Feature-Driven Economic Improvement for Network-Constrained Unit Commitment: A Closed-Loop Predict-and-Optimize Framework

Xianbang Chen, *Student Member, IEEE,* Yafei Yang, *Member, IEEE,* Yikui Liu, *Member, IEEE,* Lei Wu, *Senior Member, IEEE*

***Abstract*—As an important application in the power system operation and electricity market clearing, the network-constrained unit commitment (NCUC) problem is usually executed by Independent System Operators (ISO) in an open-looped predict-then-optimize (O-PO) process, in which an upstream prediction (e.g., on renewable energy sources (RES) and loads) and a downstream NCUC are executed in a queue. However, in the O-PO framework, a statistically more accurate prediction may not necessarily lead to a higher NCUC economics against actual RES and load realizations. To this end, this paper presents a closed-loop predict-and-optimize (C-PO) framework for improving the NCUC economics. Specifically, the C-PO leverages structures (i.e., constraints and objective) of the NCUC model and relevant feature data to train a cost-oriented RES prediction model, in which the prediction quality is evaluated via the induced NCUC cost instead of the statistical forecast errors. Therefore, the loop between the prediction and the optimization is closed to deliver a cost-oriented RES power prediction for NCUC optimization. Lagrangian relaxation is adopted to accelerate the training process, making the C-PO applicable for real-world systems. Case studies on an IEEE RTS 24-bus system and an ISO-scale 5655-bus system with real-world data show that the proposed C-PO can effectively improve the NCUC economics as compared to the traditional O-PO.**

***Index Terms*—Unit commitment, data-driven, predict-and-optimize, prescription analysis, Lagrangian relaxation.**

NOMENCLATURE

Indexes and Sets

$b/\mathcal{B}$	Index/set of branches.
$e/\mathcal{E}$	Index/set of parameters in the prediction model.
$h, h'/\mathcal{H}$	Indexes/set of historical scenarios.
$i/\mathcal{I}$	Index/set of thermal units.
$j/\mathcal{J}$	Index/set of RESs.
$k/\mathcal{K}$	Index/set of generation segments.
$n/\mathcal{N}$	Index/set of feature types.
$q/\mathcal{Q}$	Index/set of load buses.
$s, s'/\mathcal{S}$	Indices/set of training scenarios.
$t, t'/\mathcal{T}$	Indices/set of hours.
$\mathcal{I}^{cu}/\mathcal{I}^{nu}$	Set of committed/non-committed units in NCUC.

Decision Variables

C_{it}	Generation cost of unit i at hour t.
D_{it}/U_{it}	Shutdown/startup of unit i at hour t.
I_{it}	Commitment of unit i at hour t.
LS_{qt}	Load shedding of load bus q at hour t.
O^{sr}_{it}/O^{nr}_{it}	Spinning/non-spinning reserve of unit i at hour t.
P_{it}	Generation dispatch of unit i at hour t.
P^{sg}_{ikt}	Dispatch of unit i in segment k at hour t.
w_{jt}	Generation dispatch of RES j at hour t.
$z^{\star}(\tilde{\boldsymbol{w}})$	NCUC objective corresponding to RES power $\tilde{\boldsymbol{w}}$.
$.^{rd}$	Decisions in the redispatch problem.

Parameters

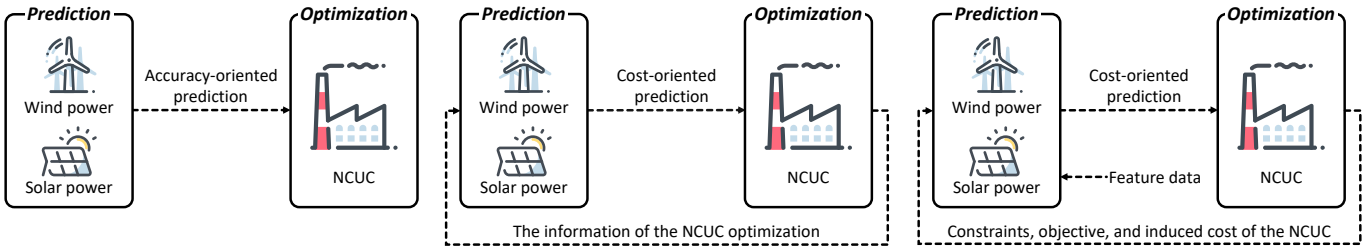
B_b	Transmission capacity of branch b.
C^{sg}_{ik}	Generation cost of unit i in segment k.
C^{su}_i/C^{sd}_i	Startup/shutdown cost of unit i.
C^{ls}	Cost of load shedding.
f^h_{tn}	Value of feature type n at hour t in scenario h.
$\mathcal{F}_b(\cdot)$	Power flow function of branch b.
L_{qt}	Electricity demand of load bus q at hour t.
N_i	No-load cost of unit i.
$\bar{O}^{sr}_i/\bar{O}^{nr}_i$	Spinning/non-spinning reserve limit of unit i.
$\bar{P}^{seg}_{ik}$	Power limit of unit i at segment k.
$\underline{P}_i$	Minimum generation of unit i.
$\bar{P}_i$	Maximum generation of unit i.
R^{su}_i/R^{sd}_i	Startup/shutdown ramping capacity of unit i.
$R^{\uparrow}_i/R^{\downarrow}_i$	Upward/downward ramping capacity of unit i.
S^{sr}_t/S^{nr}_t	System spinning/non-spinning reserve requirement at hour t.
TU_i/TD_i	Minimum on/off time requirement of unit i.
$\hat{\boldsymbol{w}}/\tilde{\boldsymbol{w}}$	Vector of predicted/actual available RES power.
$\hat{w}_{jt}/\tilde{w}_{jt}$	Predicted/actual available power of RES j at hour t.
$\lvert\cdot\rvert$	Cardinality of a set.

Manuscript received July 16, 2021; revised October 10, 2021; accepted November 13, 2021. Date of publication November 16, 2021; date of current version June 20, 2022. This work was supported in part by the U.S. National Science Foundation grant CNS-1915756 and the PSEG Foundation gift. Paper no. TPWRS-01127-2021. *(Corresponding author: Lei Wu.)*

The authors are with the ECE Department, Stevens Institute of Technology, Hoboken, NJ, 07030 USA. (E-mail: xchen130@stevens.edu, yyang125@stevens.edu, yliu262@stevens.edu, lei.wu@stevens.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2021.3128485.

Digital Object Identifier 10.1109/TPWRS.2021.3128485



(a). The traditional O-PO framework (b). The emerging C-PO framework (c). The presented feature-driven C-PO framework

Fig. 1. Illustration of different predict-then-optimize frameworks.

I. Introduction

THE network-constrained unit commitment (NCUC) problem has been widely recognized as one of the most important applications in the power system operation and electricity market clearing [1]–[4]. Specifically, the NCUC is routinely executed by Independent System Operators (ISO) to clear the day-ahead electricity market, which determines the optimal unit commitment, generation dispatch, and reserve schedules with the minimal system operation cost.

Generally, ISOs carry out NCUC in an open-looped predict-then-optimize (O-PO) framework as shown in Fig. 1(a). In O-PO, the upstream prediction step generates accuracy-oriented forecasts (e.g., on renewable energy sources (RES) and loads) based on the traditional statistical metrics, such as mean absolute error (MAE); then taking the predictions as inputs, NCUC is implemented in the downstream optimization step. However, the statistically more accurate predictions cannot necessarily enable higher NCUC economics against actual RES and load realizations. In this paper, the NCUC economics refers to the actual system cost which includes the startup and shutdown costs of generators involved in the day-ahead NCUC schedule and redispatch problems as well as the actual generation cost of generators corresponding to the final dispatch levels in the redispatch problem. It can be calculated after solving the day-ahead NCUC (based on RES predictions) and redispatch (based on given NCUC decisions and RES realizations) problems. Indeed, [5] clearly shown the asymmetric impacts of under- and over-predicted net loads on the NCUC economics. For instance, the over-predictions may merely cause RES curtailments, but the under-predictions could trigger the deployment of scarce and expensive non-spinning reserves. To this end, a biased prediction may outperform an accuracy-oriented prediction regarding the NCUC economics. As for improving NCUC economics, [5] used the tailored prediction techniques, instead of the accuracy-oriented predictions in O-PO, for properly considering the properties of NCUC.

To improve the NCUC economics of O-PO, one emerging idea is to generate cost-oriented predictions by feeding certain information of the downstream optimization (e.g., costs induced by the predictions) back to the upstream prediction and measuring the prediction quality via the induced cost instead of merely the statistical prediction accuracy. As shown in Fig. 1(b), this idea is referred to as the closed-loop predict-and-optimize (C-PO) [6]–[11]. Towards a class of newsvendor problems, [6] presented a simple yet effective data-driven method to perform the cost-oriented prediction and the optimization in a single step with the enhanced economics. However, concerning that the data-driven method in [6] could cause infeasibility when complex constraints are involved, a smart predict-then-optimize (SPO) framework was proposed to predict objective coefficients for linear programming (LP) [7] and mixed-integer linear programming (MILP) [8] problems. The core of the SPO framework is a SPO loss function which measures difference between the optimal objectives corresponding to the prediction and the actual realization. Recently, [9] and [10] introduced a concept of *predictive prescription* to formally describe the C-PO, which can directly prescribe decisions using readily available feature data. Reference [11] further presented a scalable prescription method for LP and MILP problems, which can seamlessly integrate the cost-oriented prediction and the optimization with complicated constraints. In summary, the above references [6]–[11] clearly demonstrated the potential of the cost-oriented C-PO methods in boosting the economics of real-world optimization tasks through a closed predict-and-optimize loop.

Recently, C-PO has also been applied on power system problems, including RES trading [12]–[16] and ED [17]–[21].

- Regarding the RES trading, [12] applied the decision tree-based SPO framework in [8] to derive bidding prices for RES producers with the improved profits. Using a regression tree-based C-PO method, [13] reduced the total cost of a wind farm, even though the root mean square error (RMSE) of wind power predictions is slightly worse. Based on the particle swarm optimization, [14] designed a profit-oriented approach to predict RES power, which tightly integrates the RES power prediction and trading to boost the profits of RES producers. Concerning the complexity of the method in [14], [15] adopted the easy-to-implement method in [6] with publicly accessible feature data (e.g., raw RES predictions) to drive the cost-oriented predictions. Compared to the O-PO benchmark in the Danish electricity market, this feature-driven method derives slightly under-predicted results with improved profits of RES producers. The same authors further proposed a bilevel programming-based C-PO approach in [16] for RES power forecasting and trading in the Iberian electricity market, illustrating that the feedback of the optimization can render enhanced economic benefits over the O-PO approach.

TABLE I
APPLICATIONS OF C-PO ON POWER SYSTEMS

Research	Prediction task	Optimization task	Foundation of C-PO methodology	Property of C-PO methodology	Property of optimization task
[12]	RES bidding price	RES trading	Decision tree-based SPO method in [8]	Heuristic	LP
[13]	RES generation	RES trading	Regression tree method	Heuristic	LP
[14]	RES generation	RES trading	Particle swarm optimization	Heuristic	LP
[15]	RES generation	RES trading	Data-driven method in [6]	Mathematical programming	LP
[16]	RES generation	RES trading	Bilevel programming	Mathematical programming	LP
[17]	Load and reserve	ED	Bilevel programming	Mathematical programming	LP
[18], [19]	Energy price	ED	SPO method in [7]	Mathematical programming	LP
[20]	Load	ED	Kernel learning	Heuristic	LP
[21]	Load	ED	Deep neural network	Heuristic	LP
This paper	RES generation	NCUC	A method inspired by SPO loss in [7]	Mathematical programming	MILP

- In terms of the applications on ED, [17] enhanced the C-PO method [16] to reduce the long-run ED costs for a 300-bus system by 11.4%. References [18] and [19] investigated the effectiveness of the SPO framework [7] in predicting energy prices, and highlighted the remarkable economic improvements in the LP-based ED problem. References [20] and [21] analyzed the underlying reasons for the remarkable economic improvements of the cost-oriented C-PO in ED, which indeed can derive load predictions with slightly higher forecast errors yet much lower ED costs.

Table I summarizes the C-PO applications in power systems [12]–[21], which exclusively focus on LP problems. However, extending C-PO to the MILP-based NCUC could encounter computational challenges, especially for large-scale systems. Motivated by the above observations, this paper presents a feature-driven C-PO framework with three modules to improve the NCUC economics. First, the data-processing module performs the feature and scenario selections to identify proper feature types and training scenarios. Given the features and scenarios, the cost-oriented modeling-and-training module forms an empirical risk minimization (ERM) problem based on the SPO loss [7], which is solved via a Lagrangian relaxation (LR) based decomposition to obtain a cost-oriented RES prediction model. Finally, the closed-loop predict-and-optimize module integrates the trained RES prediction model and the NCUC formulation as a feature-driven NCUC prescription model, which jointly executes the cost-oriented RES power prediction and the NCUC optimization.

It is noteworthy that this paper does not intend to present a new prediction method for replacing the statistical accuracy-oriented prediction models in O-PO. Instead, the presented C-PO utilizes the accuracy-oriented predictions as feature data to generate the cost-oriented RES power predictions for NCUC. To this end, the presented C-PO framework is compatible to and can be fairly compared with the ISO's current O-PO practice. Although only the RES power prediction is discussed in this paper, the presented C-PO framework can be extended to predict other factors such as load involved in NCUC.

The major works in this paper are summarized as follows:

- As shown in Fig. 1(c), a novel feature-driven C-PO framework is presented for improving the NCUC economics. Specifically, by leveraging the feature data, the NCUC structure (i.e., constraints and objective), and the induced NCUC cost, the C-PO can deliver cost-oriented RES power predictions for improving the NCUC economics. Unlike the SPO framework [7] which is limited to predicting objective parameters of LP problems, the presented C-PO can predict right-hand-side parameters of the MILP-based NCUC;
- For real-world large-scale systems with massive network constraints, the C-PO optimization task which embeds the NCUC structure could be intractable if solved directly [28]. Therefore, a customized LR-based decomposition approach [25] is applied to effectively solve the MILP-based C-PO model;
- With practical historical data from the Belgian ISO [22], experiments on the IEEE RTS 24-bus system show that the presented C-PO outperforms the O-PO with as much as 1.96% improvements on the daily average NCUC economics. Tests on the ISO-scale 5655-bus system further demonstrate the potential of the C-PO for large-scale NCUC problems.

The rest of the paper is organized as follows: Section II presents the procedure to evaluate the NCUC economics and adopts an illustrative example to justify our motivation to improve the NCUC economics; Section III expounds the presented C-PO framework; Section IV analyzes the experimental results; and Section V concludes this paper.

II. MOTIVATION TO IMPROVE NCUC ECONOMICS OF O-PO

This section first expounds the procedure to evaluate the NCUC economics, followed by an illustrative example showing our motivation of exploring the C-PO framework to improve the NCUC economics of the traditional O-PO framework.

A. The Process of Evaluating NCUC Economics

As shown in Fig. 2, the NCUC economics is evaluated by solving the day-ahead NCUC problem and the redispatch problem in a queue. Specifically, in the day-ahead stage, as the actual RES power availability on the dispatch day remains unknown, the NCUC is implemented using the RES predictions; after the actual RES power availability is observed,

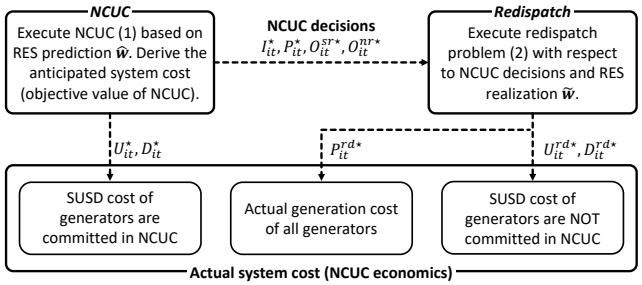


Fig. 2. Procedure to evaluate the NCUC economics.

the redispatch problem is executed to find optimal dispatch decisions with respect to the given NCUC solutions.

The compact form of the MILP-based NCUC is shown as in (1), where $\boldsymbol{x}$ represents the M-dimentional vector of binary variables in the NCUC problem; $\boldsymbol{y}$ denotes the vector of continuous variables; $\boldsymbol{c}$ and $\boldsymbol{d}$ indicate the cost vectors; $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{F}$ express the constant matrices; $\boldsymbol{g}$ is the right-hand-side vector of all constraints excluding the RES power limits; $\boldsymbol{Fy} \leq \hat{\boldsymbol{w}}$ represents the RES power limits with $\hat{\boldsymbol{w}}$ being the vector of RES power predictions. The detailed formulation of the NCUC is provided in Appendix A.

Solving the NCUC problem (1) will provide: (*i*) the optimal solutions to unit commitment $I^{\star}_{it}$, generation dispatch $P^{\star}_{it}$, spinning reserve schedule $O^{sr\star}_{it}$, and non-spinning reserve schedule $O^{nr\star}_{it}$; (*ii*) the *startup/shutdown (SUSD) cost* of generators $\boldsymbol{c}^{\top}\boldsymbol{x}^{\star}$; and (*iii*) the *anticipated system cost* $z^{\star}(\hat{\boldsymbol{w}})$ (which, as shown in Fig. 1(c), is also referred to as the *induced NCUC cost* given the RES predictions $\hat{\boldsymbol{w}}$). Here, the superscript $\star$ indicates optimal solutions of the corresponding variables.

$$z(\hat{\boldsymbol{w}}) = \min_{\boldsymbol{x},\boldsymbol{y}} \boldsymbol{c}^{\top}\boldsymbol{x} + \boldsymbol{d}^{\top}\boldsymbol{y} \tag{1.1}$$

$$s.t. \quad \boldsymbol{Ax} + \boldsymbol{By} \leq \boldsymbol{g} \tag{1.2}$$

$$\boldsymbol{Fy} \leq \hat{\boldsymbol{w}} \tag{1.3}$$

$$\boldsymbol{x} = [x_1, \cdots, x_M] \in \{0,1\}^M \tag{1.4}$$

Taking the solutions $I^{\star}_{it}$, $P^{\star}_{it}$, $O^{sr\star}_{it}$, and $O^{nr\star}_{it}$ from (1) as inputs, the redispatch problem (2) is further solved to derive the optimal dispatch decisions with respect to the actual RES realizations $\tilde{\boldsymbol{w}}$. In (2), the superscript cu/nu indicates the term corresponds to the generators committed/not committed in NCUC. The detailed formulation of (2) is provided in Appendix B.

$$z^{rd}(\tilde{\boldsymbol{w}}) = \min_{\boldsymbol{x}^{nu},\boldsymbol{y}^{cu},\boldsymbol{y}^{nu}} \boldsymbol{c}^{nu\top}\boldsymbol{x}^{nu} + \boldsymbol{d}^{cu\top}\boldsymbol{y}^{cu} + \boldsymbol{d}^{nu\top}\boldsymbol{y}^{nu} \tag{2.1}$$

$$s.t. \quad \boldsymbol{B}^{cu}\boldsymbol{y}^{cu} \leq \boldsymbol{g}^{cu} \tag{2.2}$$

$$\boldsymbol{A}^{nu}\boldsymbol{x}^{nu} + \boldsymbol{B}^{nu}\boldsymbol{y}^{nu} \leq \boldsymbol{g}^{nu} \tag{2.3}$$

$$\boldsymbol{F}^{cu}\boldsymbol{y}^{cu} + \boldsymbol{F}^{nu}\boldsymbol{y}^{nu} \leq \tilde{\boldsymbol{w}} \tag{2.4}$$

$$\boldsymbol{x}^{nu} = [x^{nu}_1, \cdots, x^{nu}_N] \in \{0,1\}^N \tag{2.5}$$

In the redispatch problem (2), (2.2) describes constraints for generators committed in NCUC (i.e., $I^{\star}_{it} = 1$). That is, a generator committed in NCUC shall remain the ON-state ($I^{rd}_{it} = I^{\star}_{it} = 1$) and can only adjust its generation P^{rd}_{it} based on the base-point generation dispatch $P^{\star}_{it}$ and spinning reserve schedule $O^{sr\star}_{it}$ (i.e., $\boldsymbol{g}^{cu}$ depends on $P^{\star}_{it}$ and $O^{sr\star}_{it}$ as shown in (36)). Constraints (2.3) and (2.5) describe limits for generators not committed in NCUC (i.e., $I^{\star}_{it} = 0$). That is, a generator not committed in NCUC can remain OFF-state ($I^{rd}_{it} = 0$), or quickly switch to ON-state ($I^{rd}_{it} = 1$) for providing generation P^{rd}_{it} within its non-spinning reserve range $O^{nr\star}_{it}$ (i.e., $\boldsymbol{g}^{nu}$ depends on $O^{nr\star}_{it}$) as shown in (37)-(38)).

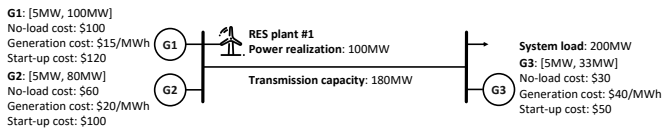


Fig. 3. Structure of a 2-bus system.

After solving the redispatch problem (2), we can obtain the *redispatch cost* $z^{rd\star}(\tilde{\boldsymbol{w}})$ which consists of: (*i*) *SUSD cost* of the generators not commitment in NCUC ($\boldsymbol{c}^{nu\top}\boldsymbol{x}^{nu\star}$) and (*ii*) *generation cost* of all the generators ($\boldsymbol{d}^{cu\top}\boldsymbol{y}^{cu\star} + \boldsymbol{d}^{nu\top}\boldsymbol{y}^{nu\star}$).

As a result, the *actual system cost* against the actual RES realization $\tilde{\boldsymbol{w}}$ is measured by the summation of $\boldsymbol{c}^{\top}\boldsymbol{x}^{\star}$ from (1) and $z^{rd\star}(\tilde{\boldsymbol{w}})$ from (2.1), which accounts for the *NCUC economics*.

B. Justification on the Motivation to Improve NCUC Economics of O-PO: An Illustrative Example

A 2-bus system shown in Fig. 3 is utilized to justify our motivation of using the C-PO for enhancing the NCUC economics of O-PO. This example conducts a single-hour NCUC study. Spinning/non-spinning reserve capacities $\bar{O}^{sr}_i/\bar{O}^{nr}_i$ of the three generators are respectively 30MW/10MW, 24MW/40MW, and 9.9MW/33MW. Spinning/non-spinning reserve requirements of system S^{sr}_t/S^{nr}_t are set as 3%/7% of the system load.

TABLE II
COMPARISON OF C-PO AND O-PO VIA THE 2-BUS SYSTEM

			Case 1		Case 2	
Method			C-PO	O-PO	C-PO	O-PO
RES power prediction/MW			130	72	90	107
MAE/MW			30	28	10	7
NCUC	G1	Dispatch/MW	50	97	90	73
		Reserve/MW	±6	±4	±6	±0
	G2	Dispatch/MW	OFF	11	OFF	OFF
		Reserve/MW	+40	±6	+40	+40
	G3	Dispatch/MW	20	20	20	20
		Reserve/MW	±0	±10	±0	±6
	Dispatch of RES/MW		130	72	90	107
	Anticipated system cost/$		1,850	2,938	2,450	2,195
Redispatch	Actual generation of G1/MW		56	93	84	73
	Actual generation of G2/MW		24	5	OFF	7
	Actual generation of G3/MW		20	20	20	20
	Actual utilized RES/MW		100	82	96	100
Actual system cost/$			**2,580**	**2,754**	**2,360**	**2,495**

Note: symbol "±" indicates the bi-directional spinning reserve O^{sr}_{it}; symbol "+" means the upward only non-spinning reserve O^{nr}_{it}.

Two cases are reported in Table II. In Case 1, C-PO delivers an over-prediction of RES power and O-PO obtains an under-prediction; while Case 2 describes an opposite situation. Both cases show that O-PO provides more accurate RES power forecasts than C-PO, but is outperformed by C-PO in terms of the NCUC economics.

Case 1: In the O-PO with an under-predicted RES power ($\hat{w}_1$ = 72MW), the total thermal generation from the NCUC solution is high ($\sum_{i=1}^{3} P_i^\star$ = 128MW). Thus, the redispatch can only partly utilize the available RES power (i.e., w_1^{rd} = 82MW but the available RES power $\tilde{w}_1$ = 100MW), due to the limited spinning-down reserve and the transmission restriction. In comparison, the C-PO obtains an over-prediction result ($\hat{w}_1$ = 130MW) and has to deploy the non-spinning reserve of G2 in the redispatch (P_2^{rd} = 24MW); however, C-PO can fully use the actual available RES power (w_1^{rd} = 100MW). Finally, C-PO achieves a lower actual system cost ($2,580 vs $2,754).

Case 2: Both methods turn on G1 and G3 in NCUC. However, O-PO dispatches G1 at 73MW, which causes the 7MW non-spinning reserve utilization of G2 in the redispatch, in despite of the full utilization of the 100MW RES power. In comparison, even though C-PO can only utilize 96MW RES power due to the limited spinning-down reserve and transmission restriction, the non-spinning reserve of G2 is not deployed in the redispatch. As a result, C-PO reaches a lower actual system cost ($2,360 vs $2,495).

This illustrative example clearly shows that the predictions of higher accuracy do not necessarily lead to better NCUC economics. The reason is that the open-looped process fails to consider the complicated structure of NCUC as well as the nonlinear dependency of the actual system cost on NCUC solutions and redispatch efforts against actual RES realizations.

In summary, for improving the NCUC economics, it is necessary to generate the cost-oriented predictions by closing the loop between the prediction and the optimization, even though the prediction accuracy might be slightly compromised.

III. Feature-Driven C-PO Framework

This section first expounds the three major steps to implement the feature-driven C-PO as shown in Fig. 4, including data-processing, cost-oriented modeling-and-training, and closed-loop predict-and-optimize modules. Finally, a rolling C-PO execution for routinely solving the daily NCUC is introduced.

A. Data-Processing Module

The prediction model will be trained via historical data. Specifically, data of a single historical scenario h include the perfect NCUC cost $z^\star(\tilde{\boldsymbol{w}}^h)$ corresponding to the actual RES realization $\tilde{\boldsymbol{w}}^h$ (i.e., solving (1) with $\tilde{\boldsymbol{w}}^h$) and a feature vector $\boldsymbol{f}$ containing $|\mathcal{N}|$ feature types. Potential feature types could include day-ahead predictions on loads and RES farms, which are readily available to ISOs in the day-ahead stage. The data-processing module executes the feature selection and training scenario selection to identify the relevant feature types and representative historical scenarios for each RES farm j.

1) Feature Selection: Generally, various feature types (e.g., region load and raw predictions of RES) are available in practice. If too few feature types are used to train the prediction model, the model may perform poorly in both the training and the practical perdition process (i.e., underfitting); on the other hand, if too many feature types are utilized for training, the model could possess many parameters and become overly complex, so that it could not be well generalized in the practical perdition process (i.e., overfitting). Therefore, feature selection is utilized to determine the proper set of relevant feature types for the prediction model, so that an accurate and generalizable prediction model can be derived while avoiding potential underfitting and overfitting issues.

Specifically, the importance of individual feature types on RES farm j is quantified via the following steps.

Step 1: For hour t of RES farm j, build a linear multivariate model (3) between its historical available RES power vector $[\tilde{w}_{tj}^1, \cdots, \tilde{w}_{tj}^{|\mathcal{H}|}]$ and feature vectors $\boldsymbol{f}_{t,1}, \cdots, \boldsymbol{f}_{t,|\mathcal{N}|}$ with the regression coefficients $\beta_{t,j,0}, \cdots, \beta_{t,j,|\mathcal{N}|}$. The regression coefficients is calculated via the least-squares method.

$$\begin{bmatrix} \tilde{w}_{t,j}^1 \\ \vdots \\ \tilde{w}_{tj}^{|\mathcal{H}|} \end{bmatrix} = \beta_{t,j,0} + \beta_{t,j,1} \begin{bmatrix} f_{t,1}^1 \\ \vdots \\ f_{t,1}^{|\mathcal{H}|} \end{bmatrix} + \cdots + \beta_{t,j,|\mathcal{N}|} \begin{bmatrix} f_{t,|\mathcal{N}|}^1 \\ \vdots \\ f_{t,|\mathcal{N}|}^{|\mathcal{H}|} \end{bmatrix} \quad (3)$$

Step 2: Calculate the average standard regression coefficient (SRC) SRC_{jn}^{avr} over all $|\mathcal{T}|$ hours via (4) for each feature type n, where σ_{tj} and σ_{tn} are respectively standard deviations of $[\tilde{w}_{tj}^1, \cdots, \tilde{w}_{tj}^{|\mathcal{H}|}]$ and $[f_{tn}^1, \cdots, f_{tn}^{|\mathcal{H}|}]$.

$$SRC_{jn}^{avr} = \sum_{t \in \mathcal{T}} \frac{|\beta_{tjn}\sigma_{tn}/\sigma_{tj}|}{|\mathcal{T}|} \quad \forall j \in \mathcal{J}, n \in \mathcal{N} \quad (4)$$

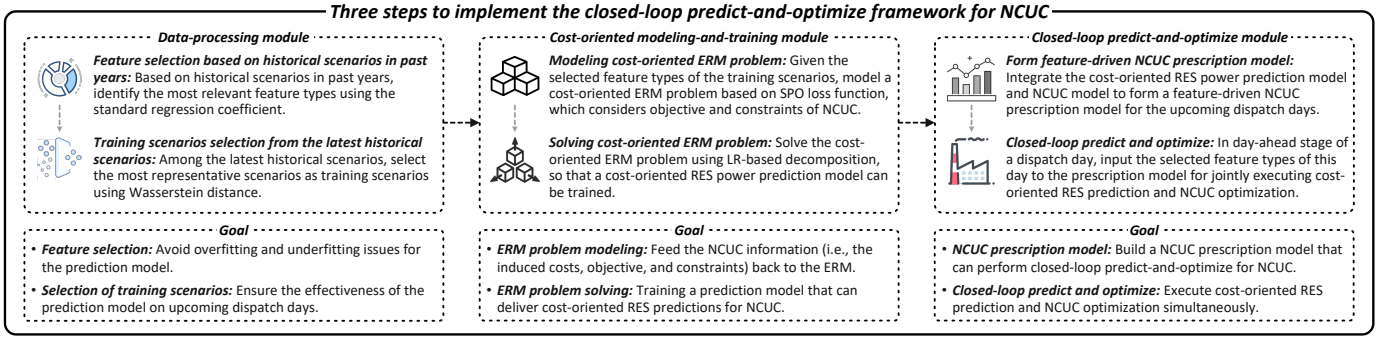


Fig. 4. The three steps to realize the feature-driven closed-loop predict-and-optimize framework for NCUC.

A larger value of SRC_{jn}^{avr} indicates a higher importance of feature type n on RES farm j. Finally, feature type n with the highest value SRC_{jn}^{avr} is selected as the relevant feature of RES farm j.

Among various feature selection methods, SRC has been widely regarded as simple [23] yet effective [27] for identifying the relationship between feature types. In this paper, we leverage its interpretability to quantify the relevance of individual feature types on the RES power output based on massive historical scenarios, ensuring that the identified feature type is reasonable and interpretable. Although certain nonlinear regression models may describe the relationship between feature types and the RES power output more accurately than the linear regression model (3), the accuracy of the linear regression model can be well guaranteed via the SRC method while maintaining the cost-oriented perdition model to be linear and consistent with the current ISO practice.

2) Selection of Training Scenarios: After selecting the relevant features, NT representative historical scenarios among the latest NH historical days shall be identified to effectively train a prediction model for the upcoming dispatch day.

Regarding this, Wasserstein distance $\mathcal{W}^{h\prime}$ [24] between the RES prediction power $\hat{\boldsymbol{w}}^{h\prime}$ and the actual RES realization $\tilde{\boldsymbol{w}}^{h\prime}$ is utilized to quantitatively assess the extremeness of the latest historical scenario h'. $\hat{\mathcal{C}}^{h\prime}$ and $\tilde{\mathcal{C}}^{h\prime}$ are empirical cumulative distribution functions of $\hat{w}^{h\prime}$ and $\tilde{w}^{h\prime}$, respectively. Then, NT representative scenarios with median values of $\mathcal{W}^{h\prime}$ are identified as training scenarios for the upcoming dispatch days.

$$\mathcal{W}^{h\prime}\left(\hat{\boldsymbol{w}}^{h\prime},\tilde{\boldsymbol{w}}^{h\prime}\right)=\int_{-\infty}^{+\infty}\left|\hat{\mathcal{C}}^{h\prime}(x)-\tilde{\mathcal{C}}^{h\prime}(x)\right|dx \quad \forall h'=1,\cdots,NH \tag{5}$$

For example, we select scenarios for the upcoming dispatch day 2020/01/07 with NH = 6 and NT = 2. First, data of the latest NH = 6 historical days 2020/01/01 to 2020/01/06 are collected. Assuming their $\mathcal{W}^{h\prime}$ are 0.63, 1.36, 0.50, 0.98, 0.37, and 0.60, the NT = 2 days 2020/01/06 ($\mathcal{W}^{h\prime}$ = 0.60) and 2020/01/01 ($\mathcal{W}^{h\prime}$ = 0.63) with the median $\mathcal{W}^{h\prime}$ values will be identified as representative training scenarios for 2020/01/07.

B. Cost-Oriented Modeling-and-Training Module

The core idea of the closed-loop framework is to train a cost-oriented RES power prediction model that is tailored for NCUC. Given the proper feature types and representative scenarios identified from the data-processing module, the cost-oriented ERM problem based on the SPO loss is formed for training the cost-oriented RES prediction model.

1) Cost-Oriented ERM Problem Modeling: The SPO loss is first defined, which is utilized to quantify the system operation cost loss due to the imperfect RES prediction information.

Definition 1 [7]: Given the RES prediction $\hat{\boldsymbol{w}}_s$ and the actual realization $\tilde{\boldsymbol{w}}_s$ of scenario s, the SPO loss $l_s^{spo}(\hat{\boldsymbol{w}}_s,\tilde{\boldsymbol{w}}_s)$ is defined as in (6), where $z^\star(\hat{\boldsymbol{w}}_s)$ and $z^\star(\tilde{\boldsymbol{w}}_s)$ are respectively optimal objective values of (1) induced by $\hat{\boldsymbol{w}}_s$ and $\tilde{\boldsymbol{w}}_s$.

$$l_s^{spo}(\hat{\boldsymbol{w}}_s,\tilde{\boldsymbol{w}}_s):=\left|z^\star(\hat{\boldsymbol{w}}_s)-z^\star(\tilde{\boldsymbol{w}}_s)\right| \tag{6}$$

By leveraging l_s^{spo} and the NCUC model (1), the ERM (7) can be formed to train a cost-oriented perdition model $\boldsymbol{H}$.

$$\begin{aligned}
&\min_{\substack{\boldsymbol{H},\\ \boldsymbol{x}_1,\cdots,\boldsymbol{x}_{|\mathcal{S}|},\\ \boldsymbol{y}_1,\cdots,\boldsymbol{y}_{|\mathcal{S}|}}} \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\{l_s^{spo}(\hat{\boldsymbol{w}}_s,\tilde{\boldsymbol{w}}_s)\}+\lambda\left\|\boldsymbol{H}\right\|_1\\
=&\min_{\substack{\boldsymbol{H},\\ \boldsymbol{x}_1,\cdots,\boldsymbol{x}_{|\mathcal{S}|},\\ \boldsymbol{y}_1,\cdots,\boldsymbol{y}_{|\mathcal{S}|}}} \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\left\{\left|(\boldsymbol{c}^\top\boldsymbol{x}_s+\boldsymbol{d}^\top\boldsymbol{y}_s)-z^\star(\tilde{\boldsymbol{w}}_s)\right|\right\}+\lambda\left\|\boldsymbol{H}\right\|_1\\
s.t.\ &\boldsymbol{A}\boldsymbol{x}_s+\boldsymbol{B}\boldsymbol{y}_s\le\boldsymbol{g} && \forall s\in\mathcal{S}\\
&\boldsymbol{F}\boldsymbol{y}_s\le\boldsymbol{H}\boldsymbol{f}_s && \forall s\in\mathcal{S}\\
&\boldsymbol{x}_s=[x_{s,1},\cdots,x_{s,M}]\in\{0,1\}^M && \forall s\in\mathcal{S}
\end{aligned} \tag{7}$$

In (7), $\lambda\left\|\boldsymbol{H}\right\|_1$ is a norm-1 regulation term for preventing $\boldsymbol{H}$ from overfitting, where λ is a hyperparameter. The regulation term $\lambda\left\|\boldsymbol{H}\right\|_1$ can be equivalently converted into a linear formulation as shown in Appendix C, leaving (7) as a MILP problem. Solving (7) is essentially the training process, which derives the trained cost-oriented perdition model $\boldsymbol{H}^\star$.

From the perspective of machine learning, the ERM (7) can be regarded as a linear regression problem, which searches for the best linear map $\boldsymbol{H}^\star$ from the feature data $\boldsymbol{f}_s$ to the cost-oriented RES power predictions for NCUC.

2) Cost-Oriented ERM Problem Solving: The ERM (7) may be intractable if solved directly, due to the coupling of $\boldsymbol{H}$ over all training scenarios. That is, the $|\mathcal{S}|$ training scenarios in set $\mathcal{S}$ share the same $\boldsymbol{H}$ via constraint $\boldsymbol{F}\boldsymbol{y}_s\le\boldsymbol{H}\boldsymbol{f}_s$, which makes the multi-scenario coupled ERM intractable. As a result, the LR-based decomposition [25] is applied to effectively solve (7), training a cost-oriented RES prediction model $\boldsymbol{H}^\star$ for NCUC.

LR decomposition: The LR decomposition can be applied to dualize certain coupling constraints and decompose the original problem into $|\mathcal{S}|$ disjoint subproblems that can be solved in parallel [25]. Following this, the original ERM (7) is decomposed into $|\mathcal{S}|$ subproblems as in (8), in which $\mu_{s,e}$ is the Lagrangian multiplier for element e of $\boldsymbol{H}_s$. The detailed decomposition process is expounded in Appendix D.

$$\begin{aligned}
&\min_{\boldsymbol{H}_s,\boldsymbol{x}_s,\boldsymbol{y}_s}\left\{\begin{array}{l}\frac{1}{|\mathcal{S}|}[|(\boldsymbol{c}^\top\boldsymbol{x}_s+\boldsymbol{d}^\top\boldsymbol{y}_s)-z^\star(\tilde{\boldsymbol{w}}_s)|+\lambda\left\|\boldsymbol{H}_s\right\|_1]\\ +\sum_{e\in\mathcal{E}}(\mu_{s,e}H_{s,e})-\sum_{s'\in\mathcal{S}}\sum_{e\in\mathcal{E}}\frac{\mu_{s',e}H_{s,e}}{|\mathcal{S}|}\end{array}\right\}\\
s.t.\ &\boldsymbol{A}\boldsymbol{x}_s+\boldsymbol{B}\boldsymbol{y}_s\le\boldsymbol{g}\\
&\boldsymbol{F}\boldsymbol{y}_s\le\boldsymbol{H}_s\boldsymbol{f}_s\\
&\boldsymbol{x}_s=[x_{s,1},\cdots,x_{s,M}]\in\{0,1\}^M
\end{aligned} \tag{8}$$

Each subproblem corresponds to a training scenario s. To this end, instead of solving (7) directly, we can parallelly tackle $|\mathcal{S}|$ single-scenario problems as in (8) with reduced computational complexity.

Subgradient-based iterative algorithm: To approach the global optimum of the original ERM (7), proper values of $\mu_{s,e}$ shall be identified. Thus, a subgradient algorithm is adopted to update $\mu_{s,e}$ iteratively via (9), until the optimality gap is small enough. In (9), α_c is a scalar step size determined by (10), where ω is a constant scalar chosen between 0 and 2 [25].

Algorithm 1: Subgradient-based iterative algorithm for solving (7)

Input: The maximum number of iterations c^{max}, the optimality gap threshold $\epsilon^\star$, the iteration counter c, Lagrangian multipliers $\mu_{s,e,c}$, and the best upper bound $UB^\star = +\infty$.
Output: The trained prediction model $\boldsymbol{H}^\star$.
begin
  while $c \le c^{max}$ **do**
    Solve $|\mathcal{S}|$ subproblems (8) to get a lower bound LB_c of (7) (i.e., summation of objective values of the $|\mathcal{S}|$ subproblems), binary decisions $\boldsymbol{x}^\star_{c,s}$, and $H_{s,e,c}$;
    Solve the original problem (7) with $\boldsymbol{x}_s = \boldsymbol{x}^\star_{c,s}$ $\forall s \in \mathcal{S}$ to get an upper bound UB_c and the prediction model $\boldsymbol{H}_c$;
    if $UB_c \le UB^\star$ **then**
      Update the best upper bound as $UB^\star = UB_c$;
    else if $UB_c > UB^\star$ **then**
      Keep the current best upper bound $UB^\star$;
    end
    Compute the optimality gap $\epsilon_c = 100\% \times (UB^\star - LB_c)/UB^\star$;
    if $\epsilon_c \le \epsilon^\star$ **then**
      $\boldsymbol{H}^\star = \boldsymbol{H}_c$;
      break
    else if $\epsilon_c > \epsilon^\star$ **then**
      Update $\mu_{s,e,c+1}$ via (9), and set $c = c + 1$;
    end
  end
end

$$\mu_{s,e,c+1} = \mu_{s,e,c} + \alpha_c \left[H_{s,e,c} - \frac{\sum_{s' \in \mathcal{S}} H_{s',e,c}}{|\mathcal{S}|} \right] \tag{9}$$

$$\alpha_c = \frac{\omega(UB^\star - LB_c)}{\sum_{e \in \mathcal{E}} [H_{s,e,c} - (\sum_{s' \in \mathcal{S}} H_{s',e,c})/|\mathcal{S}|]^2} \tag{10}$$

The detailed procedure of the subgradient-based method is described in Algorithm 1.

C. Closed-Loop Predict-and-Optimize Module

The feature-driven NCUC prescription model (11) is formulated by embedding the trained cost-oriented RES power prediction model $\boldsymbol{H}^\star$ into the original NCUC model (1).

$$\begin{aligned} z(\boldsymbol{f}) = \ & \min_{\boldsymbol{x},\boldsymbol{y}} \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y} \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \le \boldsymbol{g} \\ & \boldsymbol{F}\boldsymbol{y} \le \boldsymbol{H}^\star \boldsymbol{f} \\ & \boldsymbol{x} = [x_1, \cdots, x_M] \in \{0,1\}^M \end{aligned} \tag{11}$$

To gain insights into the feature-driven NCUC prescription model (11), it is worthwhile to compare (11) to the original NCUC model (1):

- In the objective $z(\cdot)$, $\hat{\boldsymbol{w}}$ is replaced by $\boldsymbol{f}$, thus the prescription model (11) is driven by the feature data $\boldsymbol{f}$, i.e., (11) is a *feature-driven model*.
- In the RES power constraints, the RES prediction vector $\hat{\boldsymbol{w}}$ is replaced by the cost-oriented prediction vector $\boldsymbol{H}^\star \boldsymbol{f}$, in which $\boldsymbol{H}^\star$ is trained with the feedback of NCUC structure and induced NCUC cost. Thus, (11) realizes the closed-loop idea by using the cost-oriented prediction $\boldsymbol{H}^\star \boldsymbol{f}$ tailored for NCUC.
- Recall that the ERM (7) is a linear regression problem, the cost-oriented prediction $\boldsymbol{H}^\star \boldsymbol{f}$ is essentially a weighted sum of the features $\boldsymbol{f}$.
- Solving (11) can simultaneously predict the RES power $\boldsymbol{H}^\star \boldsymbol{f}$ and prescribe the corresponding optimal NCUC decisions $\boldsymbol{x}^\star$ and $\boldsymbol{y}^\star$, thus (11) is a NCUC prescription model that can *perform predict-and-optimize instead of predict-then-optimize.*

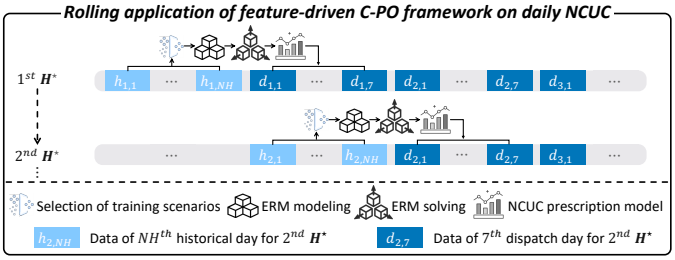


Fig. 5. A rolling-based C-PO implementation for solving daily NCUC.

D. Rolling-based C-PO Implementation for Daily NCUC

A rolling-based C-PO implementation for routinely solving the daily NCUC problems is shown in Fig. 5. Specifically, to balance computational burden and solution quality of the NCUC prescription model (11), $\boldsymbol{H}^\star$ is retrained every 7 days to update (11) on a weekly basis.

Take the 1st $\boldsymbol{H}^\star$ as an example, this $\boldsymbol{H}^\star$ corresponds to 7 upcoming dispatch days $d_{1,1}, \cdots, d_{1,7}$ and NH latest historical days $h_{1,1}, \cdots, h_{1,NH}$. Before $d_{1,1}$, NT training scenarios are selected among days $h_{1,1}, \cdots, h_{1,NH}$ to form the ERM problem (7). After solving (7) to obtain the 1st trained $\boldsymbol{H}^\star$, the 1st NCUC prescription model (11) can be built to routinely solve the NCUC tasks for the week $d_{1,1}, \cdots, d_{1,7}$. As time goes by, the above process is repeated for training the 2nd $\boldsymbol{H}^\star$ and forming the 2nd NCUC prescription model, which corresponds to the daily NCUC tasks for the 7 days $d_{2,1}, \cdots, d_{2,7}$. In case studies, the impact of different settings on NH will be evaluated; NT is uniformly set as 2 for obtaining a proper size of the ERM (7).

IV. Case Studies

In this section, an IEEE RTS 24-bus system and an ISO-scale 5655-bus system [26] are employed to evaluate the NCUC economic improvements of C-PO and the effectiveness of the LR-based decomposition in solving the ERM problem for C-PO. Practical data from the Belgian ISO [22] are properly scaled to build the RES, load, and feature data for case studies. The dataset used in the case studies can be referred to from [29], which includes: (*i*) the detailed configurations for both the IEEE RTS 24-bus system and the 5655-bus system; (*ii*) chronological RES and load data (both predictions and realizations) for the Belgian system ranging from 01/01/2018 to 12/31/2020; and (*iii*) all the feature data. All the cases are solved via Gurobi 9.1 on a 2.4 GHz PC.

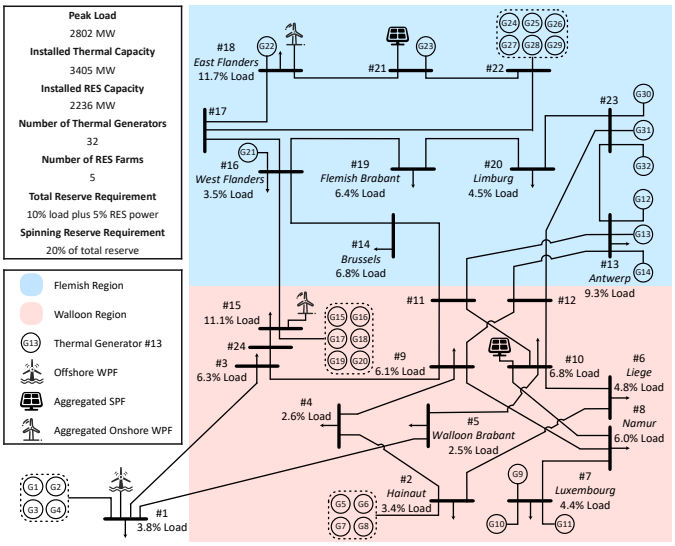


Fig. 6. The modified IEEE RTS 24-bus system.

A. Feature Data

The Belgian ISO operates a 2-region transmission grid including 11 provinces (load bus), 11 solar power farms (SPF), 4 onshore wind power farms (WPF), and 1 offshore WPF. Based on predictions of individual SPFs and WPFs, the aggregation predictions of SPFs and WPFs (i.e., summation of predictions from individual SPF/WPF farms) in each region are also calculated. Thus, we have 22 feature types including hour label, predictions of Belgian load, and raw predictions (i.e., the accuracy-oriented predictions) for 15 individual farms and 4 regional farm aggregations, which are readily available in day-ahead stage. The ISOs typically execute NCUC by solving (1) with the raw RES predictions, which is known as the O-PO.

Within the dataset [29], data in 2018 and 2019 are utilized for feature selection, and data in 2020 are employed for evaluating the C-PO framework. The feature selection step identifies the raw RES prediction as the most relevant feature type $\boldsymbol{f}$. With this, the prediction model $\boldsymbol{H}^\star$ describes the unitless weights that adjust the raw predictions $\boldsymbol{f}$ of individual RES farms to derive the cost-oriented RES power predictions $\boldsymbol{H}^\star \boldsymbol{f}$.

B. Tests on the IEEE RTS 24-Bus System

The IEEE RTS 24-bus system originally including 32 thermal generators is modified to simulate the Belgian transmission network, as shown in Fig. 6. Specifically, 15 RES farms are connected to the system as 4 aggregated farms at buses #10, #15, #18, and #21, and one offshore WPF is connected to the system at bus #1. The actual load and RES data are respectively scaled down to 22% and 39%. The O-PO and C-PO are compared via 366 daily NCUC runs in 2020. The penalty cost of load shedding C^{ls} is set as \$$10^5$/MWh. Considering the 24-bus system is a medium-size system, for the fair comparison with O-PO, the ERM problem (7) is directly solved to generate the prediction model $\boldsymbol{H}^\star$ without using the LR-based decomposition.

TABLE III
MODELS TO COMPARE

Models		Parameter setting		
		λ	NH	NT
Practical models to ISO	O-PO	-	-	-
	C-PO-1	10^5	2	2
	C-PO-2	10^5	3	2
	C-PO-3	10^5	4	2
	C-PO-4	10^5	5	2
	C-PO-5	10^5	6	2
	C-PO-6	10^5	7	2
Ideal models	C-PO-I	10^5	-	1
	P-PO-I	-	-	-

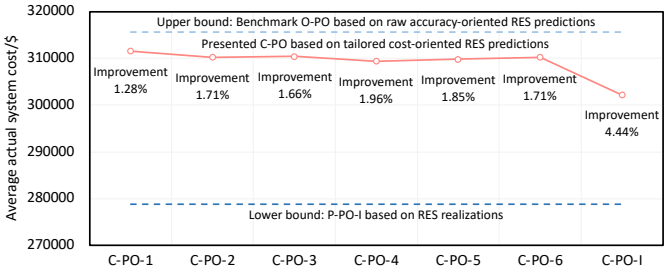


Fig. 7. Comparison of actual system costs over 01/01/2020 to 12/31/2020.

1) Overall Economics of the Entire Year: Nine models listed in Table III are compared. Specifically, in addition to the ISO's current practice O-PO and six practical C-PO models with different NH values, two ideal models C-PO-I and P-PO-I (perfect predict-then-optimize) are also presented for comparison. C-PO-I utilizes the data (i.e., anticipated system costs and features) of the actual dispatch days, instead of historical days, to train the prediction model $\boldsymbol{H}^\star$ and update the NCUC prescription model (11), which intuitively could lead to a more effective training. On the other hand, P-PO-I directly executes NCUC using the actual RES power realization, which shall achieve the best NCUC economics. It is noteworthy that these two ideal models are merely utilized to compared with the O-PO and C-PO models in terms of NCUC economics, but are not practically applicable due to the data causality issue.

Fig. 7 compares the actual system costs of all the models over the entire year, in which O-PO and P-PO-I respectively provide upper and lower bounds. Obviously, the economic gap between P-PO-I and O-PO is noteworthy. In addition, all the C-PO models can partly bridge the gap, among them the ideal C-PO-I performs the best with a 4.44% improvement over O-PO, and the best practical C-PO model is C-PO-4 with an 1.96% improvement over O-PO. The outperformances of C-POs over O-PO indicate three facts: (*i*) the proposed C-PO framework enables noticeable improvements in the NCUC economics over O-PO; (*ii*) the improvements can be further boosted when the most representative data of the dispatch days are available; and (*iii*) the data-processing module renders the practical C-PO models to be acceptably sensitive to parameter NH.

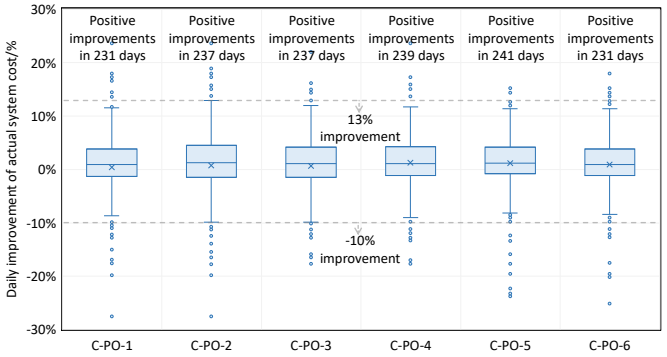


Fig. 8. Box plot of economic improvements over 01/01/2020 to 12/31/2020.

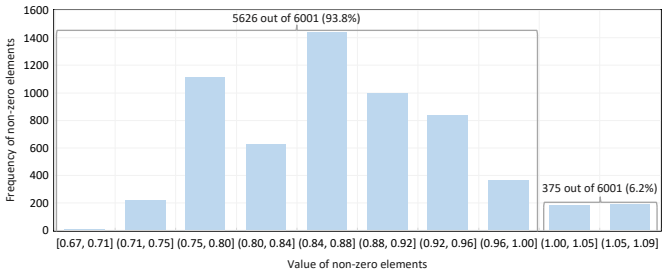


Fig. 9. Histogram on values of non-zero elements in $\boldsymbol{H}^\star$.

Fig. 8 further depicts the box plot for the daily economic improvements. According to this figure, the practical C-PO models can result in positive improvements in the NCUC economics for 231 to 241 dispatch days throughout the year. Moreover, although some outliers exist, the economic impacts mainly fall in the range of [-10%, 13%], indicating that the economic improvements caused by C-POs are balanced and stable, i.e., avoiding excessive improvements for some days accompanies with significant deteriorations for others.

Based on the analyses of Figs. 7 and 8, it is concluded that the proposed C-PO can steadily improve the NCUC economics compared to the O-PO. The highest average daily improvement of 1.96% is achieved by the practical C-PO-4 model.

2) Comparison of C-PO-4 and O-PO: To further analyze the economic improvements, results of C-PO-4 and O-PO over the entire year are compared from perspectives of costs and prediction errors. In addition, two selective dispatch days are also discussed in detail.

Cost comparison of C-PO-4 and O-PO: Table IV compares the average costs of C-PO-4 and O-PO over the entire year. Both the NCUC SUSD cost and anticipated system cost of C-PO-4 are higher than O-PO ($22,052 vs $21,919 and $285,530 vs $268,090), implying that C-PO-4 prefers to a more conservative NCUC strategy by committing more generators. That is, compared to O-PO, C-PO-4 uses more generators to support the loads, and some ON-state generators may not be utilized fully or economically, leading to a higher generation cost ($273,530 vs $269,760). However, C-PO-4 outperforms O-PO in terms of the redispatch SUSD cost ($13,828 vs $23,911) and the actual system cost ($309,410 vs $315,590). This is because the conservative NCUC strategies of C-PO-4, with enough flexibilities from the ON-state generators for dispatch adjustment, could avoid the extensive utilization of non-spinning reserves from those generators NOT committed in NCUC, enabling the improvements in the NCUC economics.

Prediction error comparison of C-PO-4 and O-PO: Table V compares the prediction errors of C-PO-4 and O-PO. In addition to MAE and RMSE, the mean over-prediction percentage error (MOPE) and the mean under-prediction percentage error (MUPE) are also reported, which are calculated as (12) and (13).

$$\text{MOPE}_t = \frac{\max\{0, \hat{w}_t - \tilde{w}_t\}}{\tilde{w}_t} \times 100\% \tag{12}$$

$$\text{MUPE}_t = \frac{\max\{0, \tilde{w}_t - \hat{w}_t\}}{\tilde{w}_t} \times 100\% \tag{13}$$

The tailored cost-oriented prediction $\boldsymbol{H}^\star \boldsymbol{f}$ in (11) is essentially determined by the trained prediction model $\boldsymbol{H}^\star$. In this system, $\boldsymbol{H}^\star$ is a sparse matrix of 120 rows and 520 columns. Fig. 9 shows the histogram of non-zero elements of all the 53 trained $\boldsymbol{H}^\star$ over the entire year, which include 6,001 non-zero elements in total. Fig. 9 shows that 93.8% of the 6,001 non-zero elements are no larger than 1. Recalling that as the identified feature types $\boldsymbol{f}$ are the raw predictions, the above results implicate that C-PO-4 tends to lower down the raw predictions via $\boldsymbol{H}^\star$. This is because C-PO-4 learns from the feedback information (i.e., induced NCUC cost, constraints, and objective) of the NCUC optimization that properly under-predicted RES power could improve NCUC economics. This analysis also explains why the NCUC SUSD and anticipated system costs of C-PO-4 in Table IV are higher.

Analysis on the two selective dispatch days: C-PO-4 and

TABLE IV
AVERAGE COSTS OVER 01/01/2020 TO 12/31/2020

Method	NCUC		Redispatch		Actual system cost/$
	SUSD cost/$	Anticipated system cost/$	SUSD cost/$	Generation cost/$	
C-PO-4	22,052	285,530	13,828	273,530	309,410
O-PO	21,919	268,090	23,911	269,760	315,590

TABLE V
PREDICTION ERRORS OVER 01/01/2020 TO 12/31/2020

Method	MAE/MW	RMSE/MW	MOPE	MUPE
C-PO-4	121.2	148.1	22.6%	12.0%
O-PO	104.3	130.0	33.9%	5.5%

TABLE VI
COMPARISONS OF 06/22/2020 AND 12/29/2020

Date	Method	SUSD cost/$		Generation cost/$	Actual system cost/$	MAE /MW
		NCUC	Redispatch			
06/22	C-PO-4	21,000	10,500	269,270	300,770	42.7
	O-PO	22,500	28,500	271,080	322,080	66.1
12/29	C-PO-4	27,000	25,500	438,810	491,310	158.9
	O-PO	27,000	30,000	437,380	494,380	137.8

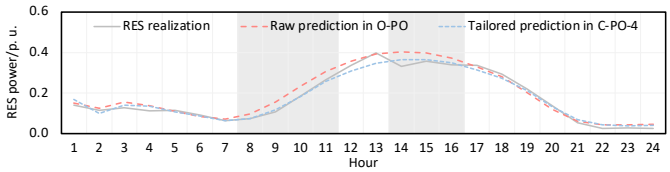


Fig. 10. Predictions and actual realization of RES power on 06/22/2020.

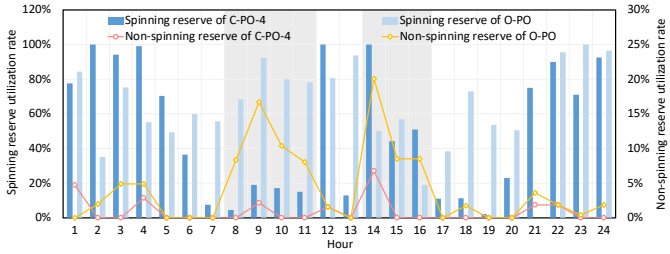


Fig. 11. Reserve deployment in the redispatch problem on 06/22/2020.

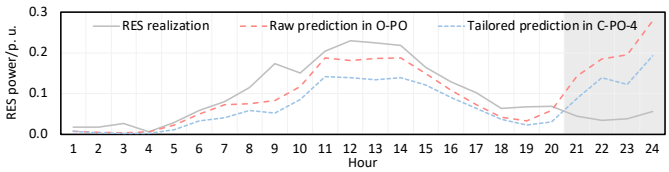


Fig. 12. Predictions and actual realization of RES power on 12/29/2020.

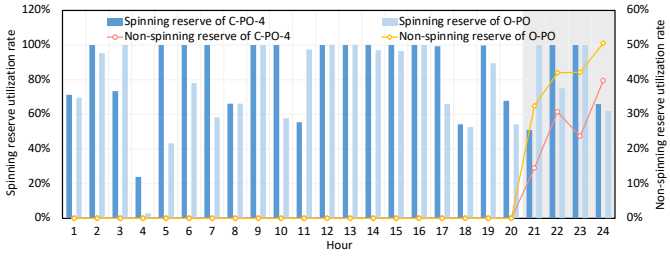


Fig. 13. Reserve deployment in the redispatch problem on 12/29/2020.

O-PO are further compared on two selective dispatch days 06/22/2020 and 12/29/2020. Specifically, on 06/22/2020, C-PO-4 achieves *better NCUC economics with a better prediction accuracy* (from perspective of MAE); while on 12/29/2020, C-PO-4 reaches *better NCUC economics with a worse prediction accuracy*. The results are listed in Table VI.

Performance on dispatch day 06/22/2020: For the dispatch day 06/22/2020, Fig. 10 depicts the predictions and actual realization of RES power in per unit values over the system peak load, and Fig. 11 sketches the reserve deployment in the redispatch problem.

Fig. 10 shows that the tailored predictions in C-PO-4 achieve higher accuracy than the raw predictions in O-PO, especially in the 7 gray-shaded hours. Table VI further shows that compared to O-PO, C-PO-4 with the tailored predictions derives a NCUC strategy with lower SUSD cost ($21,000 vs $22,500). In addition, the NCUC strategy in C-PO-4 leads to lower SUSD cost in the redispatch problem ($10,500 vs $28,500) and generation cost ($269,270 vs $271,080). This is mainly because C-PO-4 deploys less reserve in the redispatch problem than O-PO. Indeed, Fig. 11 illustrates that O-PO utilizes more spinning reserves in 16 hours (hours 1, 6-11, 13, 15, 17-20, and 22-24), and more non-spinning reserves in 13 hours (hours 2-4, 8-11, 14-16, 18, 21, and 24), especially the 7 gray-shaded hours.

As a result, C-PO-4 with the tailored cost-oriented RES predictions achieves higher prediction accuracy and lower actual system cost ($300,770 vs $322,080).

Performance on dispatch day 12/29/2020: For the dispatch day 12/29/2020, Fig. 12 depicts the predictions and actual realization of RES power in per unit values over the system peak load, and Fig. 13 sketches the reserve deployment in the redispatch problem.

As shown in Fig. 12, the tailored predictions are consistently lower than the raw predictions in all hours. Interestingly, Table VI indicates that C-PO-4 with the under-predicted RES power determines a NCUC strategy with the same SUSD cost ($27,000) as that of O-PO. Moreover, compared to O-PO, C-PO-4 has a higher generation cost ($438,810 vs $437,380), but deploys less non-spinning reserve in the redispatch problem ($25,500 vs $30,000). This is because, the actual RES realizations fall down in hours 21-24 (the gray shaded area in Fig. 12), and both O-PO and C-PO-4 over-predict the RES power in those hours; however, O-PO presents a worse over-prediction, and has to deploy more non-spinning reserves in the redispatch problem, as shown in Fig. 13.

Consequently, C-PO-4 with the tailored predictions reaches a lower actual system cost, although the RES forecast accuracy is slightly worse.

Based on the above analyses on the 24-bus system, it can be concluded that: (*i*) the statistically more accurate RES power predictions in O-PO does not necessarily lead to a lower actual system cost; (*ii*) the presented C-PO outperforms the benchmark O-PO in terms of the NCUC economics; and (*iii*) the outperformance stems from the utilization of the tailored cost-oriented RES power predictions for NCUC, which is enabled by the proposed closed-loop structure in C-PO.

C. Tests on the ISO-Scale 5655-Bus System

To show the potential of C-PO in real-world systems, C-PO is further evaluated on an ISO-scale 5655-bus system with 6,630 lines and 461 generators [26]. The 5 aggregated RES farms are connected at buses #5051, #5446, #2817, #1607, and #1043. Parameters λ, NH, and NT are set as 10^5, 7, and 2. Load and RES data are scaled up to 360% and 400%. The penalty cost of load shedding C^{ls} is set as $105/MWh. The MIP gap of Gurobi is set as 1%, and the optimality gap $\epsilon^\star$ in C-PO-LR is set as 1%.

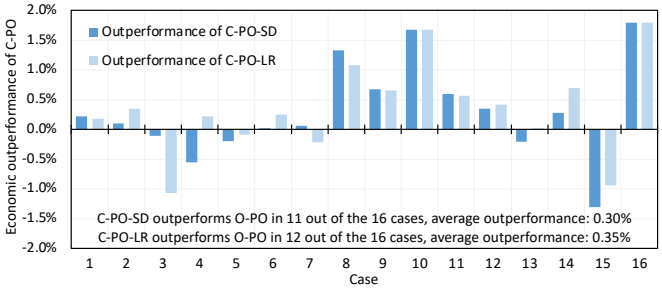


Fig. 14. Actual system cost improvements of C-POs over O-PO.

TABLE VII
SIXTEEN SELECTIVE DAYS FOR THE 5655-BUS SYSTEM STUDY

Case	Date	Characteristics of the dispatch day
1	01/21	The day in January 2020 with the highest net peak load
2	02/06	The day in February 2020 with the highest net peak load.
3	03/04	The day in March 2020 with the highest net peak load.
4	04/03	The day in April 2020 with the highest net peak load.
5	05/04	The day in May 2020 with the highest net peak load.
6	06/17	The day in June 2020 with the highest net peak load.
7	07/08	The day in July 2020 with the highest net peak load.
8	08/20	The day in August 2020 with the highest net peak load.
9	09/29	The day in September 2020 with the highest net peak load.
10	10/16	The day in October 2020 with the highest net peak load.
11	11/26	The day in November 2020 with the highest net peak load.
12	12/08	The day in December 2020 with the highest net peak load.
13	01/24	The day in 2020 with the highest RES power availability, i.e., the largest value of $\sum_{t\in\mathcal{T}}[\sum_{j\in\mathcal{J}}\tilde{w}_{jt}/\sum_{q\in\mathcal{Q}}L_{qt}]$.
14	05/11	The day in 2020 with the lowest RES power availability, i.e., the smallest value of $\sum_{t\in\mathcal{T}}[\sum_{j\in\mathcal{J}}\tilde{w}_{jt}/\sum_{q\in\mathcal{Q}}L_{qt}]$.
15	07/21	The day in 2020 that the RES power vector $[\sum_{j\in\mathcal{J}}\tilde{w}_{j,1},\cdots,\sum_{j\in\mathcal{J}}\tilde{w}_{j,\|\mathcal{T}\|}]$ has the largest range (i.e., $\max\{\sum_{j\in\mathcal{J}}\tilde{w}_{j,t}\}-\min\{\sum_{j\in\mathcal{J}}\tilde{w}_{j,t}\}$).
16	12/22	The day in 2020 that the RES power vector $[\sum_{j\in\mathcal{J}}\tilde{w}_{j,1},\cdots,\sum_{j\in\mathcal{J}}\tilde{w}_{j,\|\mathcal{T}\|}]$ has the largest standard deviation.

TABLE VIII
ACTUAL SYSTEM COSTS ON THE 5655-BUS SYSTEM

Case	Actual system cost/$\$10^6$		
	O-PO	C-PO-SD	C-PO-LR
1	8.78	8.76	8.77
2	8.20	8.19	8.17
3	7.73	7.74	7.81
4	5.05	5.08	5.04
5	5.09	5.10	5.09
6	5.92	5.92	5.91
7	6.53	6.53	6.55
8	6.32	6.23	6.25
9	6.60	6.55	6.55
10	6.52	6.41	6.41
11	7.74	7.70	7.70
12	8.46	8.43	8.42
13	9.02	9.04	9.02
14	3.06	3.05	3.04
15	4.54	4.60	4.58
16	6.92	6.77	6.79

As listed in Table VII, sixteen selective days in 2020 are studied to compare the three models including the traditional O-PO, C-PO-SD (i.e., the ERM (7) is solved directly), and C-PO-LR (i.e., the ERM (7) is solved via the LR decomposition).

Table VIII lists the actual system costs of the sixteen cases from the three models, and Fig. 14 further visualizes actual system cost improvements of C-PO-SD and C-PO-LD over O-PO. As shown in Table VIII, actual system costs of all the three models achieve the same magnitude of 10^6, indicating that the proposed C-PO approaches will not compromise the current NCUC practice with unexpected and large distortions.

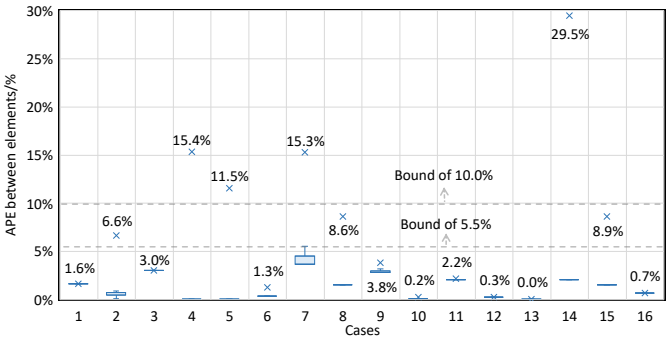


Fig. 15. Box plot of APE of non-zero $\boldsymbol{H}^\star$ values in the sixteen cases.

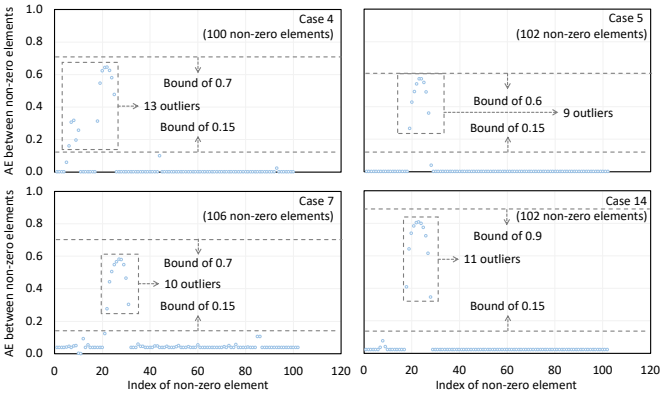


Fig. 16. Scatter plots of AE of non-zero $\boldsymbol{H}^\star$ values in Cases 4, 5, 7, and 14.

Fig. 14 further shows that both C-PO-SD and C-PO-LD outperform O-PO in about 2/3 out of the sixteen cases, in terms of the actual system cost. Indeed, the average NCUC economics improvement of the C-PO-LR over O-PO is about 0.35%. Considering the magnitude of actual daily operation costs for practical systems, a 0.35% improvement could be quite remarkable. Moreover, it shall be pointed out that C-PO-LR performs close to C-PO-SD in all cases, illustrating effectiveness of the LR decomposition approach.

As the economic performance highly depends on the trained prediction models $\boldsymbol{H}^\star$, the sparsity of $\boldsymbol{H}^\star$ (i.e., the total number and positions of non-zero elements in the matrix) and values of non-zero elements in $\boldsymbol{H}^\star$ from C-PO-SD and C-PO-LR are further compared. In terms of the sparsity, our experimental results indicate that the $\boldsymbol{H}^\star$ matrices trained by C-PO-LR are exactly the same as those trained by C-PO-SD in all cases. In terms of the values of non-zero elements, box plots of the absolute percentage error (APE) between two corresponding non-zero elements, defined as in (14), are shown in Fig. 15. According to Fig. 15, the APE intervals are fairly narrow, and most points are below the 5.5% bound. Regarding the average value, only Cases 4, 5, 7, and 14 are higher than 10%, while the small APEs in the other twelve cases indicate that C-PO-LR indeed can provide $\boldsymbol{H}^\star$ of similar quality as C-PO-SD in majority cases.

$$APE_e = \frac{|H_e^{\star,C-PO-SD} - H_e^{\star,C-PO-LR}|}{H_e^{\star,C-PO-SD}} \qquad (14)$$

TABLE IX
COMPUTATION RESULTS OF THE 5655-BUS SYSTEM

Case	Training time/s		Optimality gap	
	C-PO-SD	C-PO-LR	C-PO-SD	C-PO-LR
1	1273.6	593.2	0.32%	0.62% (4 Iterations)
2	1111.7	1029.2	0.59%	0.89% (3 Iterations)
3	1655.8	927.5	0.51%	0.69% (3 Iterations)
4	828.6	619.2	0.86%	0.64% (4 Iterations)
5	685.9	512.3	0.81%	0.69% (4 Iterations)
6	3686.1	1364.1	0.93%	0.77% (4 Iterations)
7	1581.5	1312.6	0.33%	0.35% (4 Iterations)
8	1803.8	1215.9	0.74%	0.99% (4 Iterations)
9	1266.1	1211.8	0.67%	0.17% (4 Iterations)
10	1140.8	1086.3	0.36%	0.73% (4 Iterations)
11	2632.4	1089.1	0.49%	0.82% (3 Iterations)
12	1462.7	1321.3	0.31%	0.76% (4 Iterations)
13	1436.4	834.7	0.72%	0.74% (4 Iterations)
14	1138.9	714.8	0.98%	0.89% (4 Iterations)
15	1810.2	767.6	0.87%	0.99% (4 Iterations)
16	2146.1	290.8	0.92%	0.88% (1 Iteration)

Cases 4, 5, 7, and 14 are further analyzed via the scatter plots of the absolute error (AE), as calculated in (15). The results are presented in Fig. 16. As shown in Fig. 16, in these four cases, the outliers merely account for about 1/10 AEs. The reason is that these elements in $\boldsymbol{H}^{\star}$ correspond to small-valued constants in $\boldsymbol{f}$. Thus, the resulting terms $H_e f_e$ in the ERM problem (7) are relatively small and have negligible effect on the final solution quality. Considering all the outliers are smaller than 0.9 and the remaining AEs are smaller than 0.15, it can be concluded that C-PO-LR is able to train a similar $\boldsymbol{H}^{\star}$ and derive solutions of similar quality as C-PO-SD.

$$AE_e = |H_e^{\star,C-PO-SD} - H_e^{\star,C-PO-LR}| \tag{15}$$

Table IX further compares computational performance of the sixteen cases. According to the table, both C-PO models achieve the desired optimal gap within reasonable time in all sixteen cases. Moreover, C-PO-LR significantly outperforms C-PO-SD in terms of the total computational time in all cases. This is because: (*i*) solving $|\mathcal{S}|$ subproblems (8) is computationally cheaper than solving the original ERM problem (7) directly; and (*ii*) solving $|\mathcal{S}|$ subproblems (8) in parallel significantly boosts the computation efficiency. Thus, C-PO-LR is able to provide solutions of similar quality as C-PO-SD at a much shorter time.

In summary, numerical studies on the ISO-scale 5655-bus system show that the LR-based decomposition enables the feature-driven C-PO framework to be practically applicable for the real-world NCUC tasks.

V. CONCLUSIONS

The NCUC problem, as one of the most important daily tasks for ISOs, is typically implemented in an O-PO framework. In observing that the statistically more accurate RES power predictions in O-PO may not necessarily lead to the higher NCUC economics, this paper presents a feature-driven C-PO framework for improving the NCUC economics. By effectively leveraging the readily available feature data (e.g., raw accuracy-oriented predictions), the NCUC model structure (i.e., constraints and objective), and the induced NCUC cost, C-PO can generate the tailored cost-oriented RES power predictions for the NCUC task, leading to lower actual system costs. Furthermore, an LR-based decomposition method is adopted to enable the C-PO to be practically applicable in real-world power systems. Based on the case studies on an IEEE RTS 24-bus system and an ISO-scale 5655-bus system, the following conclusions are observed:

- The proposed feature-driven C-PO can generate cost-oriented RES predictions for NCUC, although the forecast accuracy could be slightly lower in certain cases. This is because C-PO evaluates prediction quality via the difference between the NCUC costs induced by RES predictions and actual RES realizations, while considering the structure of NCUC models.
- The cost-oriented predictions, which are adaptively generated by $\boldsymbol{H}^{\star}$, enable noticeable economic improvement in C-PO compared to O-PO. According to the 24-bus studies with real-world data, the average improvement in the daily actual system cost over an entire year could be as high as 1.96%.
- Based on the 5655-bus system, the presented LR-based decomposition approach can efficiently train the cost-oriented prediction model of C-PO. The computational time indicates that the presented C-PO is applicable to the ISO practice.

Regarding the future works, several directions deserve further exploration: (*i*) how does the C-PO framework perform for more complicated optimization tasks such as the transmission-distribution coordination? and (*ii*) whether the cost-oriented idea can be extended to emission-oriented and reliability-oriented studies?

APPENDIX

A. NCUC Modeling

The NCUC problem is modeled as a MILP model (16)-(34). The objective (16) is to minimize the total operation cost including startup, shutdown, generation, and load shedding costs, subject to prevailing generator and system constraints. The generator constraints include segment-based generation cost representation (17), ramping limits (18) and (19), minimum on and off time requirements (20) and (21), generation limits (22)-(25), startup-shutdown-commitment status logic (26), spinning (27) and non-spinning reserve (28) capacities, integrality requirement (29), and RES power limitation (30). Regarding the system constraints, (31) indicates the power balance, (32) and (33) describe the system reserve requirements, and (34) represents the DC power flow-based transmission network constraints. Note that ISOs usually include certain slack variables (like LS_{qt} used in (31) and (34)) to ensure that a feasible solution to the market clearing always exists. However, load shedding is not observed in our case studies with a sufficiently large C^{ls}.

$$\min_{\boldsymbol{x},\boldsymbol{y}} \sum_{t\in\mathcal{T}} \left[\sum_{i\in\mathcal{I}} (C_i^{su} U_{it} + C_i^{sd} D_{it} + C_{it}) + \sum_{q\in\mathcal{Q}} C^{ls} LS_{qt} \right]$$
$$\boldsymbol{x} = \{\boldsymbol{I}, \boldsymbol{U}, \boldsymbol{D}\};\ \boldsymbol{y} = \{\boldsymbol{P}, \boldsymbol{LS}, \boldsymbol{O}^{sr}, \boldsymbol{O}^{nr}, \boldsymbol{w}\} \tag{16}$$

Generator Constraints:

$$C_{it} = N_i I_{it} + \sum_{k\in\mathcal{K}} C_{ik}^{sg} P_{ikt}^{sg} \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{17}$$

$$P_{it} - P_{i,t-1} \le \bar{P}_i(1 - I_{it}) + R_i^{\uparrow} I_{i,t-1} + R_i^{su}(I_{it} - I_{i,t-1}) \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{18}$$

$$P_{i,t-1} - P_{it} \le \bar{P}_i(1 - I_{i,t-1}) + R_i^{\downarrow} I_{it} + R_i^{sd}(I_{i,t-1} - I_{it}) \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{19}$$

$$\sum_{t'=t-TU_i+1}^{t} U_{it'} \le I_{it} \quad \forall t = TU_i, \cdots, |\mathcal{T}|, i \in \mathcal{I} \tag{20}$$

$$\sum_{t'=t-TD_i+1}^{t} D_{it'} \le 1 - I_{it} \quad \forall t = TD_i, \cdots, |\mathcal{T}|, i \in \mathcal{I} \tag{21}$$

$$P_{it} = \sum_{k\in\mathcal{K}} P_{ikt}^{sg} \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{22}$$

$$0 \le P_{ikt}^{sg} \le \bar{P}_{ik}^{sg} I_{it} \quad \forall t \in \mathcal{T}, i \in \mathcal{I}, k \in \mathcal{K} \tag{23}$$

$$P_{it} - O_{it}^{sr} \ge \underline{P}_i I_{it} \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{24}$$

$$P_{it} + O_{it}^{sr} \le \bar{P}_i I_{it} \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{25}$$

$$U_{it} - D_{it} = I_{it} - I_{i,t-1} \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{26}$$

$$0 \le O_{it}^{sr} \le \bar{O}_i^{sr} I_{it} \quad \forall t \in \mathcal{T}, i \in \mathcal{I}, k \in \mathcal{K} \tag{27}$$

$$O_{it}^{nr} = \bar{O}_i^{nr}(1 - I_{it}) \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{28}$$

$$I_{it}, U_{it}, D_{it} \in \{0, 1\} \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{29}$$

$$0 \le w_{jt} \le \hat{w}_{jt} \quad \forall t \in \mathcal{T}, j \in \mathcal{J} \tag{30}$$

System Constraints:

$$\sum_{i\in\mathcal{I}} P_{it} + \sum_{j\in\mathcal{J}} w_{jt} = \sum_{q\in\mathcal{Q}}(L_{qt} - LS_{qt}) \quad \forall t \in \mathcal{T} \tag{31}$$

$$\sum_{i\in\mathcal{I}} O_{it}^{sr} \ge S_t^{sr} \quad \forall t \in \mathcal{T} \tag{32}$$

$$\sum_{i\in\mathcal{I}}(O_{it}^{sr} + O_{it}^{nr}) \ge S_t^{sr} + S_t^{nr} \quad \forall t \in \mathcal{T} \tag{33}$$

$$-B_b \le \mathcal{F}_b(\boldsymbol{P}_t, \boldsymbol{w}_t, \boldsymbol{L}_t - \boldsymbol{LS}_t) \le B_b \quad \forall t \in \mathcal{T}, b \in \mathcal{B} \tag{34}$$

B. Redispatch Modeling

The redispatch problem (35)-(46) with the given solutions $I_{it}^{\star}$, $P_{it}^{\star}$, $O_{it}^{sr\star}$, and $O_{it}^{nr\star}$ from NCUC as well as the actual RES realizations $\tilde{w}_{jt}$ is utilized to evaluate the NCUC economics.

The objective of the redispatch problem (35) is to minimize startup, shutdown, and generation costs of the generators not committed in NCUC but scheduled for providing non-spinning reserves, generation costs of the generators committed in NCUC, as well as the load shedding costs. For the committed generators, (36) ensures their commitment consistency with the NCUC solution and limits their dispatch adjustment within the scheduled spinning reserve. For the non-committed generators, (37) and (38) indicate that they can be re-dispatched within the scheduled non-spinning reserve. For all generators, (39) is the segment-based generation cost; (40) and (41) limit the ramping ranges; (42) and (43) limit the generation ranges. The RES power utilization is limited by its actual realization (44). The power balance and transmission limitation are guaranteed by (45) and (46).

$$\min_{\boldsymbol{x}^{rd}, \boldsymbol{y}^{rd}} \sum_{t\in\mathcal{T}} \left\{ \sum_{i\in\mathcal{I}^{nu}} (C_i^{su} U_{it}^{rd} + C_i^{sd} D_{it}^{rd}) + \sum_{i\in\mathcal{I}} C_{it}^{rd} + \sum_{q\in\mathcal{Q}} C^{ls} LS_{qt}^{rd} \right\}$$

$$\boldsymbol{x}^{rd} = \{\boldsymbol{I}^{rd}, \boldsymbol{U}^{rd}, \boldsymbol{D}^{rd}\};\ \boldsymbol{y}^{rd} = \{\boldsymbol{P}^{rd}, \boldsymbol{LS}^{rd}, \boldsymbol{w}^{rd}\} \tag{35}$$

Generator Constraints:

$$I_{it}^{rd} = I_{it}^{\star};\ |P_{it}^{rd} - P_{it}^{\star}| \le O_{it}^{sr\star} I_{it}^{rd} \quad \forall t \in \mathcal{T}, i \in \mathcal{I}^{cu} \tag{36}$$

$$I_{it}^{rd} \in \{0, 1\};\ \underline{P}_i I_{it}^{rd} \le P_{it}^{rd} \le O_{it}^{nr\star} I_{it}^{rd} \quad \forall t \in \mathcal{T}, i \in \mathcal{I}^{nu} \tag{37}$$

$$U_{it}^{rd} - D_{it}^{rd} = I_{it}^{rd} - I_{i,t-1}^{rd} \quad \forall t \in \mathcal{T}, i \in \mathcal{I}^{nu} \tag{38}$$

$$C_{it}^{rd} = N_i I_{it}^{rd} + \sum_{k\in\mathcal{K}} C_{ik}^{sg} P_{ikt}^{rd,sg} \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{39}$$

$$P_{it}^{rd} - P_{i,t-1}^{rd} \le \bar{P}_i(1 - I_{it}^{rd}) + R_i^{\uparrow} I_{i,t-1}^{rd} + R_i^{su}(I_{it}^{rd} - I_{i,t-1}^{rd}) \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{40}$$

$$P_{i,t-1}^{rd} - P_{it}^{rd} \le \bar{P}_i(1 - I_{i,t-1}^{rd}) + R_i^{\downarrow} I_{it}^{rd} + R_i^{sd}(I_{i,t-1}^{rd} - I_{it}^{rd}) \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{41}$$

$$P_{it}^{rd} = \sum_{k\in\mathcal{K}} P_{ikt}^{rd,sg} \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{42}$$

$$0 \le P_{ikt}^{rd,sg} \le \bar{P}_{ik}^{sg} I_{it}^{rd} \quad \forall t \in \mathcal{T}, i \in \mathcal{I}, k \in \mathcal{K} \tag{43}$$

$$0 \le w_{jt}^{rd} \le \tilde{w}_{jt} \quad \forall t \in \mathcal{T}, j \in \mathcal{J} \tag{44}$$

System Constraints:

$$\sum_{i\in\mathcal{I}} P_{it}^{rd} + \sum_{j\in\mathcal{J}} w_{jt}^{rd} = \sum_{q\in\mathcal{Q}}(L_{qt} - LS_{qt}^{rd}) \quad \forall t \in \mathcal{T} \tag{45}$$

$$-B_b \le \mathcal{F}_b(\boldsymbol{P}_t^{rd}, \boldsymbol{w}_t^{rd}, \boldsymbol{L}_t - \boldsymbol{LS}_t^{rd}) \le B_b \quad \forall t \in \mathcal{T}, b \in \mathcal{B} \tag{46}$$

C. Linear Reformulation of the Norm-1 Term in (7)

To show the reformulation process of norm-1 term $\lambda \|\boldsymbol{H}\|_1$, a matrix $\boldsymbol{H}$ of R rows and C columns is taken as an example, in which rows and columns are indexed by r and c, respectively.

The norm-1 term $\lambda \|\boldsymbol{H}\|_1$ is essentially to sum the absolute values of each column and take the largest answer, as shown in (47).

$$\lambda \|\boldsymbol{H}\|_1 = \max_{c\in\{1,\cdots,C\}} \sum_{r=1}^{R} \lambda |H_{r,c}| \tag{47}$$

First, auxiliary variables $H'_{r,c}$ are introduced to handle the absolute operator as in (48). With (48), (47) can be represented as in (49).

$$H'_{r,c} \ge \lambda H_{r,c};\ H'_{r,c} \ge -\lambda H_{r,c};\ H'_{r,c} \ge 0 \quad \forall c \in \{1, \cdots, C\}, r \in \{1, \cdots, R\} \tag{48}$$

$$\lambda \|\boldsymbol{H}\|_1 = \max_{c\in\{1,\cdots,C\}} \sum_{r=1}^{R} \lambda |H_{r,c}| = \max_{c\in\{1,\cdots,C\}} \sum_{r=1}^{R} H'_{r,c} \tag{49}$$

Considering the objective of the ERM (7) is a minimization function, inequations (50) are utilized to tackle the $\max$ operator of (49) using auxiliary variable H''.

$$H'' \ge \sum_{r=1}^{R} H'_{r,c} \quad \forall c \in \{1, \cdots, C\} \tag{50}$$

Finally, $\lambda \|\boldsymbol{H}\|_1$ in the objective function of the ERM (7) can be equivalently replaced by a single variable H'' plus the linear constraints (48) and (50).

D. Decomposition of the Original ERM Problem

Recalling that the original ERM problem (7) can be represented as in (51).

$$\min_{\substack{\boldsymbol{H}, \\ \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{|\mathcal{S}|}, \\ \boldsymbol{y}_1, \cdots, \boldsymbol{y}_{|\mathcal{S}|}}} \frac{1}{|\mathcal{S}|} \sum_{s\in\mathcal{S}} \left\{ |(\boldsymbol{c}^{\top}\boldsymbol{x}_s + \boldsymbol{d}^{\top}\boldsymbol{y}_s) - z^{\star}(\tilde{\boldsymbol{w}}_s)| \right\} + \lambda \|\boldsymbol{H}\|_1$$

$$s.t. \quad \boldsymbol{A}\boldsymbol{x}_s + \boldsymbol{B}\boldsymbol{y}_s \le \boldsymbol{g} \quad \forall s \in \mathcal{S}$$

$$\boldsymbol{F}\boldsymbol{y}_s \le \boldsymbol{H}\boldsymbol{f}_s \quad \forall s \in \mathcal{S}$$

$$\boldsymbol{x}_s = [x_{s,1}, \cdots, x_{s,M}] \in \{0, 1\}^M \quad \forall s \in \mathcal{S} \tag{51}$$

We first convert (51) into an equivalent form (52).

$$\begin{aligned}
\min_{\substack{\boldsymbol{H}_1,\cdots,\boldsymbol{H}_{|\mathcal{S}|},\\ \boldsymbol{x}_1,\cdots,\boldsymbol{x}_{|\mathcal{S}|},\\ \boldsymbol{y}_1,\cdots,\boldsymbol{y}_{|\mathcal{S}|}}} \; & \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}[|(\boldsymbol{c}^\top\boldsymbol{x}_s+\boldsymbol{d}^\top\boldsymbol{y}_s)-z^\star(\tilde{\boldsymbol{w}}_s)|+\lambda\left\|\boldsymbol{H}_s\right\|_1] \\
s.t. \quad & \boldsymbol{A}\boldsymbol{x}_s+\boldsymbol{B}\boldsymbol{y}_s\le\boldsymbol{g} && \forall s\in\mathcal{S} \\
& \boldsymbol{F}\boldsymbol{y}_s\le\boldsymbol{H}_s\boldsymbol{f}_s && \forall s\in\mathcal{S} \\
& \boldsymbol{x}_s=[x_{s,1},\cdots,x_{s,M}]\in\{0,1\}^M && \forall s\in\mathcal{S} \\
& H_{s,e}=\textstyle\sum_{s'\in\mathcal{S}}H_{s',e}/|\mathcal{S}| && \forall s\in\mathcal{S},e\in\mathcal{E} \quad (52)
\end{aligned}$$

The coupling constraints on $\boldsymbol{H}$ are further dualized into the objective using Lagrangian multipliers $\mu_{s,e}$ to derive the LR problem (53).

$$\begin{aligned}
\min_{\substack{\boldsymbol{H}_1,\cdots,\boldsymbol{H}_{|\mathcal{S}|},\\ \boldsymbol{x}_1,\cdots,\boldsymbol{x}_{|\mathcal{S}|},\\ \boldsymbol{y}_1,\cdots,\boldsymbol{y}_{|\mathcal{S}|}}} \; & \left\{\begin{array}{l}\frac{1}{|\mathcal{S}|}\sum\limits_{s\in\mathcal{S}}[|(\boldsymbol{c}^\top\boldsymbol{x}_s+\boldsymbol{d}^\top\boldsymbol{y}_s)-z^\star(\tilde{\boldsymbol{w}}_s)|+\lambda\left\|\boldsymbol{H}_s\right\|_1] \\ +\sum\limits_{s\in\mathcal{S}}\sum\limits_{e\in\mathcal{E}}\mu_{s,e}\left(H_{s,e}-\sum\limits_{s'\in\mathcal{S}}H_{s',e}/|\mathcal{S}|\right)\end{array}\right\} \\
s.t. \quad & \boldsymbol{A}\boldsymbol{x}_s+\boldsymbol{B}\boldsymbol{y}_s\le\boldsymbol{g} \\
& \boldsymbol{F}\boldsymbol{y}_s\le\boldsymbol{H}_s\boldsymbol{f}_s \\
& \boldsymbol{x}_s=[x_{s,1},\cdots,x_{s,M}]\in\{0,1\}^M \quad (53)
\end{aligned}$$

The LR problem (53) can be naturally decomposed into $|\mathcal{S}|$ disjoint subproblems, each corresponding to a training scenario. With this, subproblem for scenario s can be shown as in (8).

REFERENCES

[1] T. Ding, Z. Zeng, M. Qu, J. P. S. Catalão, and M. Shahidehpour, "Two-stage chance-constrained stochastic thermal unit commitment for optimal provision of virtual inertia in wind-storage systems," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3520–3530, Jul. 2021.

[2] Y. Chen, Q. Guo, H. Sun, Z. Li, W. Wu, and Z. Li, "A distributionally robust optimization model for unit commitment based on Kullback–Leibler divergence," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5147–5160, Sep. 2018.

[3] M. Parvania and A. Scaglione, "Unit commitment with continuous-time generation and ramping trajectory models," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3169–3178, Jul. 2016.

[4] Y. Fu, Z. Li, and L. Wu, "Modeling and solution of the large-scale security-constrained unit commitment," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3524–3533, Nov. 2013.

[5] Y. Wang and L. Wu, "Improving economic values of day-ahead load forecasts to real-time power system operations," *IET Gener. Transm. Distrib.*, vol. 11, no. 17, pp. 4238–4247, Nov. 2017.

[6] G. Y. Ban and C. Rudin, "The big data newsvendor: Practical insights from machine learning," *Oper. Res.*, vol. 67, no. 1, pp. 90–108, Jan. 2019.

[7] A. N. Elmachtoub and P. Grigas, "Smart 'predict, then optimize'," *Manage. Sci.*, vol. 68, no. 1, pp. 9-26, Jan. 2022.

[8] A. Elmachtoub, J. C. N. Liang, and R. Mcnellis, "Decision trees for decision-making under the Predict-then-Optimize framework," in *Proc. 37th Intl. Conf. Mach. Learn.*, 2020, pp. 2858–2867.

[9] D. Bertsimas and N. Kallus, "From predictive to prescriptive analytics," *Manage. Sci.*, vol. 66, no. 3, pp. 1025–1044, Mar. 2020.

[10] D. Bertsimas and B. V. Parys, "Bootstrap robust prescriptive analytics," *Math. Prog.*, vol. 195, pp. 39-78, 2022.

[11] D. Bertsimas and O. S. Lami, "Holistic prescriptive analytics for continuous and constrained optimization problems," *INFORMS J. Optim.*, 2022. [Online]. Available: doi.org/10.1287/ijoo.2022.0080

[12] A. Stratigakos, A. Michiorri, and G. Kariniotakis, "A value-oriented price forecasting approach to optimize trading of renewable generation," in *Proc. 14th IEEE Madrid PowerTech*, Madrid, Spain, Jun. 2021, pp. 1–6.

[13] G. Li and H. Chiang, "Toward cost-oriented forecasting of wind power generation," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2508–2517, Jul. 2018.

[14] T. Carriere and G. Kariniotakis, "An integrated approach for value-oriented energy forecasting and data-driven decision-making application to renewable energy trading," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6933–6944, Nov. 2019.

[15] M. A. Muñoz, J. M. Morales, and S. Pineda, "Feature-driven improvement of renewable energy forecasting and trading," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3753–3763, Sep. 2020.

[16] M. A. Muñoz, S. Pineda, and J. M. Morales, "A bilevel framework for decision-making under uncertainty with contextual information," *Omega*, vol. 108, Art. no. 102575, 2022.

[17] J. D. Garcia, A. Street, T. Homem-de-Mello, and F. D. Muñoz, "Application-driven learning via joint prediction and optimization of demand and reserves requirement," 2021. [Online]. Available: https://arxiv.org/abs/2102.13273

[18] J. Mandi and T. Guns, "Interior point solving for LP-based prediction+optimisation," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 7272–7282, 2020.

[19] J.Mandi, P.J.Stuckey, and T.Guns,"Smart predict-and-optimize for hard combinatorial optimization problems," in *Proc. AAAI Conf. Art. Intel.*, vol. 34, no. 2, Apr. 2020, pp. 1603–1610.

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

[21] J. Han, L. Yan, and Z. Li, "A task-based day-ahead load forecasting model for stochastic economic dispatch," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5294-5304, Nov. 2021.

[22] OpenDataElia, Accessed: Jan. 1, 2021. [Online]. Available: https://opendata.elia.be/pages/home/

[23] Y. Zhang, X. Zhang, P. Huang, and Y. Sun, "Global sensitivity analysis for key parameters identification of net-zero energy buildings for grid interaction optimization," *Appl. Energy*, vol. 279, Dec. 2020, Art. no. 115820.

[24] D. Bertsimas and N. Mundru, "Optimization-based scenario reduction for data-driven two-stage stochastic optimization," *Oper. Res.*, 2021. [Online]. Available: doi.org/10.1287/opre.2022.2265

[25] M. L. Fisher, "An applications oriented guide to Lagrangian relaxation," *Interfaces*, vol. 15, no. 2, pp. 10–21, 1985.

[26] L. Wu, "Accelerating NCUC via binary variable-based locally ideal formulation and dynamic global cuts," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 4097–4107, Sep. 2016.

[27] Y. Wang, D. Gan, N. Zhang, L. Xie, and C. Kang, "Feature selection for probabilistic load forecasting via sparse penalized quantile regression," *J. Mod. Power Syst. Clean Energy*, vol. 7, no. 5, pp. 1200–1209, 2019.

[28] L. Wu, M. Shahidehpour, and T. Li, "Stochastic security-constrained unit commitment," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 800–811, May 2007.

[29] Closed_Loop_NCUC_Dataset-Load_RES_Feature_System, Accessed: Jul. 25, 2021. [Online]. Available: https://github.com/asxadf/Closed_Loop_NCUC_Dataset/