Constructing Uncertainty Sets from Covariates in Power Systems

Dimitris Bertsimas, *Member, IEEE,* Thodoris Koukouvinos, *Student Member, IEEE* Angelos Georgios Koulouras, *Student Member, IEEE*

***Abstract*—Robust optimization (RO) has shown promising results in dealing with the uncertainty that renewable energy sources introduce in power systems. RO is based on uncertainty sets that contain all uncertain scenarios against which we want to be robust. Despite their importance in controlling the cost and conservativeness of the RO formulations, there has been little work in determining their size. In this paper, we propose a mixed integer optimization (MIO) formulation for automating the construction of uncertainty sets, leveraging machine learning (ML). We utilize covariates to simultaneously learn a linear regression model for predicting the uncertain parameters and adjust the size of the uncertainty set based on the confidence in our predictions. In addition, we introduce optional amendments to our framework so that the uncertainty set bounds are covariate-dependent and also develop an outer approximation scheme for efficiently solving the obtained MIO problem in larger datasets. We apply our framework to uncertainty sets for wind generation for the Adaptive Robust Unit Commitment (ARUC) problem and we show that our approach gives lower probabilities of constraint violation than commonly used statistical approaches, without exhibiting a significant increase in cost.**

***Index Terms*—Robust Optimization, Uncertainty, Machine Learning, Covariates, Unit Commitment**

I. INTRODUCTION

THE increasing presence of renewable energy has created more volatile and unpredictable power grids and electricity markets. For example, the wind and solar penetration in the grid has been steadily increasing over the past few years and has created new reliability challenges. So far, deterministic approaches remain prevalent in practice, but recent advances in robust optimization (RO) offer promising results in addressing the aforementioned uncertainty [1].

RO considers a set of possible scenarios for the uncertain parameters, called the "uncertainty set", and minimizes the cost under the worst-case realization of the uncertain parameters within that uncertainty set [1]. So, the uncertainty sets are at the heart of RO, as their size and structure can have a significant impact on the solution of the RO problem. If the uncertainty set is large, we protect against a large range of realizations of the uncertain parameters, so we are more conservative and the cost increases. Similarly, if the uncertainty set is small, we may exclude scenarios that are of interest and may have to make up for them with costly subsequent actions, e.g., with additional commitments and re-dispatching.

In general, RO has been used to protect against different types of uncertainty — such as uncertainty in the load (e.g., [2]–[5]), wind generation (e.g., [6]–[8]) or both (e.g., [9]–[11]) — and it is shown to have lower average dispatch and total costs, indicating better economic efficiency and lower volatility of the total costs [12]. While RO has been successful in the unit commitment (UC) problem, its applications in power systems and the implications of the uncertainty set modeling extend beyond the reliability unit commitment problems. Specifically, the conversation around overconservativness becomes especially important in light of the utilization of RO in energy markets and in pricing uncertainty. Works like [4], [13]–[15] discuss locational marginal prices in the RO setting, that depend on the size of the uncertainty set. Another set of works considers robust bidding for generators and virtual bidders, where the uncertainty set makes market participants profitable under uncertain market prices [16], [17].

The uncertainty set is usually specified by a set of linear (box and budget sets) or conic constraints (ellipsoidal set). While the box, budget and ellipsoidal sets are part of most of the existing works in RO, there is little guidance in selecting their size [18]. For example, [3]–[5], [7], [12] only report results for different values of the hyper-parameter controlling the size of the budget uncertainty set. While this approach offers insights into the trade-off between robustness and conservativeness, it does not offer a principled framework for selecting the budget size. Similarly, the bounds of the box uncertainty set are usually selected either somewhat arbitrarily or with strong probabilistic assumptions. Of note, the goal of these works is primarily to introduce robust UC formulations, while less focus is placed on the uncertainty set construction.

Recent works have proposed data-driven approaches for constructing uncertainty sets and have suggested new structures that capture the intricate dependencies between the uncertain parameters. [10], [19], [20] include new constraints in the uncertainty set that capture the spatio-temporal correlations in wind generation. Another line of works identifies extreme scenarios in the data and constructs an uncertainty set which is an approximation of the convex hull of the historical realizations [21]–[24]. In a similar spirit, in [8], the authors use a partition and combine scheme to reduce the size of the box uncertainty set by excluding areas with few historical realizations. These works introduce new and refined uncertainty set structures that maintain system reliability at a lower cost than the existing approaches. Still, most of

D. Bertsimas, T. Koukouvinos and A. G. Koulouras are with the Operations Research Center at the Massachusetts Institute of Technology: dbertsim@mit.edu, thkoukouv@mit.edu, angkoul@mit.edu. Research partially supported by the Advanced Research Projects Agency with award number DE-AR0001282.

these works feature some size parameters that are usually not chosen in a data-driven way and are not connected to metrics, such as estimated probabilistic guarantees, that are useful to the stakeholders. Of note, works like [10], [18] use confidence intervals and statistical methods to tie the size to probability metrics, and [21] scale their sets to cover only the most recent realizations based on a rolling time window. In contrast, our approach ties the size to the error of the underlying regression model, which can also be trained by only using a rolling window approach. Therefore, our approach can work in tandem with most of these new structures, under few modifications, and supplement them by offering a solution to the problem of the size parameter tuning.

A. Contributions

In this work, we propose a data-driven framework for constructing uncertainty sets, using mixed integer optimization (MIO) and machine learning (ML), leveraging historical realizations of the uncertain parameters along with related covariates. Our main contributions are summarized as follows:

- We propose a combined MIO and regression formulation, which predicts the nominal values of the uncertain parameters and provides the optimal size of the uncertainty set, for a pre-specified probabilistic guarantee. Our formulation applies to the commonly used uncertainty sets, including the box, budget, ellipsoidal, and box-budget uncertainty sets.
- We prove that, under mild assumptions, the uncertainty sets constructed with our approach satisfy certain probabilistic guarantees controlled by a parameter in the MIO problem.
- We extend the proposed MIO formulation to covariate-dependent bounds for the uncertainty set, allowing the size to adapt with covariate realizations.
- We develop an outer approximation scheme, inspired by [25], to enhance the computational time of our approach.
- We conduct numerical experiments on the adaptive robust unit commitment (ARUC) problem, on the IEEE 14-bus and 118-bus systems, under uncertain wind penetration, using wind forecasting data from a large US Independent System Operator (ISO). We demonstrate that our approach yields uncertainty sets with good empirical probabilities of constraint violation, without a compromise in the cost.

The remainder of the paper is organized as follows: In Section II, we introduce the basic RO concepts. In Section III, we provide the MIO formulations for constructing uncertainty sets using regression models and in Section IV we extend them to covariate-dependent uncertainty set bounds. In Section V we derive an outer approximation scheme for efficiently solving the obtained MIO problems. In Section VI, we introduce the ARUC problem, and in Section VII, we present the numerical experiments. Finally, in Section VIII, we summarize our conclusions.

The notation that we use is as follows: We use bold faced characters such as $\boldsymbol{a}$ to represent vectors and capital letters such as $\boldsymbol{A}$ to represent matrices. The vector $\boldsymbol{e}_i$ denotes the i-th unit vector and $\boldsymbol{e}$ denotes the vector of ones. We define $[I]=\{1,\dots,I\}$. The ℓ_1 norm of a vector is defined as $\|\boldsymbol{x}\|_1=\sum_{i=1}^{I}|x_i|$, the ℓ_2 norm is defined as $\|\boldsymbol{x}\|_2=\sqrt{\sum_{i=1}^{I}x_i^2}$, and the ℓ_∞ norm is defined as $\|\boldsymbol{x}\|_\infty=\max_{i\in[I]}|x_i|$. Let $\boldsymbol{y}$ denote the uncertain parameters and $\boldsymbol{x}$ denote the covariates. We denote the unknown joint distribution of $(\boldsymbol{x},\boldsymbol{y})$, with $\mathbb{P}^*$, the empirical joint distribution of $(\boldsymbol{x},\boldsymbol{y})$ over the training set S with $\mathbb{P}_S$ and the loss of the regression model parametrized by $\boldsymbol{\beta}$ with $L_{\mathcal{W}}(\boldsymbol{\beta})$, where $\mathcal{W}$ denotes the distribution of the data.

II. RO and Uncertainty Set Preliminaries

In this section, we briefly introduce the concepts and preliminaries on robust constraints, probabilistic guarantees, and uncertainty sets, which are useful for the rest of the paper.

Let $\mathcal{U}$ denote an uncertainty set. We consider the following robust constraint:

$$f(\boldsymbol{s},\boldsymbol{y})\le 0,\quad \forall\boldsymbol{y}\in\mathcal{U},$$

where vectors $\boldsymbol{s}$ and $\boldsymbol{y}$ represent the optimization variables and uncertain parameters, respectively. For example, in the capacity constraint of a UC problem, $\boldsymbol{s}$ refers to the production of wind resources and $\boldsymbol{y}$ to their uncertain capacity. The robust constraint, in this case, ensures that the production does not exceed the capacity for all realizations of $\boldsymbol{y}$ in an uncertainty set $\mathcal{U}$ and it is formulated as follows:

$$\boldsymbol{s}\le\boldsymbol{y},\quad \forall\boldsymbol{y}\in\mathcal{U}.$$

In a typical RO problem, the goal is to compute the robust counterpart (RC), which is a deterministic reformulation for the worst-case realization of $\boldsymbol{y}\in\mathcal{U}$. To compute the RC, we need the following assumptions: The function f is linear or concave in $\boldsymbol{y}$ and $\mathcal{U}$ is convex.

In general, we assume that the uncertain parameters are distributed according to an *unknown* probability distribution $\mathbb{P}^*$. Hence, it is important to construct an uncertainty set that is both computationally tractable and implies a probabilistic guarantee for $\mathbb{P}^*$ [26]. For a given $\epsilon>0$, any $\boldsymbol{s}$ and any function f concave in $\boldsymbol{y}$, the uncertainty set $\mathcal{U}$ implies a probabilistic guarantee at level ϵ, if the following holds:

$$f(\boldsymbol{s},\boldsymbol{y})\le 0,\,\forall\boldsymbol{y}\in\mathcal{U}\implies\mathbb{P}^*(f(\boldsymbol{s},\boldsymbol{y})\le 0)\ge 1-\epsilon.$$

In this paper, we focus on norm-bounded uncertainty sets, where the norm difference of the uncertain parameters $\boldsymbol{y}$ from their nominal values $\hat{\boldsymbol{y}}$ is at most δ, i.e.,

$$\mathcal{U}(\hat{\boldsymbol{y}},\delta):=\{\boldsymbol{y}\in\mathbb{R}^n:\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|_\ell\le\delta\},\quad \ell\in\{1,2,\infty\}.$$

We refer to δ as the size of the uncertainty set.

III. Constructing Uncertainty Sets from Regression Models

We propose a standardized framework for constructing uncertainty sets from data. We assume that we have an m-dimensional uncertain parameter vector $\tilde{\boldsymbol{y}}$. For each $\tilde{y}_j$, we assume that we have n historical realizations $y_{1j},\dots,y_{nj}$ as well as covariates $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$, which have predictive power over the uncertain parameters. Our goal is to build an ML model for predicting $\tilde{y}_j$ from the covariates, while finding the optimal uncertainty set bounds. More precisely, our objective is to learn the coefficients of a linear regression model, while also finding the tightest possible norm-bounded uncertainty set such that $p \times 100\%$ of our predicted values in the training set fall in the uncertainty set.

A. Box Uncertainty Set

We first consider a box uncertainty set centered around the predicted values, given by:

$$\mathcal{U} = \left\{ \boldsymbol{y} \in \mathbb{R}^m : \, |y_j - \hat{y}_j| \leq \delta_j, \ \forall j \in [m] \right\},$$

whose size is parameterized by $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$. Given regression coefficients $\boldsymbol{\beta}_j$ and a realization of the covariates $\boldsymbol{x}$, we can obtain a prediction $\hat{y}_j = \boldsymbol{\beta}_j^T \boldsymbol{x}$. We define binary variables z_{ij} such that $z_{ij} = 1$ if the prediction of the j-th coordinate of the uncertain parameters, for data point i, is within δ_j from the realized value and $z_{ij} = 0$ otherwise. We then enforce the empirical probabilistic guarantee requirement that at least p of the residuals are below δ_j, i.e., $\sum_{i=1}^n z_{ij} \geq p \ n$, where p is a pre-specified parameter. The proposed MIO formulation that simultaneously learns the regression coefficients, $\boldsymbol{\beta}$, and the box uncertainty set size, $\boldsymbol{\delta}$, is as follows:

$$\min_{\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{z}} \ \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| + \frac{1}{m} \sum_{j=1}^{m} \delta_j, \tag{1a}$$

$$\text{s.t. } |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| \leq \delta_j + M(1 - z_{ij}), \quad i \in [n], j \in [m], \tag{1b}$$

$$\sum_{i=1}^{n} z_{ij} \geq p \ n, \qquad j \in [m], \tag{1c}$$

$$z_{ij} \in \{0, 1\}, \qquad i \in [n], j \in [m], \tag{1d}$$

where M is a big number. Note that Problem (1) can be solved for each dimension of the uncertain parameters separately and in parallel, speeding up computations.

In Theorem 1, we show that a box uncertainty set constructed via Problem (1) satisfies a certain probabilistic guarantee. We define $\mathcal{U}_j = \{y \in \mathbb{R} : \, |y - \hat{y}_j| \leq \delta_j, \}$, that is, the individual uncertainty set of the j-th dimension.

Theorem 1. *Suppose that $\mathcal{D}$ is a distribution over $X \times Y$ such that with probability 1 we have that $\|\boldsymbol{x}\|_2 \leq R$ and $y_j \leq Q$, $\forall j \in [m]$. Let S be the training set $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^n$. For each coordinate j of $\boldsymbol{y}$, suppose that $\boldsymbol{\beta}_j^*$, that is the optimal solution to Problem (1), satisfies $\|\boldsymbol{\beta}_j^*\|_2 \leq B$ for all j. Then, for any $\epsilon \in (0, 1)$, with probability of at least $1 - \epsilon$ over the choice of an i.i.d. sample of size $n \to \infty$, for all j*

$$\mathbb{P}^*(f(\boldsymbol{s}, \boldsymbol{y}) \leq 0) \geq p. \tag{2}$$

Proof. The MAE $\phi(a, y) = |a - y_j|$ is a ρ-Lipschitz continuous loss function and $\max_{a \in [-BR, BR]} \phi(a, y_j) \leq c$, for $c = BR + Q$ since y_j is bounded by Q. Then, using the Radamacher complexity bounds [27], for any $\epsilon \in (0, 1)$, with probability of at least $1 - \epsilon$,

$$L_{\mathcal{D}}(\boldsymbol{\beta}_j^*) \leq L_S(\boldsymbol{\beta}_j^*) + \frac{2\rho BR}{\sqrt{n}} + c\sqrt{\frac{2\log(2\epsilon)}{n}}.$$

So, for each uncertain parameter j, with probability of at least $1 - \epsilon$,

$$\begin{aligned} \mathbb{P}^*(s_j \leq y_j) &\geq \mathbb{P}^*(y_j \in \mathcal{U}_j) = \mathbb{P}^*(L_{\mathcal{D}}(\boldsymbol{\beta}_j^*) \leq \delta_j) \\ &\geq \mathbb{P}_S\left(L_S(\boldsymbol{\beta}_j^*) + \frac{2\rho BR}{\sqrt{n}} + c\sqrt{\frac{2\log(2\epsilon)}{n}} \leq \delta_j \right) \\ &= \mathbb{P}_S\left(L_S(\boldsymbol{\beta}_j^*) \leq \delta_j - \frac{2\rho BR}{\sqrt{n}} - c\sqrt{\frac{2\log(2\epsilon)}{n}} \right). \end{aligned}$$

Also, $\mathbb{P}^*(s_j \leq y_j) \geq \mathbb{P}^*(y_j \in \mathcal{U}_j) \geq \mathbb{P}_S(L_S(\boldsymbol{\beta}_j^*) \leq \delta_j) \geq p$, for $n \to \infty$, which concludes the proof. □

In many cases, the assumption of bounded $\boldsymbol{y}$ is not too strong. For example, the uncertain wind penetration in the grid is bounded by the installed capacity of wind generators. Of note, Theorem 1 connects the unknown distribution of the uncertain parameters with the empirical distribution obtained from the training data, assuming a sufficiently large training set. As a result, in practice, we can control the out of sample constraint violation with the parameter p.

B. Budget Uncertainty Set

We also consider the budget uncertainty set centered around the predicted values, given by:

$$\mathcal{U} = \left\{ \boldsymbol{y} \in \mathbb{R}^m : \, \sum_{j=1}^{m} |y_j - \hat{y}_j| \leq \Gamma \right\},$$

whose size is parameterized by Γ. We define the binary variables z_i, such that $z_i = 1$ if the sum of the residuals across all dimensions of the uncertain parameters for the i-th data point is at most Γ and $z_i = 0$ otherwise. We then enforce the empirical probabilistic guarantee requirement that at least p of the residuals are below Γ, i.e., $\sum_{i=1}^n z_i \geq p \ n$. The proposed MIO formulation that simultaneously learns the regression coefficients, $\boldsymbol{\beta}$, and the size of the budget uncertainty set, Γ, is as follows:

$$\min_{\boldsymbol{\beta}, \Gamma, \boldsymbol{z}} \ \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| + \frac{1}{m}\Gamma, \tag{3a}$$

$$\text{s.t. } \sum_{j=1}^{m} |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| \leq \Gamma + M(1 - z_i), \ i \in [n], \tag{3b}$$

$$\sum_{i=1}^{n} z_i \geq p \ n, \tag{3c}$$

$$z_i \in \{0, 1\}, \qquad i \in [n], \tag{3d}$$

where M is a big number. We next show that an uncertainty set constructed via Problem (3) satisfies a certain probabilistic guarantee.

Theorem 2. *Suppose that $\mathcal{D}$ is a distribution over $X \times Y$ such that with probability 1 we have that $\|\boldsymbol{x}\|_2 \leq R$ and*

$y_j \leq Q,\ \forall j \in [m]$. Let S be the training set $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^n$. Suppose that $\boldsymbol{\beta}_j^*$, the optimal solution to Problem (3), with $\frac{\Gamma}{m}$ instead of Γ, satisfy $\|\boldsymbol{\beta}_j^*\|_2 \leq B$ for all j. Then, for any $\epsilon \in (0,1)$, with probability of at least $1-\epsilon$ over the choice of an i.i.d. sample of size $n \to \infty$

$$\mathbb{P}^*(f(\boldsymbol{s}, \boldsymbol{y}) \leq 0) \geq 1 - m\ (1-p). \tag{4}$$

Proof. Using the same reasoning as in the proof of Theorem 1, from [27], we obtain that for any $\epsilon \in (0,1)$, with probability of at least $1-\epsilon$, we have that

$$L_{\mathcal{D}}(\boldsymbol{\beta}_i^*) \leq L_S(\boldsymbol{\beta}_i^*) + g(n),$$

where $g(n) = \frac{2\rho BR}{\sqrt{n}} + c\sqrt{\frac{2\log(2\epsilon)}{n}}$

$$\begin{aligned}
\mathbb{P}^*(f(\boldsymbol{s}, \boldsymbol{y}) \leq 0) &\geq \mathbb{P}^*(\boldsymbol{y} \in \mathcal{U}) \\
&= \mathbb{P}^*\left(\sum_{j=1}^m L_{\mathcal{D}}(\boldsymbol{\beta}_j^*) \leq \Gamma\right) \\
&\geq \mathbb{P}^*\left(\cap_{j=1}^m \left\{L_{\mathcal{D}}(\boldsymbol{\beta}_j^*) \leq \frac{\Gamma}{m}\right\}\right) \\
&\geq \sum_{j=1}^m \mathbb{P}^*\left(L_{\mathcal{D}}(\boldsymbol{\beta}_j^*) \leq \frac{\Gamma}{m}\right) - (m-1) \\
&\geq \sum_{j=1}^m \mathbb{P}_S\left(L_S(\boldsymbol{\beta}_j^*) + g(n) \leq \frac{\Gamma}{m}\right) - (m-1) \\
&\geq \sum_{j=1}^m \mathbb{P}_S\left(\sum_{j=1}^m L_S(\boldsymbol{\beta}_j^*) + g(n) \leq \frac{\Gamma}{m}\right) \\
&\quad - (m-1).
\end{aligned}$$

Also, $\mathbb{P}^*(f(\boldsymbol{s}, \boldsymbol{y}) \leq 0) \geq \mathbb{P}^*(\boldsymbol{y} \in \mathcal{U}) \geq \sum_{j=1}^m \mathbb{P}_S(L_S(\boldsymbol{\beta}_j^*) \leq \frac{\Gamma}{m} - g(n)) - (m-1) \geq mp - (m-1) = 1 - m\ (1-p)$, for $n \to \infty$, which concludes the proof. $\square$

We note that the probabilistic guarantee from Theorem 2 is meaningful only in problems where m is not very large.

C. Ellipsoidal Uncertainty Set

Our framework also applies to the ellipsoidal uncertainty set, that is,

$$\mathcal{U} = \{\boldsymbol{y} \in \mathbb{R}^m : \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2 \leq \Gamma\}.$$

In this case, we would have to modify constraint (3b) of Problem (3) as follows:

$$\|\boldsymbol{R}_i\|_2 \leq \Gamma + M(1 - z_i),\ \ i \in [n],$$

where $R_{ij} = y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij},\ \ \forall i, j$. In this case, the resulting optimization problem is mixed integer conic quadratic. Note that an ellipsoidal uncertainty set constructed with the proposed way satisfies the probabilistic guarantee of Theorem 2, for $\frac{\Gamma}{\sqrt{m}}$ instead of Γ. The proof follows from the proof of Theorem 2 and the inequality $\|\boldsymbol{R}_i\|_2 \leq \|\boldsymbol{R}_i\|_1 \leq \sqrt{m}\|\boldsymbol{R}_i\|_2$.

D. Box-Budget Uncertainty Set

Another uncertainty set that is often used in practice is the box-budget uncertainty set, which is defined as

$$\mathcal{U} = \left\{\boldsymbol{y} \in \mathbb{R}^m : \ |y_j - \hat{y}_j| \leq \delta_j, \forall j, \sum_{j=1}^m |y_j - \hat{y}_j| \leq \Gamma\right\}.$$

In this case, we require that at least p of the residuals fall within the box-budget uncertainty set. We utilize binary variables $\boldsymbol{z}, \boldsymbol{u}$ for the box and budget uncertainty sets, respectively. When a training point falls within the uncertainty set, both $\boldsymbol{z}, \boldsymbol{u}$ are one, otherwise we set them both to zero. This is captured by the constraint $\sum_{j=1}^m z_{ij} = m\ u_i$. We extend the formulations of Section III-A and Section III-B as follows:

$$\min_{\boldsymbol{\beta}, \boldsymbol{\delta}, \Gamma, \boldsymbol{z}, \boldsymbol{u}} \ \frac{1}{nm}\sum_{i=1}^n \sum_{j=1}^m |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| + \frac{1}{m}\sum_{j=1}^m \delta_j + \frac{1}{m}\Gamma, \tag{5a}$$

$$\text{s.t. } |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| \leq \delta_j + M(1 - z_{ij}), \quad i \in [n], j \in [m], \tag{5b}$$

$$\sum_{j=1}^m |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| \leq \Gamma + M(1 - u_i),\ \ i \in [n], \tag{5c}$$

$$\sum_{i=1}^n z_{ij} \geq p\ n, \qquad j \in [m], \tag{5d}$$

$$\sum_{i=1}^n u_i \geq p\ n, \tag{5e}$$

$$\sum_{j=1}^m z_{ij} = m\ u_i, \qquad i \in [n], \tag{5f}$$

$$z_{ij}, u_i \in \{0, 1\}, \qquad i \in [n], j \in [m]. \tag{5g}$$

Note that Problem (5) is still a linear MIO problem, with an increased number of binary variables compared to Problems (1) and (3). In addition, the box-budget uncertainty set constructed with Problem (5), satisfies the probabilistic guarantees of Theorem 1 and Theorem 2, simultaneously.

IV. Covariate-Dependent Uncertainty Set Bounds

While the formulations of Section III adjust the nominal prediction of the uncertain parameters based on the covariates, the size is fixed, regardless of the prediction. Since the model can be more or less confident about a prediction for different realizations of the covariates, it makes sense to have a bound that adapts accordingly. We propose that the bounds $\boldsymbol{\delta}$ be a function of the covariates, that is, $\boldsymbol{\delta}(\boldsymbol{x})$. In this case, we obtain the following formulation for a multi-dimensional box uncertainty set:

$$\min_{\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{z}} \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}|y_{ij}-\boldsymbol{\beta}_j^T\boldsymbol{x}_{ij}| + \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\delta_j(\boldsymbol{x}_{ij}), \tag{6a}$$

$$\text{s.t. } |y_{ij}-\boldsymbol{\beta}_j^T\boldsymbol{x}_{ij}| \le \delta_j(\boldsymbol{x}_{ij}) + M(1-z_{ij}), \quad i\in[n], j\in[m], \tag{6b}$$

$$\sum_{i=1}^{n} z_{ij} \ge p\ n, \qquad j\in[m], \tag{6c}$$

$$z_{ij}\in\{0,1\}, \qquad i\in[n], j\in[m]. \tag{6d}$$

In practice, we take δ_j to be an affine function of the covariates, $\delta_j(\boldsymbol{x}) = \delta_j^0 + \boldsymbol{q}_j^T\boldsymbol{x}$, where δ_j^0 and $\boldsymbol{q}_j$ are additional variables in Problem (6). Then, assuming we have solved Problem (6), for every obtained prediction in the test set $\boldsymbol{\beta}_j^T\overline{\boldsymbol{x}}$, we obtain a corresponding bound $\boldsymbol{\delta}_j(\overline{\boldsymbol{x}})$. We note that the application of the covariate-dependent bounds to the other uncertainty sets discussed in Section III is straightforward, by replacing δ with $\delta(\boldsymbol{x})$.

V. Outer Approximation Scheme for MIO

The MIO problems derived in Section III involve binary variables with dimension equal to the number of training data points, which makes them computationally intractable in larger datasets. However, in practice we often have access to more data that we can further leverage when constructing the uncertainty set. Thus, we propose an outer approximation scheme, based on the framework introduced by [25], to enhance the computational time of our approach.

A. Box uncertainty set

In this section, we derive an outer approximation scheme for solving Problem (1) in case of $m=1$. Observe that for multivariate uncertain parameters we can solve one such problem for each dimension. First, observe that Problem (1) for $m=1$, can be equivalently formulated as follows:

$$\begin{aligned} \min_{\boldsymbol{z}\in\mathcal{Z}}\ \min_{\boldsymbol{\beta},\delta} \quad & \frac{1}{n}\sum_{i=1}^{n}|y_i-\boldsymbol{\beta}^T\boldsymbol{x}_i| + \delta \\ \text{s.t.} \quad & z_i|y_i-\boldsymbol{\beta}^T\boldsymbol{x}_i| \le \delta, \quad i\in[n], \end{aligned} \tag{7}$$

where $\mathcal{Z} = \{\boldsymbol{z}\in\{0,1\}^n : \sum_{i=1}^n z_i \ge p\ n\}$. Let $f(\boldsymbol{z})$ denote the optimal value of the inner problem over $\boldsymbol{\beta}, \delta$, that is,

$$\begin{aligned} f(\boldsymbol{z}) \ = \ \min_{\boldsymbol{\beta},\delta} \quad & \frac{1}{n}\sum_{i=1}^{n}|y_i-\boldsymbol{\beta}^T\boldsymbol{x}_i| + \delta \\ \text{s.t.} \quad & z_i|y_i-\boldsymbol{\beta}^T\boldsymbol{x}_i| \le \delta, \quad i\in[n]. \end{aligned} \tag{8}$$

We then have the following outer problem:

$$\min_{\boldsymbol{z}\in\mathcal{Z}} f(\boldsymbol{z}).$$

We next derive the dual of the inner problem. We use the notation $\boldsymbol{X}$ for the matrix with i-th row equal to $\boldsymbol{x}_i^T$.

Proposition 1. *The dual of Problem* (8)*, is as follows:*

$$\begin{aligned} \max_{\boldsymbol{\mu},\boldsymbol{\theta}} \quad & \boldsymbol{y}^T diag(\boldsymbol{z})\boldsymbol{\mu} - 2\boldsymbol{\theta}^T\boldsymbol{y} + \frac{1}{n}\boldsymbol{e}^T\boldsymbol{y} \\ \text{s.t.} \quad & \sum_{i=1}^{n}\mu_i = 1, \\ & \boldsymbol{X}^T(2\boldsymbol{\theta} - \frac{1}{n}\boldsymbol{e} - diag(\boldsymbol{z})\boldsymbol{\mu}) = \boldsymbol{0}, \\ & \frac{1}{n}\boldsymbol{e} + diag(\boldsymbol{z})\boldsymbol{\mu} - \boldsymbol{\theta} \ge \boldsymbol{0}, \\ & \boldsymbol{\mu},\boldsymbol{\theta} \ge \boldsymbol{0}. \end{aligned} \tag{9}$$

Proof. Using auxiliary variables $\boldsymbol{u}$ we obtain the following primal problem

$$\begin{aligned} \min_{\boldsymbol{\beta},\boldsymbol{u},\delta} \quad & \frac{1}{n}\sum_{i=1}^{n}u_i + \delta \\ \text{s.t.} \quad & z_iu_i \le \delta, \qquad i\in[n], \\ & -\boldsymbol{u} \le \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \le \boldsymbol{u}. \end{aligned}$$

The dual objective is as follows:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\theta}) &= \inf_{\boldsymbol{\beta},\boldsymbol{u},\delta}\{\frac{1}{n}\boldsymbol{e}^T\boldsymbol{u} + \delta + \sum_{i=1}^{n}\mu_i(z_iu_i - \delta) \\ &\quad + \boldsymbol{\nu}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{u}) + \boldsymbol{\theta}^T(-\boldsymbol{u} - \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta})\} \\ &= \boldsymbol{\nu}^T\boldsymbol{y} - \boldsymbol{\theta}^T\boldsymbol{y} + \inf_{\delta}\left\{\left(1 - \sum_{i=1}^{n}\mu_i\right)\delta\right\} \\ &\quad + \inf_{\boldsymbol{u}}\left\{\left(\frac{1}{n}\boldsymbol{e} + \text{diag}(\boldsymbol{z})\boldsymbol{\mu} - \boldsymbol{\nu} - \boldsymbol{\theta}\right)^T\boldsymbol{u}\right\} \\ &\quad + \inf_{\boldsymbol{\beta}}\left\{\left(\boldsymbol{X}^T\boldsymbol{\theta} - \boldsymbol{X}^T\boldsymbol{\nu}\right)^T\boldsymbol{\beta}\right\} \\ &= \boldsymbol{\nu}^T\boldsymbol{y} - \boldsymbol{\theta}^T\boldsymbol{y}, \end{aligned}$$

as long as,

$$\begin{aligned} & 1 - \sum_{i=1}^{n}\mu_i = 0, \\ & \boldsymbol{X}^T\boldsymbol{\theta} - \boldsymbol{X}^T\boldsymbol{\nu} = \boldsymbol{0}, \\ & \frac{1}{n}\boldsymbol{e} + \text{diag}(\boldsymbol{z})\boldsymbol{\mu} - \boldsymbol{\nu} - \boldsymbol{\theta} = \boldsymbol{0}, \end{aligned}$$

The result then follows if we substitute the value of $\boldsymbol{\nu}$ from the second constraint. □

Observe that the function $f(\boldsymbol{z})$, which is defined as the dual objective, is convex as it is the point-wise maximum over affine functions of $\boldsymbol{z}$. Thus, it satisfies the following inequality:

$$f(\tilde{\boldsymbol{z}}) \ge f(\boldsymbol{z}) + \nabla f(\boldsymbol{z})^T(\tilde{\boldsymbol{z}} - \boldsymbol{z}), \quad \forall \tilde{\boldsymbol{z}}, \boldsymbol{z}. \tag{10}$$

[25] proposed to minimize $f(\boldsymbol{z})$, by iteratively minimizing valid piecewise linear lower-approximations, defined by (10).

Starting from some $\boldsymbol{z}^0$, at iteration t, the outer problem is as follows:

$$\begin{aligned} \min_{\eta,\boldsymbol{z}} \quad & \eta \\ \text{s.t.} \quad & \eta \geq f(\boldsymbol{z}^i) + \nabla_{\boldsymbol{z}} f(\boldsymbol{z}^i)^T(\boldsymbol{z}-\boldsymbol{z}^i), \qquad i \in [t-1], \\ & \sum_{j=1}^{n} z_j \geq p\ n, \\ & z \in \{0,1\}^n, \end{aligned} \tag{11}$$

where $\nabla_{\boldsymbol{z}} f(\boldsymbol{z}^i)$ denotes a sub-gradient of the function f at $\boldsymbol{z}^i$. For known $\boldsymbol{z}$, we can obtain $f(\boldsymbol{z})$ as the optimal value of Problem (9). Further, we can also compute the sub-gradient. Let $\boldsymbol{\mu}^*, \boldsymbol{\theta}^*$ denote the optimal solutions of Problem (9). Then, we can compute the i-th coordinate of the sub-gradient as follows:

$$(\nabla_{\boldsymbol{z}} f(\boldsymbol{z}))_i = \mu_i^* y_i.$$

Finally, we note that from the optimal solution $\boldsymbol{z}^*$, we can retrieve the optimal values $\boldsymbol{\beta}^*, \delta^*$ by solving the following continuous optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\beta},\delta} \quad & \frac{1}{n}\sum_{i=1}^{n} |y_i - \boldsymbol{\beta}^T \boldsymbol{x}_i| + \delta \\ \text{s.t.} \quad & z_i^* |y_i - \boldsymbol{\beta}^T \boldsymbol{x}_i| \leq \delta, \qquad i \in [n]. \end{aligned} \tag{12}$$

The proposed cutting planes approach is summarized in pseudocode in Algorithm 1.

Algorithm 1 Outer approximation scheme for Problem (1)

Input: Data $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$, initialization $\boldsymbol{z}^0$.
Output: Optimal solutions, $\boldsymbol{\beta}^*, \delta^*$.

```
t = 1
Let f(z^{t+1}) = ∞, η^{t+1} = −∞
while f(z^{t+1}) − η^{t+1} > ϵ do
    Solve Problem (11), with input z^0, ..., z^t and obtain
    optimal solutions η^{t+1}, z^{t+1}
    Solve Problem (9), with input z^{t+1}, and obtain
    f(z^{t+1}), ∇f(z^{t+1})
    t = t + 1
end while
Solve Problem (12) with input z* and obtain β*, δ*
return β*, δ*
```

The method described in Algorithm 1 was originally proposed for continuous decision variables by [28], and later extended to binary decision variables by [29], while also providing a proof of termination in a finite, yet exponential in the worst case, number of iterations. We implement Algorithm 1 using lazy callbacks, by constructing a single branch-and-bound tree and generating a new cut at a feasible solution $\boldsymbol{z}$, as originally proposed by [30].

Finally, we note that there are many options for initializing Algorithm 1. One approach is to randomly sample $p\,n$ indices, set those to 1 and the remaining ones to 0. Instead, we can fit a linear regression model and then set to 1 the indices corresponding to residuals below the p quantile and the remaining ones to 0. This can serve as a good warm-start for Algorithm 1.

B. Budget uncertainty set

In this section, we derive an outer approximation scheme for solving Problem (3), using the same methodology as for the box uncertainty set. In this case, for fixed $\boldsymbol{z}$, the inner problem is as follows:

$$f(\boldsymbol{z}) = \begin{aligned} \min_{\boldsymbol{\beta},\Gamma} \quad & \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| + \frac{1}{m}\Gamma \\ \text{s.t.} \quad & z_i \sum_{j=1}^{m} |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| \leq \Gamma, \quad i \in [n]. \end{aligned} \tag{13}$$

The constraints for the outer problem are the same as for the box uncertainty set. We next derive the dual of the inner problem.

The dual of Problem (13), is as follows:

$$\begin{aligned} \max_{\boldsymbol{\mu},\boldsymbol{\theta}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{m} (\mu_i z_i + \frac{1}{nm} - 2\theta_{ij}) y_{ij} \\ \text{s.t.} \quad & \sum_{i=1}^{n} \mu_i = \frac{1}{m}, \\ & \sum_{i=1}^{n} \left(2\theta_{ij} - \frac{1}{nm} - \mu_i z_i\right) \boldsymbol{x}_{ij} = \boldsymbol{0}, \quad j \in [m], \\ & \frac{1}{nm} + \mu_i z_i - \theta_{ij} \geq 0, \quad i \in [n], j \in [m], \\ & \mu_i, \theta_{ij} \geq 0. \end{aligned} \tag{14}$$

The function $f(\boldsymbol{z})$ is convex as it is the point-wise maximum over affine functions of $\boldsymbol{z}$, thus we can solve the problem using Algorithm 1. Let $\boldsymbol{\mu}^*, \boldsymbol{\theta}^*$ denote the optimal solutions of Problem (14). In this case the i-th coordinate of the sub-gradient can be computed as follows:

$$(\nabla_{\boldsymbol{z}} f(\boldsymbol{z}))_i = \sum_{j=1}^{m} \mu_i^* y_{ij}.$$

We can initialize Algorithm 1 as follows: We first fit a linear regression model on each data set $(\boldsymbol{X}^j, \boldsymbol{y}^j)$ and obtain coefficients $\boldsymbol{\beta}_j$. Then, we set to 1 the indices i for which $r_i = \sum_j |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| \leq q_p$, where q_p denotes the p quantile of the residuals, and set to 0 the remaining ones.

We note that Algorithm 1 also applies to the box-budget uncertainty set. In this case, using the same arguments, that is deriving the dual of Problem (5) for fixed $\boldsymbol{z}, \boldsymbol{u}$, it can be shown that the objective function is convex in both $\boldsymbol{z}, \boldsymbol{u}$, as it is the point-wise maximum over affine functions of $\boldsymbol{z}, \boldsymbol{u}$. We can then apply Algorithm 1, with a modified version of Problem (11), involving the additional constraints for the binary variables $\boldsymbol{z}, \boldsymbol{u}$ from Problem (5). Finally, we note that by leveraging strong duality in convex optimization, we can also apply Algorithm 1 for the ellipsoidal uncertainty set.

VI. Adaptive Robust Unit Commitment

In this section, we consider the ARUC problem, which we utilize in the numerical examples of Section VII. We first provide the formulation of the deterministic UC problem.

We use the following set notation:

- $\mathcal{K}$: Set of generators.

- $\mathcal{R}$: Set of wind resources.
- $\mathcal{T}$: Set of time periods.
- $\mathcal{K}_b$: Set of generators located at bus b.
- $\mathcal{R}_b$: Set of wind resources located at bus b.

We use the following parameters:

- p_i^{min}, p_i^{max}: Min/max production levels of generator i.
- RU_{it}, RD_{it}: Ramp-up/down rates of generator i at time t.
- F_l: Capacity of transmission line l.
- K_{lb}: Line flow distribution factor for transmission line l at bus b.
- D_{bt}: Power demand at bus b during time t.
- F_{it}, C_{it}: Operating/production cost for generator i at time t.

We use the following variables:

- $x_{it} \in \{0,1\}$: If generator i is on at time t.
- $u_{it} \in \{0,1\}$: If generator i is turned on at time t.
- $v_{it} \in \{0,1\}$: If generator i is turned down at time t.
- $p_{it} \geq 0$: Production of generator (or wind resource) i at time t.

A. Deterministic UC Formulation

The deterministic UC formulation is based on [10], [31]. The problem formulation is as follows:

$$\min_{\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{p}} \quad \sum_{i\in\mathcal{K}}\sum_{t\in\mathcal{T}} F_{it}x_{it} + C_{it}p_{it} \tag{15a}$$

$$\text{s.t.} \quad (\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}) \in \mathcal{X}, \tag{15b}$$

$$\sum_{i\in\mathcal{K}} p_{it} + \sum_{i\in\mathcal{R}} p_{it} \geq \sum_b D_{bt}, \quad t\in\mathcal{T}, \tag{15c}$$

$$\sum_b K_{lb}\left(\sum_{i\in\mathcal{K}_b} p_{it} + \sum_{i\in\mathcal{R}_b} p_{it} - D_{bt}\right) \leq F_l, \tag{15d}$$

$$-\sum_b K_{lb}\left(\sum_{i\in\mathcal{K}_b} p_{it} + \sum_{i\in\mathcal{R}_b} p_{it} - D_{bt}\right) \leq F_l, \tag{15e}$$

$$p_i^{\min}x_{it} \leq p_{it} \leq p_i^{\max}x_{it}, \quad i\in\mathcal{K}, t\in\mathcal{T}, \tag{15f}$$

$$p_{it}^{\min} \leq p_{it} \leq p_{it}^{\max}, \quad i\in\mathcal{R}, t\in\mathcal{T}, \tag{15g}$$

$$-RD_{it} \leq p_{it} - p_{i(t-1)} \leq RU_{it}, \quad i\in\mathcal{K}, t\in\mathcal{T}. \tag{15h}$$

The set $\mathcal{X}$ includes the constraints $x_{i(t-1)} - x_{it} + u_{it} \geq 0$, $x_{it} - x_{i(t-1)} + v_{it} \geq 0$, $x_{it} - x_{i(t-1)} \leq x_{i\tau}$, $\tau \in [t+1, \min\{t+\text{MinUp}_i - 1, T\}]$, $x_{i(t-1)} - x_{it} \leq 1 - x_{i\tau}$, $\tau \in [t+1, \min\{t+\text{MinDw}_i - 1, T\}]$. The first two are the turn-on and turn-off constraints. In particular, a generator is turned on at time t, if $x_{i(t-1)} = 0$ and $x_{it} = 1$. Similarly, a generator is turned off at time t, if $x_{i(t-1)} = 1$ and $x_{it} = 0$. The last two constraints capture the minimum up and minimum down times of each generator. For example, if a generator is turned on at time t, then it must remain on at least for the next MinUp periods. Constraint (15c) is the energy balance equation that matches the system level supply and load at each time period. Constraints (15d) and (15e) follow the dc power flow model in [32] to confine the power transmission below the line capacity F_l. Constraints (15f) and (15g) indicate that committed generators and wind resources can be dispatched within a range. Constraint (15h) is the ramp rate constraint, i.e., the speed at which a generator can increase or decrease its production level is bounded in a range.

B. ARUC with Linear Decision Rules

Following the approach of [31], we utilize linear decision rules (LDR) in order to obtain a computationally tractable approximation of Problem (15). LDR often achieve near-optimal performance, see [1]. Let z_{it} denote the uncertain capacity of wind resource i at time t (denoted as $p_{it}^{\max}$ in Problem (15)), $\boldsymbol{z}_t$ denote the vector of capacities of all wind resources at time t and $\hat{\boldsymbol{z}}_t$ denote the vector of predicted capacities. Let $\boldsymbol{r}_t = \boldsymbol{z}_t - \hat{\boldsymbol{z}}_t$. We then impose the decision rule

$$p_{it}(\boldsymbol{r}_t) = p_{it}^0 + \boldsymbol{w}_{it}^T\boldsymbol{r}_t.$$

We consider the box-budget uncertainty set $\mathcal{Z}_t = \{\boldsymbol{r}_t : |r_{t_i}| \leq \delta_{t_i},\ \sum_{i=1}^m |r_{t_i}| \leq \Gamma_t\}$, with box bounds $\boldsymbol{\delta}_t$ for all wind resources at time t. The final RC is obtained by following the methodology from [1], that is deriving the dual for each constraint, and it is formulated as follows:

$$\min_{\substack{\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\\ \boldsymbol{p},\boldsymbol{p}^0,\boldsymbol{w}}} \quad \sum_{i\in\mathcal{K}}\sum_{t\in\mathcal{T}} F_{it}x_{it} + C_{it}p_{it} \tag{16a}$$

$$\text{s.t.} \quad (\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}) \in \mathcal{X}, \tag{16b}$$

$$\sum_{i\in\mathcal{K}} p_{it} + \sum_{i\in\mathcal{R}} p_{it}^0 - \boldsymbol{\delta}_t^T\boldsymbol{\lambda}_t^{\text{pb}} - \Gamma_t\mu_t^{\text{pb}} \geq \sum_b D_{bt}, \forall t, \tag{16c}$$

$$\sum_b K_{lb}\left(\sum_{i\in\mathcal{K}_b} p_{it} + \sum_{i\in\mathcal{R}_b} p_{it}^0 - D_{bt}\right) + \boldsymbol{\delta}_t^T\boldsymbol{\lambda}_{lt}^{\text{tr, max}} + \Gamma_t\mu_{lt}^{\text{tr, max}} \leq F_l, \quad \forall l,t, \tag{16d}$$

$$-\sum_b K_{lb}\left(\sum_{i\in\mathcal{K}_b} p_{it} + \sum_{i\in\mathcal{R}_b} p_{it}^0 - D_{bt}\right) + \boldsymbol{\delta}_t^T\boldsymbol{\lambda}_{lt}^{\text{tr, min}} + \Gamma_t\mu_{lt}^{\text{tr, min}} \leq F_l, \quad \forall l,t, \tag{16e}$$

$$p_i^{\min}x_{it} \leq p_{it} \leq p_i^{\max}x_{it}, \quad \forall i,t, \tag{16f}$$

$$p_{it}^0 + \boldsymbol{\delta}_t^T\boldsymbol{\lambda}_{it}^{\text{cap, max}} + \Gamma_t\mu_{it}^{\text{cap, max}} \leq \boldsymbol{e}_i^T\hat{z}_t, \quad \forall i,t, \tag{16g}$$

$$p_{it}^0 - \boldsymbol{\delta}_t^T\boldsymbol{\lambda}_{it}^{\text{cap, min}} - \Gamma_t\mu_{it}^{\text{cap, min}} \geq p_{it}^{\min}, \quad \forall i,t, \tag{16h}$$

$$-RD_{it} \leq p_{it} - p_{i(t-1)} \leq RU_{it}, \quad \forall i,t, \tag{16i}$$

$$\left(\boldsymbol{\lambda}_t^{\text{pb}}, \mu_t^{\text{pb}}, \boldsymbol{\nu}_t^{\text{pb}}, \boldsymbol{\pi}_t^{\text{pb}}\right) \in \mathcal{F}_t^{\text{pb}}, \tag{16j}$$

$$\left(\boldsymbol{\lambda}_{it}^{\text{cap, max}}, \mu_{it}^{\text{cap, max}}, \boldsymbol{\nu}_{it}^{\text{cap, max}}, \boldsymbol{\pi}_{it}^{\text{cap, max}}\right) \in \mathcal{F}_{it}^{\text{cap, max}}, \tag{16k}$$

$$\left(\boldsymbol{\lambda}_{it}^{\text{cap, min}}, \mu_{it}^{\text{cap, min}}, \boldsymbol{\nu}_{it}^{\text{cap, min}}, \boldsymbol{\pi}_{it}^{\text{cap, min}}\right) \in \mathcal{F}_{it}^{\text{cap, min}}, \tag{16l}$$

$$\left(\boldsymbol{\lambda}_{lt}^{\text{tr, max}}, \mu_{lt}^{\text{tr, max}}, \boldsymbol{\nu}_{lt}^{\text{tr, max}}, \boldsymbol{\pi}_{lt}^{\text{tr, max}}\right) \in \mathcal{F}_{lt}^{\text{tr, max}}, \tag{16m}$$

$$\left(\boldsymbol{\lambda}_{lt}^{\text{tr, min}}, \mu_{lt}^{\text{tr, min}}, \boldsymbol{\nu}_{lt}^{\text{tr, min}}, \boldsymbol{\pi}_{lt}^{\text{tr, min}}\right) \in \mathcal{F}_{lt}^{\text{tr, min}}. \tag{16n}$$

We note that for dual variables $(\boldsymbol{\lambda}_t, \mu_t, \boldsymbol{\nu}_t, \boldsymbol{\pi}_t)$, we have $\mathcal{F}_t =$

$$\left\{(\boldsymbol{\lambda}_t, \mu_t, \boldsymbol{\nu}_t, \boldsymbol{\pi}_t) \geq \boldsymbol{0} : \ \boldsymbol{\nu}_t - \boldsymbol{\pi}_t = \boldsymbol{a}, \boldsymbol{\nu}_t + \boldsymbol{\pi}_t = \boldsymbol{\lambda}_t + \mu_t\boldsymbol{e}\right\},$$

where the vector $\boldsymbol{a}$ contains the constraint coefficients that multiply the uncertain parameters.

Additional reserve constraints can be added with small changes. In addition, the spinning reserve requirements in Problem (16) can be implicitly satisfied by the uncertainty set, see [2]. While we utilize a UC setup for the experiments, we note that our method is not dependent on the underlying RO problem. For example, we could utilize covariates and a regression model for predicting the market prices of the Day-Ahead Market and set up a robust bidding problem [16].

VII. Numerical Experiments

In this section, we apply our method to wind forecasting models from a vendor that caters to a large US ISO. We use 24-hour forecasts for the operating day, for a selection of 5 wind resources, and 5 forecasting models, along with realizations of the available capacity. The training set includes the three days prior to the day of interest, which is the test set. The relatively short time horizon of the training set allows training on days that are close to the test day. We conduct numerical experiments on the IEEE 14-bus system (https://icseg.iti.illinois.edu/ieee-14-bus-system) as well as the IEEE 118-bus system (https://icseg.iti.illinois.edu/ieee-118-bus-system). We focus on the box-budget uncertainty set that is often used in the ARUC problem [12], [31]. The transmission line flow limits are set to 120 MW.

We compare our approach with a statistical method for constructing the box bounds, which sets the uncertainty set bounds to be the 100 $p\%$ confidence interval (CI) under a normal distribution [18]. The standard deviation is obtained as the empirical standard deviation of the 5 forecasting models predictions at each wind resource, at each time period. We note that the approach from [18] results in extremely conservative budget bounds, thus we utilize the budget size Γ obtained by our method.

A. IEEE 14-bus system

We first consider the IEEE 14-bus system, which contains 5 units, 11 loads, and 18 transmission lines. We place 5 wind generators on the grid at random. Refer to Table I and Fig. 1, for the corresponding locations. The nominal capacity of each wind resource, for each time period, is obtained as the average of the predictions of the 5 forecasting models. We scaled both the capacity of the 5 thermal units and the loads to $4\times$ their nominal values to obtain a wind penetration of 33% and avoid infeasibilities. So, this example relies strongly on wind power.

TABLE I
Wind resource capacities and locations on the IEEE 14-bus system.

Wind Resource	Bus	Nominal Capacity (MW)
1	8	109.8
2	3	120.6
3	3	88.4
4	2	47.5
5	11	100.1

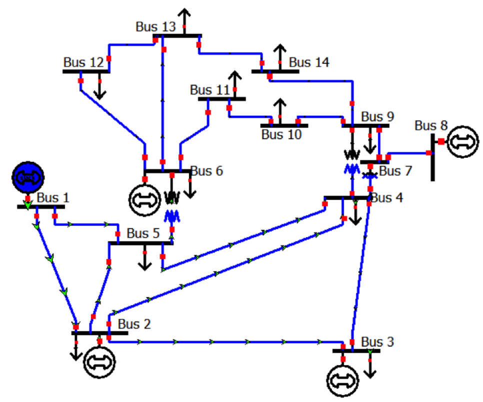

Fig. 1. The IEEE 14-bus system grid.

We report the objective, the probability of constraint violation and the actual constraint violation (in MW) under the deterministic and the adaptive robust models in Tables II, III and IV, respectively. The first column contains the results for the deterministic problem (Det). The second column reports the results for the ARUC problem under the confidence interval box bounds (CI). Further, in columns 3 and 4, we provide results for Problem (5), with constant (Const. MIO) and covariate-dependent (Cov. MIO) bounds. We note that for the latter, we utilize a linear function, see Section IV.

TABLE II
Objective values in \$ $\times 10^3$ for nominal and robust UC.

p	Det	CI	Const. MIO	Cov. MIO
0.70	894.2	929.3	949.0	953.6
0.75	894.2	933.5	962.8	962.4
0.80	894.2	940.0	967.9	970.5
0.85	894.2	945.6	963.4	951.4
0.90	894.2	953.0	980.3	966.7
0.95	894.2	964.5	992.1	998.9

From Table II, we observe that in most cases the objective increases, as we increase the probabilistic guarantee p, because we protect against more realizations of the uncertainty. In addition, the covariate-dependent bounds sometimes have lower cost for the same probabilistic guarantee.

TABLE III
Probability of constraint violation in the capacity constraints of wind resources for the nominal and robust UC models.

p	Det	CI	Const. MIO	Cov. MIO
0.70	0.30	0.08	0.00	0.03
0.75	0.30	0.07	0.02	0.08
0.80	0.30	0.04	0.00	0.03
0.85	0.30	0.06	0.01	0.05
0.90	0.30	0.04	0.00	0.05
0.95	0.30	0.03	0.02	0.02

From Table III, we observe that our approach provides better empirical probabilities of constraint violation than both the deterministic problem and the confidence intervals approach, across the board. We further notice that our approach with constant bounds provides slightly better probabilistic guaran-

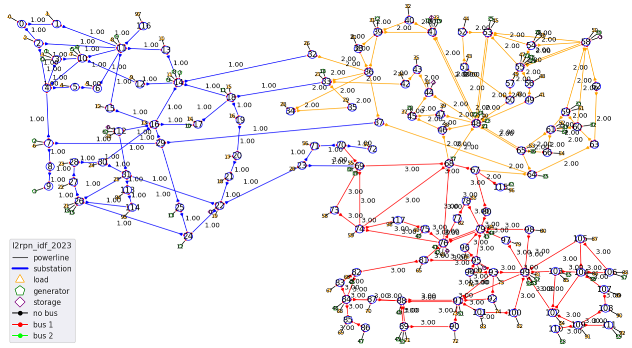

Fig. 2. The IEEE 118-bus system grid.

tees than the covariate-dependent bounds, while being more conservative.

TABLE IV
CONSTRAINT VIOLATION IN THE CAPACITY CONSTRAINTS OF WIND RESOURCES FOR THE NOMINAL AND ROBUST UC MODELS.

p	**Det**	**CI**	**Const. MIO**	**Cov. MIO**
0.70	431.9	66.5	0.0	10.3
0.75	431.9	48.1	12.1	52.3
0.80	431.9	40.7	0.0	24.5
0.85	431.9	32.4	8.6	60.2
0.90	431.9	25.3	0.0	26.4
0.95	431.9	11.3	5.9	15.9

Table IV reports the difference between the realized capacity and the scheduled dispatch of wind resources, when the dispatch exceeds the realized capacity. We notice that our method achieves lower actual constraint violation than both the deterministic problem as well as the robust problem with confidence intervals. So, when using RO, we need fewer reserves in the Real-Time Market, especially under our approach.

B. IEEE 118-bus system

In this section, we consider the IEEE 118-bus system, which contains 54 units, 91 loads, and 186 transmission lines. We place 5 wind generators on the grid, whose bus locations are provided in Table V. Refer to Figure 2, for the corresponding locations. Again, the capacity of each wind resource, for each time period, is obtained as the average of 5 forecast models. The wind penetration in this example is 11%.

TABLE V
WIND RESOURCE CAPACITIES AND LOCATIONS ON THE IEEE 118-BUS SYSTEM.

Wind Resource	**Bus**	**Nominal Capacity (MW)**
1	8	109.8
2	3	120.6
3	3	88.4
4	2	47.5
5	11	100.1

We report the objective, the probability of constraint violation and the actual constraint violation (in MW) under the deterministic and the adaptive robust models in Tables VI, VII and VIII, respectively.

TABLE VI
OBJECTIVE VALUES IN \$ $\times 10^6$ FOR NOMINAL AND ROBUST UC.

p	**Det**	**CI**	**Const. MIO**	**Cov. MIO**
0.70	3.329	3.390	3.381	3.400
0.75	3.329	3.393	3.388	3.406
0.80	3.329	3.397	3.393	3.414
0.85	3.329	3.394	3.402	3.390
0.90	3.329	3.353	3.416	3.417
0.95	3.329	3.417	3.415	3.442

TABLE VII
PROBABILITY OF CONSTRAINT VIOLATION IN THE CAPACITY CONSTRAINTS OF WIND RESOURCES FOR THE NOMINAL AND ROBUST UC MODELS.

p	**Det**	**CI**	**Const. MIO**	**Cov. MIO**
0.70	0.28	0.11	0.04	0.04
0.75	0.28	0.03	0.01	0.08
0.80	0.28	0.11	0.01	0.03
0.85	0.28	0.01	0.02	0.04
0.90	0.28	0.07	0.00	0.02
0.95	0.28	0.01	0.00	0.00

From Tables VI and VII, we observe that our approach achieves lower empirical probabilities of constraint violation than both the deterministic problem and the robust problem with confidence intervals, generally, without a compromise in the cost. Further, from Table VIII we observe that our approach reduces the amount of violation compared to the deterministic and robust problems, reducing the need for reserves. Specifically, if we use the bounds corresponding to the 95% probabilistic guarantee, we would not need any spinning reserves for wind, since there is no violation in any of the capacity constraints, across all time periods. This holds in both the 14-bus and 118-bus system examples.

TABLE VIII
CONSTRAINT VIOLATION IN THE CAPACITY CONSTRAINTS OF WIND RESOURCES FOR THE NOMINAL AND ROBUST UC MODELS.

p	**Det**	**CI**	**Const. MIO**	**Cov. MIO**
0.70	371.7	407.2	42.0	60.3
0.75	371.7	18.9	2.1	216.4
0.80	371.7	77.1	11.0	21.8
0.85	371.7	5.8	8.7	49.9
0.90	371.7	35.5	0.0	34.2
0.95	371.7	9.6	0.0	0.0

Of note, all optimization problems are implemented using Julia 1.5.3 and the Julia package `JuMP.jl` version 0.21.6. A gap of 0.05% is obtained for Problem (16) on the IEEE 118-bus system within 20 minutes. In Problem (5), we take $M = 10^4$.

C. Improving Scalability via the Outer Approximation Scheme

In this section, we illustrate how we can speed-up our method using the outer approximation scheme that we developed in Section V. We construct a box uncertainty set by solving Problem (1), either directly or via the outer approximation scheme outlined in Algorithm 1. We fix $p = 0.9$ and vary the number of days used in the training set from 3 to 8.

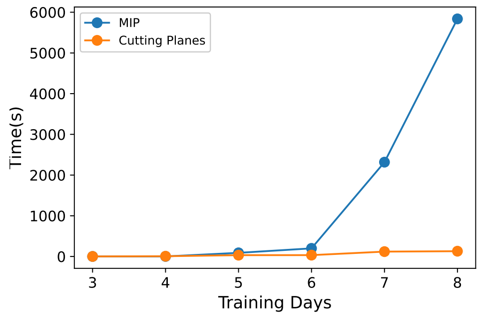

Fig. 3. Computational time with number of days in the training set.

In Figure 3, we see a significant speed-up when using Algorithm 1 and 7 or 8 training days. In practice, we can use Algorithm 1 to leverage more data during training.

VIII. Conclusion

In summary, we developed a framework for constructing uncertainty sets by utilizing covariates to predict the uncertain parameters and find the optimal size of the uncertainty set, for a pre-specified in sample probabilistic guarantee. We showed that the uncertainty sets constructed by our approach satisfy certain out of sample probabilistic guarantees. Moreover, we applied our method to the ARUC problem under uncertain wind resource capacity and we showed that our method exhibits small constraint violations at a similar cost compared to statistical approaches.

References

[1] D. Bertsimas and D. den Hertog, *Robust and adaptive optimization*. Dynamic Ideas, 2022.

[2] R. Jiang, M. Zhang, G. Li, and Y. Guan, "Two-stage network constrained robust unit commitment problem," *European Journal of Operational Research*, vol. 234, no. 3, pp. 751–762, 2014.

[3] Y. An and B. Zeng, "Exploring the modeling capacity of two-stage robust optimization: Variants of robust unit commitment model," *IEEE transactions on Power Systems*, vol. 30, no. 1, pp. 109–122, 2014.

[4] H. Ye, Y. Ge, M. Shahidehpour, and Z. Li, "Uncertainty marginal price, transmission reserve, and day-ahead market clearing with robust unit commitment," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 1782–1795, 2016.

[5] M. I. Alizadeh, M. P. Moghaddam, and N. Amjady, "Multistage multiresolution robust unit commitment with nondeterministic flexible ramp considering load and wind variabilities," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 872–883, 2017.

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

[7] P. Xiong and P. Jirutitijaroen, "A linear decision rule approach for robust unit commitment considering wind power generation," in *2014 Power Systems Computation Conference*, pp. 1–7, IEEE, 2014.

[8] M. Zhang, J. Fang, X. Ai, B. Zhou, W. Yao, Q. Wu, and J. Wen, "Partition-combine uncertainty set for robust unit commitment," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 3266–3269, 2020.

[9] C. Lee, C. Liu, S. Mehrotra, and M. Shahidehpour, "Modeling transmission line constraints in two-stage robust unit commitment problem," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1221–1231, 2013.

[10] C. Li, J. Zhao, T. Zheng, and E. Litvinov, "Data-driven uncertainty sets: Robust optimization with temporally and spatially correlated data," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, 2016.

[11] H. Jin, H. Sun, Q. Guo, and J. Wu, "Robust unit commitment considering reserve from grid-scale energy storage," in *2016 IEEE 8th International Power Electronics and Motion Control Conference (IPEMC-ECCE Asia)*, pp. 246–251, IEEE, 2016.

[12] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE transactions on power systems*, vol. 28, no. 1, pp. 52–63, 2012.

[13] Z. Zhao, Y. Liu, L. Guo, L. Bai, and C. Wang, "Locational marginal pricing mechanism for uncertainty management based on improved multi-ellipsoidal uncertainty set," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 4, pp. 734–750, 2021.

[14] X. Fang, B.-M. Hodge, E. Du, C. Kang, and F. Li, "Introducing uncertainty components in locational marginal prices for pricing wind power and load uncertainties," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 2013–2024, 2019.

[15] R. Mieth and Y. Dvorkin, "Distribution electricity pricing under uncertainty," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 2325–2338, 2019.

[16] M. Rahimiyan and L. Baringo, "Strategic bidding for a virtual power plant in the day-ahead and real-time markets: A price-taker robust optimization approach," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2676–2687, 2015.

[17] H. Yang, S. Zhang, D. Qiu, J. Zhao, M. Lai, Z. Y. Dong, and Z. H. Dong, "Distributionally robust optimal bidding of controllable load aggregators in the electricity market," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1089–1091, 2017.

[18] Y. Guan and J. Wang, "Uncertainty sets for robust unit commitment," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1439–1440, 2013.

[19] A. Wasilkoff, P. Andrianesis, and M. Caramanis, "Day-ahead estimation of renewable generation uncertainty set for more efficient market clearing," in *2023 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, 2023.

[20] 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 Transactions on Power Systems*, 2022.

[21] F. Golestaneh, P. Pinson, and H. B. Gooi, "Polyhedral predictive regions for power system applications," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 693–704, 2018.

[22] M. Awadalla and F. Bouffard, "Flexibility characterization of sustainable power systems in demand space: A data-driven inverse optimization approach," *IEEE Transactions on Power Systems*, 2024.

[23] A. Velloso, A. Street, D. Pozo, J. M. Arroyo, and N. G. Cobos, "Two-stage robust unit commitment for co-optimized electricity markets: An adaptive data-driven approach for scenario-based uncertainty sets," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 958–969, 2019.

[24] Z. Lin, H. Chen, Q. Wu, T. Ji, and Q. Wu, "Extreme scenarios based data-adaptive probability uncertainty set for distributionally robust transmission expansion planning," *CSEE Journal of Power and Energy Systems*, 2023.

[25] D. Bertsimas, R. Cory-Wright, and J. Pauphilet, "A unified approach to mixed-integer optimization problems with logical constraints," *SIAM Journal on Optimization*, vol. 31, no. 3, pp. 2340–2367, 2021.

[26] D. Bertsimas, V. Gupta, and N. Kallus, "Data-driven robust optimization," *Mathematical Programming*, vol. 167, pp. 235–292, 2018.

[27] P. L. Bartlett and S. Mendelson, "Rademacher and gaussian complexities: Risk bounds and structural results," *J. Machine Learning Res.*, vol. 3, no. Nov, pp. 463–482, 2002.

[28] J. E. Kelley, Jr, "The cutting-plane method for solving convex programs," *Journal of the society for Industrial and Applied Mathematics*, vol. 8, no. 4, pp. 703–712, 1960.

[29] M. A. Duran and I. E. Grossmann, "An outer-approximation algorithm for a class of mixed-integer nonlinear programs," *Mathematical programming*, vol. 36, pp. 307–339, 1986.

[30] I. Quesada and I. E. Grossmann, "An lp/nlp based branch and bound algorithm for convex minlp optimization problems," *Computers & chemical engineering*, vol. 16, no. 10-11, pp. 937–947, 1992.

[31] P. Xiong, P. Jirutitijaroen, and C. Singh, "A distributionally robust optimization model for unit commitment considering uncertain wind power generation," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 39–49, 2016.

[32] S. Wang, S. Shahidehpour, D. S. Kirschen, S. Mokhtari, and G. Irisarri, "Short-term generation scheduling with transmission and environmental constraints using an augmented lagrangian relaxation," *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1294–1301, 1995.