{u}^{\mathrm{up}}$ as presented in Algorithm 1, generating the polytope $\mathcal{U}(\boldsymbol{H})$. We assume that the normalized uncertainty at time t can be represented by uncertainty from $t-L(L\ge1)$ to $t-1$ linearly with a certain error. For simplicity and without loss of generality, we set $L=1$. Let $\varepsilon_{n,t}$ denote the uncertainty for the nth renewable base at time t. $\varepsilon_{n,t}/C_n$ can be expressed as

$$\frac{\varepsilon_{n,t}}{C_n}=\sum_{n'=1}^{N^{\mathrm{r}}}k_{n,n',t}\frac{\varepsilon_{n',t-1}}{C_{n'}}+w_{n,t},\quad \forall n\in\mathcal{N}^{\mathrm{r}},t=1,\ldots,T \tag{3}$$

where $k_{n,n',t}$ denotes the factor of $\varepsilon_{n',t-1}/C_{n'}$ to $\varepsilon_{n,t}/C_n$, and $w_{n,t}$ denotes the error. $k_{n,n',t}$ can be easily estimated by multiple linear regression. Let $r^{\mathrm{up}}_{n,t}\ge0$ denote the maximum residual of the linear regression, and $r_{n,t}^{\text{low}} \geq 0$ denote the absolute value of the minimum residual. The correlation constraints corresponding to Equation (3) can be formulated as follows:

$$-\lambda \cdot r_{n,t}^{\text{low}} \leq w_{n,t} \leq \lambda \cdot r_{n,t}^{\text{up}},\ \forall n \in \mathcal{N}^{\text{r}}, t = 1, \ldots, T \tag{4}$$

where parameter $\lambda \in [0, 1]$ can be adjusted to control the size of the VPUS. A smaller λ means fewer uncertainty points are covered in VPUS, and all historical data is included when $\lambda = 1$. Then, we can recast (4) in the form of (2) by defining $\boldsymbol{A}_t \in \mathbb{R}^{N^{\text{r}} \times D}$, $\boldsymbol{u}_t^{\text{low}}, \boldsymbol{u}_t^{\text{up}} \in \mathbb{R}^{N^{\text{r}}}$ according to (5), (6) and (7), respectively.

$$\boldsymbol{u}_t^{\text{low}} = \lambda \left[r_{1,t}^{\text{low}}, \ldots, r_{N^{\text{r}},t}^{\text{low}} \right], \forall t = 1, \ldots, T \tag{6}$$

$$\boldsymbol{u}_t^{\text{up}} = \lambda \left[r_{1,t}^{\text{up}}, \ldots, r_{N^{\text{r}},t}^{\text{up}} \right], \forall t = 1, \ldots, T \tag{7}$$

Next, we have

$$\boldsymbol{A} = \left[\boldsymbol{A}_1^{\top}, \ldots, \boldsymbol{A}_T^{\top} \right]^{\top}, \boldsymbol{A} \in \mathbb{R}^{N^{\text{r}}T \times D} \tag{8}$$

$$\boldsymbol{u}^{\text{low}} = \left[\boldsymbol{u}_1^{\text{low}}, \ldots, \boldsymbol{u}_T^{\text{low}} \right]^{\top}, \boldsymbol{u}^{\text{low}} \in \mathbb{R}^{N^{\text{r}}T} \tag{9}$$

$$\boldsymbol{u}^{\text{up}} = \left[\boldsymbol{u}_1^{\text{up}}, \ldots, \boldsymbol{u}_T^{\text{up}} \right]^{\top}, \boldsymbol{u}^{\text{up}} \in \mathbb{R}^{N^{\text{r}}T} \tag{10}$$

It is worth mentioning that traditional box-bound limitations can also be considered in $\mathcal{U}(\boldsymbol{H})$. Hence, we have $\boldsymbol{A} \in \mathbb{R}^{h \times D}$ and $\boldsymbol{u}^{\text{low}}, \boldsymbol{u}^{\text{up}} \in \mathbb{R}^{h}$, where $h = N^{\text{r}}T + D$.

For illustrative purpose, we consider a 2-dimensional uncertainty of the nth renewable base at $t-1$ and t in Fig. 1. The green bound and blue bound in Fig. 1 denote correlations' constraints and box-bound limitations, respectively. The VPUS $\mathcal{U}(\boldsymbol{H})$ we proposed is the yellow area in Fig. 1 which can respect the correlations' constraints and box-bound limitations simultaneously.

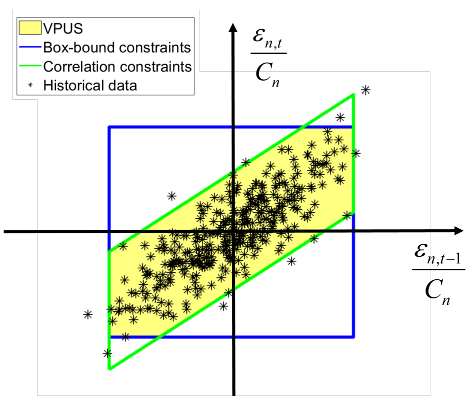

Fig. 1: Illustration of 2-dimensional variable polyhedral uncertainty set (VPUS).

Algorithm 1: Data-Driven Polytope Generation

Get historical data for N^{r} renewable bases;
for $t \in \{1, \ldots, T\}$ **do**
  for $n \in \mathcal{N}^r$ **do**
    Estimate $k_{n,n',t}$ in (3) by multiple linear regression (using historical data), then get $r_{n,t}^{\text{low}}, r_{n,t}^{\text{up}}$ in (4);
  end
  Formulate $\boldsymbol{A}_t, \boldsymbol{u}_t^{\text{low}}, \boldsymbol{u}_t^{\text{up}}$ via (5), (6) and (7), respectively;
end
Formulate $\boldsymbol{A}, \boldsymbol{u}^{\text{low}}, \boldsymbol{u}^{\text{up}}$ via (8), (9) and (10), respectively;
Reformulate $\boldsymbol{A}, \boldsymbol{u}^{\text{low}}, \boldsymbol{u}^{\text{up}}$ with box-bound constraints;
return $\boldsymbol{A}, \boldsymbol{u}^{\text{low}}, \boldsymbol{u}^{\text{up}}$ for VPUS $\mathcal{U}(\boldsymbol{H})$;

III. Renewable Planning Model With Variable Polyhedral Uncertainty Set

In this section, we first introduce the compact form of the VPUS-based renewable planning model in power systems. A general VPUS-based SAP is then proposed to solve the problem. As a side note, it is applicable in the multistage and distributionally robust literature.

A. Planning Model

Without loss of generality, the renewable planning model in power systems with VPUS $\mathcal{U}(\boldsymbol{H})$ can be rewritten in a compact form

$$\text{(P)} \quad \min_{\boldsymbol{x}, \boldsymbol{H}} \quad \boldsymbol{c}^{\top}\boldsymbol{x} + \boldsymbol{f}^{\top} \cdot \text{diag}(\boldsymbol{H}) \tag{11a}$$

$$\text{s.t.} \quad \boldsymbol{M}\boldsymbol{x} + \boldsymbol{N} \cdot \text{diag}(\boldsymbol{H}) \leq \boldsymbol{b} \tag{11b}$$

$$\boldsymbol{B}\boldsymbol{x} + \boldsymbol{F} \cdot \text{diag}(\boldsymbol{H}) + \boldsymbol{C}\boldsymbol{y}(\boldsymbol{\varepsilon}) + \boldsymbol{E}\boldsymbol{\varepsilon} \leq \boldsymbol{d}, \quad \forall \boldsymbol{\varepsilon} \in \mathcal{U}(\boldsymbol{H}) \tag{11c}$$

where $\boldsymbol{x}$ represents the planning decision of flexible resources and operational decision variable. $\boldsymbol{H}$ is the same as that in Equation (2), denoting the installed capacities of renewable bases, and diag$(\cdot)$ represents the vector formed by the diagonal elements of the matrix. The objective (11a) is to minimize the total cost, including both investment and operation costs. $\boldsymbol{f}^{\top} \cdot \text{diag}(\boldsymbol{H})$ denotes the investment cost of renewables. The investment cost of flexible resources and operational cost are modeled in $\boldsymbol{c}^{\top}\boldsymbol{x}$. $\boldsymbol{y}(\boldsymbol{\varepsilon})$ is a function of $\boldsymbol{\varepsilon}$ that maps uncertainty to re-dispatch decisions.

The traditional affine policy is intractable for solving problem (P), as nonlinear terms are introduced by the VPUS $\mathcal{U}(\boldsymbol{H})$. When uncertainty correlation is ignored, SAP [1] is proposed

$$\boldsymbol{A}_t = \begin{bmatrix} \boldsymbol{0}_{N^{\text{r}} \times (t-1)N^{\text{r}}} & \begin{matrix} -k_{1,1,t} & \ldots & -k_{1,N^{\text{r}},t} \\ \vdots & \ddots & \vdots \\ -k_{N^{\text{r}},1,t} & \ldots & -k_{N^{\text{r}},N^{\text{r}},t} \end{matrix} & \boldsymbol{I}_{N^{\text{r}}} & \boldsymbol{0}_{N^{\text{r}} \times (T-t)N^{\text{r}}} \end{bmatrix}, \forall t = 1, \ldots, T \tag{5}$$

TABLE I: Renewable Planning Results.

	λ (p.u.)	Cost ($\times 10^7$\$)	Fuel ($\times 10^7$\$)	PV (MW)	Wind (MW)	ES (MWh)	Renewable Ratio (%)	Infeasibility Ratio (%)
Case 1: VBUS	NA	2.8353	2.4291	1,410	1,739	5,344	8.25	0
Case 2: Stochastic	NA	2.4427	2.0036	3,845	5,456	191	24.32	12.4
Case 3:VPUS	1.0	2.7244	2.3687	3,101	860	4,131	10.53	0
Case 3:VPUS	0.9	2.5635	2.2066	1,394	5,035	685	16.65	0
Case 3:VPUS	0.7	2.4802	2.2471	4,396	1,294	260	15.12	0.4
Case 3:VPUS	0.5	2.4246	2.0044	4,863	4,381	301	24.25	8.8

to address computational challenges. Nevertheless, it does not apply once the uncertainty correlation is considered in VPUS. Next, a new SAP method based on the VPUS like Equation (2) will be proposed to solve the problem (P).

B. Modified Surrogate Affine Policy

First, we introduce the surrogate variable $\boldsymbol{\eta} \in \mathbb{R}^D$ and the surrogate uncertainty set

$$\tilde{\mathcal{U}} = \{\boldsymbol{\eta} \in \mathbb{R}^D | -\boldsymbol{u}^{\text{low}} \leq \boldsymbol{A}\boldsymbol{\eta} \leq \boldsymbol{u}^{\text{up}}\}, \tag{12}$$

where $\boldsymbol{A}, \boldsymbol{u}^{\text{low}}, \boldsymbol{u}^{\text{up}}$ in Equation (12) are the same as them in Equation (2). The surrogate uncertainty set $\tilde{\mathcal{U}}$ is constant. Let $\hat{\boldsymbol{G}}$ denote the surrogate affine policy, then the uncertainty and re-dispatch policy can be recast as

$$\boldsymbol{\varepsilon} = \boldsymbol{H}\boldsymbol{\eta}, \hat{\boldsymbol{y}}(\boldsymbol{\eta}) = \hat{\boldsymbol{G}}\boldsymbol{\eta}. \tag{13}$$

Next, we substitute Equation (13) into Equation (11c). Then, Equation (11c) is recast as

$$\boldsymbol{B}\boldsymbol{x} + \boldsymbol{F} \cdot \text{diag}(\boldsymbol{H}) + \boldsymbol{C}\hat{\boldsymbol{G}}\boldsymbol{\eta} + \boldsymbol{E}\boldsymbol{H}\boldsymbol{\eta} \leq \boldsymbol{d}, \quad \forall \boldsymbol{\eta} \in \tilde{\mathcal{U}}. \tag{14}$$

The jth row of Equation (14) can be written as

$$d_j \geq \max_{\boldsymbol{\eta} \in \tilde{\mathcal{U}}} \quad (\boldsymbol{C}\boldsymbol{G} + \boldsymbol{E}\boldsymbol{H})_j \boldsymbol{\eta} + \boldsymbol{B}_j \boldsymbol{x} + \boldsymbol{F}_j \cdot \text{diag}(\boldsymbol{H}), \quad \forall j \in \mathcal{J} \tag{15}$$

where $(\cdot)_j$ denotes the jth row of the matrix/vector. $\mathcal{J}$ is the set of rows in Equation (14). The maximization problem (15) is feasible and bounded, as the surrogate uncertainty set $\tilde{\mathcal{U}}$ is a polytope. Thus, the strong duality of (15) holds according to the weak form of Slater's condition. Following the strong duality, Equation (15) is recast as

$$\begin{cases} \boldsymbol{C}\hat{\boldsymbol{G}} + \boldsymbol{E}\boldsymbol{H} + \boldsymbol{\pi}^{\text{low}}\boldsymbol{A} - \boldsymbol{\pi}^{\text{up}}\boldsymbol{A} = \boldsymbol{0} & \text{(16a)} \\ \boldsymbol{B}\boldsymbol{x} + \boldsymbol{F} \cdot \text{diag}(\boldsymbol{H}) - \boldsymbol{d} + \boldsymbol{\pi}^{\text{low}}\boldsymbol{u}^{\text{low}} + \boldsymbol{\pi}^{\text{up}}\boldsymbol{u}^{\text{up}} \leq \boldsymbol{0} & \text{(16b)} \\ \boldsymbol{\pi}^{\text{low}}, \boldsymbol{\pi}^{\text{up}} \geq \boldsymbol{0} & \text{(16c)} \end{cases}$$

where $\boldsymbol{x}, \hat{\boldsymbol{G}}, \boldsymbol{\pi}^{\text{low}}, \boldsymbol{\pi}^{\text{up}}, \boldsymbol{H}$ are variables. Finally, the model (P) with the VPUS $\mathcal{U}(\boldsymbol{H})$ is recast as

$$\text{(SAP-P)} \quad \min_{\boldsymbol{x}, \boldsymbol{H}, \hat{\boldsymbol{G}}} \quad \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{f}^\top \cdot \text{diag}(\boldsymbol{H})$$
$$\text{s.t.} \quad \text{(11b)}, \text{(16a)} - \text{(16c)}.$$

(SAP-P) is a mixed-integer linear programming (MILP) or linear programming problem. If the planning model involves integer variables, such as the storage deployment indicator, then it is an MILP problem. It can be solved efficiently using off-the-shelf solvers.

After optimal $\hat{\boldsymbol{G}}^\star$ and $\boldsymbol{H}^\star$ are attained, the re-dispatch can be expressed as

$$\hat{\boldsymbol{y}}^\star(\boldsymbol{\eta}) = \hat{\boldsymbol{G}}^\star \boldsymbol{\eta} = \hat{\boldsymbol{G}}^\star (\boldsymbol{H}^\star)^{-1} \boldsymbol{\varepsilon}. \tag{17}$$

It should be emphasized that the optimization model (SAP-P) does not include any $\boldsymbol{H}^{-1}$ term. Therefore, $C_n = 0$ remains an option for renewable planning. The solution $\boldsymbol{H}^*$ obtained from (SAP-P) can thus be non-invertible. In this case, items associated with $\{n | C_n = 0\}$ can be removed in $\boldsymbol{H}^*$ for (17).

IV. Case Study

Case studies are carried out with a simplified transmission system. There is one PV farm, one wind farm, an ES system, an aggregated unit and a load point in the system, and transmission constraints are not considered. Renewable generation and load data are obtained from real historical data. The installed capacities of the aggregated unit is 8.5 GW. The investment cost of ES, PV farm, and wind farm are set as \$123/MWh, \$85/MW and \$137/MW, respectively, which have been amortized to the daily cost. The fuel cost is set to \$40/MWh. As we consider the seasonal characteristics of renewable generation, a year is represented by four typical seasonal days. Thus, we construct a VPUS for each season individually using corresponding seasonal data (90 days × 24 hours each season).

The following cases are performed to validate the effectiveness of the proposed VPUS-based model:

- Case 1: Affine policy-based approach with VBUS.
- Case 2: Stochastic programming-based approach.
- Case 3: Affine policy-based approach with VPUS.

The objective of the planning model is to minimize the investment and operational costs. The decision variables include the installed capacities of PV, wind, and ES. In Case 2, 50 representative scenarios are generated in the stochastic approaches, with correlations captured by the covariance matrix derived from historical data. In Case 3, λ is the parameter controlling the size of the uncertainty set. A smaller λ indicates that the polytope covers a smaller uncertainty region.

The planning results are shown in Table I. The columns "Cost" and "Fuel" represent the total and fuel costs, respectively. "PV", "Wind", and "ES" denote their installed capacities. "Renewable Ratio" denotes the proportion of renewable generation in total electricity generation. 1,000 scenarios are randomly generated to assess the feasibility of planning results from various methods. "Infeasibility Ratio" refers to the percentage of scenarios in which renewable curtailment occurs.

When $\lambda = 1$, Case 3 reduces the total cost by 3.9% (3.9% $\approx$ (2.8353-2.7244)/2.8353) compared to Case 1. It has a higher renewable ratio and reduced ES. This shows the VPUS-based model leads to more economical planning results, as the temporal-spatial correlations mitigate the impact of uncertainty. Thus, more renewable can be accommodated, and unnecessary ES or thermal power are reduced.

It is also observed that affine-based approaches (Cases 1 and 3) do not have infeasible scenarios when $\lambda = 1$. In contrast, Case 2 (stochastic approaches) has 12.4% infeasible scenarios and a lower total cost. This is because stochastic programming based approaches cannot include all possible scenarios in the model. Furthermore, λ can be adjusted to control the VPUS size, balancing conservativeness and robustness. We observe that the total cost of Case 3 ($\lambda = 0.5$) is slightly lower (i.e., 0.74% $\approx$ (2.4427-2.4246)/2.4427) than Case 2. Meanwhile, Case 3 ($\lambda = 0.5$) has a lower infeasibility ratio (8.8% vs. 12.4%), indicating better planning results. This is mainly because the PDF information of Case 2 is not perfect, and extreme scenarios are seldom considered. This results in Case 2 and Case 3 ($\lambda = 0.5$) achieving a similar renewable ratio and fuel cost, but 36.5% (36.5% $\approx$ (301-191)/301) more ES against extreme scenarios is deployed in Case 3 ($\lambda = 0.5$). Thus, Case 3 ($\lambda = 0.5$) has a 3.6% lower ratio of infeasible scenarios than Case 2. The results provide two insights. Firstly, the uncertainty set is crucial to the planning results. Secondly, the proposed VPUS can achieve a good tradeoff between economics (or conservativeness) and robustness. It can outperform stochastic approaches in both cost and robustness when λ is set properly.

References

[1] H. Ye, "Surrogate affine approximation based co-optimization of transactive flexibility, uncertainty, and energy," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4084–4096, 2018.

[2] L. Zhang, H. Ye, Y. Ge, and Z. Li, "Ramping-based variable-timescale co-optimization for distribution planning and operation," *IEEE Trans. Power Syst.*, vol. 40, no. 3, pp. 2519–2531, 2024.

[3] R. A. Jabr, "Adjustable robust OPF with renewable energy sources," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4742–4751, 2013.

[4] A. Lorca and X. A. Sun, "Multistage robust unit commitment with dynamic uncertainty sets and energy storage," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 1678–1688, 2016.

[5] B. Zhou, J. Fang, X. Ai, Y. Zhang, W. Yao, Z. Chen, and J. Wen, "Partial-dimensional correlation-aided convex-hull uncertainty set for robust unit commitment," *IEEE Trans. Power Syst.*, vol. 38, no. 3, pp. 2434–2446, 2023.