# 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) immunizes against uncertainties in power systems through uncertainty sets that control the robustness and conservativeness of the underlying optimization problem. Despite earlier work in their structure and properties, there are few suggestions on calibrating their size. In this paper, we propose a method to determine (predict) the uncertainty set size using machine learning models and mixed-integer optimization (MIO), leveraging historical data that consist of covariates or features. In essence, we utilize covariates to simultaneously predict the uncertain parameters and construct an uncertainty set around the nominal prediction based on the confidence in the prediction itself. 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 underlying MIO problem in larger datasets. We apply our framework to uncertainty sets for available wind resource capacity in the Adaptive Robust Unit Commitment (ARUC) problem. We show that our approach gives lower probabilities of constraint violation than commonly used statistical approaches, without necessarily exhibiting an increase in the 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 overconservativeness becomes especially important in light of the utilization of RO in energy markets and in pricing uncertainty. Works like [4], [13]–[16] 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 [17], [18].

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 [19]. 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], [20], [21] 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 [22]–[25]. 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], [19] use confidence intervals and statistical methods to tie the size to probability metrics, and [22] 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. Similar to linear regression, we use historical data to do predictions and adjust the uncertainty set bounds. The historical data consist of *covariates* or *features*, like wind speed, humidity, temperature etc., and a target variable, like available wind power.

The main benefits of this approach are threefold. First, it links the size of the uncertainty set to tangible metrics, which are palatable to the stakeholders, like the error of the model that predicts the uncertain parameters and the probability of constraint violation. Therefore, we can see the dependency of the size on these metrics. For example, an inaccurate model with a high training error will yield larger bounds for the same probability of violation, compared to a very accurate model. This is also a more transparent take, compared to tuning the size manually, which is valuable to high-stake decision-making, like grid operations. Second, this approach makes few assumptions on the distribution of the data. Third, it provides a systematic and automatic procedure, reducing the need for expertise. All that, while it remains generic enough to be useful in any robust optimization problem, with the ability to incorporate extra constraints and attributes, based on the preference and expertise of the user.

### 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 [26], 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 $[I] = \{1, \ldots, I\}$. The $\ell_1$ norm of a vector is defined as $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{I} |x_i|$, the $\ell_2$ norm is defined as $\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{I} x_i^2}$, and the $\ell_\infty$ 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:

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

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:

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

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 $\mathbb{P}^*$ [27]. 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 $\epsilon$, if the following holds:

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

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 $\delta$, i.e.,

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

We refer to $\delta$ 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 a 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:

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

whose size is parameterized by $\boldsymbol{\delta} = (\delta_1,\dots,\delta_m)$. 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 $z_i$ such that $z_i = 1$ if the prediction for data point $i$ falls in the uncertainty set. Specifically, $z_i = 1$, if each coordinate $j$ of data point $i$ is within $\delta_j$ from the realized value and $z_i = 0$ otherwise. We then enforce the empirical probabilistic guarantee requirement that at least $p$ 100 % of the residuals are in the uncertainty set, i.e., $\sum_{i=1}^n z_i \geq p\ n$, 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:

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

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

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

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

where $M$ is a big number.

In Theorem 1, we show that a box uncertainty set constructed via Problem (1) satisfies a certain probabilistic guarantee.

**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 $\rho$-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 [28], 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 $1-\epsilon$,

$$\begin{aligned}\mathbb{P}^*(f(\boldsymbol{s},\boldsymbol{y}) \leq 0) &\geq \mathbb{P}^*(\boldsymbol{y} \in \mathcal{U}) = \mathbb{P}^*(L_{\mathcal{D}}(\boldsymbol{\beta}_j^*) \leq \delta_j,\ \forall j)\\ &\geq \mathbb{P}_S\left(L_S(\boldsymbol{\beta}_j^*) + \frac{2\rho BR}{\sqrt{n}} + c\sqrt{\frac{2\log(2\epsilon)}{n}} \leq \delta_j, \forall j\right)\\ &= \mathbb{P}_S\left(L_S(\boldsymbol{\beta}_j^*) \leq \delta_j - \frac{2\rho BR}{\sqrt{n}} - c\sqrt{\frac{2\log(2\epsilon)}{n}}, \forall j\right).\end{aligned}$$

Also, $\mathbb{P}^*(f(\boldsymbol{s},\boldsymbol{y}) \geq \mathbb{P}^*(\boldsymbol{y} \in \mathcal{U}) \geq \mathbb{P}_S(L_S(\boldsymbol{\beta}_j^*) \leq \delta_j, \forall j) \geq p$, for $n \to \infty$, which concludes the proof. $\square$

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 $\Gamma$. 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 $\Gamma$ and $z_i = 0$ otherwise. We then enforce the empirical probabilistic guarantee requirement that at least $p$ of the residuals are below $\Gamma$, 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, $\Gamma$, is as follows:

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

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

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

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

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 $\|\boldsymbol{x}\|_2\le R$ and $y_j\le Q$, $\forall j\in[m]$. 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 $\Gamma$, satisfy $\|\boldsymbol{\beta}_j^*\|_2\le B$ 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 $n\to\infty$*

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

*Proof.* Using the same reasoning as in the proof of Theorem 1, from [28], we obtain that for any $\epsilon\in(0,1)$, with probability of at least $1-\epsilon$, we have that

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

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

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

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

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

### C. Box-Budget Uncertainty Set

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

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

In this case, we require that at least $p$ 100 % of the residuals fall within the box-budget uncertainty set. We extend the formulations of Section III-A and Section III-B as follows:

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

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

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

$$\sum_{i=1}^{n} z_i\ge p\ n, \tag{5d}$$

$$z_i\in\{0,1\}, \qquad i\in[n]. \tag{5e}$$

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.

### D. Ellipsoidal Uncertainty Set

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

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

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

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

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 $\Gamma$. The proof follows from the proof of Theorem 2 and the inequality $\|\boldsymbol{R}_i\|_2\le\|\boldsymbol{R}_i\|_1\le\sqrt{m}\|\boldsymbol{R}_i\|_2$.

## IV. Covariate-Dependent Uncertainty Set Bounds

While the formulations of Section III predict the nominal value of the uncertain parameters based on the covariates, the size is fixed. This means that we have the same buffer of uncertainty around the prediction, regardless of the prediction itself.

For example, if we consider uncertainty in the available capacity of a wind resource, then we obtain a prediction and a fixed interval around that, which we believe that it contains the true value. However, we do not really want to have the same size of interval in all cases. Specifically, if we predict very low available wind capacity, then we are not interested in large deviations, because we are going to schedule it to a small amount anyway. If we predict very large wind, we would expect to have more slack in our predictions, since we may end up scheduling significantly more than the quantity that is available during the operating day.

In this section, we provide bounds that adapt to the input covariates, borrowing ideas from gradient boosting estimators. In that way, the covariates adapt to the error of the estimator, similar to gradient boosting. Then, we go from a size that is based on the model error on average, to a size that is based on the model error at each individual data point.

### A. Parallels with Gradient Boosting and Residual Fitting

In gradient boosting and residual fitting methods, the learning algorithm iteratively fits new models to the residuals or errors of previous models, to progressively provide a more accurate prediction. The main idea is to construct each new model to predict either the negative gradient of the loss function or the loss function itself, which are based on the current predictions. Therefore, the additional models act as corrections to the initial prediction.

If we use only two models, then the first estimator minimizes the prediction error from the target variable $\boldsymbol{y}$, while the second estimator predicts the error of the first estimator. This is exactly our goal — an estimate of the error at each data point, so we can set the bounds accordingly.

In the new formulation that we introduce below, the role of the bounds is similar to the second estimator in a gradient boosting setting, i.e., they follow the error of the first estimator. So, when the error is large, the bounds are larger, while the bounds are smaller, if we predict that the estimator was accurate. The modified formulation offers more degrees of freedom, since the bounds are not the same for all data points but change based on the information that we have about that data.

### B. Covariate-Dependent Method

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:

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

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

$$\sum_{i=1}^{n} z_i \ge p\ n, \tag{6c}$$

$$z_i \in \{0,1\}, \qquad i\in[n]. \tag{6d}$$

In practice, we take $\delta_j$ to be an affine function of the covariates, $\delta_j(\boldsymbol{x}) = \delta_j^0 + \boldsymbol{q}_j^T\boldsymbol{x}$, where $\delta_j^0$ 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 $\delta$ with $\delta(\boldsymbol{x})$.

Note that $\delta_j(\boldsymbol{x})$ mimics the error of the initial linear regression model, evoking the connections with gradient boosting. Specifically, at optimality, $\delta_j^*(\boldsymbol{x}) = |y_{ij} - \boldsymbol{\beta}_j^T\boldsymbol{x}_{ij}|$ for some data point $i$, which is the largest in-sample residual of the model that meets the probabilistic guarantee.

## 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 [26], 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). First, observe that Problem (1) can be equivalently formulated as follows:

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

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 $\boldsymbol{\beta},\boldsymbol{\delta}$, that is,

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

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:*

$$\begin{aligned} \max_{\boldsymbol{\mu},\boldsymbol{\theta}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{m}\mu_{ij}y_{ij}z_i + \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}y_{ij} - 2\sum_{i=1}^{n}\sum_{j=1}^{m}\theta_{ij}y_{ij} \\ \text{s.t.} \quad & \sum_{i=1}^{n}\mu_{ij} = \frac{1}{m}, \quad \forall j, \\ & \sum_{i=1}^{n}\left(2\theta_{ij} - \frac{1}{nm} - \mu_{ij}z_i\right)\boldsymbol{x}_{ij} = \boldsymbol{0}, \quad \forall j, \\ & \frac{1}{nm} + \mu_{ij}z_i - \theta_{ij} \ge 0, \\ & \boldsymbol{\mu},\boldsymbol{\theta} \ge \boldsymbol{0}. \end{aligned} \tag{9}$$

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:

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

[26] 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:

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

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, $\boldsymbol{\beta}^*, \delta^*$.

1: $t = 1$
2: Let $f(\boldsymbol{z}^{t+1}) = \infty, \eta^{t+1} = -\infty$
3: **while** $f(\boldsymbol{z}^{t+1}) - \eta^{t+1} > \epsilon$ **do**
4: Solve Problem (11), with input $\boldsymbol{z}^0, \ldots, \boldsymbol{z}^t$ and obtain optimal solutions $\eta^{t+1}, \boldsymbol{z}^{t+1}$
5: Solve Problem (9), with input $\boldsymbol{z}^{t+1}$, and obtain $f(\boldsymbol{z}^{t+1})$, $\nabla f(\boldsymbol{z}^{t+1})$
6: $t = t + 1$
7: **end while**
8: Solve Problem (12) with input $\boldsymbol{z}^*$ and obtain $\boldsymbol{\beta}^*, \delta^*$
9: **return** $\boldsymbol{\beta}^*, \delta^*$

The method described in Algorithm 1 was originally proposed for continuous decision variables by [29], and later extended to binary decision variables by [30], while also providing a proof of convergence in a finite, yet exponential in the worst case, number of iterations. In addition, [26] offer insights on improving the convergence through heuristics. 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 [31].

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. Covariate-Dependent Bound*

If we decide to use the covariate dependent bound, then we have the problem

$$\begin{aligned} \min_{\boldsymbol{z} \in \mathcal{Z}} \min_{\boldsymbol{\beta}, \boldsymbol{\delta}^0, \boldsymbol{q}} \quad & \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| + \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left(\delta_j^0 + \boldsymbol{q}_j^T \boldsymbol{x}_{ij}\right) \\ \text{s.t.} \quad & z_i |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| \leq \delta_j^0 + \boldsymbol{q}_j^T \boldsymbol{x}_{ij}, \quad i \in [n], j \in [m]. \end{aligned} \tag{13}$$

In this case, the dual problem that defines $f(\boldsymbol{z})$ is as follows

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

We can still apply Algorithm 1 by replacing Problem (9) with Problem (14).

### *C. 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:

$$f(\boldsymbol{z}) = \begin{aligned} \min_{\boldsymbol{\beta}, \Gamma} \quad & \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| + \frac{1}{m} \Gamma \\ \text{s.t.} \quad & z_i \sum_{j=1}^m |y_{ij} - \boldsymbol{\beta}_j^T \boldsymbol{x}_{ij}| \leq \Gamma, \quad i \in [n]. \end{aligned} \tag{15}$$

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 (15), is as follows:

$$\begin{aligned}
\max_{\boldsymbol{\mu},\boldsymbol{\theta}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{m}(\mu_i z_i + \frac{1}{nm} - 2\theta_{ij})y_{ij} \\
\text{s.t.} \quad & \sum_{i=1}^{n}\mu_i = \frac{1}{m}, \\
& \sum_{i=1}^{n}\left(2\theta_{ij} - \frac{1}{nm} - \mu_i z_i\right)\boldsymbol{x}_{ij} = \boldsymbol{0}, \quad j \in [m], \\
& \frac{1}{nm} + \mu_i z_i - \theta_{ij} \geq 0, \quad i \in [n], j \in [m], \\
& \mu_i, \theta_{ij} \geq 0.
\end{aligned} \tag{16}$$

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 (16). In this case the $i$-th coordinate of the subgradient 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.

The proposed outer approximation scheme no longer applies to the ellipsoid uncertainty set, since the same arguments about convexity of the function $f(\boldsymbol{z})$ do not hold in this case.

## 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], [32]. The problem formulation is as follows:

$$\min_{\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{p}} \quad \sum_{i\in\mathcal{K}}\sum_{t\in\mathcal{T}} F_{it}x_{it} + C_{it}p_{it} \tag{17a}$$

$$\text{s.t.} \quad (\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) \in \mathcal{X}, \tag{17b}$$

$$\sum_{i\in\mathcal{K}} p_{it} + \sum_{i\in\mathcal{R}} p_{it} = \sum_{b} D_{bt}, \quad t \in \mathcal{T}, \tag{17c}$$

$$\sum_{b} K_{lb}\left(\sum_{i\in\mathcal{K}_b} p_{it} + \sum_{i\in\mathcal{R}_b} p_{it} - D_{bt}\right) \leq F_l, \tag{17d}$$

$$-\sum_{b} K_{lb}\left(\sum_{i\in\mathcal{K}_b} p_{it} + \sum_{i\in\mathcal{R}_b} p_{it} - D_{bt}\right) \leq F_l, \tag{17e}$$

$$p_i^{\min} x_{it} \leq p_{it} \leq p_i^{\max} x_{it}, \quad i \in \mathcal{K}, t \in \mathcal{T}, \tag{17f}$$

$$p_{it}^{\min} \leq p_{it} \leq p_{it}^{\max}, \quad i \in \mathcal{R}, t \in \mathcal{T}, \tag{17g}$$

$$-RD_{it} \leq p_{it} - p_{i(t-1)} \leq RU_{it}, \quad i \in \mathcal{K}, t \in \mathcal{T}. \tag{17h}$$

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 (17c) is the energy balance equation that matches the system level supply and load at each time period. Constraints (17d) and (17e) follow the dc power flow model in [33] to confine the power transmission below the line capacity $F_l$. Constraints (17f) and (17g) indicate that committed generators and wind resources can be dispatched within a range. Constraint (17h) 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 [32], we utilize linear decision rules (LDR) in order to obtain a computationally tractable approximation of Problem (17) under wind capacity uncertainty. LDR often achieve near-optimal performance, see [1]. Let $p_{it}^{\max}$ denote the uncertain capacity of wind resource $i$

at time $t$, and $\bar{p}_{it}^{\max}$ the forecasted or expected capacity. The deviation from the expected capacity is $r_{it} = p_{it}^{\max} - \bar{p}_{it}^{\max}$ and is uncertain. We then impose the decision rule

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

Before we proceed with the formulation, for completeness, we provide the robust counterpart under different uncertainty sets for the capacity constraint. Details are available at [1]. In the robust version of Problem (17), the capacity constraint is

$$p_{it}(\boldsymbol{r}_t) \leq (\bar{p}_{it}^{\max} + r_{it}), \forall \boldsymbol{r}_t \in \mathcal{Z}_t, \tag{18}$$

where $r_{it}$ is the deviation from the expected available capacity and $\mathcal{Z}_t$ is the uncertainty set at time $t$. Therefore, we require that the production $p_{it}(\boldsymbol{r}_t)$ is less than or equal to the realized capacity $(p_i^{\max} + r_{it})$ for all scenarios in the uncertainty set. Then, using linear decision rules,

$$p_{it}^0 + \boldsymbol{w}_{it}^T \boldsymbol{r}_t \leq (\bar{p}_{it}^{\max} + r_{it}), \forall \boldsymbol{r}_t \in \mathcal{Z}_t, \tag{19}$$

or $p_{it}^0 + (\boldsymbol{w}_{it} - \boldsymbol{e}_i)^T \boldsymbol{r}_t \leq \bar{p}_{it}^{\max}, \forall \boldsymbol{r}_t \in \mathcal{Z}_t$. The robust counterpart then is

$$p_{it}^0 + \max_{\boldsymbol{r}_t \in \mathcal{Z}_t} (\boldsymbol{w}_{it} - \boldsymbol{e}_i)^T \boldsymbol{r}_t \leq \bar{p}_{it}^{\max}. \tag{20}$$

We next provide the derivation of the robust counterpart by following the methodology from [1]. In the following notation, $\boldsymbol{\alpha}$ denotes the constraint coefficients multiplying the uncertain parameters, i.e., $\boldsymbol{\alpha} = \boldsymbol{w}_{it} - \boldsymbol{e}_i$.

*a) Box Uncertainty Set:* Assuming the uncertainty set is $\mathcal{Z}_t = \{\boldsymbol{r}_t : |r_{it}| \leq \delta_{it}\}$, we obtain

$$p_{it}^0 + \boldsymbol{\delta}_t^T \boldsymbol{\lambda}_{it}^{\text{cap, max}} \leq \bar{p}_{it}^{\max}, \tag{21}$$

where $\left(\boldsymbol{\lambda}_{it}^{\text{cap, max}}, \mu_{it}^{\text{cap, max}}, \boldsymbol{\nu}_{it}^{\text{cap, max}}, \boldsymbol{\pi}_{it}^{\text{cap, max}}\right) \in \mathcal{F}_t =$

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

*b) Box-Budget Uncertainty Set:* Assuming the uncertainty set is $\mathcal{Z}_t = \{\boldsymbol{r}_t : |r_{it}| \leq \delta_{it}, \ \|\boldsymbol{r}_t\|_1 \leq \Gamma_t\}$, we obtain

$$p_{it}^0 + \boldsymbol{\delta}_t^T \boldsymbol{\lambda}_{it}^{\text{cap, max}} + \Gamma_t \mu_{it}^{\text{cap, max}} \leq \bar{p}_{it}^{\max}, \quad \forall i, t, \tag{22}$$

where $\left(\boldsymbol{\lambda}_{it}^{\text{cap, max}}, \mu_{it}^{\text{cap, max}}, \boldsymbol{\nu}_{it}^{\text{cap, max}}, \boldsymbol{\pi}_{it}^{\text{cap, max}}\right) \in \mathcal{F}_t =$

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

*c) Convex Hull Uncertainty Set:* Assuming the uncertainty set is the convex hull of $K$ historical realizations $\boldsymbol{r}_{kt}$ for $k = 1, ..., K$ of the uncertain parameter at time $t$, $\mathcal{Z}_t = \left\{(\boldsymbol{r}_t, \boldsymbol{\zeta}_t) : \boldsymbol{r}_t = \sum_{k=}^{K} \zeta_{kt} \boldsymbol{r}_{kt}, \ \sum_{k=1}^{K} \zeta_{kt} = 1, \ \boldsymbol{\zeta}_t \geq \boldsymbol{0}\right\}$, we obtain

$$p_{it}^0 + \lambda_t^{\text{cap, max}} \leq \bar{p}_{it}^{\max}, \tag{23}$$

where $\lambda_t^{\text{cap, max}} \in \mathcal{F}_t = \left\{\lambda_t : \boldsymbol{\alpha}^T \boldsymbol{r}_{kt} \leq \lambda_t, \forall k \in [K]\right\}$. Note that LDR are optimal for the convex hull uncertainty set [1].

The complete formulation for the box-budget uncertainty set is

$$
\begin{aligned}
\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} && \text{(24a)} \\
\text{s.t.} \quad & (\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) \in \mathcal{X}, && \text{(24b)} \\
& \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, && \text{(24c)} \\
& \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) \\
& \quad + \boldsymbol{\delta}_t^T \boldsymbol{\lambda}_{lt}^{\text{tr, max}} + \Gamma_t \mu_{lt}^{\text{tr, max}} \leq F_l, \quad \forall l, t, && \text{(24d)} \\
& - \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) \\
& \quad + \boldsymbol{\delta}_t^T \boldsymbol{\lambda}_{lt}^{\text{tr, min}} + \Gamma_t \mu_{lt}^{\text{tr, min}} \leq F_l, \quad \forall l, t, && \text{(24e)} \\
& p_i^{\min} x_{it} \leq p_{it} \leq p_i^{\max} x_{it}, \quad \forall i, t, && \text{(24f)} \\
& p_{it}^0 + \boldsymbol{\delta}_t^T \boldsymbol{\lambda}_{it}^{\text{cap, max}} + \Gamma_t \mu_{it}^{\text{cap, max}} \leq \bar{p}_{it}^{\max}, \quad \forall i, t, && \text{(24g)} \\
& 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, && \text{(24h)} \\
& - RD_{it} \leq p_{it} - p_{i(t-1)} \leq RU_{it}, \quad \forall i, t, && \text{(24i)} \\
& \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}}, && \text{(24j)} \\
& \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}}, && \text{(24k)} \\
& \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}}, && \text{(24l)} \\
& \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}}, && \text{(24m)} \\
& \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}}. && \text{(24n)}
\end{aligned}
$$

Additional reserve constraints can be added with small changes. In addition, the spinning reserve requirements in Problem (24) 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 [17].

## 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.

In this case, the covariates are the predictions of the 5 forecasting models. So, the coefficients of the linear regression model are weights we assign to the prediction of each model and the individual models are combined to provide more accurate predictions. We seek to provide larger weights to more accurate models and smaller weights to less accurate models. The weights are determined in an optimal way by leveraging the historical data and finding the combination that

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2025.3530410



minimizes the training error. The models themselves range from simple linear models to complicated physical models. If we did not have access to these models, we would use more traditional features like weather and seasonal data to predict the available wind capacity.

The training set includes the two days prior to the day of interest, which serves as the test set. The short time horizon allows the training set to be representative of conditions in the test day, so that the coefficients or weights we learn are close to optimal for the test day. So, the sample size is 2 days $\times$ 24 hours $\times$ 5 wind generators. 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). 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 [19]. The standard deviation is obtained as the empirical standard deviation of the 5 forecasting model predictions at each wind resource, at each time period. We note that the approach from [19] results in extremely conservative budget bounds, thus we utilize only the box uncertainty set.

We also compare our approach to an uncertainty set constructed by taking the convex hull (CH) of the historical data [22]–[24]. We obtain the historical deviations $\boldsymbol{r}$ from the forecasted capacity, using the average of the 5 forecasting models as the prediction and taking the difference from the corresponding realization. If the difference is positive, i.e. we underestimated the amount of available capacity for a resource $i$ at some time period $t$, we set it to zero. Otherwise, the robust model could schedule more power than the available capacity. We take the approach of [24] and include data from $K$ historical days prior to the operating day.

This is an excellent alternative to our method, if covariates are not available, especially considering the optimality of LDR with this uncertainty set. It has shown good performance in [24] but comes with a possible bottleneck, because it does not provide a data-driven way for controlling the conservativeness and selecting which historical data will be included in the set. In this case, we select $K = 6$, as opposed to the two training days in our method, based on the probability of constraint violation on the test day, which in practice is not known a priori. Therefore, the results for this set are the best possible for this set of historical data and robust formulations.

We use the following metrics. The first is the objective of the deterministic problem and the objective of the robust problem, which corresponds to the worst-case cost for protecting against all wind capacity realizations in the uncertainty set. We also report the amount that was scheduled over the realized capacity for all wind resources. This constraint violation is defined as $\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{R}_t}(p^*_{it}-p^{\max}_{it})$, for the units $\mathcal{R}_t$ that where initially scheduled above the realized available capacity at time $t$. In the deterministic problem, $p^*_{it}$ is the optimal value of the wind production variables. In the robust problem, $p^*_{it} = p^*_{it}(\boldsymbol{r})$, where $\boldsymbol{r}$ is the realized uncertainty in the operating day. This excess production will be made up by the reserves in the operating day. Similarly, we report the empirical probability of constraint violation for the wind capacity constraints. It is defined as the number of renewable capacity constraints that are violated over the total number of renewable capacity constraints across all generators and time periods. We find that the mean squared errors (MSEs) of the underlying models on the test day are in the same range across different uncertainty sets. Therefore, most of the differences in the results can be attributed to the bounds themselves.

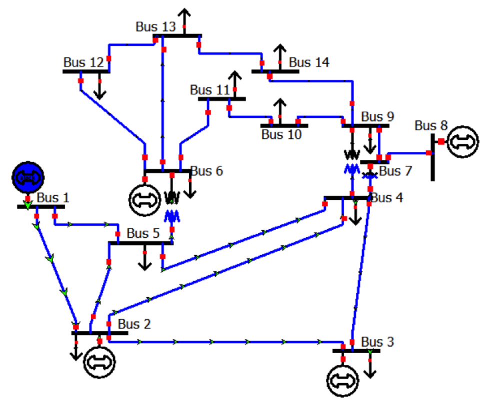


Fig. 1. The IEEE 14 bus system grid.

### 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. The realized wind capacity is, on aggregate, 130 % of the total intra-day load.

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 |

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. The second column reports the results for the ARUC problem under the confidence interval approach. The third column reports the results under the convex hull uncertainty set. Further, in column 4, we provide results for Problem (1), with constant (Con. Box) bounds. We also report the results for Problem (5) under constant (Con. Budg.) and covariate-dependent bounds (Cov. Budg.) that are a linear function of the covariates.

TABLE II
OBJECTIVE VALUES FOR NOMINAL AND ROBUST UC.

| $p$ | Det | CI | CH | Con. Box | Con. Budg. | Cov. Budg. |
|---|---|---|---|---|---|---|
| 0.70 | 25,705 | 44,439 | 57,917 | 57,125 | 48,054 | 43,405 |
| 0.75 | 25,705 | 46,417 | 57,917 | 60,293 | 50,500 | 49,259 |
| 0.80 | 25,705 | 49,721 | 57,917 | 70,927 | 53,195 | 52,866 |
| 0.85 | 25,705 | 52,289 | 57,917 | 77,383 | 62,169 | 57,123 |
| 0.90 | 25,705 | 57,737 | 57,917 | 104,234 | 75,804 | 63,572 |
| 0.95 | 25,705 | 63,893 | 57,917 | 122,965 | 103,760 | 82,633 |
| 0.99 | 25,705 | 75,392 | 57,917 | 170,478 | 117,673 | 91,917 |

In Table II, we observe that the objective increases, as we increase the probabilistic guarantee $p$, because we protect against more realizations of the uncertainty. The box uncertainty is more conservative than the box-budget set, while the covariate-dependent bounds 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 | CH | Con. Box | Con. Budg. | Cov. Budg. |
|---|---|---|---|---|---|---|
| 0.70 | 0.28 | 0.10 | 0.10 | 0.06 | 0.01 | 0.08 |
| 0.75 | 0.28 | 0.11 | 0.10 | 0.03 | 0.01 | 0.06 |
| 0.80 | 0.28 | 0.12 | 0.10 | 0.03 | 0.00 | 0.03 |
| 0.85 | 0.28 | 0.10 | 0.10 | 0.01 | 0.00 | 0.02 |
| 0.90 | 0.28 | 0.06 | 0.10 | 0.00 | 0.01 | 0.03 |
| 0.95 | 0.28 | 0.03 | 0.10 | 0.01 | 0.01 | 0.02 |
| 0.99 | 0.28 | 0.03 | 0.10 | 0.00 | 0.00 | 0.01 |

From Table III, we observe that our approach provides better empirical probabilities of constraint violation than both the deterministic problem, the CI and the CH, across the board. We further notice that our approach with constant bounds provides slightly better probabilistic guarantees than the covariate-dependent bounds, while being more conservative. These probabilities and violations are calculated based on the test day, so they are not necessarily monotonic. In principle, we expect them to be monotonic, but there is some small volatility depending on the test day. Still, there is an overall monotonic pattern.

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

| $p$ | Det | CI | CH | Con. Box | Con. Budg. | Cov. Budg. |
|---|---|---|---|---|---|---|
| 0.70 | 186.6 | 222.3 | 309.9 | 47.2 | 0.1 | 266.8 |
| 0.75 | 186.6 | 453.7 | 309.9 | 36.6 | 8.6 | 40.3 |
| 0.80 | 186.6 | 724.3 | 309.9 | 23.2 | 0.0 | 28.0 |
| 0.85 | 186.6 | 268.6 | 309.9 | 11.8 | 0.0 | 49.4 |
| 0.90 | 186.6 | 95.8 | 309.9 | 0.0 | 11.6 | 102.5 |
| 0.95 | 186.6 | 15.4 | 309.9 | 4.3 | 2.7 | 8.9 |
| 0.99 | 186.6 | 2.2 | 309.9 | 0.0 | 0.0 | 11.2 |

Table IV reports the difference between the realized capacity and the scheduled dispatch of wind resources, when the dispatch exceeds the realized capacity. This is roughly equivalent to the amount of reserves that we have to deploy in the operating day to make up for the wind forecast errors. We notice that our method achieves lower constraint violation than both the deterministic problem as well as the robust problem with confidence intervals or the convex hull.

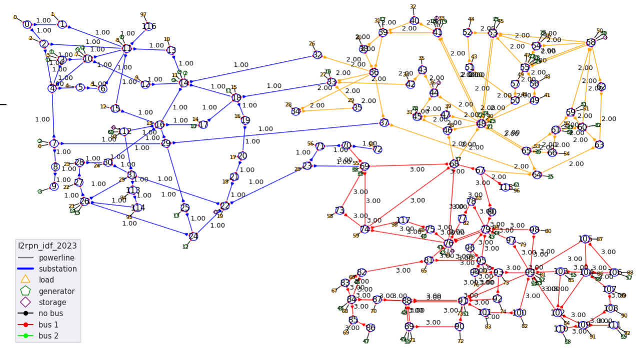


Fig. 2. The IEEE 118 bus system grid.

The CI and CH approaches are less conservative than our approach for the same in-sample guarantee in some cases, and provide good probabilistic guarantees on the test day. However, they miss the mark on the absolute constraint violation, which arguably is of most interest, since it dictates the amount of reserves and ad hoc expenses in the Real-Time Market. The combination of small probabilities of violation and larger absolute violations shows that the bounds are generally well calibrated, but may be off for a few generators or time periods. Our method is consistent in both metrics, offering more reliable bounds, albeit at a larger cost. However, when we consider more complicated sets, such as the box-budget set, and more complicated bounds, such as the adjustable bounds, we get a win-win in terms of reducing conservativeness and maintaining reliability.

## 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 %. Note that the following results are not directly comparable to the 14-bus system, since the same generators are placed at different buses, while the wind penetration is different as well.

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 | CH | Con. Box | Con. Budg. | Cov. Budg. |
|---|---|---|---|---|---|---|
| 0.70 | 3.329 | 3.363 | 3.393 | 3.395 | 3.389 | 3.386 |
| 0.75 | 3.329 | 3.367 | 3.370 | 3.398 | 3.398 | 3.392 |
| 0.80 | 3.329 | 3.372 | 3.370 | 3.409 | 3.392 | 3.397 |
| 0.85 | 3.329 | 3.378 | 3.370 | 3.421 | 3.415 | 3.401 |
| 0.90 | 3.329 | 3.384 | 3.370 | 3.453 | 3.436 | 3.422 |
| 0.95 | 3.329 | 3.396 | 3.370 | 3.475 | 3.469 | 3.442 |
| 0.99 | 3.329 | 3.418 | 3.370 | 3.532 | 3.497 | 3.459 |

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

| $p$ | Det | CI | CH | Con. Box | Con. Budg. | Cov. Budg. |
|---|---|---|---|---|---|---|
| 0.70 | 0.23 | 0.11 | 0.16 | 0.03 | 0.00 | 0.06 |
| 0.75 | 0.23 | 0.13 | 0.16 | 0.02 | 0.00 | 0.04 |
| 0.80 | 0.23 | 0.08 | 0.16 | 0.01 | 0.00 | 0.04 |
| 0.85 | 0.23 | 0.10 | 0.16 | 0.01 | 0.00 | 0.03 |
| 0.90 | 0.23 | 0.07 | 0.16 | 0.00 | 0.00 | 0.03 |
| 0.95 | 0.23 | 0.03 | 0.16 | 0.00 | 0.00 | 0.03 |
| 0.99 | 0.23 | 0.02 | 0.16 | 0.00 | 0.00 | 0.02 |

The takeaways in the 118-bus system are consistent with the results of the 14-bus system, albeit with less pronounced differences and violations, since the wind penetration is lower in the grid.

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

| $p$ | Det | CI | CH | Con. Box | Con. Budg. | Cov. Budg. |
|---|---|---|---|---|---|---|
| 0.70 | 124.4 | 363.3 | 215.4 | 10.3 | 0.0 | 106.6 |
| 0.75 | 124.4 | 351.5 | 215.4 | 13.4 | 0.0 | 23.5 |
| 0.80 | 124.4 | 229.2 | 215.4 | 5.8 | 0.0 | 45.4 |
| 0.85 | 124.4 | 104.1 | 215.4 | 3.3 | 0.0 | 47.8 |
| 0.90 | 124.4 | 75.7 | 215.4 | 0.0 | 0.0 | 23.5 |
| 0.95 | 124.4 | 7.5 | 215.4 | 0.0 | 0.0 | 16.7 |
| 0.99 | 124.4 | 1.4 | 215.4 | 0.0 | 0.0 | 25.4 |

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 $0.1\%$ is obtained for Problem (24) 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 6.

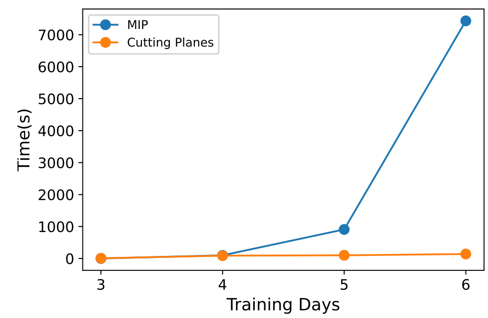


Fig. 3. Computational time with number of days in the training set.

From Figure 3, we observe a significant speed-up when using Algorithm 1 and 5 or 6 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 the size of the uncertainty set. We showed that the uncertainty sets constructed by our approach satisfy certain probabilistic guarantees. Moreover, we applied our method to the ARUC problem, under uncertainty in the wind resource capacity and we showed that our method exhibits small constraint violations compared to statistical approaches. The results improved further with covariate-dependent bounds, which allow the size of the uncertainty set to adapt with the realizations of covariates.

## 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] D. Bertsimas and A. G. Koulouras, "Marginal pricing in adaptive robust unit commitment under load and capacity uncertainty," *IEEE Transactions on Power Systems*, 2024.

[17] 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.

[18] 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.

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

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

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

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

**Dimitris Bertsimas** (M'11) received the B.S. degree in electrical engineering and computer science from the National Technical University of Athens, Greece, the M.S. degree in operations research, and the Ph.D. in applied mathematics and operations research from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. He is currently the Boeing Leaders for Global Operations Professor of Management, a Professor of Operations Research, and the Associate Dean with the Master of Business Analytics, MIT. Since 1988, he has been a Faculty Member at MIT. He has coauthored more than 200 scientific papers and six graduate-level textbooks. His research interests include optimization, stochastic systems, machine learning, and their application, focusing on robust optimization, statistics, healthcare, transportation, and finance. He was the recipient of numerous research awards, including the Harold Larnder Prize (2016), Philip Morse Lectureship prize (2013), William Pierskalla Best Paper Award in health care (2013), Best Paper Award in Transportation (2013), Farkas Prize (2008), Erlang Prize (1996), SIAM Prize in Optimization (1996), the Bodossaki Prize (1998), and Presidential Young Investigator Award (1991–1996), and was also the recipient of the Samuel M. Seegal prize (1999) and The Jamieson Prize (2013) for the recognition of his educational contributions. He is also the Member of the National Academy of Engineering and an INFORMS Fellow.

**Thodoris Koukouvinos** received the B.S. in applied mathematics from the National Technical University of Athens, Greece, and is a Ph.D. student at the with the Operations Research Center, Massachussetts Institute of Technology (MIT), Cambridge, MA, USA. He has received the A.G. Leventis Foundation scholarship for graduate studies.

**Angelos Georgios Koulouras** (S'18) received the B.Sc. degree with highest honors in electrical and computer engineering from The University of Texas at Austin (UT Austin), Austin, TX, USA, in 2020. He is currently working toward the Ph.D. degree in operations research with the Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA, USA. He is supported by the Onassis Foundation scholarship for graduate studies. He is a Member of IEEE Eta Kappa Nu (HKN). He was the recipient of the Unrestricted Endowed Presidential Scholarship at UT Austin, in 2019.