Joint Chance-constrained Unit Commitment: Statistically Feasible Robust Optimization with Learning-to-Optimize Acceleration

Jinhao Liang, Wenqian Jiang, Chenbei Lu *and* Chenye Wu

***Abstract*—Renewable energy penetration increases the power grid's operational uncertainty, threatening the economic effectiveness and reliability of the grid. In this paper, we examine how uncertainty affects unit commitment (UC), a classical electricity market procedure. Stochastic programming has helped handle uncertainty for UC and performed well with distribution knowledge, but the lack of such information in practice deteriorates the effectiveness. Such a dilemma becomes more pronounced when dealing with joint chance constraints solely based on samples. We introduce statistical feasibility into UC and develop robust sample-based algorithms employing appropriate uncertainty sets to hedge uncertainty without distribution dependence. We also propose a learn-to-optimize acceleration method to convexify UC. Furthermore, we construct an optimization kernel to boost computational efficiency.**

***Index Terms*—Unit Commitment, Chance-constrained Programming, Robust Optimization, Statistical Feasibility**

I. Introduction

Unit commitment (UC), a classical procedure commonly performed by Independent System Operators (ISOs), determines a generator dispatch schedule that minimizes system costs of meeting forecast demand, considering various constraints associated with generating resources and system reliability [1]. And now, UC faces a new challenge from renewable energy integration. The inherent variability and intermittency of renewable energy sources result in large forecasting errors [2], introducing significant uncertainty to the UC and complicating the solution process.

Classical methods for addressing uncertainty in UC include chance-constrained programming (CCP) and robust optimization (RO) [3]. CCP leverages distribution information to model uncertainty [4]. However, such information may be hard to obtain in practice. Moreover, the UC typically involves complicated joint chance constraints (CCs), which deter distribution estimation. The biased distribution information may lead to unreliable solutions [5]. Unlike CCP, RO employs uncertainty sets to capture uncertainty, only requiring the observed samples [6]. While RO guarantees the best performance in the worst case, it can be overly conservative. Distributionally robust optimization (DRO) addresses the limitations observed in RO and CCP, while it still requires distribution information and may not effectively handle joint CCs [6].

Furthermore, UC is an NP-hard mixed-integer program (MIP) [3], resulting in high computational costs. Meanwhile, ISOs need to perform UC multiple times within the strict clearing time requirements in the daily markets for favorable solutions [7]. Thus, ensuring the computational efficiency of UC becomes a critical challenge.

To overcome the above issues, we propose statistically feasible robust UC algorithms for favorable solutions. We further combine machine learning (ML) and optimization techniques for acceleration.

A. Related Works

Stochastic optimization has enabled power system operation. In [8], Wang *et al.* formulate UC as a CCP to maximize wind power utilization with high probability. In [9], Sundar *et al.* propose a chance-constrained $N-1$ secure UC to handle wind fluctuations and component outages. In [10], Li *et al.* examine the robust operation of a hybrid AC/DC microgrid using a two-stage RO method. In [11], Hou *et al.* utilize data-driven RO to address wind and photovoltaic power uncertainties in UC. Gu *et al.* integrate CCP and RO for emission-aware economic dispatch in [12].

DRO further improves the operational effectiveness. Xiong *et al.* propose a DRO model to address UC with volatile wind power generation in [13]. Duan *et al.* put forward a data-driven DRO method for UC in [14]. More recently, distributionally robust UC has adopted statistical distances, such as Kullback-Leibler divergence [15], [16], and Wasserstein-metric [17], [18]. To improve tractability, Zhou *et al.* design an ambiguity set using unimodality, mode skewness, and second-order moment information in [19]. However, the literature on UC with DRO rarely addresses joint CCs [20]. Few studies utilize a conservative approximation to obtain solutions for joint CCs in DRO [21]–[23]. Other studies convexify the joint chance-constrained problem, whose effectiveness typically depends on relatively strong assumptions [24], [25]. In contrast, our proposed method can effectively solve the joint CCs by transforming the original problem into equivalent robust constraints. Furthermore, our approach solely relies on observed samples, unlike DRO and CCP. Table I compares our approach with its rivals, highlighting our advantages.

We are not the first to consider dealing joint CCs with samples. Our work is inspired by [28], which integrates data into RO for high-quality solutions with theoretical guarantee. We further the research line by introducing statistical feasibility into UC, an NP-hard problem, which is more challenging than linear programming (LP) considered in closely related works

J. Liang, W. Jiang and C. Wu are with School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen. C. Lu is with the Institute for Interdisciplinary Information Sciences, Tsinghua University.

TABLE I: Stochastic Optimization Approach Comparison.

Category	Related Work	DAF[a]	VRG[b]	JCCs[c]	SFG[d]
CCP	[8], [9]	✗	✗	UA[e]	✗
RO	[1], [2], [11]	✓	✓	✓	✗
DRO	[13], [15], [18]	✓	✓	UA	✗
SG	[26], [27]	✓	✗	✓	✗
RSRO	This Work	✓	✓	✓	✓

[a] DAF: Distribution assumption-free;
[b] VRG: Violation rate guarantee; [c] JCCs: Handling the joint CCs;
[d] SFG: Statistical feasibility guarantee; [e] UA: Under assumptions.

TABLE II: UC Acceleration Approach Comparison.

Category	Related Work	LCS[a]	FG[b]	BV[c]	LC[d]	OS[e]
RL	[39]–[41]	✗	✓	✗	✗	✗
Supervised Learning	[31]	✗	✗	✗	✗	✗
	[32]–[35]	✓	✓	✗	✗	✗
	[7], [36]–[38]	✓	✓	✓	✗	✗
	This work	✓	✓	✓	✓	✓

[a] LSS: Large Scale System; [b] FG: Feasibility Guarantee;
[c] BV: Binary Variables; [d] LC: Load Characteristic;
[e] OS: Optimization Structure.

[5], [6]. First, the high-dimensional uncertainty in transmission line constraints and reserve requirements needs to be tackled. Additionally, binary variables lead to non-convexities in UC, complicating the uncertainty set design. By overcoming these obstacles, our study advances statistical feasibility in power system operations. Furthermore, to overcome the computational burden caused by binary variables [3], we use ML and optimization to enhance UC's computational performance, addressing an issue not encountered in [5], [6].

Optimization and ML are two major tools to boost computational efficiency of power system operations. Zhai *et al.* identify over 85% redundant constraints analytically for fast computation in [29]. In [30], Lu *et al.* customize an optimization kernel for economic dispatch to speed up the solving process with specific problem properties efficiently. Learn-to-optimize (L2O) literature is increasing on designing learning models for conventional optimization problems. The initial idea of L2O for UC is to replace MIP methods with ML models. Yang *et al.* establish a mapping from loads to outputs of units based on an expert system in [31]. Nonetheless, it is challenging to learn both the unit status and generation level with feasibility simultaneously. Thus, another strategy uses ML to screen redundant constraints based on historical data to simplify UC, e.g., [32]–[35]. However, these works still need to solve MIP, which is the main hurdle in solving UC. To this end, Ramesh *et al.* identify the unit status with feasibility in [7]. In [36], Wu *et al.* offer a neural network-based SCUC algorithm to obtain the unit status. Sang *et al.* introduce a provably robust L2O approach for obtaining the convex SCUC in [37]. An online learning method to fix the unit status is proposed in [38]. Furthermore, reinforcement learning algorithms are recommended for UC [39], [40], but their implementations in large-scale UC are quite challenging. To the best of our knowledge, we are the first to embed load characteristics into ML models, enabling the investigation of the relationship between load and unit status. Furthermore, we customize an optimization kernel for UC to obtain the solutions without repeatedly solving the optimization problem to achieve remarkable computational performance. Table II compares our approach with existing methods, highlighting our core features.

B. Our Contributions

Towards investigating how to solve joint CCs in UC solely based on observed samples with performance guarantee, our major contributions are as follows:

- *Statistical Feasibility Customized for UC*: We extend statistical feasibility into MIP and formulate the statistically feasible UC. To effectively solve UC, we develop a sample-based uncertainty set construction algorithm and reconstruct it considering the specific form of constraints, yielding less conservative solutions.
- *Clustering-based L2O Approach Customized for UC*: Binary variables lead to high computational cost in solving UC. Hence, we consider the load characteristics to aid the unit status identification from an L2O perspective. By identifying all unit status, we transform the MIP-based UC (MIP UC) into the LP-based UC (LP UC), which effectively relieves the computational burden.
- *Optimization Kernel Design for Real-World Applications*: Motivated by the properties of LP UC, we exploit the optimization problem structure and customize an optimization kernel for UC to boost its computational efficiency.

II. System Model

In this section, we first revisit the deterministic UC and propose a stochastic UC. Then, we analyze the shortcomings of classical UC and incorporate ramping services for remedy.

A. Deterministic UC

Consider a bulk grid with demands $\mathcal{M}$, generators $\mathcal{N}$, and renewable energy generators $\mathcal{W}$. Under a sequence of constraints, the ISO calculates the unit status and generation levels of all units across T time slots to minimize system costs:

$$\min \ \sum\nolimits_{t=1}^{T}\Big(\sum\nolimits_{i\in\mathcal{N}}\left(C_i\left(g_i^t\right)+S_i^t\right)\Big)$$

$$\text{s.t. } G^t=\sum\nolimits_{i\in\mathcal{N}}g_i^t,\ W^t=\sum\nolimits_{i\in\mathcal{W}}w_i^t,\ \forall t, \tag{1a}$$

$$D^t=\sum\nolimits_{i\in\mathcal{M}}d_i^t,\ G^t+W^t=D^t,\ \forall t, \tag{1b}$$

$$u_i^t g_i^{\min}\le g_i^t\le u_i^t g_i^{\max},\ \forall i\in\mathcal{N},\forall t, \tag{1c}$$

$$g_i^t-g_i^{t-1}\le(u_i^t-u_i^{t-1})\mathrm{SU}^i+u_i^{t-1}\mathrm{RU}^i+(1-u_i^t)g_i^{\max},\ \forall i\in\mathcal{N},\forall t, \tag{1d}$$

$$g_i^{t-1}-g_i^t\le(u_i^{t-1}-u_i^t)\mathrm{SD}^i+u_i^t\mathrm{RD}^i+(1-u_i^{t-1})g_i^{\max},\ \forall i\in\mathcal{N},\forall t, \tag{1e}$$

$$u_i^t-u_i^{t-1}\le u_i^\tau,\ \forall i\in\mathcal{N},\forall t,\ \tau\in[t+1,\min\{t+T_i^{\mathrm{on}}-1,T\}], \tag{1f}$$

$$u_i^{t-1}-u_i^t\le 1-u_i^\tau,\ \forall i\in\mathcal{N},\forall t,\ \tau\in[t+1,\min\{t+T_i^{\mathrm{off}}-1,T\}], \tag{1g}$$

$$0\le \mathrm{SR}_i^t\le\min\{u_i^t g_i^{\max}-g_i^t,\ (u_i^t-u_i^{t-1})\mathrm{SU}^i$$

$$+ u_i^{t-1}\mathrm{RU}^i\},\ \forall i \in \mathcal{N}, \forall t, \tag{1h}$$

$$(G^t + W^t + \sum\nolimits_{i\in\mathcal{N}} \mathrm{SR}_i^t) - D^t \geq R^t,\ \forall t, \tag{1i}$$

$$\boldsymbol{f}^t = \boldsymbol{H}_g(\boldsymbol{g}^t + \boldsymbol{w}^t) - \boldsymbol{H}_d\boldsymbol{d}^t, \quad \forall t, \tag{1j}$$

$$-\bar{f}_i \leq f_i^t \leq \bar{f}_i,\ \forall i \in \mathcal{A}, \forall t, \tag{1k}$$

$$S_i^t \geq o_i \cdot (u_i^t - u_i^{t-1}),\ S_i^t \geq 0,\ \forall i \in \mathcal{N}, \forall t, \tag{1l}$$

$$u_i^t \in \{0,1\}, \quad \forall i \in \mathcal{N}, \forall t, \tag{1m}$$

where g_i^t, u_i^t, SR_i^t and S_i^t are decision variables: g_i^t (with vector form $\boldsymbol{g}^t = [g_i^t, \forall i \in \mathcal{N}]$) denotes the generation of generator i at time t; u_i^t (with vector form $\boldsymbol{u}^t = [u_i^t, \forall i \in \mathcal{N}]$) characterizes unit status of generator i at time t ($u_i^t = 1$ indicates generator i is on at time t, and 0 otherwise); SR_i^t (with vector form $\mathbf{SR}^t = [\mathrm{SR}_i^t, \forall i \in \mathcal{N}]$) is the amount of spinning reserve of unit i at time t; S_i^t describes the startup cost of unit i at time t. Other parameters are defined as follows: G^t is the total generation at time t; w_i^t (with the vector form $\boldsymbol{w}^t = [w_i^t, \forall i \in \mathcal{M}]$) denotes the renewable energy generator i at time t; W^t indicates the renewable energy generation at time t (predicted value is used for day-ahead clearing); d_i^t (with the vector form $\boldsymbol{d}^t = [d_i^t, \forall i \in \mathcal{M}]$) is demand i at time t; D^t characterizes the total net demand at time t (predicted value is used for day-ahead clearing); $g_i^{\min}$, $g_i^{\max}$ are minimal and maximal output limits of unit i; RU^i, RD^i denote ramp up and down limits of generator i; SU^i, SD^i indicate startup and shutdown capability of unit i; T_i^{on} (T_i^{off}) is minimum up (down) time of unit i; R^t denotes reserve requirement at time t; $\boldsymbol{H}_g, \boldsymbol{H}_d$ are the shift factor matrix for generators and demands, respectively; f_i^t (with the vector form $\boldsymbol{f}^t = [f_i^t, \forall i \in \mathcal{A}]$) indicates power flow across transmission line i at time t; $\bar{f}_i$ (with the vector form $\bar{\boldsymbol{f}} = [\bar{f}_i, \forall i \in \mathcal{A}]$) characterizes the capacity of transmission line i; $C_i(\cdot)$ is the production cost function of generator i; and o_i denotes the startup cost of generator i.

Eqs. (1a) and (1b) enforce the power balance; constraint (1c) ensures the scheduled generation does not violate units' status and capacity; constraints (1d) and (1e) limit the ramping up and down rates for each generator [42]; constraints (1f) and (1g) guarantee the minimum up and down time periods; constraints (1h) and (1i) describe the system reserve requirements for each period [43]; constraints (1j) and (1k) describe the upper and lower limited capacities on transmission lines; Eq. (1l) formulates the startup cost for each generator; constraint (1m) describes the feasible region of u_i^t.

Remark: Problem (1) is deterministic as we utilize the predicted value in decision-making. However, the unpredictability of renewable energy generation [44], [45] requires the ISOs to respond to the renewable generation prediction error. Next, we characterize the uncertainty of renewable energy generation and formulate stochastic UC.

B. Stochastic UC

To characterize the uncertainty, we model the renewable energy generation w_i^t and demand d_i^t as random variables:

$$w_i^t = \bar{w}_i^t + \xi_{\mathrm{w},i}^t, \quad d_i^t = \bar{d}_i^t + \xi_{\mathrm{d},i}^t,\ \forall t, \tag{2}$$

where $\bar{w}_i^t$ and $\bar{d}_i^t$ denote the predicted renewable energy generation and predicted demand; $\xi_{\mathrm{w},i}^t$ and $\xi_{\mathrm{d},i}^t$ characterize prediction errors. This allows us to develop a mathematical formulation for joint chance-constrained UC as follows:

$$\min\ \sum\nolimits_{t=1}^{T} \Big(\sum\nolimits_{i\in\mathcal{N}} \left(C_i\left(g_i^t\right) + S_i^t\right) + J(M^t)\Big)$$

$$\text{s.t.} G^t = \sum\nolimits_{i\in\mathcal{N}} g_i^t,\ W^t = \sum\nolimits_{i\in\mathcal{W}} (\bar{w}_i^t + \xi_{\mathrm{w},i}^t), \tag{3a}$$

$$D^t = \sum\nolimits_{i\in\mathcal{M}} (\bar{d}_i^t + \xi_{\mathrm{d},i}^t), \quad \forall t, \tag{3b}$$

$$M^t = G^t + \sum\nolimits_{i\in\mathcal{W}} (\bar{w}_i^t + \xi_{\mathrm{w},i}^t) - \sum\nolimits_{i\in\mathcal{M}} (\bar{d}_i^t + \xi_{\mathrm{d},i}^t), \forall t, \tag{3c}$$

$$J(M^t) = \gamma_{\mathrm{w}}\mathbb{E}[(M^t)^+] + \gamma_{\mathrm{d}}\mathbb{E}[(-M^t)^+],\ \forall t, \tag{3d}$$

$$\boldsymbol{B}^t(\mathbf{SR}^t; \boldsymbol{\xi}_{\mathrm{w}}^t, \boldsymbol{\xi}_{\mathrm{d}}^t) = G^t + (\bar{W}^t + \boldsymbol{\xi}_{\mathrm{w}}^t) + \sum\nolimits_{i\in\mathcal{N}} \mathrm{SR}_i^t - (\bar{D}^t + \boldsymbol{\xi}_{\mathrm{d}}^t),\ \forall t, \tag{3e}$$

$$\mathbb{P}\left(\cap_{t=1}^{T}\{\boldsymbol{B}^t(\mathbf{SR}^t; \boldsymbol{\xi}_{\mathrm{w}}^t, \boldsymbol{\xi}_{\mathrm{d}}^t) \geq R^t\}\right) \geq 1-\rho,\ \forall t, \tag{3f}$$

$$\boldsymbol{f}^t(\boldsymbol{g}^t; \boldsymbol{\xi}_{\mathrm{w}}^t, \boldsymbol{\xi}_{\mathrm{d}}^t) = \boldsymbol{H}_g(\boldsymbol{g}^t + \bar{\boldsymbol{w}}^t + \boldsymbol{\xi}_{\mathrm{w}}^t) - \boldsymbol{H}_d(\bar{\boldsymbol{d}}^t + \boldsymbol{\xi}_{\mathrm{d}}^t), \forall t, \tag{3g}$$

$$\mathbb{P}\left(\cap_{t=1}^{T}\{f_i^t(\boldsymbol{g}^t; \boldsymbol{\xi}_{\mathrm{w}}^t, \boldsymbol{\xi}_{\mathrm{d}}^t) \leq \bar{f}_i\}\right) \geq 1-\rho, \forall i \in \mathcal{A}, \tag{3h}$$

$$\mathbb{P}\left(\cap_{t=1}^{T}\{f_i^t(\boldsymbol{g}^t; \boldsymbol{\xi}_{\mathrm{w}}^t, \boldsymbol{\xi}_{\mathrm{d}}^t) \geq -\bar{f}_i\}\right) \geq 1-\rho, \forall i \in \mathcal{A}, \tag{3i}$$

$$S_i^t \geq o_i \cdot (u_i^t - u_i^{t-1}),\ S_i^t \geq 0,\ \forall i \in \mathcal{N}, \forall t, \tag{3j}$$

$$u_i^t \in \{0,1\},\ \forall i \in \mathcal{N}, \forall t, \tag{3k}$$

(1c)- (1h),

where operator $(\cdot)^+ = \max(\cdot, 0)$.

The parameters in Problem (3) are defined as follows: M^t characterizes generation-demand mismatch (positive yields load curtailment, negative suggests renewable generation curtailment); γ_{w}, γ_{d} denote unit cost for curtailing renewable generation and demand, respectively; $J(M^t)$ is the cost of curtailment for renewable and load; $\bar{W}^t$, $\bar{D}^t$ indicate predicted renewable generation and demand at time t; $\boldsymbol{\xi}_{\mathrm{w}}^t, \boldsymbol{\xi}_{\mathrm{d}}^t$ are vector forms of the prediction errors at time t (i.e., $\boldsymbol{\xi}_{\mathrm{w}}^t = [\xi_{\mathrm{w},i}^t, \forall i \in \mathcal{W}]$); $\bar{\boldsymbol{w}}^t$, $\bar{\boldsymbol{d}}^t$ are vector forms of the predicted renewable energy generation and demands (i.e., $\bar{\boldsymbol{w}}^t = [\bar{w}_i^t, \forall i \in \mathcal{W}]$); $1-\rho$ denotes stability requirement for CCs. Note that we use the notations $\boldsymbol{B}^t(\mathbf{SR}^t; \boldsymbol{\xi}_{\mathrm{w}}^t, \boldsymbol{\xi}_{\mathrm{d}}^t)$ and $\boldsymbol{f}^t(\boldsymbol{g}^t; \boldsymbol{\xi}_{\mathrm{w}}^t, \boldsymbol{\xi}_{\mathrm{d}}^t)$ to highlight that they are functions of decision and random variables.

Remark: Constraint (3f) denotes the system reserve requirements, allowing ISOs to mitigate uncertainty. We note that $\mathbf{SR}^t$ is the sole decision variable in constraint (3f). Meanwhile, $\mathbf{SR}^t$ is associated with binary variables $\boldsymbol{u}^t$ and other parameters in (1h) are predetermined. To simplify, we consider the maximum value of $\mathbf{SR}^t$ and assume the ramping up capability is enough. As a result, $\boldsymbol{u}^t$ dictates the probability $\mathbb{P}\left(\cap_{t=1}^{T}\{\boldsymbol{B}^t(\mathbf{SR}^t; \boldsymbol{\xi}_{\mathrm{w}}^t, \boldsymbol{\xi}_{\mathrm{d}}^t) \geq R^t\}\right)$. This probability will change discretely with $\boldsymbol{u}^t$, which means ISOs cannot continuously adjust the reserve in the system. Thus, ISOs may have to commit more units to satisfy the constraint, even if the gap with the expected reserve is small, which is unfavorable. A simple example illustrates this dilemma.

Example: In a 3-bus system shown in Fig. 1, bus 1 requests 30MW demand, with a local wind turbine W1. Generators G1 and G2 are located at buses 2 and 3, respectively. R^t is assigned to 3MW. We assume the same fuel costs for G1 and G2 with different capacities and startup costs. Specifically,

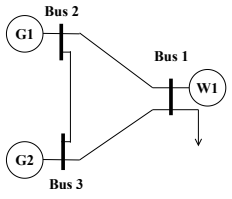


Fig. 1: Topology.

TABLE III: Results Illustration.

u_1	u_2	Probability	Fixed Cost ($)	Startup Cost ($)
0	0	-	-	-
0	1	0.05	200	400
1	0	0.65	100	200
1	1	0.99	300	600

the capacity of G1 is 25MW, with a fixed cost and startup cost of \$100 and \$200, respectively, while G2's capacity is 15MW, and its corresponding costs are twice as much as G1's. For simplicity, we consider the single time slot UC and the predicted renewable generation is 10MW. We assume the renewable generation prediction error follows Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, with the mean μ being 0 and standard deviation σ being 5MW. Table III shows the probability of satisfying (3f) when $\boldsymbol{u}^t$ is assigned to different values.

We notice that the probability of satisfying the system reserve requirements can only take three values in this system. ISOs must commit all generators to satisfy the constraint, which increases system costs. This motivates us to investigate how to find cost-effective ways to maintain reserve. The increasing penetration of renewable energy highlights the central role of system reserve in handling uncertainty in the power system. However, inflexible traditional reserves may cause economic losses, necessitating a more flexible reserve.

C. Incorporating ramping services to Aid UC

Ramping services, more flexible than traditional reserves, help meet reserve requirements in a cost-effective manner. We use the ramp capability product (RCP), a ramping service, to aid the UC. RCP was introduced by Midcontinent Independent System Operator to manage both expected variations and unexpected uncertainty in "net load" [46]. The distinct feature of RCP is that it co-optimizes energy and ancillary services in day-ahead operations [47]. UC with RCP can help ISOs balance costs and stability more easily. The new formulation incorporating RCP into UC is as follows:

$$\begin{aligned}
\min\ & \sum_{t=1}^{T}\Big(\sum\nolimits_{i\in\mathcal{N}}\left(C_i\left(g_i^t\right)+S_i^t\right)+J(M^t)+H(r^t)\Big)\\
\text{s.t. }& \boldsymbol{B}^t(\mathbf{SR}^t;\boldsymbol{\xi}_{\mathrm{w}}^t,\boldsymbol{\xi}_{\mathrm{d}}^t)=G^t+(\bar{W}^t+\boldsymbol{\xi}_{\mathrm{w}}^t)+\\
&\qquad\sum\nolimits_{i\in\mathcal{N}}\mathrm{SR}_i^t-(\bar{D}^t+\boldsymbol{\xi}_{\mathrm{d}}^t)+r^t,\ \forall t, &\text{(4a)}\\
&\mathbb{P}\left(\cap_{t=1}^{T}\{\boldsymbol{B}^t(\mathbf{SR}^t,r^t;\boldsymbol{\xi}_{\mathrm{w}}^t,\boldsymbol{\xi}_{\mathrm{d}}^t)\geq R^t\}\right)\geq 1-\rho,\forall t, &\text{(4b)}\\
&\text{(1c)- (1h), (3a)-(3d), (3g) - (3k),}
\end{aligned}$$

where r^t ($\boldsymbol{r}=[r^t,\forall t]$) decides the RCP bought at time t, and $H(\cdot)$ describes the cost of the RCP.

III. Statistically Feasible UC

In this section, we explain the definition of statistical feasibility and formulate a statistically feasible UC.

A. Statistical Feasibility

In practice, we usually cannot acquire the exact distribution of random variable $\boldsymbol{\xi}$. However, there are observed samples $\mathcal{D}_{\boldsymbol{\xi}}$ ($D_{\boldsymbol{\xi}}=\{\boldsymbol{\xi}_i,\forall i\in\mathcal{S}\}$, where $\boldsymbol{\xi}_i$ denotes one sample). Specifically, $\boldsymbol{\xi}_i$ is a vector of prediction error with dimension T.

Before formulating a statistically feasible UC, we first introduce the definition of statistical feasibility.

Definition 1 (Statistical Feasibility [28]): For a chance-constrained problem:

$$\min f(x),\quad \text{s.t. }\mathbb{P}(g(x;\theta)\leq b)\geq 1-\epsilon, \tag{5}$$

where $f(x)$ indicates the objective function, x is the decision variable, θ describes the random variable characterizing the uncertainty, $1-\epsilon$ denotes the stability requirement for satisfying the constraint. Given a dataset $\mathcal{D}_\theta$, an algorithm produces a solution $\hat{x}$. This algorithm enjoys a statistical feasibility guarantee if it can ensure its solution $\hat{x}$ is feasible for the constraint with a specific confidence level $1-\zeta$, i.e.,

$$\mathbb{P}_{\mathcal{D}_\theta}\left(\mathbb{P}_\theta\left(g\left(\hat{x}\left(D_\theta\right);\theta\right)\leq b\right)\geq 1-\epsilon\right)\geq 1-\zeta, \tag{6}$$

where $\hat{x}(\mathcal{D}_\theta)$ is the solution based on dataset $\mathcal{D}_\theta$.

We integrate the collected dataset into (5) to obtain (6), which means we adopt the notation $\mathbb{P}_{\mathcal{D}_\theta}$ to denote the probability taken over the dataset $\mathcal{D}_\theta$. Specifically, statistically feasible constraint guarantees any solution of (6) is feasible for (5) with confidence level $1-\zeta$ over $\mathcal{D}_\theta$, which eliminates the reliance on distribution information. Notably, although the form of the nested CC is similar to ours, its outer constraint still relies on distribution information, while we directly depend on the collected samples.

We customize statistical feasibility to UC:

$$\begin{aligned}
\min\ & \sum_{t=1}^{T}\Big(\sum\nolimits_{i\in\mathcal{N}}\left(C_i\left(g_i^t\right)+S_i^t\right)+J(M^t)+H(r^t)\Big)\\
\text{s.t.}\ & \mathbb{P}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}},\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{d}}}}\Big(\mathbb{P}_{\boldsymbol{\xi}_{\mathrm{w}},\boldsymbol{\xi}_{\mathrm{d}}}\big(\cap_{t=1}^{T}\{\boldsymbol{B}^t(\mathbf{SR}^t,r^t;\boldsymbol{\xi}_{\mathrm{w}}^t,\boldsymbol{\xi}_{\mathrm{d}}^t)\\
&\qquad\geq R^t\}\big)\geq 1-\rho\Big)\geq 1-\delta, &\text{(7a)}\\
&\mathbb{P}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}},\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{d}}}}\Big(\mathbb{P}_{\boldsymbol{\xi}_{\mathrm{w}},\boldsymbol{\xi}_{\mathrm{d}}}\left(\cap_{t=1}^{T}\{f_i^t(\boldsymbol{g}^t;\boldsymbol{\xi}_{\mathrm{w}}^t,\boldsymbol{\xi}_{\mathrm{d}}^t)\leq\bar{f}_i\}\right)\\
&\qquad\geq 1-\rho\Big)\geq 1-\delta,\ \forall i\in\mathcal{A}, &\text{(7b)}\\
&\mathbb{P}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}},\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{d}}}}\Big(\mathbb{P}_{\boldsymbol{\xi}_{\mathrm{w}},\boldsymbol{\xi}_{\mathrm{d}}}\left(\cap_{t=1}^{T}\{f_i^t(\boldsymbol{g}^t;\boldsymbol{\xi}_{\mathrm{w}}^t,\boldsymbol{\xi}_{\mathrm{d}}^t)\geq-\bar{f}_i\}\right)\\
&\qquad\geq 1-\rho\Big)\geq 1-\delta,\ \forall i\in\mathcal{A}, &\text{(7c)}\\
&\text{(1c)- (1h), (3a)-(3d), (3g) - (3k),(4a),}
\end{aligned}$$

where $1-\delta$ is the required confidence level for the inside CCs.

Remark: Problem (7) is also a chance-constrained UC with joint CCs, which is challenging to solve even with exact distribution information. However, RO can utilize uncertainty sets to characterize uncertainty, eliminating the reliance on distribution information and allowing us to obtain feasible solutions with only observed samples. This motivates us to solve the problem (7) from an RO perspective.

B. Statistically Feasible UC with RO Technique

The robust UC is formulated as follows:

$$\begin{aligned}
\min\ & \sum_{t=1}^{T}\Big(\sum\nolimits_{i\in\mathcal{N}}\left(C_i\left(g_i^t\right)+S_i^t\right)+J(M^t)+H(r^t)\Big)\\
\text{s.t. }& \cap_{t=1}^{T}\{\boldsymbol{B}^t(\mathbf{SR}^t,r^t;\boldsymbol{\xi}_{\mathrm{w}}^t,\boldsymbol{\xi}_{\mathrm{d}}^t)\geq R^t\},
\end{aligned}$$

$$\forall \boldsymbol{\xi}_{\mathrm{w}} \in \mathcal{U}^{r}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}},\ \forall \boldsymbol{\xi}_{\mathrm{d}} \in \mathcal{U}^{r}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{d}}}}, \tag{8a}$$
$$\cap_{t=1}^{T} \{ f_i^t(\boldsymbol{g}^t; \boldsymbol{\xi}_{\mathrm{w}}^t, \boldsymbol{\xi}_{\mathrm{d}}^t) \leq \bar{f}_i \},$$
$$\forall \boldsymbol{\xi}_{\mathrm{w}} \in \mathcal{U}^{\mathrm{ub},i}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}},\ \forall \boldsymbol{\xi}_{\mathrm{d}} \in \mathcal{U}^{\mathrm{ub},i}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{d}}}}, \tag{8b}$$
$$\cap_{t=1}^{T} \{ f_i^t(\boldsymbol{g}^t; \boldsymbol{\xi}_{\mathrm{w}}^t, \boldsymbol{\xi}_{\mathrm{d}}^t) \geq -\bar{f}_i \},$$
$$\forall \boldsymbol{\xi}_{\mathrm{w}} \in \mathcal{U}^{\mathrm{lb},i}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}},\ \forall \boldsymbol{\xi}_{\mathrm{d}} \in \mathcal{U}^{\mathrm{lb},i}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{d}}}}, \tag{8c}$$
(1c)- (1h), (3a)-(3d), (3g) - (3k),(4a),

where $\mathcal{U}^{r}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}}, \mathcal{U}^{\mathrm{ub},i}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}}, \mathcal{U}^{\mathrm{lb},i}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}}$ are the the uncertainty sets constructed based on dataset $\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}$; and $\mathcal{U}^{r}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{d}}}}, \mathcal{U}^{\mathrm{ub},i}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{d}}}}, \mathcal{U}^{\mathrm{lb},i}_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{d}}}}$ are the the uncertainty sets constructed based on dataset $\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{d}}}$;

RO seeks to build the smallest uncertainty set, which can contain all observed samples and ensure the solution satisfies the uncertainty set. However, RO typically becomes too conservative because uncertainty sets handle rare outliers the same as regular data points. The excessive pursuit of robustness can increase costs, even exceeding the cost of violating constraints. Consequently, our analysis minimizes costs and controls the violation level of constraints when designing uncertainty sets.

C. Bridging CCP and RO

To bridge the gap between CCP and RO, we construct a proper uncertainty set $\mathcal{U}$. For RO, solutions are feasible if the uncertain parameter θ is in $\mathcal{U}$. Hence, the stability requirement in (5) is $\mathbb{P}(\theta \in \mathcal{U}) \geq 1-\epsilon$. This indicates that the stability requirement is associated with the interval of uncertainty set. We can use observed samples to construct an uncertainty set that satisfies the stability requirements, equivalent to the chance-constrained UC.

Remark: After formulating the statistically feasible robust UC, we can solve problem (8) for a statistically feasible solution to problem (7). Theoretically, the solution could provide the statistical feasibility guarantee by the uncertainty set constructed based on samples instead of the random variable's actual distribution. However, two challenges remain: 1) constructing a proper uncertainty set that yields non-conservative solutions is challenging; 2) we need to overcome the uncertainty in the objective function.

D. Analysis for Statistically Feasible RO

We compare traditional optimization methodologies (e.g., CCP, DRO, and RO) with ours to illustrate the improvement of our proposed approach.

The CCP directly assumes the distribution type of θ, as shown in (5). According to the DRO, the distribution of θ belongs to an ambiguity set $\mathcal{P}$. The CCs should be satisfied for any distribution $P \in \mathcal{P}$. Mathematically:

$$\inf_{x,\theta} \left\{ f(x) \middle| \inf_{P \in \mathcal{P}} \mathbb{P}_{\theta} \left(g(x;\theta) \leq b \right) \geq 1-\epsilon \right\}. \tag{9}$$

DRO approaches differ in ambiguity set definitions and solution strategies [13]–[19]. While DRO relieves the reliance on random variable distribution compared to CCP, defining the ambiguity set still necessitates certain assumptions about distribution, such as single-peaked, sub-gaussian distribution, and lower and upper bounds. After specifying the ambiguity set,

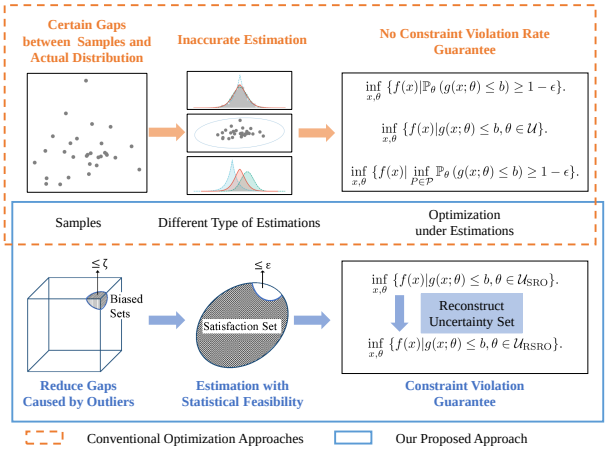


Fig. 2: Our Approach v.s. Its Conventional Rivals.

DRO estimates parameters such as mean, variance, skewness, and modality [19]. Specifically, compared with our proposed approach, the DRO has three limitations: 1) inability to handle joint CCs due to ambiguity set's complexity; 2) ambiguity set estimation without capturing sample collection uncertainty; and 3) computational burden when dealing with large sample sets.

In general, estimating the real distribution $\mathcal{F}_{\theta}$ only with samples is challenging. Even when all samples are obtained from a standard normal distribution, they may yield varying estimates [6]. With an inaccurate distribution estimation of θ, CCP cannot handle the CC with a good performance guarantee. The reason is twofold: 1) overemphasizing severe circumstances might lead to conservative solutions; 2) constraints can be violated. This highlights a certain gap between sample-based estimated distributions and actual distributions. Similar challenges arise with other optimization methods like DRO [19], [48]. While RO guarantees the best performance in the worst case without distribution requirements, it can be overly conservative. As shown in Fig. 2, samples will lead to inaccurate estimations, i.e., distribution for CC, uncertainty set for RO, and ambiguity set for DRO. By solving the inaccurate optimization, we cannot guarantee the violation rate at an appropriate level due to uncertainty in collected samples. Since we often cannot control how the sample set is composed, this motivates us to formally propose a framework for capturing the uncertainty of the sample sets.

We use tractable geometric shapes compatible with RO to find a high-probability zone with statistical feasibility. Specifically, we design a framework to integrate data into RO based on calibrating geometric shapes of uncertainty sets and splitting data to achieve finite-sample non-parametric statistical feasibility. Since our formulation considers the sample set's uncertainty, it is more practical, quantifing the gap between the estimated distribution based on samples and the actual distribution. It remains robust and theoretically guarantees that its solutions are feasible under a stability requirement for a specific constraint for any realization of the uncertain parameters. Specifically, the robustness in our approach is not solely about worst-case performance but rather about the balance between costs and stability.

For statistically feasible RO, the cube on the lower left corner in Fig. 2 characterizes all possible sample sets. Any point in the cube is a possible sample set with a number of samples, i.e., $\mathcal{D}_\theta = \{\boldsymbol{\theta}^{(i)}, \forall i\}$. We guarantee that except for a small part of biased sets (bounded by ζ, potentially brings huge estimation error), any other sample set allows us to accurately guarantee that based on the sample set, the original CC in (5) can be satisfied.

Remark: Although our proposed approach differs from classical RO in constructing uncertainty sets, it still can be categorized as RO because it takes an uncertainty set perspective. The key is to build a tractable uncertainty set that meets requirements with high probability. This follows the RO concept of optimizing against uncertainty. Also, it guarantees robustness. Any solution from our technique meets constraints with the specified confidence level.

IV. Solution Methodology

In this section, we employ a sample-oriented technique to approximate the objective function. Then, we develop an uncertainty set construction method and mitigate RO's conservativeness by reconstructing it using the form of constraints.

A. Objective Function Approximation

The curtailment costs $J(M^t)$ bring uncertainty to the objective function. We can decompose the uncertainty as follows:

$$\begin{aligned} J(M^t) =\ & \gamma_{\mathrm{w}}\mathbb{E}[(M^t)^+] + \gamma_{\mathrm{d}}\mathbb{E}[(-M^t)^+] \\ =\ & \gamma_{\mathrm{w}}\mathbb{E}\left[\left(G^t + \bar{W}^t + \boldsymbol{\xi}_{\mathrm{w}}^t - \left(\bar{D}^t + \boldsymbol{\xi}_{\mathrm{d}}^t\right)\right)^+\right] \\ & + \gamma_d\mathbb{E}\left[\left(\bar{D}^t + \boldsymbol{\xi}_{\mathrm{d}}^t - \left(G^t + \bar{W}^t + \boldsymbol{\xi}_{\mathrm{w}}^t\right)\right)^+\right]. \end{aligned} \tag{10}$$

Due to the hardness of obtaining exact distribution information about prediction error, we introduce sample average approximate (SAA), a popular approach to approximate the objective function [49]. SAA employs samples randomly selected from the dataset to approximate the expectation of $J(M^t)$. And selected samples make up a dataset $\mathcal{D}_{\mathrm{saa}}$ with size n_{saa}.

We calculate the sample average of $J(M^t)$ as follows:

$$J(M^t) = \sum\nolimits_{\boldsymbol{\xi}(k)\in\mathcal{D}_{\mathrm{saa}}} \widetilde{J}(M^t, \boldsymbol{\xi}(k))/n_{\mathrm{saa}}, \tag{11}$$

where

$$\begin{aligned} \widetilde{J}(M^t, \boldsymbol{\xi}(k)) = & \gamma_{\mathrm{w}}\left(G^t + \bar{W}^t + \boldsymbol{\xi}_{\mathrm{w}}^t(k) - \left(\bar{D}^t + \boldsymbol{\xi}_{\mathrm{d}}^t(k)\right)\right)^+ \\ & + \gamma_{\mathrm{d}}\left(\bar{D}^t + \boldsymbol{\xi}_{\mathrm{d}}^t(k) - \left(G^t + \bar{W}^t + \boldsymbol{\xi}_{\mathrm{w}}^t(k)\right)\right)^+. \end{aligned} \tag{12}$$

B. Statistically Feasible Robust UC

The major hurdle in solving the statistically feasible robust UC is constructing a non-conservative uncertainty set based on historical data. The desirable uncertainty set should satisfy constraints (8a)-(8c) to meet the stability requirement with the lowest system costs. This motivates us to construct an uncertainty set in two phases. In the subsequent analysis, we only construct $\mathcal{U}^r_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}}$ because the process to construct other uncertainty sets follows exactly the same routine.

We divide the dataset $\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}$ into two parts $\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}},1}$ and $\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}},2}$ with size z_1 and z_2, respectively. These two datasets are used in the two phases.

1) Shape Estimation: The first phase is parameter estimation of uncertainty sets. For better tractability and performance, we adopt ellipsoids to approximate the uncertainty set.

For ellipsoidal uncertainty set $\mathcal{U}^r_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}}$:

$$\mathcal{U}^r_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}} = \{\boldsymbol{\xi}_{\mathrm{w}} | (\boldsymbol{\xi}_{\mathrm{w}} - \boldsymbol{\mu}_{\mathrm{w}})^{\mathsf{T}} \boldsymbol{M}^{-1} (\boldsymbol{\xi}_{\mathrm{w}} - \boldsymbol{\mu}_{\mathrm{w}}) \le s^r_{\mathrm{w}}\}, \tag{13}$$

where $\boldsymbol{\mu}_{\mathrm{w}}$ is a vector with the same size of $\boldsymbol{\xi}_{\mathrm{w}}$, $\boldsymbol{M}$ is a symmetric matrix of $\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}},1}$, and s^r_{w} is a scalar. $\boldsymbol{\mu}$ and $\boldsymbol{M}$ influence the shape of uncertainty set, and s^r_{w} characterizes the size of the ellipsoid.

We estimate $\boldsymbol{\mu}_{\mathrm{w}}$ by sample mean of $\boldsymbol{\xi}^i_{\mathrm{w}}$ in $\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}},1}$:

$$\boldsymbol{\mu}_{\mathrm{w}} = {z_1}^{-1} \sum\nolimits_{i=1}^{z_1} \boldsymbol{\xi}^i_{\mathrm{w}}. \tag{14}$$

For a symmetric matrix, we use the sample covariance matrix for estimation since it could provide the geometric information of $\boldsymbol{\xi}_{\mathrm{w}}$:

$$\boldsymbol{M}_{\mathrm{w}} = (z_1 - 1)^{-1} \sum\nolimits_{i=1}^{z_1} \left(\boldsymbol{\xi}^i_{\mathrm{w}} - \boldsymbol{\mu}_{\mathrm{w}}\right)^T \left(\boldsymbol{\xi}^i_{\mathrm{w}} - \boldsymbol{\mu}_{\mathrm{w}}\right). \tag{15}$$

Following the same routine, we can also obtain the symmetric matrix $\boldsymbol{M}_{\mathrm{d}}$ for load uncertainty.

2) Size Calibration: Although we have obtained the approximated parameters in the first phase, the uncertainty set's size is unknown. We want to design a proper scalar s^r_{w} to calibrate the uncertainty set's size so that it covers $1-\rho$ portion of dataset $\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}$. Intuitively, larger s covers more historical data and yields a more conservative solution. We seek the minimal uncertainty set, which could guarantee statistical feasibility.

Before setting s^r_{w}, we define a transformation function:

$$y(\boldsymbol{\xi}_{\mathrm{w}}) = (\boldsymbol{\xi}_{\mathrm{w}} - \boldsymbol{\mu}_{\mathrm{w}})^{\mathsf{T}} {\boldsymbol{M}_{\mathrm{w}}}^{-1} (\boldsymbol{\xi}_{\mathrm{w}} - \boldsymbol{\mu}_{\mathrm{w}}). \tag{16}$$

We can calculate $y(\boldsymbol{\xi}_{\mathrm{w}})$ for every $\boldsymbol{\xi}^j_{\mathrm{w}} \in \mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}},2}$ to get the scalar $\hat{s}_j$. And we sort the values of scalar $[\hat{s}_1, \hat{s}_2, ..., \hat{s}_{z_2}]$ in ascending order. The index j^* satisfies:

$$j^* = \min\left\{r : \sum\nolimits_{j=0}^{r-1} C^j_{z_2} (1-\rho)^j (\rho)^{z_2 - j} \ge 1 - \delta\right\}. \tag{17}$$

The uncertainty set fulfilling statistical feasibility is produced by selecting $s^r_{\mathrm{w}} = \hat{s}_{j^*}$ as the desired scalar.

Theorem 1 (Theorem 1 in [28]): Consider a dataset $\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}$, we divide it into two parts whose sizes are z_1 and z_2, respectively. The uncertainty set $\mathcal{U}^r_{\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}}$ constructed based on $\mathcal{D}_{\boldsymbol{\xi}_{\mathrm{w}}}$ satisfies the statistical feasibility if $z_2 \ge \log\delta \log^{-1}(1-\rho)$, where $1-\rho$ describes the stability requirements and $1-\delta$ indicates the confidence level.

This theorem indicates that our method does not depend on the dimension of the random variable $\boldsymbol{\xi}$. Thus, our strategy can work well with high-dimensional joint distribution.

The robust constraint should be transformed into a deterministic constraint for efficient calculation. The robust constraint (8a) can be rewritten as linear constraints:

$$\boldsymbol{B}^t(\mathbf{SR}^t, r^t; \boldsymbol{\xi}^t_{\mathrm{w}}, \boldsymbol{\xi}^t_{\mathrm{d}}) \ge R^t + \alpha^r_{\mathrm{w,t}} + \alpha^r_{\mathrm{d,t}}, \forall t, \tag{18}$$

where $\alpha^t_{\mathrm{w}}, \alpha^t_{\mathrm{d}}$ satisfies:

$$\alpha^r_{\mathrm{w,t}} = -\mu^t_{\mathrm{w}} + \sqrt{s^r_{\mathrm{w}}} ||(\boldsymbol{M}^t_{\mathrm{w}})^{\frac{1}{2}}||_2, \ \alpha^r_{\mathrm{d,t}} = \mu^t_{\mathrm{d}} + \sqrt{s^r_{\mathrm{d}}} ||(\boldsymbol{M}^t_{\mathrm{d}})^{\frac{1}{2}}||_2.$$

Algorithm 1: Statistically Feasible RO (SRO)

Input: $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w}}$, $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d}}$, $1-\rho$, and $1-\delta$;
Divide the dataset $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w}}$ into $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w},1}$, $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w},2}$;
Divide the dataset $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d}}$ into $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d},1}$, $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d},2}$;
Obtain the approximated objective function by Eq. (11);
Calculate $\boldsymbol{\mu}_\mathrm{w}$, $\boldsymbol{\mu}_\mathrm{d}$, $\boldsymbol{M}_\mathrm{w}$, and $\boldsymbol{M}_\mathrm{d}$;
Calculate $y(\boldsymbol{\xi}_\mathrm{w}^j)$ and $y(\boldsymbol{\xi}_\mathrm{d}^j)$ for all $\boldsymbol{\xi}_\mathrm{w}^j \in \mathcal{D}_{\boldsymbol{\xi}_\mathrm{w},2}$, $\boldsymbol{\xi}_\mathrm{d}^j \in \mathcal{D}_{\boldsymbol{\xi}_\mathrm{d},2}$;
Sort all $y(\boldsymbol{\xi}_\mathrm{w}^j)$ and $y(\boldsymbol{\xi}_\mathrm{d}^j)$ in ascending order;
Calculate j^* by Eq. (17) and obtain s_w^r and s_d^r;
for $j \in \mathcal{A}$ **do**
  Obtain $s_{\mathrm{w,t}}^{\mathrm{ub},i}$, $s_{\mathrm{w,t}}^{\mathrm{lb},i}$, $s_{\mathrm{d,t}}^{\mathrm{ub},i}$, and $s_{\mathrm{d,t}}^{\mathrm{lb},i}$;
end
for $t \leftarrow 1$ **to** T **do**
  Calculate $\alpha_{\mathrm{w,t}}^r$ and $\alpha_{\mathrm{d,t}}^r$;
  for $i \in \mathcal{A}$ **do**
    Calculate $\alpha_{\mathrm{w,t}}^{\mathrm{ub},i}$, $\alpha_{\mathrm{w,t}}^{\mathrm{lb},i}$, $\alpha_{\mathrm{d,t}}^{\mathrm{ub},i}$, and $\alpha_{\mathrm{d,t}}^{\mathrm{lb},i}$;
  end
end
Solve problem (21) and return solutions;

$\boldsymbol{M}_\mathrm{w}^t$ is the submatrix of $\boldsymbol{M}_\mathrm{w}$ corresponding to dimension $\boldsymbol{\xi}_\mathrm{w}^t$ in $\boldsymbol{\xi}_\mathrm{w}$. We take $\alpha_{\mathrm{w,t}}^r$ as an example to introduce the detailed transformation:

$$\alpha_{\mathrm{w,t}}^r = \max_{\boldsymbol{\xi}_\mathrm{w}} \ \xi_\mathrm{w}^t \text{ s.t. } (\boldsymbol{\xi}_\mathrm{w}-\mu_\mathrm{w})^\mathsf{T}\boldsymbol{M}_\mathrm{w}^{-1}(\boldsymbol{\xi}_\mathrm{w}-\mu_\mathrm{w}) \le s_\mathrm{w}^r. \quad (19)$$

Slater's condition guarantees the strong duality condition:

$$\alpha_{\mathrm{w,t}}^r = \max_{\lambda}\min_{\boldsymbol{\xi}_\mathrm{w}} \xi_\mathrm{w}^t + \lambda\left[(\boldsymbol{\xi}_\mathrm{w}-\mu_\mathrm{w})^\mathsf{T}\boldsymbol{M}_\mathrm{w}^{-1}(\boldsymbol{\xi}_\mathrm{w}-\mu_\mathrm{w}) - s_\mathrm{w}^r\right] \quad \text{s.t. } \lambda \ge 0. \quad (20)$$

Because the covariance matrix $\boldsymbol{M}$ is positive definite, we can solve the quadratic programming and derive the objective explicitly to characterize $\alpha_{\mathrm{w,t}}^r$. Similarly, we can obtain the $\alpha_{\mathrm{d,t}}^r$, $\alpha_{\mathrm{w,t}}^{\mathrm{ub},i}$, $\alpha_{\mathrm{w,t}}^{\mathrm{lb},i}$, $\alpha_{\mathrm{d,t}}^{\mathrm{ub},i}$, and $\alpha_{\mathrm{d,t}}^{\mathrm{lb},i}$. Such transformations yield the following formulation:

$$\begin{aligned}
&\min \sum_{t=1}^T \left(\sum_{i\in\mathcal{N}}\left(C_i\left(g_i^t\right)+S_i^t\right)+J(M^t)+H(r^t)\right)\\
&\text{s.t.} \boldsymbol{B}^t(\mathbf{SR}^t, r^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) \ge R^t + \alpha_{\mathrm{w,t}}^r + \alpha_{\mathrm{d,t}}^r, \ \forall t, && (21a)\\
&\quad f_i^t(\boldsymbol{g}^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) \le \bar{f}_i - \alpha_{\mathrm{w,t}}^{\mathrm{ub},i} - \alpha_{\mathrm{d,t}}^{\mathrm{ub},i}, \ \forall i \in \mathcal{A}, \ \forall t, && (21b)\\
&\quad f_i^t(\boldsymbol{g}^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) \ge -\bar{f}_i + \alpha_{\mathrm{w,t}}^{\mathrm{lb},i} + \alpha_{\mathrm{d,t}}^{\mathrm{lb},i}, \ \forall i \in \mathcal{A}, \ \forall t, && (21c)\\
&\quad \text{(1c)- (1h), (3a)-(3d), (3g) - (3k),(4a).}
\end{aligned}$$

This problem is now tractable, and Algorithm 1 illustrates the procedure of constructing an uncertainty set.

C. Uncertainty Set Reconstruction

Although we have attempted to construct the smallest uncertainty set, the obtained solutions can still be too conservative. Here, we demonstrate the limits of Algorithm 1 and reconstruct the uncertainty set for remedy.

Conservativeness stems from the objective of constructing an uncertainty set that approximates ellipsoid and polyhedron forms, not specific constraints. Revisiting the statistical feasibility definition, it holds $\mathbb{P}(g(x;\theta) \le b) \ge \mathbb{P}(\theta \in \mathcal{U})$. However, the gap between $\mathbb{P}(g(x;\theta) \le b)$ and $\mathbb{P}(\theta \in \mathcal{U})$ is hard to control for Algorithm 1. This steers us toward reducing the gap. In the last stage, the uncertainty set is constructed without optimization problem information. The uncertainty set reconstruction can benefit from the constraint's form.

We first analyze the solution and constraints. It is clear that the solutions satisfy the following:

$$\mathbb{P}\left(\cap_{t=1}^T\{\boldsymbol{B}^t(\mathbf{SR}^t, r^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) \ge R^t\}\right) \ge 1-\rho, \quad (22)$$

if $\mathcal{U}_{\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w}}}^r = \{\boldsymbol{\xi}_\mathrm{w} | \cap_{t=1}^T\{\boldsymbol{B}^t(\mathbf{SR}^t, r^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) \ge R^t\}\}$ and $\mathcal{U}_{\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d}}}^r = \{\boldsymbol{\xi}_\mathrm{d} | \cap_{t=1}^T \{\boldsymbol{B}^t(\mathbf{SR}^t, r^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) \ge R^t\}\}$.

By closing the gap, we can make the solution less conservative. Due to uncertainty, the optimal solution cannot be directly obtained. However, Algorithm 1 can be employed since its conservative approach effectively satisfies Eq. (22).

Before reconstructing the uncertainty set, we first derive the form of the reconstructed uncertainty set. Specifically, we take the reconstructed uncertainty set for reserve requirements $\mathcal{U}_{\mathrm{recon}}^r$ as an example:

$$\mathcal{U}_{\mathrm{recon}}^r = \{\boldsymbol{\xi}_\mathrm{w}, \boldsymbol{\xi}_\mathrm{d} | \cap_{t=1}^T \{\boldsymbol{B}^t(\hat{\mathbf{SR}}^t, \hat{r}^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) \ge R^t - s_{\mathrm{recon}}^r\}\},$$

where $\hat{\mathbf{SR}}^t$ and $\hat{r}^t$ are the solutions returned by Algorithm 1.

We split the dataset into two parts following the same routine in Algorithm 1. The first portion calculates $\hat{\mathbf{SR}}^t$ and $\hat{r}^t$, while the second estimates s_{recon}^r. Specifically, two phases are needed to reconstruct an uncertainty set, much like its construction.

1) Shape Estimation: We run Algorithm 1 based on the first part of the dataset to obtain the solution.

2) Size Calibration: Due to the form of the reconstructed uncertainty set, we utilize historical data to calibrate its size and the solution from the previous phase for shrinking. Specifically, we introduce the transformation function:

$$h_r(\boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) = \max_t \ -\boldsymbol{B}^t(\hat{\mathbf{SR}}^t, \hat{r}^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) + R^t. \quad (23)$$

We utilize $h_r(\boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t)$ to transform samples from $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w},2}$ and $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d},2}$ to obtain the scalar $\hat{s_j}$. Then, we sort them following the routine in Algorithm 1 and choose the specific $s = \hat{s_{j^*}}$.

We also need to transform the robust constraint into a deterministic constraint for an efficient solving. The original constraint is

$$\cap_{t=1}^T\{\boldsymbol{B}^t(\mathbf{SR}^t, r^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) \ge R^t\}, \forall \boldsymbol{\xi}_\mathrm{w} \in \mathcal{U}_{\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w}}}, \forall \boldsymbol{\xi}_\mathrm{d} \in \mathcal{U}_{\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d}}},$$

which is equivalent to:

$$\boldsymbol{B}^t(\mathbf{SR}^t, r^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) - \boldsymbol{B}^t(\hat{\mathbf{SR}}^t, \hat{r}^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) \ge s_{\mathrm{recon}}^r, \forall t. \quad (24)$$

The reconstructed uncertainty set for transmission line constraints can be obtained following the same routine of $\mathcal{U}_{\mathrm{recon}}^r$ as illustrated in [6]. This leads to a tractable problem:

$$\begin{aligned}
&\min \sum_{t=1}^T \left(\sum_{i\in\mathcal{N}}\left(C_i\left(g_i^t\right)+S_i^t\right)+J(M^t)+H(r^t)\right)\\
&\text{s.t. } \boldsymbol{B}^t(\mathbf{SR}^t, r^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) - \boldsymbol{B}^t(\hat{\mathbf{SR}}^t, \hat{r}^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) \ge s_{\mathrm{recon}}^r, \forall t, && (25a)\\
&\quad f_i^t(\boldsymbol{g}^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) - f_i^t(\hat{\boldsymbol{g}^t}; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) + s_{\mathrm{recon}}^{\mathrm{ub},i} \le 0, \forall t, \forall i \in \mathcal{A}, && (25b)\\
&\quad f_i^t(\boldsymbol{g}^t; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) - f_i^t(\hat{\boldsymbol{g}^t}; \boldsymbol{\xi}_\mathrm{w}^t, \boldsymbol{\xi}_\mathrm{d}^t) - s_{\mathrm{recon}}^{\mathrm{lb},i} \ge 0, \forall t, \forall i \in \mathcal{A}, && (25c)\\
&\quad \text{(1c)- (1h),(3a)-(3d), (3g) - (3k),(4a),}
\end{aligned}$$

where $s_{\mathrm{recon}}^{\mathrm{ub}}$ and $s_{\mathrm{recon}}^{\mathrm{lb}}$ are scalers generated by the reconstructed uncertainty set for transmission line constraints.

Algorithm 2: Reconstructed SRO (RSRO)

Input: $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w}}$, $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d}}$, $1-\rho$, and $1-\delta$;
Divide the dataset $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w}}$ into $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w},1}$, $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w},2}$;
Divide the dataset $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d}}$ into $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d},1}$, $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d},2}$;
Obtain the solutions $\hat{\mathbf{SR}}^t$,$\hat{r}^t$, and $\hat{\boldsymbol{g}}^t$ by Algorithm1 with $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{w},1}$ and $\mathcal{D}_{\boldsymbol{\xi}_\mathrm{d},1}$;
Calculate $h_r(\boldsymbol{\xi}_\mathrm{w}^j, \boldsymbol{\xi}_\mathrm{d}^j)$ for all $\boldsymbol{\xi}_\mathrm{w}^j \in \mathcal{D}_{\boldsymbol{\xi}_\mathrm{w},2}$ and $\boldsymbol{\xi}_\mathrm{d}^j \in \mathcal{D}_{\boldsymbol{\xi}_\mathrm{d},2}$ based on (23);
Sort all $h_r(\boldsymbol{\xi}_\mathrm{w}^j, \boldsymbol{\xi}_\mathrm{d}^j)$ in ascending order;
Calculate j^* by Eq. (17) and obtain s^r_{recon};
for $j \in \mathcal{A}$ **do**
  Calculate $h_{\mathrm{ub}}(\boldsymbol{\xi}_\mathrm{w}^j, \boldsymbol{\xi}_\mathrm{d}^j)$ and $h_{\mathrm{lb}}(\boldsymbol{\xi}_\mathrm{w}^j, \boldsymbol{\xi}_\mathrm{d}^j)$ for all $\boldsymbol{\xi}_\mathrm{w}^j \in \mathcal{D}_{\boldsymbol{\xi}_\mathrm{w},2}$ and $\boldsymbol{\xi}_\mathrm{d}^j \in \mathcal{D}_{\boldsymbol{\xi}_\mathrm{d},2}$ based on;
  Sort all $h_{\mathrm{ub}}(\boldsymbol{\xi}_\mathrm{w}^j, \boldsymbol{\xi}_\mathrm{d}^j)$ and $h_{\mathrm{lb}}(\boldsymbol{\xi}_\mathrm{w}^j, \boldsymbol{\xi}_\mathrm{d}^j)$ in ascending order;
  Calculate j^* by Eq. (17) and obtain $s^{\mathrm{ub},i}_{\mathrm{recon}}$ and $s^{\mathrm{lb},i}_{\mathrm{recon}}$;
end
Solve problem (25) and return solutions;

Algorithm 2 details how to reconstruct the uncertainty set. **Remark**: We can obtain favorable solutions based on the proposed methods. However, the high computational complexity of MIP renders the problem computationally challenging. In practice, UC must be solved frequently under different conditions within a preset time. To make our methods practical, MIP-solving running time should be reduced.

V. Accelerating for Real-World Applications

In this section, we first develop a clustering-based unit status identifier to fix unit status, allowing us to obtain LP UC formulation to relieve the computational burden. We then customize an optimization kernel for UC to improve its computational efficiency, inspired by the features of LP UC.

A. Learning to Identify Unit Status

We first discuss how to obtain favorable solutions in UC from the L2O perspective. Two factors determine whether L2O works well: the quality of the training set (especially whether previous solved instances are optimal) and whether similar UC inputs result in the same unit status. For the first factor, there is extensive existing knowledge about previous instances solved by commercial solvers. For the second factor, we assume that UC's unit status will be the same if given similar inputs like the predicted load. This assumption is intuitive and has been widely adopted in the literature [32]. By satisfying these two factors, we could obtain favorable solutions by developing a clustering-based unit status identifier.

We concentrate on combining clustering and ML models that can learn the mapping between the input, including predicted load and renewable generation, and optimal solution $\boldsymbol{u}^*$ collected from Algorithm 2 for day-ahead operations. We first perform clustering on the historical data to extract patterns that can aid in better classification. Specifically, for the N_h historical demand samples $\boldsymbol{d}^{\mathrm{in}} \in \mathbf{R}^{n_\mathrm{d} \times T}$ and renewable generation samples $\boldsymbol{w}^{\mathrm{in}} \in \mathbf{R}^{n_\mathrm{w} \times T}$, we employ k-means algorithms with Euclidean distance to divide the set of N_h samples $\boldsymbol{d}^{\mathrm{in}}$ and $\boldsymbol{w}^{\mathrm{in}}$ into N_{cluster} disjoint subgroups. By applying this approach, we obtain the clustering model:

$$\mathcal{S} = \mathcal{F}^{\mathrm{cluster}}(\boldsymbol{d}^{\mathrm{in}}, \boldsymbol{w}^{\mathrm{in}}), \tag{26}$$

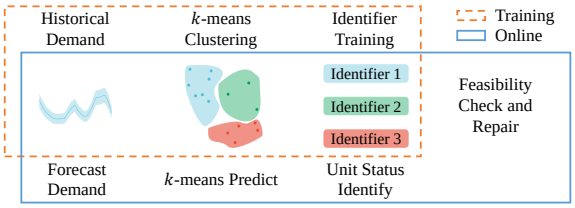


Fig. 3: Learning to Identify Unit Status.

where $\mathcal{S}$ represents the indices of subgroups for incoming demand samples, and $\mathcal{F}^{\mathrm{cluster}}$ denotes the clustering model.

Consider a dataset of input-output pairs $(\mathcal{X}, \mathcal{Y})$. $\mathcal{X}$ represents all inputs, i.e., $\mathcal{X} = \{\boldsymbol{d}_i^{\mathrm{in}}, \boldsymbol{w}_i^{\mathrm{in}}, \forall i\}$. The output ground truth set is $\mathcal{Y} = \{\boldsymbol{u}_i, \forall i\}$. Then, $\{\boldsymbol{d}_i^{\mathrm{in}}, \boldsymbol{w}_i^{\mathrm{in}}; \boldsymbol{u}_i\}$ forms an input-output pair. We employ LightGBM, a powerful gradient boosting framework [50], to train a unit status identifier $\mathcal{F}_i^{\mathrm{identify}}$ for each subgroup $i \in \mathcal{S}$. After training, each incoming sample should first be predicted by clustering model $\mathcal{F}^{\mathrm{cluster}}$ and obtain its indices i^* of subgroups. Then, the corresponding identifier $\mathcal{F}_{i^*}^{\mathrm{identify}}$ is employed. Specifically, our proposed method can handle a large number of combinations of unit status because we build a specific unit status identifier for each unit, which makes the prediction results independent of the number of units. Finally, we utilize the feasibility check and repair method mentioned in [36], [37] to guarantee the feasibility of the predicted unit status. During online implementation, we first ascertain the cluster to which the upcoming instance belongs and then use the corresponding unit status identifier to forecast the unit status. We also check and repair the feasibility of output following the similar method in [37]:

$$\boldsymbol{u}^p = \arg\min_{\boldsymbol{u}} \|\boldsymbol{u} - \boldsymbol{u}^o\|_2 \quad \text{s.t. (1f)-(1g)}, \tag{27}$$

where $\boldsymbol{u}^p$ is feasible solution, and $\boldsymbol{u}^o$ is the output of identifier.

By identifying the unit status via the proposed L2O approach, we can obtain LP UC formulation. Furthermore, the value of $\boldsymbol{r}$ can also be identified based on the form of Eq. (4a). As a result, we can eliminate the startup costs, RCP costs, and constraints that are solely associated with $\boldsymbol{u}$ and $\boldsymbol{r}$ in the subsequent analysis, thus simplifying the problem and improving computational efficiency. In summary, the L2O algorithm is presented in Fig. 3.

Remark: It remains a theoretical challenge to ensure that learning algorithms find the global best solution. Hence, we exclusively seek solutions with low-performance losses and computational performance improvements. To mitigate the limitations of learning algorithms, we first examine the feasibility of constraints and repair the solution if needed. More specifically, we address the UC after fixing all unit status, ensuring all constraints are met, and yielding feasible solutions for day-ahead operations. We also maintain reserve requirement and transmission line constraint violation rates at a predetermined level, offering a mechanism to manage potential system risks.

B. Optimization Kernel for Efficient UC

Inspired by the properties of LP UC, we exploit the optimization problem structure and customize an optimization

kernel for UC in advance to further improve its computational efficiency. We show that the generation profile g_i^t need not to be calculated for each generator, as the overall generation $g^t = \sum_{i\in\mathcal{N}} g_i^t$ can be solved directly. This approach can significantly improve computational performance. Specifically, we adopt the notion of parametric programming to achieve it.

The key idea of parametric programming is to replace the optimization problem with a parameterized function with the parameter g^t, which is

$$\begin{aligned} &\text{MinUC}(g^t) = \min\ C\left(g^t\right) + J(M^t) \\ &\text{s.t. } C\left(g^t\right) = \sum\nolimits_{i\in\mathcal{N}} C_i\left(g_i^t\right) \\ &\qquad \text{(1c) - (1e), (3a)- (3d), (4a), (25a)- (25b).} \end{aligned} \tag{28}$$

We can observe that the objective function now only explicitly contains the total generation g^t. In order to transform the optimization problem into one with only a variable g^t, we need to solve the following two problems: determining g^t's feasible region and finding $C\left(g^t\right)$'s analytical form.

The feasible region of g^t is determined by a number of linear constraints. Denoting the feasible region of total generation $\boldsymbol{g}^t$ by $\mathcal{A}(g^t)$:

$$g_{\min}^t = \inf\ \mathcal{A}(g^t), \quad g_{\max}^t = \sup\ \mathcal{A}(g^t). \tag{29}$$

It is direct to indicate that $g^t \in [g_{\min}^t, g_{\max}^t]$ by the continuity of the feasible region. $C\left(g^t\right)$ has the following form:

$$C\left(g^t\right) := \ \min\ C\left(g^t\right) + J(M^t) \quad \text{s.t. (1c)- (1e).} \tag{30}$$

In order to decide $C\left(g^t\right)$'s analytical form, we first deduce the following lemma to uncover its basic structure:

Lemma 1: The generation function $C\left(g^t\right)$ is continuous, piecewise linear, and convex in g^t over $[g_{\min}^t, g_{\max}^t]$.

This lemma can be proved using the same routine as the proof of Theorem 2 in [51]. These properties yield an efficient algorithm (Curve Generation Algorithm) [30], [52] to generate the curve of $C\left(g^t\right)$. Once we have the knowledge of $C\left(g^t\right)$, we can easily transform the original problem into a simple optimization with a single decision variable:

$$\min\ C\left(g^t\right) + J(M^t) \quad \text{s.t. } g_{\min}^t \le g^t \le g_{\max}^t. \tag{31}$$

VI. Numerical Studies

In this section, we compare popular benchmarks with our proposed approaches to highlight the superiority of ours. All experiments are solved using Gurobi 9.5.2 [53] on a laptop with an Intel Core i5-12500H CPU and 16G RAM.

A. Experimental Setup

We utilize three systems to evaluate the effectiveness of different methods, including a 39-bus system (10 generators, 46 transmission lines) [54], whose generator information is from [55], a modified 118-bus system [56] (54 generators, 188 transmission lines), and a modified 300-bus system [57] (69 generators, 411 transmission lines).

We obtain actual and predicted wind power data from the Belgian Electricity System Operator (Elia) [58], covering January 2016 to August 2021 with the resolution of one hour.

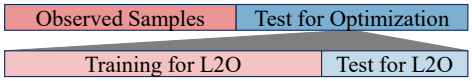


Fig. 4: Dataset Splitting.

Elia also provides actual and forecast loads during the same period. For wind power and load data, sample and test sets comprise 50% and 50% of the total data, respectively. The sample set constructs the uncertainty set, and the test set evaluates different methodologies. The stability requirement $1-\rho$ and confidence level $1-\delta$ are set at 0.95.

To evaluate the acceleration performance, we leverage the solved instances in the previous part to train the unit status identifier and test the effectiveness of accelerating algorithms. The corresponding training and testing instances contain 80% and 20% of all the solved instances, respectively. Fig. 4 visualizes the dataset splitting process. We set the number of clusters in the L2O as 5, and the reserve requirement R^t as 5% of demand in every period. We set the load and renewable energy generation curtailment costs to be \$300/MWh and \$20/MWh. The cost of RCP is assumed to be piecewise linear. Specifically, the capacity for each segment of the RCP has been set at 33MW, with the prices for each phase delineated as \$8.8/MWh, \$9.2/MWh, and \$9.5/MWh, respectively.

B. Competing Methods

We compare the following approaches for UC:

CCP: After decomposing the joint CCs, we solve the problem and assume the uncertainty follows the Gaussian distribution.

RO: Classical robust optimization (CRO), statistically feasible robust optimization (SRO), and reconstructed SRO (RSRO) are used. RSRO is accelerated by transforming MIP UC to LP UC and customizing the optimization kernel, which are named RSRO_L and RSRO_K, respectively.

Scenario Generation (SG): SG creates scenarios to solve UC.

DRO: DRO presumes that the actual distribution belongs to an ambiguity set. We employ the benchmark in [19].

Deterministic Optimization (DO): We utilize the actual demand to conduct DO as the perfect approach.

To evaluate the performance of different methods, we focus on two metrics: system costs and maximum constraint violation rate. Specifically, we evaluate the average additional costs compared to the perfect approach (in percentage) and term this gap the regret percentage (REP) [5]. The smaller the gap, the less the extra cost of the method to handle the uncertainty.

C. Evaluation on 39-bus System

We first consider the 39-bus system, in which the maximum wind power output is 50% of total conventional generation capacity. Specifically, the real output of wind turbines is much less than the maximum output.

Table IV compares the performance of different methodologies, indicating that RSRO satisfies the stability requirements at the lowest cost, with an REP of 8.92%. RSRO surpasses its competitors because it achieves the lowest cost and close to predetermined violation rates. Although CCP is less expensive than other methods, its violation rate is over triple

TABLE IV: Performance Evaluation of Optimization Methods.

	CCP	CRO	SRO	RSRO	SG	DRO
REP (%)	5.83	42.95	24.80	**8.92**	13.45	11.43
Violation Rate (%)	18.0	0	0	**2.8**	1.0	1.4
CPU Time (s)	**3.97**	4.32	4.41	9.90	44.46	4.66

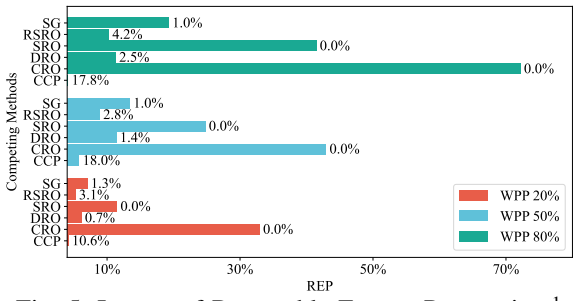


Fig. 5: Impact of Renewable Energy Penetration.[1]

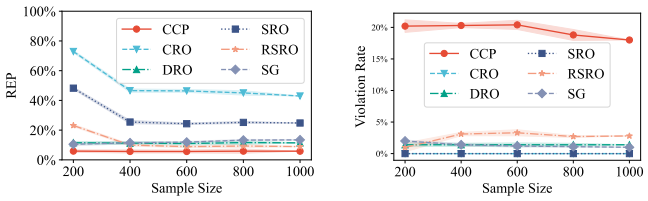
the stability requirements. Conversely, CRO has the highest REP of 42.95% due to its nature. SRO and DRO have mostly mitigated RO's conservativeness, but they do not fully exploit the potential of stability requirements. SG is conservative since it creates constraints from samples whose performance may be impaired by many extreme scenarios. And SG is less efficient than others, taking 44.46s running time.

1) Sensitivity Analysis for Renewable Energy Penetration: We study the sensitivity of different approaches to a gradual increase in renewable energy penetration. Fig. 5 illustrates the performance of different approaches based on the 39-bus system when the wind power penetration rate is 20%, 50%, and 80%, respectively. Interestingly, REP increases as renewable energy penetration increases. This is because the higher renewable energy penetration rates increase uncertainty, which leads to additional costs to hedge against the uncertainty. RSRO consistently outperforms its rivals in stability requirements and system costs as the renewable energy penetration rate changes. DRO ranks second best, offering acceptable REP and violation rate, but is more conservative than RSRO. Although CRO, SRO, and SG are conservative, the conservativeness of SG is less than that of CRO and SRO as the renewable energy penetration rate increases. CCP provides the lowest cost, but it cannot always satisfy CCs.

2) Sensitivity Analysis for Sample Size: All strategies address uncertainty for future instances using historical samples, which means sample size is key to these approaches' effectiveness. Our previous experiments use sufficient samples to estimate the parameters. In this part, we select a specific number of examples from the sample set to evaluate the performance of all methods as the sample size increases. Furthermore, we employ Monte Carlo simulation in sample size analysis to demonstrate the performance. Specifically, we randomly select a specific number of samples five times from the total sample. Other parameters are the same as in previous simulations. Fig. 6 illustrates all methods' costs and violation rates with the sample size increases.

Interestingly, sample size affects some approaches more

[1]WPP represents the percentage of the maximum wind power output relative to the total generator's capacity. The percentage to the right of the bars denotes violation rates.



(a) Impact on System Costs (b) Impact on Violation Rate

Fig. 6: Impact of Sample Size.

than others. When the sample size is 200, RSRO fails. However, it can better exploit the potential of CCs and provide a less conservative solution with an increasing yet condition-satisfying violation rate as the sample size increases. SRO and CRO are conservative but their performance improves with more samples. The performance of DRO slightly improves as the sample size increases. This indicates that CCP cannot perform better even with sufficient data. In contrast, SG's cost increases with the sample size while the violation rate remains relatively constant. This arises from SG's capability to generate enough constraints to yield feasible solutions from a limited set of samples. However, larger samples provide more constraints, resulting in a more conservative solution.

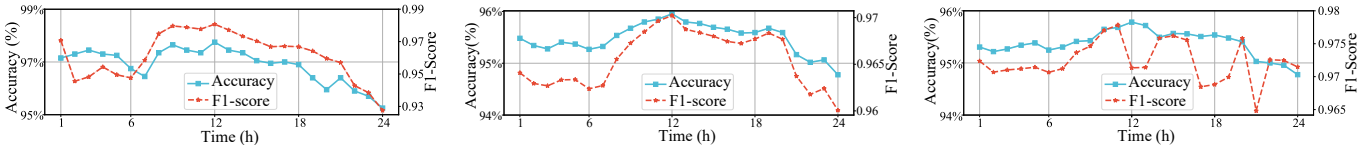
3) Performance of Accelerating Algorithms: We adopt the L2O approach and optimization kernel to accelerate the solving process. Using the solved instances from the previous part, we train the identifier and test L2O methods on 80% and 20% of the data, respectively.

Fig. 7(a) implies the average accuracy and F1-score for each time point in the system. We offer a strategy that achieves high accuracy ($> 95\%$) and F1-score (> 0.93) at all times. Despite RSRO's remarkable performance, almost twice as much time consumption than rivals impedes practical uses. Hence, we embed ML into our method for acceleration. Our proposed solution reduces running time with tolerable performance loss, as shown in Table V. Specifically, RSRO_L reduces the running time by half, and RSRO_K reduces the running time more significantly due to avoiding repeatedly solving optimization problems. Furthermore, the violation rates of RSRO_L and RSRO_K still satisfy the stability requirements after embedding ML.

Fig. 8 illustrates the distribution of time consumption. CCP, SRO, RSRO, and CRO consume similar time since they share a similar number of constraints. SG takes roughly six times longer running time than RSRO. RSRO_L, based on LP, requires less running time than MIP-based techniques due to its lower complexity. RSRO_K boosts computational efficiency by constructing an optimization kernel for upcoming instances instead of solving an optimization problem. Although the ML algorithms cannot guarantee the optimal solution due to the complexity of UC and the nature of ML, their system costs are still lower than other methods.

D. Evaluation on 118-bus System

We also evaluate the performance of our approach based on the 118-bus system. This subsection compares all methods, including RSRO_L and RSRO_K, on the test set.



(a) 39-bus System (b) 118-bus System (c) 300-bus System

Fig. 7: Transformation Performance from MIP UC into the LP UC.

TABLE V: Results on 39-bus Sys.

	REP (%)	Violation Rate (%)	CPU Time (s)
CCP	6.69	22.0	3.95
CRO	42.45	0	4.17
SRO	24.54	0	4.23
RSRO	**9.53**	**2.5**	8.92
SG	13.84	1.0	43.43
DRO	11.90	1.5	4.57
RSRO_L	10.23	2.0	4.09
RSRO_K	10.23	2.0	**0.002**

TABLE VI: Results on 118-bus Sys.

	REP (%)	Violation Rate (%)	CPU Time (s)
CCP	1.05	10.0	10.31
CRO	13.99	0	14.91
SRO	5.11	0	11.71
RSRO	**1.47**	**2.0**	25.21
SG	2.07	0.5	930.52
DRO	1.76	0.5	16.79
RSRO_L	1.55	1.0	20.65
RSRO_K	1.55	1.0	**0.002**

TABLE VII: Results on 300-bus Sys.

	REP (%)	Violation Rate (%)	CPU Time (s)
CCP	0.07	9.50	14.18
CRO	-	-	-
SRO	2.50	0	22.32
RSRO	**0.26**	**1.5**	301.39
SG	0.46	0.5	1693
DRO	0.34	0.5	32.87
RSRO_L	0.29	1.0	239.99
RSRO_K	0.29	1.0	**0.002**

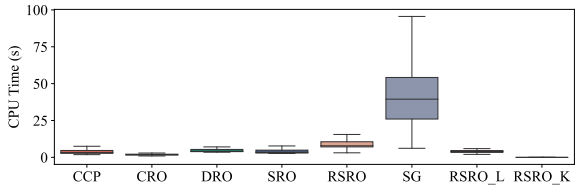


Fig. 8: Time Consumption.

Fig. 7(b) shows the transformation accuracy of the L2O approach. The transformation produces remarkable accuracy and F1-score performance over time. Table VI shows the results for the 118-bus system: RSRO offers a superior balance between REP and violation rate. Interestingly, RSRO_L and RSRO_K record similar REP values of RSRO. This suggests that the accelerated methods retain most of the effectiveness of RSRO. In contrast, CRO has a significantly higher REP value of 13.99%. For time-sensitive applications that require rapid response, SG and DRO may be too computationally intensive.

E. Evaluation on 300-bus System

We conduct a simulation on the 300-bus system to seek the practical performance of our proposed approach. Fig. 7(c) shows the transformation accuracy of the L2O approach, implying a similar performance as in the smaller system. Table VII demonstrates that the RSRO exhibits superior REP and an appropriate violation rate again. However, its computational burden is notably high. RSRO_K, benefiting from ML and optimization kernel, shines in terms of speed without much compromise in REP and violation rate. When immediate results are crucial, RSRO_K proves to be an optimal choice. Notably, the lack of results for the CRO strategy suggests that it might be unsuitable for such a large system.

F. Computational Analysis for L2O

The number of unit status combinations in 24 hours is vast, which may lengthen training. Table VIII illustrates the running time breakdown for this task. Due to the nature of the tree-based model, we can train the unit status identifier in a limited time. Inference time is longer than training time because we make predictions for each unit in each incoming sample to be solved, resulting in a large loop and long inference time.

TABLE VIII: Running Time for Unit Status Identification.

	39-bus	118-bus	300-bus
Clustering Time (s)	0.22	0.58	0.83
Training Time (s)	6.43	23.34	27.79
Inference Time (s)	136.17	783.31	937.88

VII. Conclusion

This paper develops statistically feasible robust optimization for UC, relieving conservativeness without distribution requirements. We also accelerate our methods for practical applications based on ML algorithms and optimization techniques. Our RSRO approach enjoys the best REP and comes closest to meeting the stability requirement. Our accelerating algorithm can decrease the computational time significantly. The numerical studies based on the 39-bus system, 118-bus system, and 300-bus system verify our methods' effectiveness.

Much more can be done. For instance, we construct the uncertainty set based on statistical metrics, ML can aid uncertainty set construction. Also, end-to-end ML may prompt unit status identification. Furthermore, the impact of data quality on the results remains unknown.

References

[1] D. Bertsimas, E. Litvinov, X. A. Sun, *et al.*, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 52–63, 2013.

[2] C. Zhao, J. Wang, J.-P. Watson, *et al.*, "Multi-stage robust unit commitment considering wind and demand response uncertainties," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2708–2717, 2013.

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

[4] B. Liu, *Theory and Practice of Uncertain Programming*. Springer, 2009.

[5] W. Jiang, C. Lu, and C. Wu, "Robust scheduling of thermostatically controlled loads with statistically feasible guarantees," *IEEE Trans. Smart Grid*, 2023.

[6] C. Lu, N. Gu, W. Jiang, *et al.*, "Sample-adaptive robust economic dispatch with statistically feasible guarantees," *IEEE Trans. Power Syst.*, 2023.

[7] A. V. Ramesh and X. Li, "Feasibility layer aided machine learning approach for day-ahead operations," *IEEE Trans. Power Syst.*, 2023.

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

[9] K. Sundar, H. Nagarajan, L. Roald, *et al.*, "Chance-constrained unit commitment with n-1 security and wind uncertainty," *IEEE Control Netw. Syst.*, vol. 6, no. 3, pp. 1062–1074, 2019.

[10] Z. Li, Y. Xu, S. Fang, *et al.*, "Robust coordination of a hybrid ac/dc multi-energy ship microgrid with flexible voyage and thermal loads," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 2782–2793, 2020.

[11] W. Hou and H. Wei, "Data-driven robust day-ahead unit commitment model for hydro/thermal/wind/photovoltaic/nuclear power systems," *Int. J. Electr. Power Energy Syst.*, vol. 125, p. 106427, 2021.

[12] N. Gu, H. Wang, J. Zhang, *et al.*, "Bridging chance-constrained and robust optimization in an emission-aware economic dispatch with energy storage," *IEEE Trans. Power Syst.*, vol. 37, no. 2, pp. 1078–1090, 2022.

[13] P. Xiong, P. Jirutitijaroen, and C. Singh, "A distributionally robust optimization model for unit commitment considering uncertain wind power generation," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 39–49, 2017.

[14] C. Duan, L. Jiang, W. Fang, *et al.*, "Data-driven affinely adjustable distributionally robust unit commitment," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1385–1398, 2018.

[15] Y. Chen, Q. Guo, H. Sun, *et al.*, "A distributionally robust optimization model for unit commitment based on kullback–leibler divergence," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5147–5160, 2018.

[16] O. Yurdakul, F. Sivrikaya, and S. Albayrak, "Kullback-leibler divergence-based distributionally robust unit commitment under net load uncertainty," in *2021 IEEE Madrid PowerTech*, pp. 1–6, 2021.

[17] R. Zhu, H. Wei, and X. Bai, "Wasserstein metric based distributionally robust approximate framework for unit commitment," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2991–3001, 2019.

[18] X. Zheng and H. Chen, "Data-driven distributionally robust unit commitment with wasserstein metric: Tractable formulation and efficient solution method," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4940–4943, 2020.

[19] A. Zhou, M. Yang, X. Zheng, *et al.*, "Distributionally robust unit commitment considering unimodality-skewness information of wind power uncertainty," *IEEE Trans. Power Syst.*, pp. 1–12, 2022.

[20] S. Ghosal and W. Wiesemann, "The distributionally robust chance-constrained vehicle routing problem," *Oper. Res.*, vol. 68, no. 3, pp. 716–732, 2020.

[21] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations," *Oper. Res.*, vol. 58, no. 4-part-1, pp. 902–917, 2010.

[22] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969–996, 2007.

[23] W. Xie, S. Ahmed, and R. Jiang, "Optimized bonferroni approximations of distributionally robust joint chance constraints," *Math. Program.*, vol. 191, no. 1, pp. 79–112, 2022.

[24] R. Jagannathan, "Chance-constrained programming with joint constraints," *Oper. Res.*, vol. 22, no. 2, pp. 358–372, 1974.

[25] W. Xie and S. Ahmed, "On deterministic reformulations of distributionally robust joint chance constrained optimization problems," *SIAM J. Optim.*, vol. 28, no. 2, pp. 1151–1182, 2018.

[26] E. Du, N. Zhang, C. Kang, *et al.*, "Scenario map based stochastic unit commitment," *IEEE Trans. Power Syst*, vol. 33, no. 5, pp. 4694–4705, 2018.

[27] M. A. Velasquez, N. Quijano, A. I. Cadena, *et al.*, "Distributed stochastic economic dispatch via model predictive control and data-driven scenario generation," *Int. J. Electr. Power Energy Syst.*, vol. 129, p. 106796, 2021.

[28] L. J. Hong, Z. Huang, and H. Lam, "Learning-based robust optimization: Procedures and statistical guarantees," *Manage. Sci.*, vol. 67, no. 6, pp. 3447–3467, 2021.

[29] Q. Zhai, X. Guan, J. Cheng, *et al.*, "Fast identification of inactive security constraints in scuc problems," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1946–1954, 2010.

[30] C. Lu, W. Jiang, and C. Wu, "Effective end-to-end learning framework for economic dispatch," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 4, pp. 2673–2683, 2022.

[31] N. Yang, C. Yang, L. Wu, *et al.*, "Intelligent data-driven decision-making method for dynamic multisequence: An e-seq2seq-based scuc expert system," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 3126–3137, 2022.

[32] G. Dalal, E. Gilboa, S. Mannor, *et al.*, "Unit commitment using nearest neighbor as a short-term proxy," in *2018 PSCC*, pp. 1–7, 2018.

[33] S. Pineda, J. M. Morales, and A. Jiménez-Cordero, "Data-driven screening of network constraints for unit commitment," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3695–3705, 2020.

[34] F. Mohammadi, M. Sahraei-Ardakani, D. N. Trakas, *et al.*, "Machine learning assisted stochastic unit commitment during hurricanes with predictable line outages," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5131–5142, 2021.

[35] A. S. Xavier, F. Qiu, and S. Ahmed, "Learning to solve large-scale security-constrained unit commitment problems," *INFORMS J. Comput.*, vol. 33, no. 2, pp. 739–756, 2021.

[36] T. Wu, Y.-J. Angela Zhang, and S. Wang, "Deep learning to optimize: Security-constrained unit commitment with uncertain wind power generation and besss," *IEEE Trans. Sustain. Energy*, vol. 13, no. 1, pp. 231–240, 2022.

[37] L. Sang, Y. Xu, and H. Sun, "Ensemble provably robust learn-to-optimize approach for security-constrained unit commitment," *IEEE Trans. Power Syst.*, 2022.

[38] Q. Gao, Z. Yang, W. Li, *et al.*, "Online learning of stable integer variables in unit commitment using internal information," *IEEE Trans. Power Syst.*, vol. 38, no. 3, pp. 2947–2950, 2023.

[39] F. Li, J. Qin, and W. X. Zheng, "Distributed q -learning-based online optimization algorithm for unit commitment and dispatch in smart grid," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 4146–4156, 2020.

[40] P. de Mars and A. O'Sullivan, "Applying reinforcement learning and tree search to the unit commitment problem," *Appl. Energy*, vol. 302, p. 117519, 2021.

[41] A. Ajagekar and F. You, "Deep reinforcement learning based unit commitment scheduling under load and wind power uncertainty," *IEEE Trans. Sustain. Energy*, vol. 14, no. 2, pp. 803–812, 2023.

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

[43] F. Aminifar, M. Fotuhi-Firuzabad, and M. Shahidehpour, "Unit commitment with probabilistic spinning reserve and interruptible load considerations," *IEEE Trans. Power Syst*, vol. 24, no. 1, pp. 388–397, 2009.

[44] C. Lu, J. Liang, W. Jiang, *et al.*, "High-resolution probabilistic load forecasting: A learning ensemble approach," *J. Frankl. Inst.*, vol. 360, no. 6, pp. 4272–4296, 2023.

[45] Y. Wang, N. Zhang, Y. Tan, *et al.*, "Combining probabilistic load forecasts," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3664–3674, 2019.

[46] MISO, "Ramp product enhancements." Available online: https://cdn.misoenergy.org/20221201%20MSC%20Item%2006%20Ramp%20Product%20Enhancements627169.pdf (accessed December 15, 2022.)., 2022.

[47] S. Sreekumar, S. Yamujala, K. C. Sharma, *et al.*, "Flexible ramp products: A solution to enhance power system flexibility," *Renew. Sust. Energ. Rev.*, vol. 162, p. 112429, 2022.

[48] P. Zhao, C. Gu, D. Huo, *et al.*, "Two-stage distributionally robust optimization for energy hub systems," *IEEE Trans. Ind. Inform.*, vol. 16, no. 5, pp. 3460–3469, 2020.

[49] Y. Long, L. H. Lee, and E. P. Chew, "The sample average approximation method for empty container repositioning with uncertainties," *Eur. J. Oper. Res.*, vol. 222, no. 1, pp. 65–75, 2012.

[50] G. Ke, Q. Meng, T. Finley, *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," *NIPS*, vol. 30, 2017.

[51] C. Wu, G. Hug, and S. Kar, "Risk-limiting economic dispatch for electricity markets with flexible ramping products," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 1990–2003, 2016.

[52] C. Wu, S. Bose, A. Wierman, *et al.*, "A unifying approach to assessing market power in deregulated electricity markets," in *2013 IEEE PES GM*, pp. 1–5, 2013.

[53] Gurobi, "Gurobi solver 9.5.2." Available online: https://www.gurobi.com/ (accessed on October 10, 2022.)., 2022.

[54] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, 2011.

[55] H. Pandžić, T. Qiu, and D. S. Kirschen, "Comparison of state-of-the-art transmission constrained unit commitment formulations," in *2013 IEEE PES GM*, pp. 1–5, 2013.

[56] I. I. of Technology Power Group, "Ieee 118-bus test system," 2003. http://motor.ece.iit.edu/data/.

[57] W. Huang, X. Pan, M. Chen, *et al.*, "Deepopf-v: Solving ac-opf problems efficiently," *IEEE Trans. Power Syst*, vol. 37, no. 1, pp. 800–803, 2022.

[58] Elia, "Open data elia." Available online: https://opendata.elia.be/pages/home/ (accessed on November 20, 2022.)., 2022.