Predict-and-Optimize Robust Unit Commitment with Statistical Guarantees via Weight Combination

Rui Xie

Joint with Yue Chen and Pierre Pinson

Aug 5, 2025

R. Xie, Y. Chen, and P. Pinson, “Predict-and-optimize robust unit commitment with statistical guarantees via weight combination,” arXiv:2411.03138, 2024.

Subject: Unit commitment (UC) under renewable and demand uncertainty

Goal:

- Enhance **out-of-sample performance**
- Ensure **robustness**

Idea: Bridge the gap between **data** and **optimization**

Approach

- Data-driven robust optimization (RO) with **statistical guarantees**
  - ✓ Consider the randomness of data-driven uncertainty set construction

Constructed based on N samples

$$\mathbb{P}^N[\Pr[u \in \mathcal{U}] \geq 1 - \varepsilon] \geq 1 - \delta$$

Uncertainty

- Integrated forecasting and optimization (**predict-and-optimize**)
  - ✓ Different from traditional prediction that minimizes the forecast error
  - ✓ Optimize the performance of the final strategy

Related Works

Optimization methods under uncertainty

- Stochastic programming (SP): Requires accurate probability distributions
- Traditional robust optimization (RO): Lacks **statistical guarantees**
- Distributionally robust optimization (DRO): Offers statistical guarantees but needs substantial data
- Data-driven RO with **dimension-free** statistical guarantee [1]: Not adapted to two-stage RO

"**Predict-and-optimize**" [2]:

- Utilized in UC [3,4]
- Lacks integration with RO and theoretical robustness guarantees

[1] L. J. Hong, Z. Huang, and H. Lam, "Learning-based robust optimization: Procedures and statistical guarantees," Management Science, 2021.
[2] A. N. Elmachtoub and P. Grigas, "Smart "predict, then optimize"," Management Science, vol. 68, no. 1, pp. 9–26, 2022.
[3] X. Chen, Y. Yang, Y. Liu, and L. Wu, "Feature-driven economic improvement for network-constrained unit commitment: A closed-loop predict-and-optimize framework," IEEE Transactions on Power Systems, 2022.
[4] H. Wu, D. Ke, L. Song, S. Liao, J. Xu, Y. Sun, and K. Fang, "A novel stochastic unit commitment characterized by closed-loop forecast-and-decision for wind integrated power systems," IEEE Transactions on Power Systems, 2024.

- Goal: Solve a two-stage chance-constrained UC problem

Day-ahead dispatch cost

$$\min_{x\in\mathcal{X},\eta} f(x)+\eta$$

Intraday redispatch cost

$$\text{s.t. } \Pr\Big[\min_{y\in\mathcal{Y}(x,u)} h(y)\le\eta\Big]\ge 1-\varepsilon$$

- The **accurate** distribution is unknown
- Consider a **conservative approximation** **(Lemma 1)** with $\Pr[u\in\mathcal{U}]\ge 1-\varepsilon$

$$\min_{x\in\mathcal{X}} f(x)+\max_{u\in\mathcal{U}}\min_{y\in\mathcal{Y}(x,u)} h(y)$$

⬇ Linearized cost function and constraints

Two-stage robust UC $\min_{x\in\mathcal{X}} C^T x+\max_{u\in\mathcal{U}}\min_{y:Ay\ge Bx+Du+E} F^T y$

Two-Stage UC Under Uncertainty

香港中文大學 The Chinese University of Hong Kong

- Pre-dispatch feasible region

Power balance

$$\mathcal{X}=\{x=(\theta_{gt},\theta_{gt}^{\pm},p_{gt},r_{gt}^{\pm};g\in\mathcal{G},t\in\mathcal{T})\mid$$

$$\sum_{g\in\mathcal{G}}p_{gt}=\sum_{i\in\mathcal{I}}\hat{u}_{it},\forall t\in\mathcal{T}, \quad (7a)$$

Network constraint

$$-S_l\le\sum_{g\in\mathcal{G}}\pi_{gl}p_{gt}-\sum_{i\in\mathcal{I}}\pi_{il}\hat{u}_{it}\le S_l,\forall l\in\mathcal{L},\forall t\in\mathcal{T}, \quad (7b)$$

Generator states

$$\theta_{gt},\theta_{gt}^{+},\theta_{gt}^{-}\in\{0,1\},\forall g\in\mathcal{G},\forall t\in\mathcal{T}, \quad (7c)$$

$$\sum_{\tau=t}^{t+T_g^{+}-1}\theta_{g\tau}\ge T_g^{+}\theta_{gt}^{+},1\le t\le T-T_g^{+}+1,\forall g\in\mathcal{G}, \quad (7d)$$

$$\sum_{\tau=t}^{T}(\theta_{g\tau}-\theta_{gt}^{+})\ge 0,T-T_g^{+}+2\le t\le T,\forall g\in\mathcal{G}, \quad (7e)$$

$$\sum_{\tau=t}^{t+T_g^{-}-1}(1-\theta_{g\tau})\ge T_g^{-}\theta_{gt}^{-},1\le t\le T-T_g^{-}+1,\forall g\in\mathcal{G}, \quad (7f)$$

$$\sum_{\tau=t}^{T}(1-\theta_{g\tau}-\theta_{gt}^{-})\ge 0,T-T_g^{-}+2\le t\le T,\forall g\in\mathcal{G}, \quad (7g)$$

$$\theta_{gt}-\theta_{g(t-1)}=\theta_{gt}^{+}-\theta_{gt}^{-},\forall g\in\mathcal{G},\forall t\in\mathcal{T}, \quad (7h)$$

$$\theta_{gt}^{+}+\theta_{gt}^{-}\le 1,\forall g\in\mathcal{G},\forall t\in\mathcal{T}, \quad (7i)$$

Reserve

$$0\le r_{gt}^{+}\le R_g^{+}\theta_{gt},0\le r_{gt}^{-}\le R_g^{-}\theta_{gt},\forall g\in\mathcal{G},\forall t\in\mathcal{T}, \quad (7j)$$

$$\underline{P}_g\theta_{gt}+r_{gt}^{-}\le p_{gt}\le\overline{P}_g\theta_{gt}-r_{gt}^{+},\forall g\in\mathcal{G},\forall t\in\mathcal{T}, \quad (7k)$$

Ramp constraints

$$(p_{gt}+r_{gt}^{+})-(p_{g(t-1)}-r_{g(t-1)}^{-})\le K_g^{+}\theta_{g(t-1)}+K_g^{U}\theta_{gt}^{+},\forall g\in\mathcal{G},2\le t\le T, \quad (7l)$$

$$-(p_{gt}-r_{gt}^{-})+(p_{g(t-1)}+r_{g(t-1)}^{+})\le K_g^{-}\theta_{gt}+K_g^{D}\theta_{gt}^{-},\forall g\in\mathcal{G},2\le t\le T\}. \quad (7m)$$

- Pre-dispatch cost

$$f(x)=\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\left(o_g^{+}\theta_{gt}^{+}+o_g^{-}\theta_{gt}^{-}+\rho_g p_{gt}+\gamma_g^{+}r_{gt}^{+}+\gamma_g^{-}r_{gt}^{-}\right)$$

Startup/shutdown Generation Reserve

- Redispatch cost

$$h(y)=\sum_{t\in\mathcal{T}}\sum_{g\in\mathcal{G}}\left(\rho_g^{+}p_{gt}^{+}+\rho_g^{-}p_{gt}^{-}\right),$$

- Redispatch feasible region

$$\mathcal{Y}(x,u)=\{y=(p_{gt}^{\pm};g\in\mathcal{G},t\in\mathcal{T})\mid$$

Power balance

$$\sum_{g\in\mathcal{G}}(p_{gt}+p_{gt}^{+}-p_{gt}^{-})=\sum_{i\in\mathcal{I}}u_{it},\forall t\in\mathcal{T}, \quad (9a)$$

Network constraint

$$-S_l\le\sum_{g\in\mathcal{G}}\pi_{gl}(p_{gt}+p_{gt}^{+}-p_{gt}^{-})-\sum_{i\in\mathcal{I}}\pi_{il}u_{it}\le S_l,\forall l\in\mathcal{L},\forall t\in\mathcal{T}, \quad (9b)$$

Reserve

$$0\le p_{gt}^{+}\le r_{gt}^{+},0\le p_{gt}^{-}\le r_{gt}^{-},\forall g\in\mathcal{G},\forall t\in\mathcal{T}\}. \quad (9c)$$

⬇

$$\min_{x\in\mathcal{X}} C^T x+\max_{u\in\mathcal{U}}\min_{y:Ay\ge Bx+Du+E} F^T y$$

Two-stage robust UC

Data-driven uncertainty set and statistical guarantees

- What we want: An uncertainty set s.t. $\mathbb{P}^N[\Pr[u \in \mathcal{U}] \geq 1-\varepsilon] \geq 1-\delta$
- What we have: Day-ahead **prediction** $\hat{u}$ and historical **forecast error** $e_{1:N}$
- Assumption: Day-ahead forecast errors are **i.i.d. continuous** random variables
- Steps
  - ✓ Divide the historical forecast error data into two **disjoint** groups N_1 and N_2
  - ✓ N_1 determines the ellipsoid's **shape** and center
  - ✓ N_2 determines the **size** (to include enough data points in the ellipsoid)
- **Theorem 1**: The optimal solution $x_0 := x^*_{\mathcal{U}_1}$ satisfies

$$\mathbb{P}^N\left[\Pr\left[O \leq O_{x_0} \leq O_{\mathcal{U}_1}\right] \geq 1-\varepsilon\right] \geq 1-\delta$$

Optimal value of the chance-constrained problem

Performance of the obtained solution

Optimal value of the RO problem

Robust UC with Statistical Guarantees

Uncertainty set reconstruction: To reduce conservativeness

- Leverage data and UC problem information
- **Lemma 1** (The **best** uncertainty set): If (x^*, η^*) is optimal in the chance-constrained problem, then $O = O_{\mathcal{U}^*}$, where

$$\mathcal{U}^* = \left\{u \middle| \min_{y \in \mathcal{Y}(x^*, u)} h(y) \le \eta^*\right\}, \Pr[u \in \mathcal{U}^*] \ge 1 - \varepsilon$$

- Approximate $\mathcal{U}^*$ using

$$(x, \eta) = (x_0, O_{x_0} - f(x_0))$$

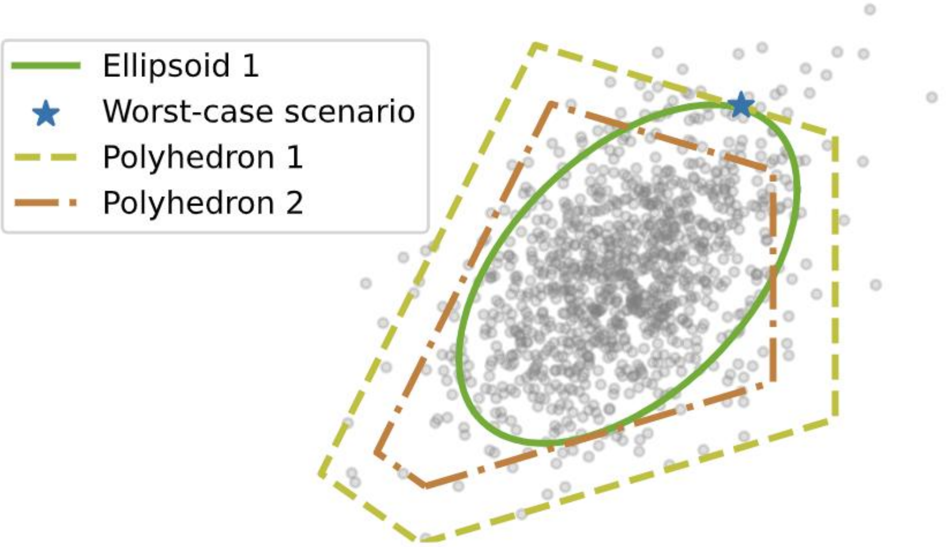
An obtained solution

Estimated performance in the historical dataset

- **Theorem 2** (Statistical guarantee):

$$\mathbb{P}^N\left[\Pr\left[O \le O_{x_1} \le O_{\mathcal{U}_2} \le f(x_0) + \beta\right] \ge 1 - \varepsilon\right] \ge 1 - \delta$$

Improve performance



Solution algorithm

- Uncertainty sets are either ellipsoidal or polyhedral
- Use the **C&CG algorithm** to solve two-stage RO problems

Algorithm 2: Solution of robust unit commitment

Input: Parameters of (11); ε; δ; $\mathcal{U}_0$; $\hat{u}$; $e_{1:N}$; N_2.
Output: UC strategy x_1.

Two datasets

1 Let $N_1 \leftarrow N - N_2$.
2 Divide $e_{1:N}$ into $e^{(1)}_{1:N_1}$ and $e^{(2)}_{1:N_2}$.

Construction

3 Calculate μ and Σ according to (14).
4 Let $\alpha \leftarrow \max\{(e_n^{(1)} - \mu)^\top \Sigma^{-1} (e_n^{(1)} - \mu) \mid n = 1, 2, \ldots, N_1\}$.
5 Let $\mathcal{U}_1' \leftarrow \{u \in \mathcal{U}_0 | (u - \hat{u} - \mu)^\top \Sigma^{-1} (u - \hat{u} - \mu) \le \alpha\}$.

C&CG

6 Solve problem (11) with $\mathcal{U} = \mathcal{U}_1'$ using the C&CG algorithm and obtain the optimal solution x_0.

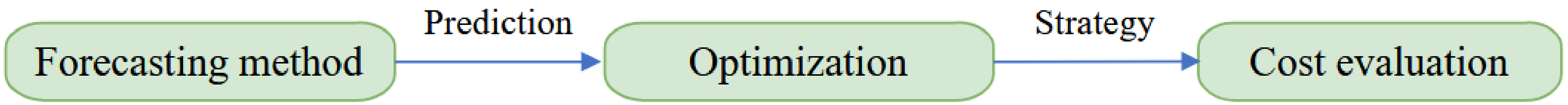
Reconstruction

7 Let $b_n \leftarrow \min_{y: Ay \ge Bx_0 + D(\hat{u} + e_n^{(2)}) + E} F^\top y$, for $n = 1, 2, \ldots, N_2$.
8 Arrange $b_n, n = 1, 2, \ldots, N_2$ from small to large and get $b_n', n = 1, 2, \ldots, N_2$.
9 Let $n^* \leftarrow \min\{n | \sum_{m=0}^{n-1} C_{N_2}^m (1-\varepsilon)^m \varepsilon^{N_2 - m} \ge 1 - \delta\}$.
10 Let $\beta \leftarrow b'_{n^*}$.
11 Let $\mathcal{U}_2 \leftarrow \{u \in \mathcal{U}_0 | \exists y, \text{ s.t. } Ay \ge Bx_0 + Du + E, F^\top y \le \beta\}$.

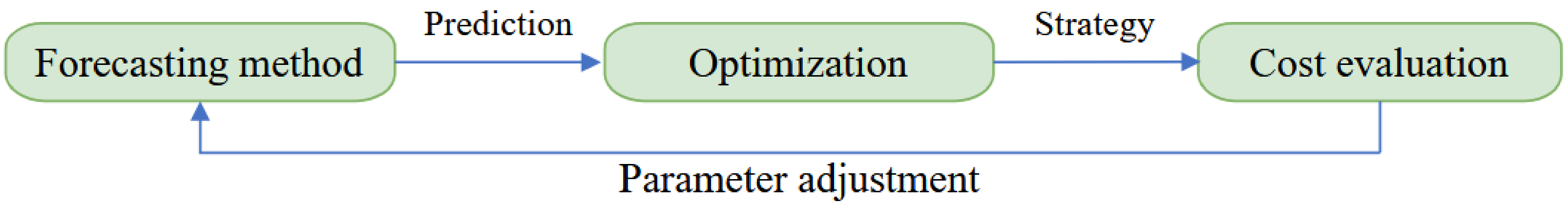
C&CG

12 Solve problem (11) with $\mathcal{U} = \mathcal{U}_2$ using the C&CG algorithm and return the optimal solution x_1.

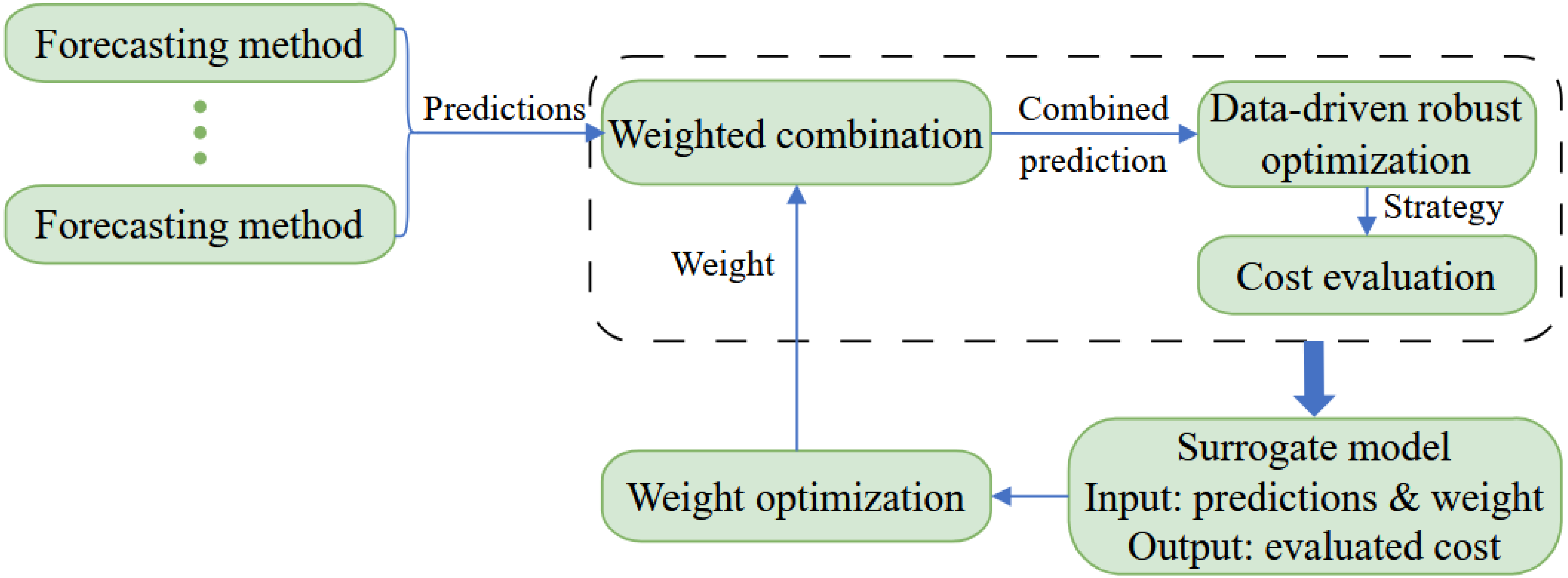
Traditional predict-then-optimize framework:



Existing predict-and-optimize framework:

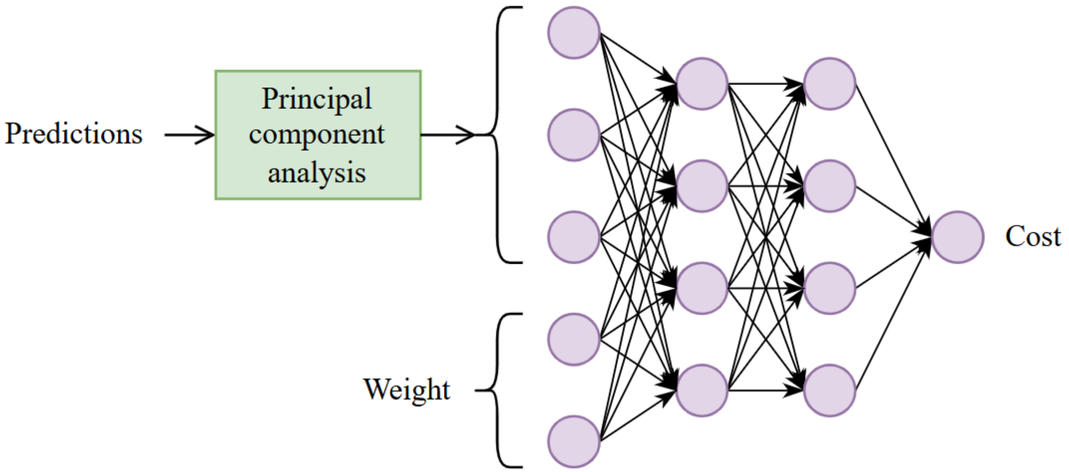


Proposed integrated forecasting and optimization framework:



Integrated Forecasting and Optimization Framework

- Construct a multilayer perceptron (MLP)-based surrogate model to **speed up** the weight optimization



Multilayer perceptron with ReLU activation

Algorithm 1: Surrogate model training

Input: Uncertainty samples $u_1, u_2, \ldots, u_{N_U}$; weight samples $w_1, w_2, \ldots, w_{N_W}$; dimension N_P of the PCA output; number of layers N_L in the MLP; number of units N_{Li} in layer i, where $i = 1, 2, \ldots, N_L$.

Output: PCA parameters $P_1, P_2, \ldots, P_{N_P}$; MLP weight parameter W_{ij} and bias parameter B_{ij} for $i = 0, 1, \ldots, N_L$ and $j = 1, 2, \ldots, N_{Li}$.

```
Compute PCA parameters P_1, P_2, ..., P_{N_P} using uncertainty samples u_1, u_2, ..., u_{N_U}.
for n_U = 1 to N_U do
    Compute PCA outputs d_{n_U 1} = P_1 u_{n_U}, d_{n_U 2} = P_2 u_{n_U}, ..., d_{n_U N_P} = P_{N_P} u_{n_U}.
    for n_W = 1 to N_W do
        Compute Û(w_{n_W}) = Σ_{m∈M} w_{n_W}^{(m)} Û^{(m)} and E^{(1:N)}(w_{n_W}) = U^{(1:N)} − Û^{(1:N)}(w_{n_W}).
        Based on Û(w_{n_W}) and E^{(1:N)}(w_{n_W}), apply Algorithm 2 to obtain the uncertainty set U(w_{n_W}) and the optimal solution x*_U(w_{n_W}).
        Using x*_U(w_{n_W}), compute the evaluated cost I(u_{n_U}, w_{n_W}) as described in Section II-C.
    end
end
Use the input data (d_{n_U 1}, d_{n_U 2}, ..., d_{n_U N_P}, w_{n_W}) and the output data I(u_{n_U}, w_{n_W}) for n_U = 1, 2, ..., N_U and n_W = 1, 2, ..., N_W to train the MLP with ReLU activation, and obtain the parameters W_{ij} and B_{ij} for i = 1, 2, ..., N_L and j = 1, 2, ..., N_{Li}.
```

MILP-based weight optimization

- ReLU activation function

$$v = \max\{0, s\} \Leftrightarrow \begin{cases} 0 \le v \le Mz \\ s \le v \le s + M(1 - z) \\ z \in \{0,1\} \end{cases}$$

Overview of the proposed framework

- Step 1 (**offline** training):
  - Train the surrogate model using historical data on uncertainty.
- Step 2 (**daily optimization** for UC):
  - Solve the MILP problem to obtain the optimal weight and the combined prediction.
  - Solve the data-driven RO problem and obtain the final strategy.

Advantages compared to existing predict-and-optimize frameworks

- Convex combination prevents severe deviations of predictions
- Surrogate model allows for rapid weight optimization
- Data-driven RO provides statistical guarantees and ensures out-of-sample performance

04 Case Studies: IEEE 30-Bus System

香港中文大學 The Chinese University of Hong Kong

Prediction data

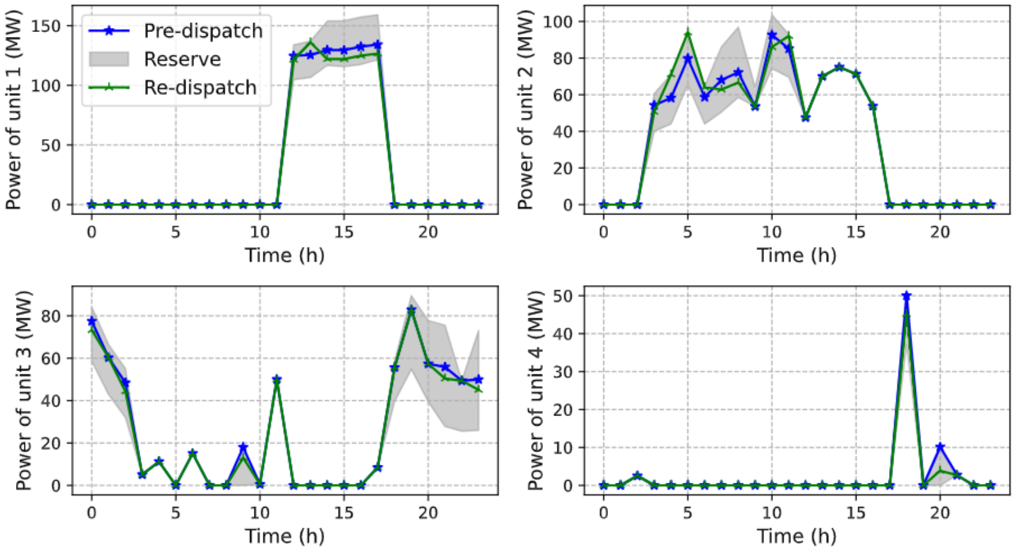
- Three forecasting methods
  - M1: BiLSTM
  - M2: Federated learning
  - M3: Using subprofiles
  - C1: Combine M1-M3 to minimize MSE
  - C2: Combine M1-M3 to optimize performance
- Combining predictions can **enhance accuracy**

TABLE I
AVERAGE FORECAST ERRORS OF DIFFERENT METHODS

	Method	RMSE	MAE
	M1	84.39	54.64
	M2	80.93	52.37
	M3	80.44	55.23
Minimize MSE	C1	76.14	51.26
Proposed	C2 (30-bus)	76.95	52.29
	C2 (118-bus)	78.72	53.80

Modified IEEE 30-bus system

- 6 generators and 2 wind farms
- Benchmark
  - $\varepsilon = \delta = 5\%, N_1 = 212, N_2 = 124$
  - 3 components in PCA
  - 2 hidden layers
  - Each layer has 16 units
  - $w = (0.28, 0.23, 0.49)$



Method comparison

TABLE II
SETTINGS OF UNIT COMMITMENT METHODS FOR COMPARISON

	Method	Statistical guarantee	Integrated forecasting and optimization	Uncertainty set reconstruction
	SP	×	×	×
100%	RO1	×	×	×
95%	RO2	×	×	×
	P1	✓	✓	×
	P2	✓	×	✓
	PSO	✓	✓	✓
	EXACT	×	✓	×
	Proposed	✓	✓	✓

TABLE III
UNIT COMMITMENT RESULTS OF DIFFERENT METHODS IN THE MODIFIED IEEE 30-BUS SYSTEM

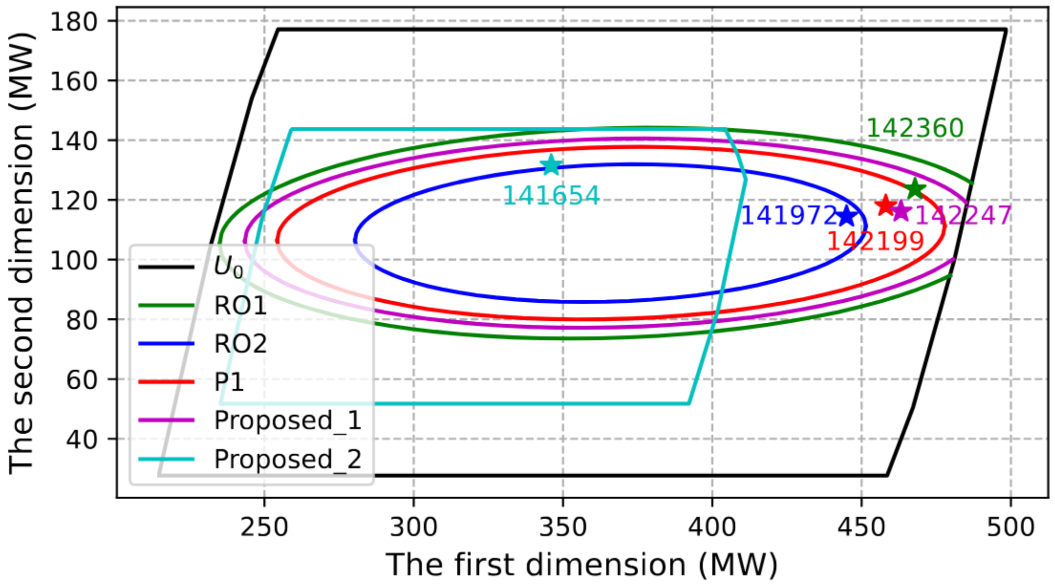
Method	Objective ($)	Feasible rate	Total cost ($)	Time (s)
SP	84832	88%	82985	218
RO1	106810	100%	92652	143
RO2	97350	97%	90468	94
P1	97848	98%	89149	124
P2	90122	98%	88318	147
PSO	89111	98%	88154	10000
Proposed	89725	98%	88243	121

- SP lacks robustness
- Traditional data-driven RO does not have statistical guarantees
- The proposed method and PSO have the **lowest** objective value and test total cost among methods that have **statistical guarantees**
- The proposed surrogate model **accelerates weight optimization** compared with PSO
- MILP-based exact solution method is not applicable to the UC problem due to generator state variables

Case Studies: A special case of two random loads

Project the uncertainty sets onto two dimensions

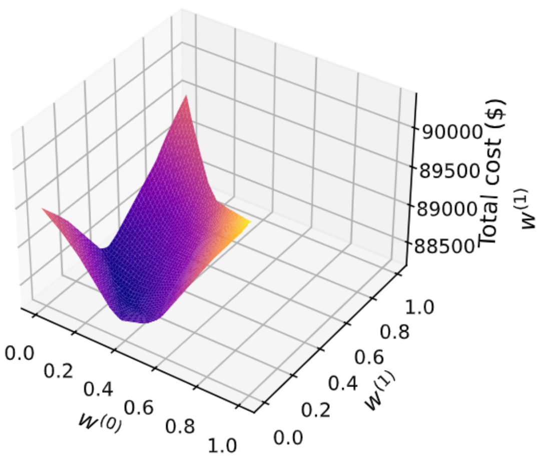
- Bound: $\mathcal{U}_0$
- RO1: 100% data points
- RO2: 95% data points
- P1: $\mathbb{P}^N[\Pr[u \in \mathcal{U}] \geq 95\%] \geq 95\%$
- **RO2 ⊂ P1 ⊂ RO1**
- Proposed_1, Proposed_2: The first and second uncertainty sets in the proposed method
- Proposed_2 excludes some **high-cost scenarios** in Proposed_1, but includes other regions to ensure the statistical guarantee

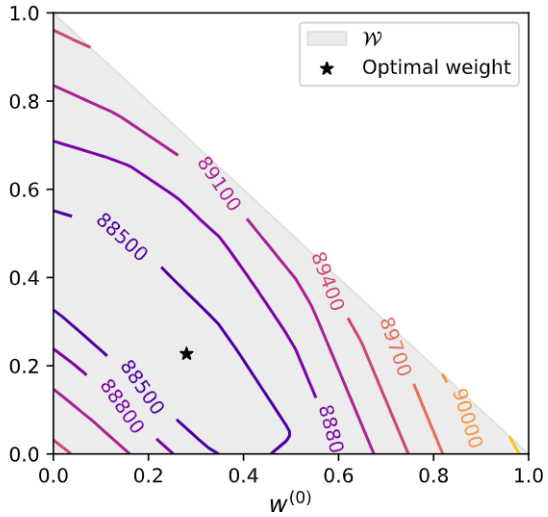


Case Studies: Sensitivity Analysis

$$\mathbb{P}^N[\Pr[u \in \mathcal{U}] \geq 1 - \varepsilon] \geq 1 - \delta$$

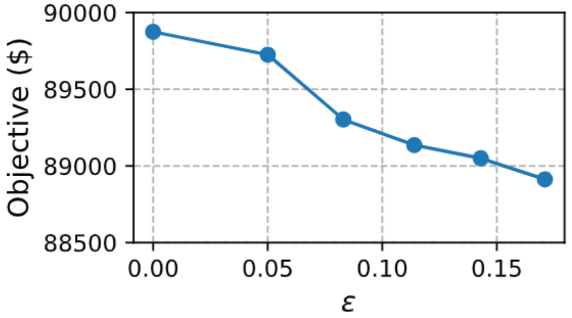
- Impact of weight w

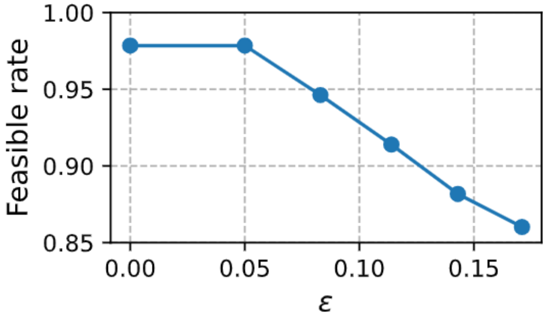




$$w^{(0)} + w^{(1)} + w^{(2)} = 1$$
$$w^{(0)}, w^{(1)}, w^{(2)} \geq 0$$

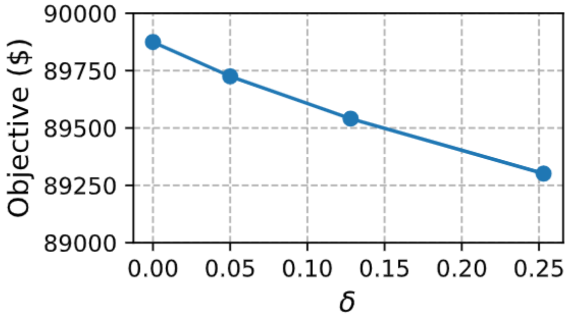
- Impacts of ε

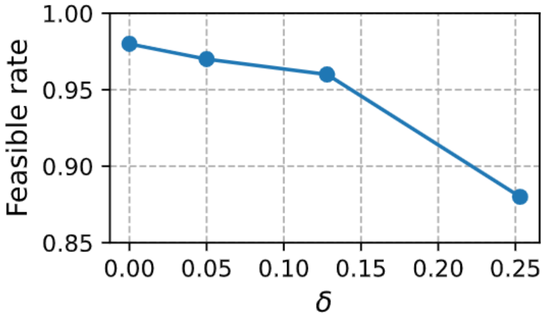




Test feasible rate $\geq 1 - \varepsilon$

- Impacts of δ





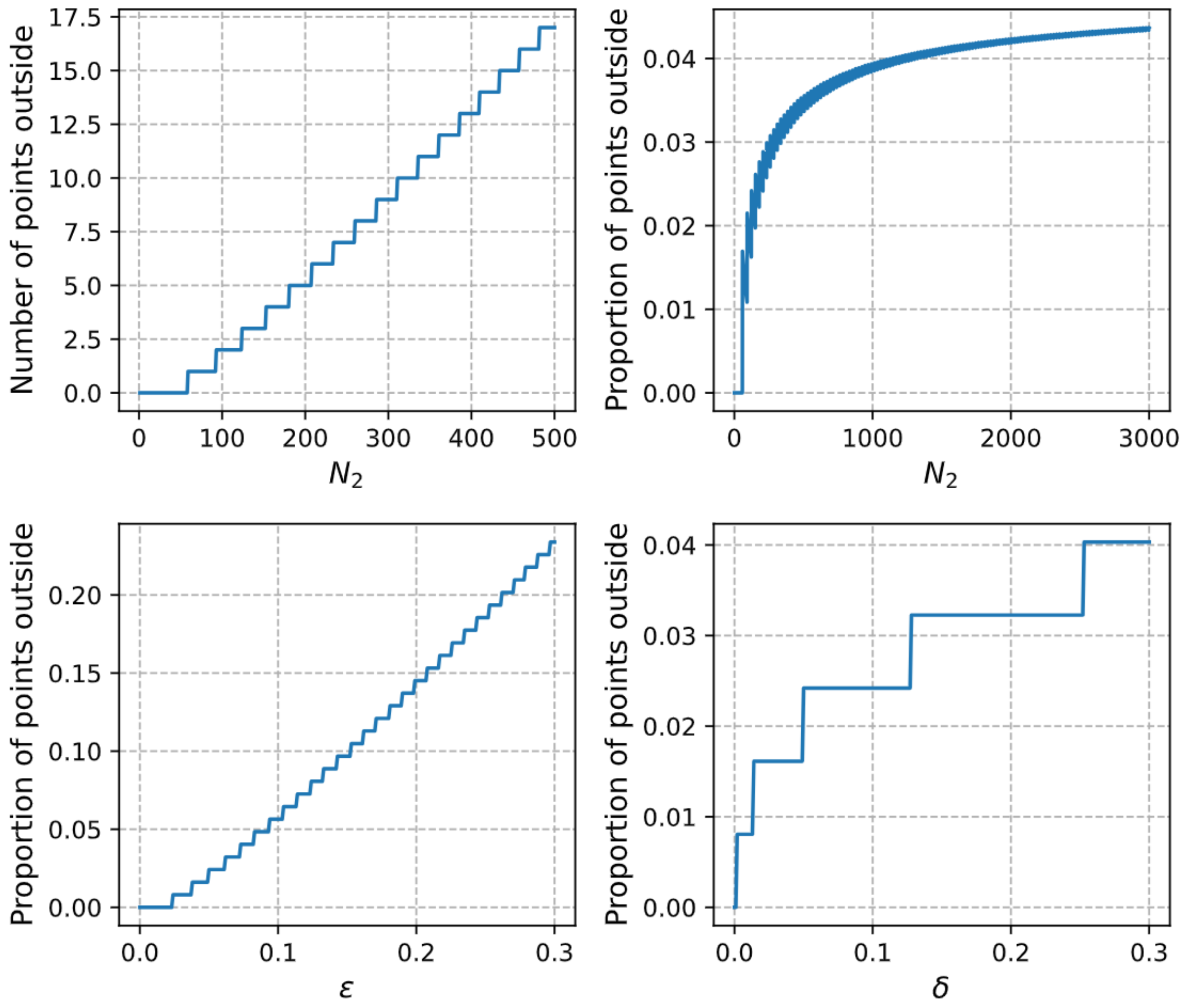
When $\delta \leq 0.13$, test feasible rate $\geq 1 - \varepsilon$

Requirements for statistical guarantee

- p: Proportion of points outside the uncertainty set
  - ✓ p $\uparrow$ ε ($N_2 \uparrow \infty$)
  - ✓ p $\uparrow$ ($\varepsilon \uparrow$) and p $\leq \varepsilon$
  - ✓ p $\uparrow$ ($\delta \uparrow$)

$$n^* = \min\{n \mid \sum_{m=0}^{n-1} C_{N_2}^{m} (1-\varepsilon)^m \varepsilon^{N_2-m} \geq 1-\delta\}$$

$$\mathbb{P}^N[\Pr[u \in \mathcal{U}] \geq 1-\varepsilon] \geq 1-\delta$$



Case Studies: IEEE 118-Bus System

- **Modified IEEE 118-bus system**
  - 54 generators and 2 wind farms
  - $\varepsilon = \delta = 5\%,\ N_1 = 212,\ N_2 = 124$

TABLE II
SETTINGS OF UNIT COMMITMENT METHODS FOR COMPARISON

	Method	Statistical guarantee	Integrated forecasting and optimization	Uncertainty set reconstruction
	SP	×	×	×
100%	RO1	×	×	×
95%	RO2	×	×	×
	P1	✓	✓	×
	P2	✓	×	✓
	Proposed	✓	✓	✓

TABLE IV
UNIT COMMITMENT RESULTS OF DIFFERENT METHODS IN MODIFIED IEEE 118-BUS SYSTEM

Method	Objective ($)	Feasible rate	Total cost ($)	Time (s)
SP	2061915	84%	2055123	1638
RO1	2096428	100%	2069461	256
RO2	2086813	100%	2064276	634
P1	2092717	100%	2068213	1304
P2	2066737	100%	2058488	1076
Proposed	2065591	100%	2056996	763

TABLE V
COMPUTATION EFFICIENCY IN THE MODIFIED IEEE 118-BUS SYSTEM

Number of random loads	25	21	17	13
Number of iterations	23	21	21	18
Computation time (s)	763	600	551	391

05 Conclusion

- Developed a **predict-and-optimize** two-stage robust UC method with **statistical guarantees**
  - Predict-and-optimize integration
  - Statistical guarantee

- Case studies show that the proposed method
  - **Balances** robustness and out-of-sample performance
  - **Outperforms** traditional SP and RO methods
  - It has satisfactory **scalability**

Thank You!

Rui Xie

Department of Mechanical and Automation Engineering

The Chinese University of Hong Kong