Unit Commitment under Uncertainty using Data-Driven Optimization with Clustering Techniques

Ning Zhao, Fengqi You

Cornell University, Ithaca, New York, USA
nz225@cornell.edu

This paper proposes a novel robust unit commitment (UC) framework with data-driven disjunctive uncertainty sets for volatile wind power outputs, assisted by machine learning techniques. To flexibly identify the uncertainty space based on wind power forecast error data with disjunctive structures, the uncertainty data are grouped using K-means and density-based spatial clustering of applications with noise following the optimal cluster number determined by the Calinski-Harabasz index. The disjunctive uncertainty sets are constructed accordingly as the union of multiple basic uncertainty sets, including conventional box and budget uncertainty sets, and data-driven uncertainty sets using Dirichlet process mixture model, principal component analysis coupled with kernel density estimation, and support vector clustering. The problem is formulated into a two-stage adaptive robust UC model with data-driven disjunctive uncertainty sets and with a multi-level optimization structure. To facilitate the solution process, a tailored decomposition-based optimization algorithm is developed. The effectiveness of the proposed framework is illustrated using an application on the IEEE 39-bus system. The proposed approach can reduce the price of robustness by 8-38 % compared to the conventional "one-set-fits-all" approaches. Benchmarking with stochastic programming indicates that the proposed framework can achieve the same or better economic performance with over 30 % less computational time.

1. Introduction

To deal with climate change and fossil fuel depletion, the United States is planning to use wind energy to provide 20 % electricity in 2030 and 35 % in 2050 (Zhao and You, 2020). Considering the intermittent and unpredictable nature of wind power, it would benefit the unit commitment (UC) decision-making process to include the wind power forecast uncertainties in terms of ensuring power systems reliability and reducing the economic costs (Ning et al., 2019). Robust UC has gained attention in recent years, owing to its robustness, flexible conservatism control, computational efficiency, and effective utilization of large-scale uncertainty data (Chen et al., 2018). The robust UC problem is generally formulated into a two-stage adaptive robust optimization (ARO) model, in which the first stage determines the on-off decisions and the second stage provides power output and dispatch decisions according to the worst case (Bertsimas et al., 2013). However, for uncertainty of the wind power forecast errors, they tend to have complex and disjunctive data structures, and studies found that it could be more appropriate to depict such uncertainty using more general distributions, such as the Gaussian mixture model (Wang et al., 2017). The finding indicates that the traditional "one-set-fits-all" approach for robust UC that constructs a single uncertainty set based on all uncertainty data may not flexibly and accurately capture the uncertainty space of the wind power forecast errors. From a machine learning perspective, clustering is a useful tool in detecting the disjunctive structure of a data set. There exists a knowledge gap to systematically handle the uncertain wind power forecast errors with disjunctive data structures in robust UC problems by integrating machine learning techniques with the robust UC model. To fill the knowledge gap, a novel ARO framework with data-driven disjunctive uncertainty sets is proposed in this work on UC with uncertain wind power forecast errors, which are independent and identically distributed (i.i.d.) (Bludszuweit et al., 2008), and a corresponding decomposition-based optimization algorithm is developed. An application on the IEEE 39-bus system is presented to illustrate the effectiveness of the proposed framework.

2. Data-driven robust unit commitment framework with disjunctive uncertainty sets

A two-stage ARO framework with data-driven disjunctive uncertainty sets is proposed in this work for UC with uncertain wind power forecast errors. The flowchart of the proposed framework is presented in Figure 1. First, the optimal number of uncertainty data clusters is determined using the Calinski-Harabasz index. Next, the uncertainty data are clustered accordingly using machine learning techniques. The data-driven disjunctive uncertainty sets are then constructed following the clustering results. The ARO problem is formulated by incorporating the data-driven disjunctive uncertainty sets, and the proposed ARO problem is then solved iteratively using a tailored decomposition-based optimization algorithm.
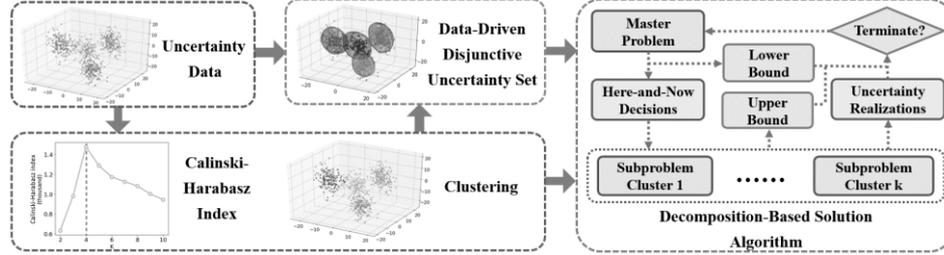

Figure 1: Flowchart of the proposed ARO framework with disjunctive uncertainty sets.

2.1 Adaptive robust unit commitment model

The mathematical formulation of the two-stage ARO model for the UC problem is presented as follows. The objective function of minimizing the total UC cost is presented by (1) (Jurković et al., 2015). Note that the optimization model has a two-stage structure. The "here-and-now" decisions in the first stage represent the commitments of generators that should be determined 24 h ahead, including online status, start-up status, and shutdown status of generators, and they should be made with respect to any realization of the uncertainties. The "wait-and-see" decisions in the second stage depict the economic dispatch process, and these decisions are determined after the uncertainty is realized. These decision variables consist of the power output of generators, power dispatch of a wind farm, and slack variables for balance constraints. The constraints are shown in (2) and (3), including the logic relations of generators, the minimum uptime and downtime of generators, the minimum and maximum power outputs of a generator, the ramping rates, the energy balance of the system, the capacities of transmission lines, the maximum outputs from wind farms that equal to the summation of forecasted wind power generation and the uncertain forecast errors, the initial commitment status and power outputs of generators, and the feasible region of decision variables. The constraint on the maximum outputs from wind farms holds for all realizations of the uncertainty to guarantee the robustness of solutions.

$$\min_{x_{i,t}, u_{i,t}, v_{i,t}} \sum_{i \in I} \sum_{t \in T - \{t_0\}} \left(C_i^G x_{i,t} + C_i^S u_{i,t} + C_i^D v_{i,t} \right) + \max_{\omega_{j,t} \in \mathbf{U}} \min_{p_{i,t}, w_{b,t}, q_t^+, q_t^-, q_{k,t}^f} \left(\begin{array}{l} \sum_{i \in I} \sum_{t \in T - \{t_0\}} C_i^c p_{i,t} + \sum_{t \in T - \{t_0\}} C^+ q_t^+ \\ + \sum_{t \in T - \{t_0\}} C^- q_t^- + \sum_{k \in K} \sum_{t \in T - \{t_0\}} C^f q_{k,t}^f \end{array} \right) \quad (1)$$

$$\text{s.t. } u_{i,t} - v_{i,t} = x_{i,t} - x_{i,t-1}, \forall i \in I, t \in T - \{t_0\}$$

$$u_{i,t} + v_{i,t} \leq 1, \forall i \in I, t \in T - \{t_0\}$$

$$\sum_{\tau = t - UT_i + 1}^{t} u_{i,\tau} \leq x_{i,t}, \forall i \in I, t \geq UT_i$$

$$\sum_{\tau = t - DT_i + 1}^{t} v_{i,\tau} \leq 1 - x_{i,t}, \forall i \in I, t \geq DT_i$$

$$p_{i,t} \geq P_i^{\min} x_{i,t}, \forall i \in I, t \in T - \{t_0\}$$

$$p_{i,t} \leq P_i^{\max} x_{i,t}, \forall i \in I, t \in T - \{t_0\}$$

$$p_{i,t-1} - p_{i,t} \leq RD_i x_{i,t} + SD_i v_{i,t}, \forall i \in I, t \in T - \{t_0\}$$

$$p_{i,t} - p_{i,t-1} \leq RU_i x_{i,t-1} + SU_i u_{i,t}, \forall i \in I, t \in T - \{t_0\}$$

$$\sum_{i \in I} p_{i,t} + \sum_{b \in B} w_{b,t} + q_t^+ - q_t^- = \sum_{b \in B} D_{b,t}, \forall t \in T - \{t_0\}$$

(2)

$$\sum_{b\in B} SF_{b,k}\left(\sum_{i\in I} MAP_{b,i}\, p_{i,t} + \sum_{j\in J} WF_{b,j}\, w_{j,t} - D_{b,t}\right) - q_{k,t}^{f} \le CAP_{k}, \forall k,t$$

$$\sum_{b\in B} SF_{b,k}\left(\sum_{i\in I} MAP_{b,i}\, p_{i,t} + \sum_{j\in J} WF_{b,j}\, w_{j,t} - D_{b,t}\right) + q_{k,t}^{f} \ge -CAP_{k}, \forall k,t$$

$$w_{j,t} \le FORE_{j,t} + \omega_{j,t}, \forall j \in J, t \in T-\{t_0\}, \omega_{j,t} \in U \tag{3}$$

$$p_{i,t=t_0} = PI_{i}, \forall i \in I$$

$$x_{i,t=t_0} = XI_{i}, \forall i \in I$$

$$x_{i,t}, u_{i,t}, v_{i,t} \in \{0,1\}, \forall i \in I, t \in T-\{t_0\}$$

$$p_{i,t}, w_{b,t}, q_{t}^{+}, q_{t}^{-}, q_{l,t}^{f} \ge 0, \forall i \in I, b \in B, k \in K, t \in T-\{t_0\}$$

2.2 Construction of data-driven disjunctive uncertainty sets assisted by clustering techniques

A novel data-driven disjunctive uncertainty set is developed in this work. The uncertainty data samples are first clustered using machine learning techniques. Subsequently, disjunctive uncertainty sets are formulated as the union of multiple uncertainty sets that are constructed using either traditional or data-driven methods based on the data grouping results, as shown in Eq(4), where L is the set of uncertainty data clusters.

$$\mathbf{U} = \bigcup_{l} \mathbf{U}_{l} \tag{4}$$

For data clustering, two machine learning techniques, namely K-means and density-based spatial clustering of applications with noise (DBSCAN), are used to group the uncertainty data according to the optimal number of clusters. K-means is an unsupervised, non-deterministic, and centroid-based clustering algorithm, where k represents the number of clusters. K-means iteratively updates the data assignments and terminates when the clustering results remain unchanged between iterations. The key criterion of DBSCAN requires the neighborhood of each data sample in a group within a given radius containing more data samples than a given number (Khan et al., 2014). Note that DBSCAN can detect outliers and filter the noise, while K-means can be sensitive to outliers.

To determine the optimal number of clusters, both the elbow method and the Calinski-Harabasz index can be used (Caliński and Harabasz, 1974). For the elbow method, the uncertainty data are first grouped using K-means for a range of numbers of clusters (k) and calculate the intra-cluster variation (W). As W monotonically decreases with the increase of k, the goal of the elbow method is to choose a value of k with a relatively low W, and the decreasing trend of W under higher values of k becomes less conspicuous. In addition to the group compactness considered in the elbow method, the Calinski-Harabasz index also includes the distribution of all groups which is represented by the inter-cluster variation (B). As the optimal number of data groups should have a good balance between the compactness and the distribution, a higher value of Calinski-Harabasz index indicates a better result.

Based on the clustering results, five types of basic uncertainty set can be constructed, including the traditional box and budget uncertainty sets, and three data-driven uncertainty sets using a variational inference algorithm for Dirichlet process mixture model (DPMM) (Ning and You, 2017), principal component analysis (PCA) coupled with kernel density estimation (KDE) (Ning and You, 2018), and support vector clustering (SVC) (Shang and You, 2017). DPMM is a Bayesian nonparametric model, and in Bayesian statistics, model parameters are handled as random variables. For each set of grouped data samples, a variational inference algorithm is applied to determine the posterior distribution of the modeling parameters, based on which the data-driven uncertainty set is constructed using both l_1 and l_∞ norms. For the approach using PCA coupled with KDE (Zhao et al., 2019), PCA is first applied on the uncertainty data to obtain their orthogonal principal components, then each uncertainty data is projected onto the principal components. Then, the distributional information of the projected data is extracted using KDE, and the uncertainty sets could be constructed by integrating the distributional information with the principal components. As for the SVC-based uncertainty sets, SVC with a piecewise linear kernel is applied as the generalized intersection kernel, which could facilitate the solution process, in terms of the computational tractability and the convex formulation of the data-driven uncertainty sets.

2.3 Construction of data-driven disjunctive uncertainty sets assisted by clustering techniques

The two-stage robust UC with the proposed disjunctive uncertainty sets have an objective function with a multi-level structure, semi-infinite constraints, and non-convex uncertainty sets. To facilitate the solution process, a decomposition-based optimization algorithm is developed. The developed algorithm conducts a partial enumeration over a subset of the disjunctive uncertainty set, considering that the maximization over the entire

disjunctive uncertainty set U should be achieved on an extreme point of a basic uncertainty set U_l. Specifically, a master problem and a set of subproblems are iteratively solved in the algorithm. The master problem optimizes the UC decisions under multiple optimality cuts that correspond to a partial enumeration of the extreme points of the basic uncertainty sets, and its optimal objective value provides a lower bound (LB) for the original minimization problem. Then, the values of first-stage decision variables are fixed following the optimal solutions of the master problem and develop a set of subproblems to investigate economic dispatch under the worst case. Note that each basic uncertainty set U_l corresponds to an individual subproblem, so a set of subproblems is developed for the disjunctive uncertainty sets. In the subproblem, the "min-max-min" structure of the ARO problem is simplified to a "max-min" structure. To reformulate the subproblem into a single-level maximization problem that can be solved directly by off-the-shelf solvers, the classical Karush-Kuhn-Tucker (KKT) conditions and the big-M method are used. After solving the set of subproblems, an upper bound (UB) of the original problem can be found as the largest optimal objective value across all basic uncertainty sets, as well as a feasible schedule under the determined first-stage decisions according to the worst case. A set of additional optimality cuts is then generated based on the uncertainty realization in the worst case and is updated in the master problem for the next iteration. The algorithm terminates when the relative optimality gap is below the tolerance level ξ.

3. Application on the IEEE 39-bus system

To illustrate the effectiveness of the proposed framework with disjunctive uncertainty sets, an application on the IEEE 39-bus system is investigated. The system includes 39 buses, 10 generators, and 46 lines (Kazarlis et al.,1996). Three wind farms exist at buses 30, 32, and 38, and the wind power forecast data are scaled down based on (Argonne National Laboratory, 2009). 800 uncertainty data samples for wind forecast error are generated from a Gaussian mixture model. The two-stage robust UC problem is coded using Pyomo in Python on a PC with an Intel i7-8700 @ 3.20 GHz CPU and 32GB RAM, running on a 64-bit Windows 10 Enterprise operating system. The reformulated master problem and subproblem using the optimization algorithm developed in Section II are solved using Gurobi 9.1 with an optimality tolerance of 10^{-6}.
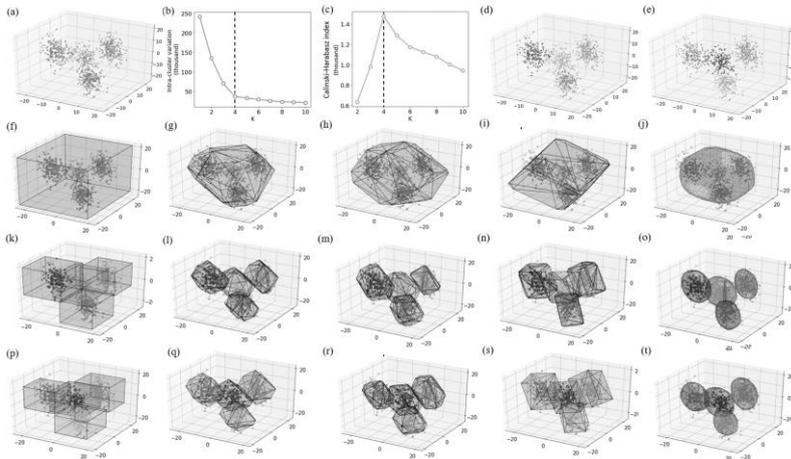

Figure 2: Uncertainty data, determination of optimal group number, clustering results, and uncertainty sets for application on IEEE 39-bus system. (a) Uncertainty data samples. (b) Intra-cluster variation. (c) Calinski-Harabasz index. (d) K-means results. (e) DBSCAN results. (f) Conventional box. (g) Conventional budget. (h) Conventional DPMM. (i) Conventional PCA & KDE. (j) Conventional SVC. (k) K-means + box. (l) K-means + budget. (m) K-means + DPMM. (n) K-means + PCA & KDE. (o) K-means + SVC. (p) DBSCAN + box. (q) DBSCAN + budget. (r) DBSCAN + DPMM. (s) DBSCAN + PCA & KDE. (t) DBSCAN + SVC

Figure 2 presents the uncertainty data, the data clustering results, and the uncertainty sets. Wind forecast error data samples for the three wind farms are shown in Figure 2a. Based on the results from both the elbow method and the Calinski-Harabasz index method in Figure 2b-c, the optimal number of data groups is 4. Next, K-means and DBSCAN are applied individually to group the data into 4 data groups, as shown in Figure 2d-e. Both the conventional and the proposed approaches are used for uncertainty set construction, and the results are shown in Figure 2f-t. The proposed disjunctive uncertainty sets could more accurately capture the uncertainty space based on the data samples compared to the conventional approach, so the proposed approach is expected to result in less conservative solutions.

The reformulated master problems have up to 730 integer variables, 7,136 continuous variables, and 13,662 constraints, while the subproblems have up to 6,514 integer variables, 7581 continuous variables, and 26,921 constraints. The solution time under conventional uncertainty sets ranges from 80 s to 221 s, and the problems with the proposed framework take 225-242 s to solve. The different solution time results from the different numbers of inner iterations that equal to the cluster number. The conventional uncertainty sets lead to only one subproblem, while the disjunctive uncertainty sets correspond to 4 subproblems.

The optimal objective values are listed in Table 1. Higher values indicate more conservative solutions. For comparison, the minimum cost for the deterministic case with no uncertainties is 426,443 USD. The price of robustness (PoR) is applied to measure the level of additional cost for the robust cases compared to deterministic planning. PoR is defined as the difference between the optimal objective values of the robust and deterministic cases divided by the optimal objective value of the deterministic case. Using the proposed disjunctive uncertainty sets, the PoR is reduced by 28-38 %, 7.6 %, 23-29 %, 27-31 %, and 21-22 % for the problems with box, budget, DPMM, PCA coupled with KDE, and SVC uncertainty sets, respectively, indicating that the proposed approach can provide less-conservative solutions under the same level of conservatism. Also, the disjunctive uncertainty sets using DBSCAN tend to have lower optimal costs than that using K-means, suggesting that DBSCAN may handle the outliers of the uncertainty data more efficiently.

Table 1: Optimal objective values under uncertainty sets with different clustering and construction approaches.

minimum cost (USD)	Box	Budget	DPMM	PCA & KDE	SVC
No Clustering	461,020	441,155	445,741	455,089	444,191
K-Means	451,178	440,041	441,277	447,271	440,508
DBSCAN	447,712	440,041	440,092	446,121	440,217

To benchmark the performance of the proposed approach, the UC costs under the optimal solutions from both the ARO models and the conventional two-stage stochastic UC models are simulated. The optimal solutions of the two-stage stochastic UC are obtained through 30 and 100 random scenarios based on the existing uncertainty data, denoted respectively as SP 30 and SP 100. The stochastic UC problems tend to be more computationally demanding, as obtaining SP 30 and SP 100 solutions take 342 and 1,194 CPUs, which are over 40 % longer than the solution time using the proposed framework. The two-stage stochastic UC problem with 800 scenarios that are formulated based on all uncertainty data cannot be solved within one hour, and the lower and upper bounds returned from the solver after one CPU hour are 427,677 and 429,768 USD, respectively.

We further simulated UC costs following the optimal solutions from ARO and SP methods throughout 100 random scenarios based on the uncertainty data. Note that the scenarios for UC simulation are different from the scenarios for stochastic UC. We found SP 30 solution cannot handle the systems contingencies effectively, as two scenarios with UC costs over 480,000 USD, and its average simulated cost is higher than that of the SP 100 solution. ARO solution with conventional box uncertainty set has the highest average cost, but it could hedge against the contingencies more effectively compared to the SP 30 solution. For the proposed approach, the solution from K-means-based disjunctive box uncertainty sets shows a lower average simulated UC cost than ARO Box with the same level of conservatism. The differences of average costs between SP 100 and ARO solutions based on the proposed disjunctive uncertainty sets are below 0.05 %, while obtaining solution SP 100 requires around 400 % longer computational time.

4. Conclusions

This paper proposed a novel robust UC framework with data-driven disjunctive uncertainty sets for the volatile wind power generation, which integrated the two-stage adaptive robust UC model with machine learning techniques to flexibly capture the uncertainty regions of the wind power forecast errors with disjunctive structures. The uncertainty data were grouped following the optimal number of data groups determined by the Calinski-Harabasz index, and the data-driven disjunctive uncertainty sets were constructed as the union of multiple basic uncertainty sets, including conventional box and budget uncertainty sets, and data-driven uncertainty sets using DPMM, PCA coupled with KDE, and SVC. The resulting problem was formulated as a two-stage robust UC model with the proposed disjunctive uncertainty sets, and a tailored decomposition-based optimization algorithm was developed to facilitate the solution process. An application on the IEEE 39-bus systems were performed. The results presented that the price of robustness reduced by 8-38 % under the proposed framework, compared to the traditional approach. Benchmarking with SP indicated that the proposed approach could achieve the same or better economic performance with over 30 % less computational time. One future research extension would be to illustrate the scalability of the proposed framework on large-scale applications.

Nomenclature

Sets and Indices
b/i/j – Index for buses/generators/wind farms.
k/l/t – Index for transmission lines/clusters/periods.
Parameters
C_i^C/C_i^G – Variable operating cost/no-load cost
C_i^S/C_i^D – Startup/shutdown cost of generator i.
C^+, C^- – Penalty for violating energy balances.
C^f – Penalty for violating transmission capacities.
CAP_k – Capacity of transmission line k.
$D_{b,t}$ – Load demand for bus b at period t.
DT_i/UT_i Minimum downtime/uptime of generator i.
$FORE_{j,t}$ – Forecasted wind power generation.
$MAP_{b,i}$ – Indicator for existence of unit i at bus b.
P_i^{max}/P_i^{min} – Maximum/minimum power output.
PI_i/XI_i – Initial power output/online status.

RD_i/RU_i – Shutdown/start-up ramp rates.
SD_i/SU_i – Maximum ramp-down/ramp-up limits.
$SF_{b,k}$ – Power transfer distribution factor.
$WF_{b,j}$ – Indicator for wind farm j at bus b.
$\omega_{j,t}$ – Uncertain wind power forecast error.
Binary Decision Variables
$u_{i,t}$ – Startup status of generator i at period t.
$v_{i,t}$ – Shutdown status of generator i at period t.
$x_{i,t}$ – Commitment status of generator i at period t.
Continuous Decision Variables
$p_{i,t}$ – Power output of generator i at period t.
$q_{k,t}^f$ – Slack variables in transmission constraints.
q_t^+, q_t^- – Slack variables in balance constraints.
$w_{j,t}$ – Power dispatch of wind farm j at time t.

References

Argonne National Laboratory, 2009, Wind Power Forecasting: State-of-the-Art 2009.

Bertsimas D., Litvinov E., Sun X. A., Zhao J., Zheng T., 2013, Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem, IEEE Transactions on Power Systems, 28, 52-63.

Bludszuweit H., Dominguez-Navarro J. A., and Llombart A., 2008, Statistical Analysis of Wind Power Forecast Error, IEEE Transactions on Power Systems, 23, 983-991.

Caliński, T., Harabasz, J., 1974, A dendrite method for cluster analysis, Communications in Statistics, 3, 1-27.

Chen Y., Guo Q., Sun H., Li Z., Wu W., Li Z., 2018, A Distributionally Robust Optimization Model for Unit Commitment Based on Kullback–Leibler Divergence, IEEE Transactions on Power Systems, 33, 5147-5160.

Kazarlis S. A., Bakirtzis A. G., Petridis V., 1996, A genetic algorithm solution to the unit commitment problem, IEEE Transactions on Power Systems, 11, 83-92.

Jurković, K., Pandšić, H., Kuzle, I., 2015, Review on unit commitment under uncertainty approaches, 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 1093-1097.

Khan, K., Rehman, S. U., Aziz, K., Fong, S., Sarasvady, S., 2014, DBSCAN: Past, present and future, in The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014), 232-238.

Ning C., You F., 2017, Data-driven adaptive nested robust optimization: General modeling framework and efficient computational algorithm for decision making under uncertainty, AIChE Journal, 63, 3790-3817.

Ning C., You F., 2018, Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods, Computers & Chemical Engineering, 112, 190-210.

Ning C., You F., 2018, Data-Driven Stochastic Robust Optimization: General Computational Framework and Algorithm Leveraging Machine Learning for Optimization under Uncertainty in the Big Data Era. Computers & Chemical Engineering, 111, 115-133.

Ning C., You F., 2019, Optimization under Uncertainty in the Era of Big Data and Deep Learning: When Machine Learning Meets Mathematical Programming. Computers & Chemical Engineering, 125, 434-448.

Ning C., You F., 2019, Data-Driven Adaptive Robust Unit Commitment under Wind Power Uncertainty: A Bayesian Nonparametric Approach. IEEE Transactions on Power Systems, 34, 2409-2418.

Shang C., Huang X., You F., 2017, Data-driven robust optimization based on kernel learning, Computers & Chemical Engineering, 106, 464-479.

Wang Y., Hu Q., Meng D., Zhu P., 2017, Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model, Applied Energy, 208, 1097-1112.

Zhao, L., Ning C., You F., 2019, Operational Optimization of Industrial Steam Systems under Uncertainty Using Data-Driven Adaptive Robust Optimization. AIChE Journal, 65, e16500.

Zhao, N., You F., 2020, Can renewable generation, energy storage and energy efficient technologies enable carbon neutral energy transition? Applied Energy, 279, 115889.

Zhao, N., You F., 2021, New York State's 100% renewable electricity transition planning under uncertainty using a data-driven multistage adaptive robust optimization approach with machine-learning. Advances in Applied Energy, 2, 100019.