Conformalized Decision Risk Assessment

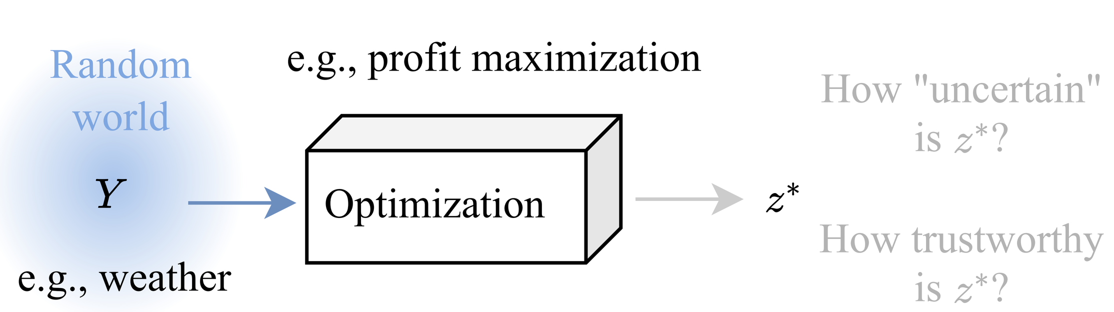
Wenbin Zhou[1], Agni Orfanoudaki[2], Shixiang Zhu[1]

[1]*Carnegie Mellon University*, [2]*University of Oxford*

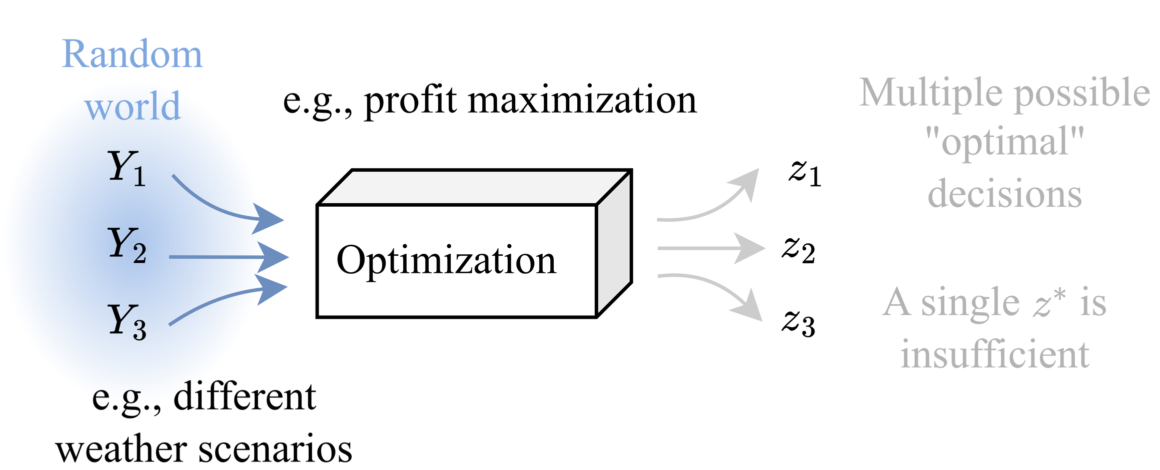
Intro: DM Under Uncertainty

Challenges:

(1) Uncertainty is not explicitly communicated



(2) Decisions could be brittle under multi-modality



Prospect: Collaborative Mechanism:

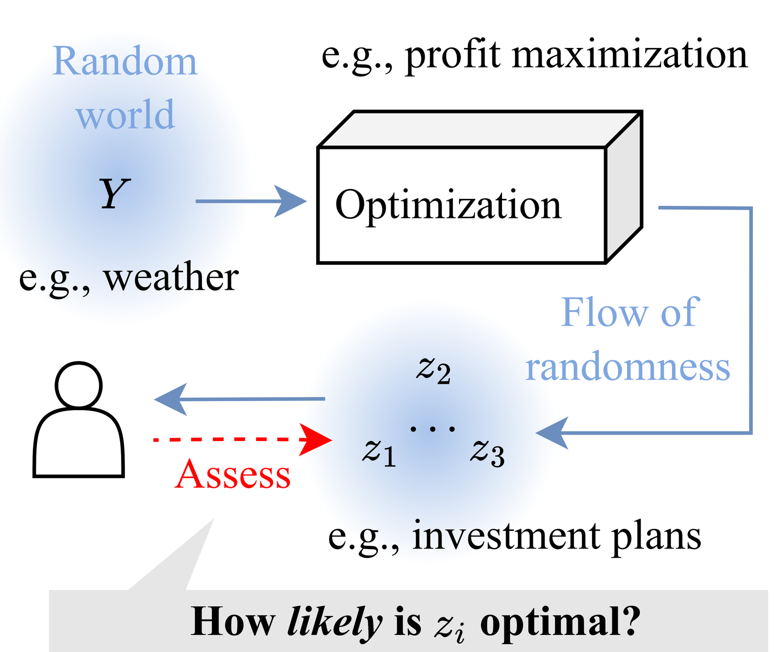


Figure: Given environment uncertainty and an optimization model, we aim to quantify the risk associated with adopting a specific decision.

Contributions: (i) Introduce "decision risk assessment"; (ii) Developed an algorithm that is statistically sound and efficient; (iii) Empirical validations of the framework.

Setup: Decision Risk Assessment

Similar to **predict-then-optimize** problems, define:

- $X \in \mathcal{X}$: observed *covariates* associated with Y;
- $Y \in \mathcal{Y}$: random *outcome* variable that serve as objective function parameters;
- θ: known parameters within the optimization problem.

Novel Objective

Given the optimization problem:

$$\pi(Y;\theta) := \arg\min_{z \in \mathbb{R}^d} \{g(z,Y;\theta) \mid z \in \mathcal{Z}(\theta)\}.$$

We aim to estimate some risk measure $\alpha(z)$ such that:

$$\mathbb{P}\{z \in \pi(Y;\theta)\} \geq 1-\alpha(z), \quad \forall z \in \mathbb{R}^d. \quad (1)$$

Framework Overview

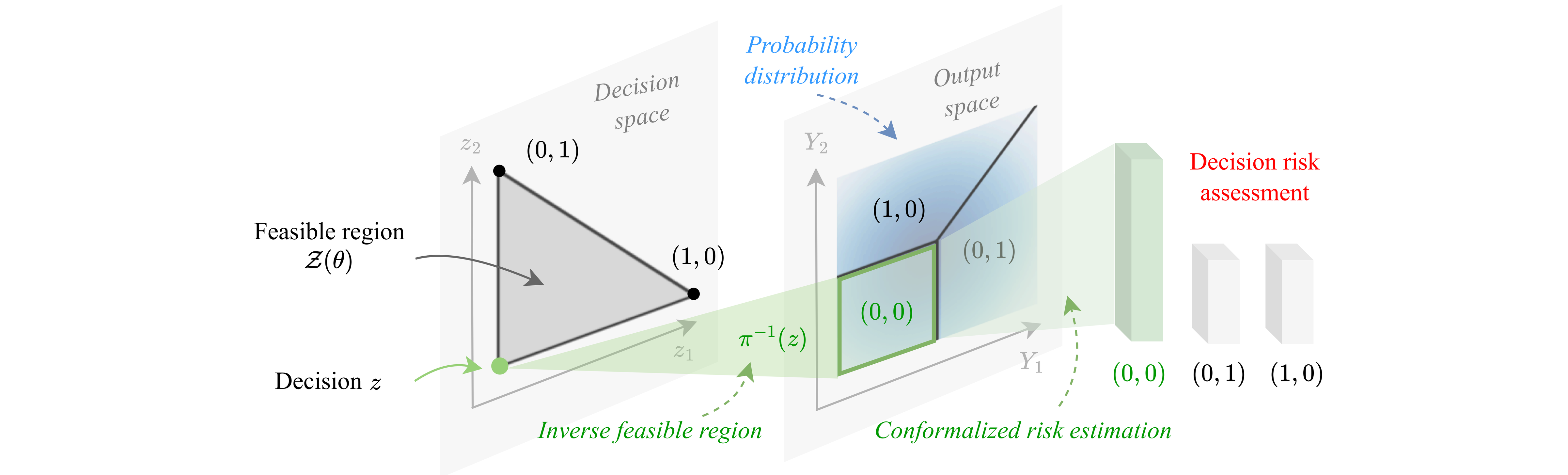


Figure: Figure 2: The overall architecture of the proposed framework, consisting of two steps: (i) Map the prespecified decision $z \in \mathcal{Z}(\theta)$ to its inverse feasible region $\pi^{-1}(z;\theta) \subseteq \mathcal{Y}$. (ii) Assess the risk certificate (i.e., coverage probability) via conformalized risk estimation over $\pi^{-1}(z;\theta)$ using data of Y.

Algorithmic Details

Step 1: Reformulation with Inverse Feasible Region:

Proposition 1: Define *inverse feasible region*

$$\pi^{-1}(z;\theta) = \bigcap_{z' \in \mathcal{Z}(\theta)} \{y \in \mathcal{Y} \mid g(y,z) \leq g(y,z')\}$$

Then, the objective (1) can be reformulated as:

$$\mathbb{P}\{z \in \pi(Y;\theta)\} \equiv \mathbb{P}\left\{Y \in \pi^{-1}(z;\theta)\right\}. \quad (2)$$

Continuing the derivation on the RHS of (2):

$$\mathbb{P}\left\{Y \in \pi^{-1}(z;\theta)\right\} \overset{(i)}{\geq} \mathbb{P}\{Y \in \mathcal{C}(X;\alpha(z))\} \overset{(ii)}{\geq} 1-\alpha(z),$$

So in Step 2, we only need to construct a set $\mathcal{C}(X;\alpha(z))$ that can jointly satisfy (i) and (ii).

Step 2: Generative Conformal Prediction: The set is defined as

$$\mathcal{C}^{(k)}(x_{n+1};\alpha) = \left\{y \in \mathbb{R}^d \mid \|y - \hat{y}^{(k)}_{n+1}\|_2 \leq h(\alpha, \mathcal{D}_{\text{cal}})\right\}$$

where $h(\alpha,\mathcal{D}_{\text{cal}})$ is the *conformalized radius*, and $\hat{y}^{(k)}_{n+1}$ is the k-th generated outcome drawn from a fitted generative model $f(Y|X)$. Then, the k-th risk estimator is defined as

$$\hat{\alpha}^{(k)}(z) = \min_{\alpha \in [0,1]} \left\{\alpha \mid \mathcal{C}^{(k)}(x_{n+1};\alpha) \subseteq \pi^{-1}(z;\theta)\right\}$$

These K estimators are averaged to obtain the final risk estimators.

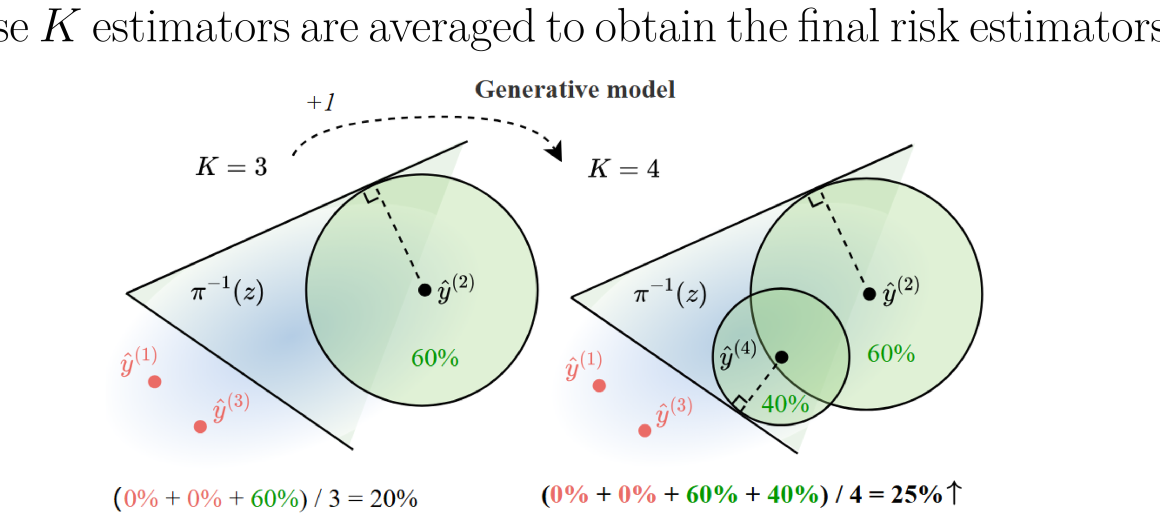


Figure: Illustration of the generative conformal prediction procedure

Theoretical Guarantee

Let r_i denote the calibrated nonconformity scores:

$$r_i = \|\hat{y}_i - y_i\|_2, \quad y_i \in \mathcal{D}_{\text{cal}}.$$

Theorem 1 The estimator $\hat{\alpha}(z)$ satisfies the following:

- When $h(\alpha,\mathcal{D}_{\text{cal}}) = \hat{Q}(\lceil (n+1)(1-\alpha) \rceil / n)$, where $\hat{Q}$ denotes the quantile function of $\{r_i\}$, then
$$\mathbb{P}\{z \in \pi(Y;\theta)\} \geq 1 - \mathbb{E}_{X,\mathcal{D}}[\hat{\alpha}(z)] - \epsilon, \quad \forall z \in \mathbb{R}^d,$$
where ϵ is a small constant induced by nonexchangeability.
- When $h(\alpha,\mathcal{D}_{\text{cal}}) = \sum_{i=1}^n r_i/(\alpha(n+1)-1)$, then
$$\mathbb{P}\{z \in \pi(Y;\theta)\} \geq 1 - \mathbb{E}_{X,\mathcal{D}}[\hat{\alpha}(z)], \quad \forall z \in \mathbb{R}^d,$$

Key: The estimator upper-bounds the true risk in expectation.

Efficient Computation: Separable Objective

Suppose there exist (potentially nonlinear) feature mappings:

$$\phi: \mathcal{X} \to \mathbb{R}^d, \quad \psi: \mathcal{Z} \to \mathbb{R}^d,$$

such that the objective function can be written as $g(z,Y;\theta) = \phi(Y)^\top \psi(z)$, then the estimator has a **closed-form** solution:

Theorem 2: The risk estimator can be simplified to

$$\hat{\alpha}(z) = 1 - \sum_{k=1}^{K} \left[w^{(k)}(z) \cdot \prod_{u \in \mathcal{E}} \mathbf{1}\left\{\phi(\hat{y}^{(k)}_{n+1})^\top (\psi(z) - u) \leq 0\right\}\right],$$

where $w^{(k)}(z)$ is the *conformalized weight*, defined as:

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{\|\phi(\hat{y}^{(k)}_i) - \phi(y_i)\| \leq \min_{u \in \mathcal{E}\setminus\psi(z)} \frac{|\phi(\hat{y}^{(k)}_{n+1})^\top(\psi(z)-u)|}{\|\phi(z)-u\|}\right\}$$

and $\mathcal{E}$ denotes the set of extreme points of $\{\psi(z) : z \in \mathcal{Z}(\theta)\}$.

Key: The algorithm is computationally efficient under certain instances or under separable function approximation.

Numerical Results

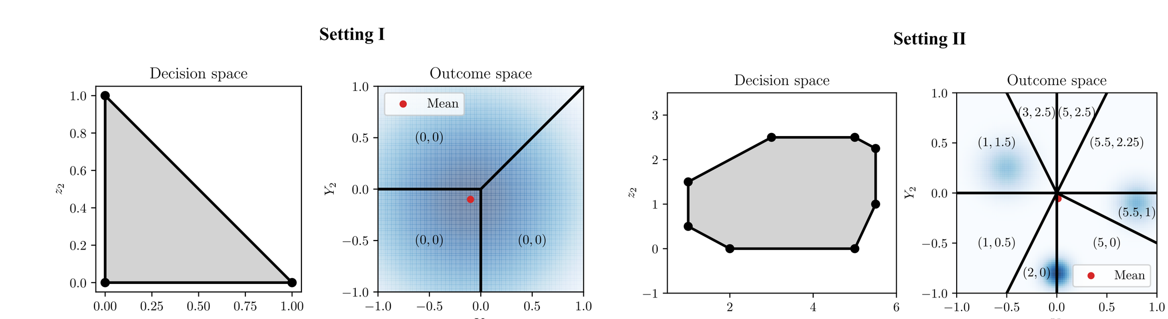


Figure: Pictorial representation of the optimization settings I and II. We illustrate the feasible region (gray shaded) in the decision space and the corresponding inverse feasible regions (cones) in the outcome space.

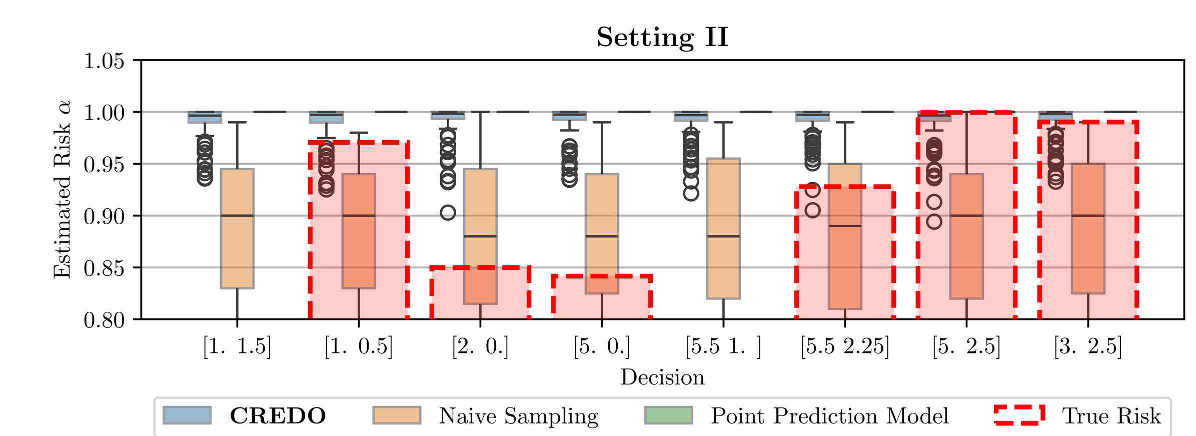


Figure: Estimated risk for three ablation models of CREDO over different decisions. Without the conformalized procedure, the `Naive Sampling` approach is prone to violating the conservativeness requirement.

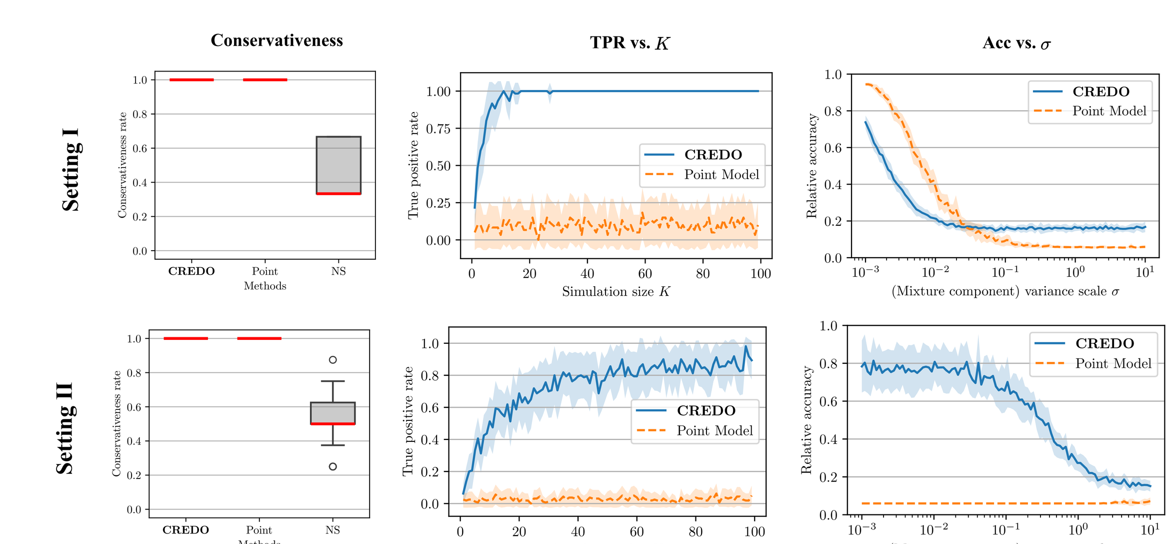


Figure: Three performance metric evaluation results of the ablation models. From left to right columns: Conservativeness of different ablation models; True positive rate (TPR) versus generative sample size K; Relative accuracy versus variance scale σ

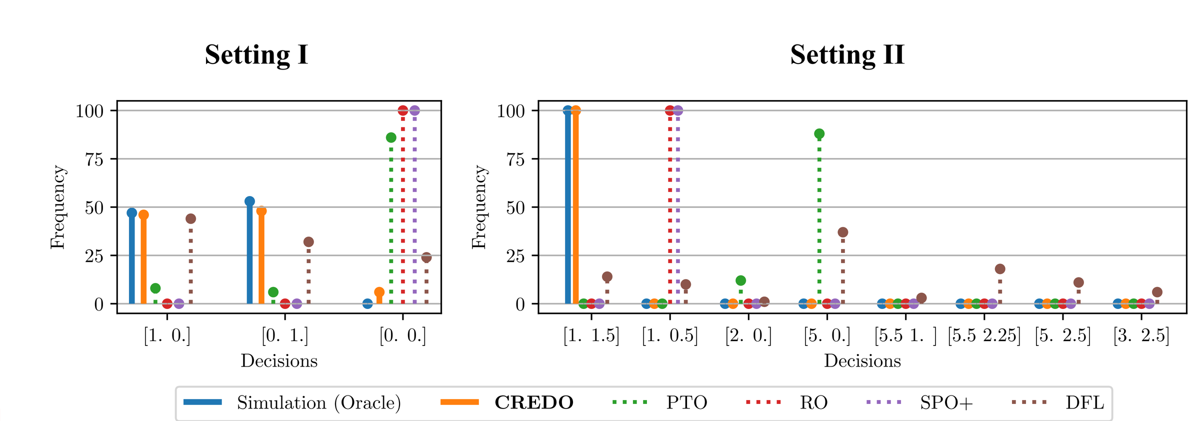


Figure: Frequency of each selected decision over 100 repeated trials, compared across all baseline methods. The left panel corresponds to Setting I, and the right to Setting II.

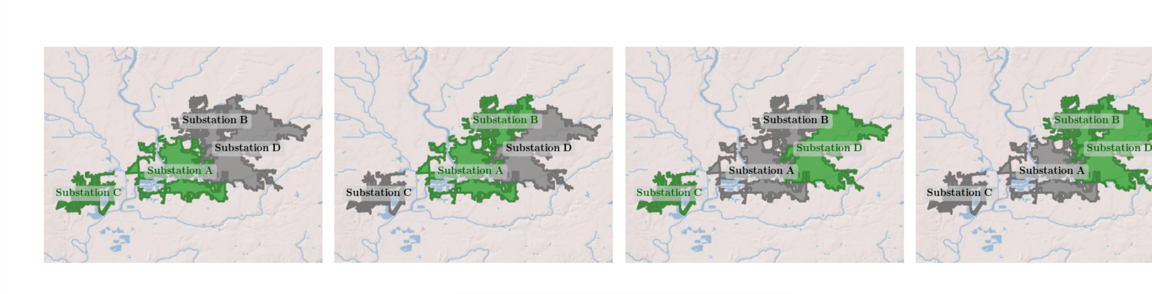


Figure: Top four candidate upgrade decisions with the lowest estimated risks (left to right) in our real data experiment recommended by `CREDO`. Each shaded region represents the span of a substation network.