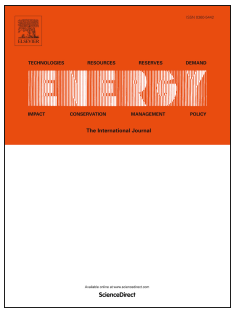
Accepted Manuscript



A data-driven approach for multi-objective unit commitment under hybrid uncertainties

Min Zhou, Bo Wang, Tiantian Li, Junzo Watada

PII: S0360-5442(18)31762-6

DOI: 10.1016/j.energy.2018.09.008

Reference: EGY 13702

To appear in: *Energy*

Received Date: 7 March 2018

Revised Date: 11 August 2018

Accepted Date: 1 September 2018

Please cite this article as: Zhou M, Wang B, Li T, Watada J, A data-driven approach for multi-objective unit commitment under hybrid uncertainties, *Energy* (2018), doi: 10.1016/j.energy.2018.09.008.

A data-driven approach for multi-objective unit commitment under hybrid uncertainties

Min Zhou[a], Bo Wang[a,*], Tiantian Li[a], Junzo Watada[b]

[a]*School of Management and Engineering, Nanjing University, Nanjing 210093, China*
[b]*Graduate School of Information, Production and Systems, Waseda University, Kitakyushu 808-0135, Japan*

Abstract

Recent years, renewable energy has taken growing penetration in power systems due to the energy shortage and environmental concerns. As a result, system operators encounter increasing difficulties in solving unit commitment optimization. In this paper, a data-driven unit commitment model is proposed to handle the hybrid uncertainties of wind power and future load. First, a non-parameter kernel density method is utilized to represent the above hybrid uncertainties, and a novel bandwidth selection strategy for the above method is then proposed to capture the inherent correlation between uncertainty representation and unit commitment. Second, a Monte Carlo simulation is developed to integrate the hybrid uncertainties into Value-at-Risk to get a comprehensive system reliability measurement. Third, considering that system operators might be interested in the inherent conflict between reliability and economy, minimizing operation costs and maximizing system reliability are taken as two objectives in the model. To get more practical schedules, the transmission line constraint is considered as well when building the mathematical model. Additionally, by integrating the reinforcement learning mechanism, a novel multi-objective particle swarm optimization algorithm is proposed to solve the complicated nonlinear model. Finally, several experiments were performed to demonstrate the effectiveness of this research.

Keywords: Multi-objective unit commitment, non-parameter kernel density method, hybrid uncertainties, reinforcement learning-based particle swarm optimization algorithm.

Acronyms

UC	Unit commitment
SOs	System operators
SUC	Stochastic unit commitment
EENS	Expected energy not served
LOLP	Loss of load probability
VALL	Value at lost load
RUC	Robust unit commitment
VaR	Value-at-Risk
DD-MOUC	Data-driven multi-objective UC
NPKDM	Non-parameter kernel density method
TLC	Transmission line constraint
MILP	Mixed integer linear programming
LR	Lagrangian relaxation
PSO	Particle swarm optimization
MOPSO	Multi-objective PSO
DT-MOPSO	Dominance time-based MOPSO
TV-MOPSO	Time-varying MOPSO
RL-MOPSO	Reinforcement learning-based MOPSO

*Corresponding author.

AgMOPSO	External archive guided MOPSO
PL	Priority list
DPL	Descending order of priority list

1. Introduction

Unit commitment (UC) is an important control process in power systems, which aims to minimize operation costs while maintaining system reliability. Usually, system operators (SOs) schedule the on/off status and output of each unit to achieve an optimal UC solution. Recent years, UC optimization has been playing an significant role in power systems due to economical and environmental concerns. However, challenges do exist when SOs try to find satisfying UC solutions. In the literature, these challenges mainly stem from the uncertainties of unpredictable renewable generation and unforseen load fluctuation [1].

In order to represent the above uncertainties, parameter estimation has been proposed by researchers [2]. The idea of parameter estimation assumes that the estimated uncertainties follow a certain distribution [3]. In [4], the wind speed was assumed to follow Weibull distribution. Then a modified maximal likelihood method was proposed to identify the parameters of the distribution. Similarly in [5], Weibull distribution was utilized to estimate a wind speed frequency distribution. Instead of the maximal likelihood method, a regression method was applied as the parameter estimator. In [6], the future load was supposed to follow Gaussian distribution. Then a seven step approximation was used to incorporate the future load uncertainties into the UC model.

Although parameter estimation has been recommended as an effective tool for uncertainty representation, problems and defects do exist. In some real applications, SOs are in doubt about determining the exact distribution as there is no theoretical guidance for it. Besides, the obtained distribution may not fit the historical data well. To solve the above problems, some data-driven approaches have been proposed by researchers in recent years. For example, Bagheri et al. applied a non-parametric estimation method to construct the reference distribution of future load. Then L_1 and L_∞ probability metrics were applied to build the confidence set based on the obtained distribution [7]. Weng et al. utilized time series analysis and supervised learning to explore the periodic pattern of power system operation [8]. Zhao and Guan et al. utilized K bins to represent the wind power distribution [9]. Then a confidence set was constructed based on the L_1 and L_∞ norm of the empirical error. Several experiments demonstrate that the estimation error of the proposed method vanishes with the increasing size of data. In addition, Blanco et al. described the uncertainties of wind power by a set of scenarios [10]. To mitigate the high computation burden brought by a huge number of scenarios, clustering techniques were utilized to assemble similar scenarios, based on which, a distributional robust optimization was utilized to improve the uncertainty description of conventional UC. Compared with parameter estimation, data-driven approaches make no hypothesis about the estimated uncertainties, thus relieving the errors caused by the assumptions of unpractical distributions.

Although the above methods improve the conventional parameter estimation from the perspectives of accuracy and universality, there is still one major concern: the above methods cannot provide exact mathematical expressions of the obtained distributions, which could be a main drawback. However, when evaluating system reliability, SOs may be interested in the statistics of the obtained distributions, some of which need to be performed on the exact mathematical expression. Based on the above consideration, the non-parameter kernel density method (NPKDM) has been used as an effective tool of uncertainty representation in recent years. For example, Rosenblatt et al. utilized a diffusion-based NPKDM to estimate the distribution of wind speed [11]. Then the bandwidth selection was performed based on the diffusion partial differential equation. Similarly in [12], NPKDM with a novel bandwidth selection method was utilized to model the distribution of wind speed. Then several statistic tests demonstrate that the method can get more satisfying results than conventional parameter estimations. However, the above two studies aim to minimize the estimation error, thus ignoring the correlation between uncertainty representation and UC optimization. In reality, uncertainty representation should serve for the economy and reliability aspects of UC. As a result, it is significant to investigate the impact of NPKDM and its bandwidth selection strategy on UC optimization.

To handle UC optimization under uncertainties, various stochastic and robust UC models have been proposed. Stochastic UC (SUC) aims to optimize the expected performance of the system under a number of scenarios [13]. In [14], stochastic outages of units and transmission lines were modeled based on scenario trees in the Monte Carlo simulation. Then value at lost load (VALL) was introduced as a measurement of active and inactive power loss. In

[15], the uncertainties of renewable generation were captured by a set of prediction intervals. Thereafter, scenarios were generated from the prediction intervals by scenario generation techniques. Schulze et al. utilized scenario generation techniques to represent the uncertainties of wind power forecast errors. Then several experiments were performed to calculate the added value of stochastic methods over deterministic ones [16]. Then concepts such as expected energy not served (EENS) and loss-of-load probability (LOLP) were introduced as reliability measurements of UC [17]. Although SUC has been recommended as an effective tool, it ignores system robustness. To mitigate the above defct, robust UC (RUC) has been studied extensively in recent years. For example, An and Zeng developed an expanded two-stage RUC model which considers the weighted summation of performance under manifold uncertainty sets [18]. Lee constructed a RUC model while considering full transmission constraints. Afterwards, novel acceleration techniques were proposed to solve the above model [19]. Jiang proposed an effective method to adjust the conservativeness of RUC by deciding the number of periods that deviates largely to the expected value [20].

Although the modeling complexity and computation burden of RUC are much more acceptable than SUC, the obtained solutions are usually conservative. Based on the above analysis, it could be reasonable to consider stochastic and robust features in a comprehensive way. Therefore, a unified reliability measurement based on Value-at-Risk (VaR) is applied to evaluate the reliability of UC schedules in this study. In reality, VaR has been popularly used in engineering problems including UC [21]. In this research, wind power and future load uncertainties are integrated into VaR to achieve a comprehensive measurement of system reliability.

Considering that SOs might be interested in the inherent conflict between reliability and economy, this paper develops a data-driven multi-objective UC model (DD-MOUC) which optimizes the operation costs and system reliability simultaneously. Then among the Pareto-optimal solutions, SOs can select the one that satisfies them the most. Besides, with the growing of environmental concerns, the emission cost is included in the operation costs. Moreover, to achieve more practical UC schedules, the transmission line constraint (TLC) is also considered when building the mathematical model. Experiments were performed to exemplify the above considerations in Section V.

As a complicated nonlinear multi-objective model, the proposed DD-MOUC is difficult to be solved by conventional methods such as mixed integer linear programming (MILP) and Lagrangian relaxation (LR). Recent years, multi-objective particle swarm optimization (MOPSO) algorithms have been used as a feasible solution to multi-objective problems. For example, Zhu proposed an external archive guided MOPSO (AgMOPSO) in which the swarm leaders are selected from the external archive [22]. Hu proposed a parallel cell coordinate system to obtain the tradeoff between exploration and exploitation by getting feedback from the evolutionary environment [23]. Wang utilized a binary method for MOPSO to solve multi-objective UC problems. Specially, a distance based strategy was proposed to update the personal best of each particle [24]. Tripathi *et al*. proposed a time-varying MOPSO (TV-MOPSO) in which the inertia weight and learning rates of the particles vary with time [26]. However, the global best (Gbest) in TV-MOPSO is selected based on a density measurement, which brings high computation requirements to the algorithm. To further mitigate the computation burden, Wang et al. proposed a dominance time-based MOPSO (DT-MOPSO) in which the dominance times of each non-dominated solution is regarded as the fitness value [27]. Then Gbest is decided by a fitness value-based roulette.

Nevertheless, meta-heuristic algorithms include the above MOPSOs are easy to get trapped in a local optimization, especially when the UC problem contains a large number of units. To further mitigate the local convergence problem, reinforcement learning mechanism is integrated into the evolvement of particles to improve existing MOPSOs. As an effective optimization method, reinforcement learning aims to find an optimal strategy in an unknown environment by learning from the experiences of taking different actions. Basically, the reinforcement learning mechanism can adjust the evolvement of the particles by rewarding well behaved actions and punishing those with poor performance. In the literature, a reinforcement learning-based particle swarm optimization algorithm has been proposed to solve single-objective problems [28]. In this research, the above idea is further extended to multi-objective optimization and proposed a novel reinforcement learning-based MOPSO (RL-MOPSO). In the algorithm, particles act in a similar way to the agents of reinforcement learning. The detailed knowledge of RL-MOPSO is provided in Section IV and its effectiveness is proved by a series of experiments in Section V.

The rest of this paper is organized as follows: Section II introduces the related knowledge on NPKDM including the proposed bandwidth selection strategy. In Section III, the objective functions and constraints are first provided, then construct the mathematical model of DD-MOUC. Section IV develops RL-MOPSO to solve the proposed model. Then in Section V, a series of experiments were performed to exemplify the feasibility of the model and algorithm. Finally, Section VI summarizes our conclusions.

2. Preliminaries

This section first explains how NPKDM represents wind power and future load uncertainties based on historical data. Then a brief introduction to the proposed bandwidth selection strategy is provided. Finally, χ^2 test is introduced to ensure the accuracy of NPKDM.

2.1. Non-parameter kernel density method

As introduced, the wind power and future load uncertainties are addressed by NPKDM in this research. Therefore, some necessary concepts of this method are provided in the following part.

Let $X_1 X_2, \cdots, X_n$ denote a set of data samples. Then the estimated distribution $f_n(x)$ can be expressed by the following function:

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - X_i}{h}), \tag{1}$$

where n is the number of samples, h is called as the bandwidth, $K(*)$ indicates a symmetric kernel function which must obey the following constraints:

$$\begin{cases} \int K(x)dx = 1 \\ \int xK(x) = 0. \\ \int x^2 K(x) > 0 \end{cases} \tag{2}$$

Mathematically, it has been proved that $f_n(x)$ can infinitely approximate the real distribution $f(x)$ if n is sufficiently large. Suppose that the Gaussian kernel function is used to represent $K(*)$, as follows:

$$K(u) = \frac{1}{\sqrt{2\pi}} exp(\frac{-u^2}{2}). \tag{3}$$

Then the estimated distribution $f_n(x)$ can be described as below:

$$f_n(x) = \frac{1}{\sqrt{2\pi}nh} \sum_{i=1}^{n} exp(-\frac{(x - X_i)^2}{2h^2}). \tag{4}$$

2.2. Bandwidth selection strategy

The bandwidth selection of NPKDM has an significant impact on the estimation result. Basically, the obtained distribution would be unsmooth if a small bandwidth is utilized. By contrast, the using of a large bandwidth will make the obtained kernel density curve excessively homogeneous. As introduced before, most of existing studies choose the optimal bandwidth by minimizing the estimation error. For example, Qin et al. utilized two different Gaussian kernel functions to estimate the distribution of wind speed individually. Then the optimal bandwidth was obtained by minimizing the integration of the square difference considering the two kernel density estimators [12]. In [11], a random value was added to the original data samples to ensure a smooth distribution curve. Then the diffusion partial differential equation was applied to calculate the optimal bandwidth.

Nevertheless, the above two methods ignore the inherent correlation between uncertainty representation and UC decision. In reality, the future load as well as wind power forecast should serve for the economy and reliability aspects of UC. Therefore, the optimization on bandwidth is further performed by considering the above aspects in this study. And, the bandwidth value is optimized together with the UC problem by the proposed RL-MOPSO, which will be explained in Section IV.

2.3. χ^2 test

To ensure the accuracy of NPKDM, χ^2 test is utilized as a statistical constraint. Suppose that the data samples are divided into k groups, then the χ^2 statistic is defined as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(F_i - E_i)^2}{E_i}, \tag{5}$$

where F_i and E_i are the observed and expected frequency of group i. And, E_i can be calculated as follows:

$$E_i = \int_{x_1}^{x_2} f_n(x)dx, \tag{6}$$

where x_1 and x_2 are the lower and upper bounds of group i.

In the χ^2 test constraint, a bandwidth value is unacceptable at a given confidence level α if the obtained χ^2 is larger than the critical value $\chi^2_{1-\alpha,k-1}$. Therefore, the χ^2 test constraint can be formed as below:

$$\chi^2 \leq \chi^2_{1-\alpha,k-1}. \tag{7}$$

3. Modeling

The proposed DD-MOUC optimizes the operation costs and system reliability simultaneously subjects to several constraints. To ensure the feasibility of UC schedules, TLC is considered as well in this study. The detailed mathematical model is presented in the following part.

3.1. Notation

First, the notation used in the subsequent problems is introduced.

Indices and sets

- t: Index of each scheduling period
- i: Index of each unit
- j: Index of each load point
- i_w: Index of each wind turbine
- n: Index of each bus
- k: Index of each transmission line
- s: Index of each particle
- T: Time horizon
- I: Set of units
- J: Set of load points
- I_w: Set of wind turbines
- N: Set of buses
- K: Set of transmission lines
- S: Set of particles

Constants

- SU_i: Startup cost of unit i ($)
- SD_i: Shutdown cost of unit i ($)
- a_i, b_i, c_i: Generation cost coefficients of unit i
- a_i^E, b_i^E, c_i^E: Emission cost coefficients of unit i
- p_i^{max}: Maximal output of unit i (MW)
- p_i^{min}: Minimal output of unit i (MW)
- $p_{i_w}^{rated}$: Rated power of wind turbine i_w (MW)
- RR_i: Maximal ramp rate of unit i (MW/hour)
- $T_{i,up}$: Minimal on time of unit i (hour)
- $T_{i,down}$: Minimal off time of unit i (hour)
- $h_{k,n}$: Change of power flow in branch k given a unit increase at bus n
- C_k: Capacity of transmission line k (MW)
- ξ_j^{te}: Expected output of load point j at t (MW)

Uncertainties

- ξ_j^t: Uncertain load of load point j at t (MW)
- $p_{i_w}^t$: Uncertain output of wind turbine i_w at t (MW)

Functions

- $G(u_i^t, p_i^t)$: Generation cost function of unit i at t
- $E(u_i^t, p_i^t)$: Emission cost function of unit i at t
- $\mathrm{VaR}_{(1-\beta)}^t$: VaR-based reliability measurement

Variables

- u_i^t: On/off status of unit i at t
- p_i^t: Output of unit i at t (MW)
- x_i^t: Startup action of unit i at t
- y_i^t: Shutdown action of unit i at t
- $(1-\beta)$: Confidence level on VaR
- SR_i^t: Spinning reserve of unit i at t (MW)
- $T_{i,on}^t$: Periods that unit i has been kept on at t (hour)
- $T_{i,off}^t$: Periods that unit i has been kept off at t (hour)
- f_k^t: Power flow of transmission line k at t (MW)

3.2. Objective functions

In this study, minimizing operation costs and maximizing system reliability are the two objectives in the mathematical model.

3.2.1. Operation costs with emission concerns

The operation costs are composed of the transition, generation and emission costs over the scheduling horizon:

$$\min \sum_{t\in T}\sum_{i\in I}[(x_i^t SU_i + y_i^t SD_i) + G(u_i^t, p_i^t) + Q(u_i^t, p_i^t)], \tag{8}$$

where x_i^t and y_i^t indicate the startup and shutdown actions of unit i at time period t, which can be calculated as follows:

$$\begin{aligned} x_i^t &= u_i^t(1-u_i^{t-1}),\\ y_i^t &= u_i^{t-1}(1-u_i^t). \end{aligned} \tag{9}$$

$G(u_i^t, p_i^t)$ represents the generation cost of unit i at output p_i^t, which can be approximately calculated by the following quadratic function:

$$G(u_i^t, p_i^t) = u_i^t[a_i + b_i p_i^t + c_i(p_i^t)^2], \tag{10}$$

where a_i, b_i and c_i are coefficients determined by the attributes of unit i.

$Q(u_i^t, p_i^t)$ measures the emission cost of unit i at time period t, which can be calculated as below:

$$Q(u_i^t, p_i^t) = u_i^t[a_i^E + b_i^E p_i^t + c_i^E(p_i^t)^2], \tag{11}$$

where a_i^E, b_i^E and c_i^E are emission coefficients of unit i. Besides, it is assumed that the transition cost contains the emission cost incurred during startup and shutdown.

3.2.2. Value-at-Risk-Based reliability measurement

VaR of an investment indicates the greatest loss under a predefined confidence level [29]. Suppose ζ represents the loss of an investment with probability density function $f(\zeta)$, then the VaR of ζ under confidence level $(1-\beta)$ is defined as follows:

$$\mathrm{VaR}_{(1-\beta)}\{\zeta\} = sup\left\{\lambda | \left\{\int_{\lambda}^{+\infty} f(\zeta)d\zeta \geq \beta\right\}\right\}, \tag{12}$$

where β is a predefined threshold between 0 and 1.

Suppose ζ represents the imprecise load, then $\mathrm{VaR}_{(1-\beta)}$ can be formed as a robust reliability measurement which evaluates the largest energy not served under the predefined confidence level. In addition, SOs can adjust the $(1-\beta)$ value to obtain different tradeoffs between economy and reliability.

Based on the above analysis, the VaR-based reliability measurement is demonstrated as follows:

$$\min \sum_{t\in T} \mathrm{VaR}_{(1-\beta)}\left\{\sum_{j\in J}\xi_j^t - \sum_{i_w\in I_w} p_{i_w}^t - \sum_{i\in I} u_i^t p_i^t - \sum_{i\in I} SR_i^t\right\}. \tag{13}$$

Equation (13) indicates that load shedding will happen if the generation amount and spinning reserve of thermal units and wind turbines cannot satisfy system load. It is difficult to calculate Equation (13) as ξ_j^t and $p_{i_w}^t$ are random variables with different probability density functions. Thus, a Monte Carlo simulation is used to realize the calculation of $\mathrm{VaR}_{1-\beta}$, the detailed steps of which is provided in Section IV.

3.3. Constraints

There are several constraints need to be considered while optimizing the above objectives.

3.3.1. Flexible power demand requirement

In the conventional UC optimization, the total generation at time period t must be equal to the exact forecast load:

$$\sum_{i\in I} u_i^t p_i^t + \sum_{i_w\in I_w} p_{i_w}^t = \sum_{j\in J} \xi_j^t. \tag{14}$$

Considering that the wind power and future load are represented by random variables in this study, the above demand requirement is modified to an inequality constraint as follows:

$$\sum_{i\in I} u_i^t p_i^t + \sum_{i_w\in I_w} p_{i_w}^t \in (\xi_L^t, \xi_U^t). \tag{15}$$

Here, ξ_L^t and ξ_U^t are lower and upper bounds of ξ^t. Equation (15) shows that the total generation amount of thermal units and wind turbines at each period is limited by a predefined interval.

To ensure the realization of Equation (15), the following constraint is used:

$$\sum_{i\in I} u_i^t p_i^{max} + \sum_{i_w\in I_w} p_{i_w}^{rated} \geq \xi_U^t. \tag{16}$$

Equation (16) ensures that the supply capability of thermal units and wind turbines should exceed the upper bound of uncertain load.

3.3.2. Ramp rate limits

The ramp rate constraint can be formed as below:

$$-RR_i \leq u_i^t(p_i^t - p_i^{t-1}) \leq RR_i. \tag{17}$$

Equation (17) indicates that the change of output between two consecutive periods is constrained by the maximal ramp rate.

3.3.3. Spinning reserve limits

The spinning reserve of unit i at time period t is limited by the remained capacity and ramp rate.

$$\begin{aligned} SR_i^t &\leq u_i^t(p_i^{max} - p_i^t), \\ SR_i^t &\leq u_i^t RR_i. \end{aligned} \tag{18}$$

When unit i is off ($u_i^t = 0$), the spinning reserve is at zero. Otherwise, SR_i^t equals the minimum value of the remained capacity and ramp rate.

3.3.4. Generation limits

$$u_i^t p_i^{min} \leq p_i^t \leq u_i^t p_i^{max}. \tag{19}$$

Equation (19) establishes the correlation between the on/off status and output of each unit. When unit i is on ($u_i^t = 1$), the output p_i^t is limited by the minimal and maximal generation amount. Otherwise, p_i^t equals zero.

3.3.5. Transmission line constraint

The power flow of transmission line k at time period t is limited by its transmission capacity:

$$-C_k < \sum_{n\in N}(\sum p_i^t h_{kn} + \sum p_{i_w}^t h_{kn} - \sum \xi_j^{te} h_{kn}) < C_k. \tag{20}$$

Here, the impact of bus n on transmission line k is measured by h_{kn}. The power flow of each line is calculated by the total impact of units, wind turbines and load points in all buses. To ensure system security, the power flow of each line should be constrained by the transmission capacity.

3.3.6. Minimum on/off time

The state of unit i should keep unchanged within the minimum on/off time:

$$\begin{aligned} T_{i,on}^{t} &\geq T_{i,up}, \\ T_{i,off}^{t} &\geq T_{i,dowm}. \end{aligned} \tag{21}$$

3.3.7. χ^2 test limits

As mentioned , the accuracy of NPKDM should be examined by the χ^2 test:

$$\chi^2 \leq \chi^2_{1-\alpha,k-1}. \tag{22}$$

3.4. Mathematical model of data-driven multi-objective unit commitment

Based on the above knowledge, the mathematical model of the proposed DD-MOUC can be constructed as problem (23).

$$\begin{aligned} &\min \sum_{t\in T}\sum_{i\in I}[(x_i^t SU_i + y_i^t SD_i) + G(u_i^t, p_i^t) + Q(u_i^t, p_i^t)] \\ &\min \sum_{t\in T} \mathrm{VaR}_{(1-\beta)}\left\{\sum_{j\in J}\xi_j^t - \sum_{i_w\in I_w} p_{i_w}^t - \sum_{i\in I} u_i^t p_i^t - \sum_{i\in I} SR_i^t\right\} \\ &s.t. \\ &\sum_{i\in I} u_i^t p_i^t + \sum_{i_w\in I_w} p_{i_w}^t \in (\xi_L^t, \xi_U^t) \\ &\sum_{i\in I} u_i^t p_i^{max} + \sum_{i_w\in I_w} p_{i_w}^{rated} \geq \xi_U^t \\ &-RR_i \leq u_i^t(p_i^t - p_i^{t-1}) \leq RR_i \\ &u_i^t p_i^{min} \leq p_i^t \leq u_i^t p_i^{max} \\ &-C_k < \sum_{n\in N}(\sum p_i^t h_{kn} + \sum p_{i_w}^t h_{kn} - \sum \xi_j^{te} h_{kn}) < C_k \\ &T_{i,on}^t \geq T_{i,up} \\ &T_{i,off}^t \geq T_{i,dowm} \\ &\chi^2 \leq \chi^2_{1-\alpha,k-1} \\ &0 < \alpha, \beta < 1. \end{aligned} \tag{23}$$

DD-MOUC aims to find a set of optimal UC solutions which minimize operation costs and maximize system reliability. With the growing of environmental concerns, this model can be easily adjusted to take emission concerns as another objective or as a constraint.

4. Solution Algorithm

In this section, a RL-MOPSO algorithm which improves the existing ones is proposed as the solution.

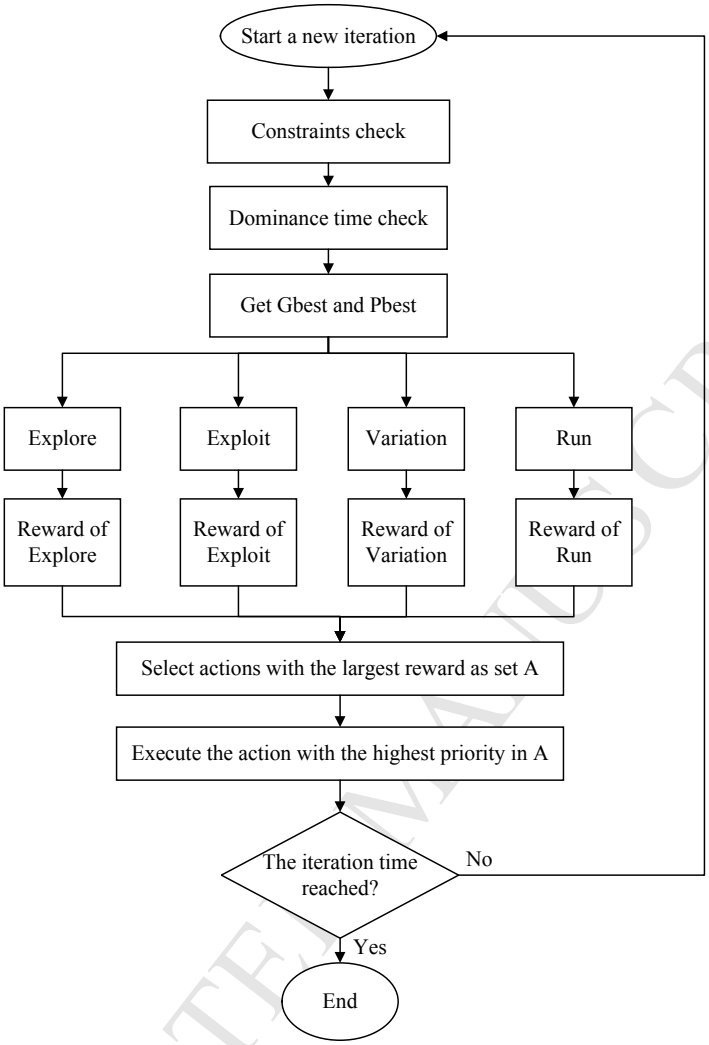


Figure 1: Reinforcement learning-based multi-objective particle swarm optimization algorithm

4.1. Particle swarm optimization algorithm

Particle swarm optimization (PSO) was first proposed by Kennedy in 1995 [30]. In the algorithm, each particle evolves by learning from Gbest and its personal best (Pbest). The velocity and position of particle s are updated as below:

$$v_s \leftarrow w * v_s + c1 * Rand * (Pbest_s - P_s) + c2 * Rand * (Gbest - P_s), \tag{24}$$

$$P_s \leftarrow P_s + v_s, \tag{25}$$

where v_s is the velocity of particle s with an upper and lower bounds of v_{max} and v_{min}, w is the inertia weight, $c1$ and $c2$ are the learning rates, $Rand$ is a random value between 0 and 1, $Pbest_s$ is the best position that particle s has ever found, and $Gbest$ is the global best reflects the best position that all particles have found in the past iterations. It is noted that the parameters change as different actions are taken in RL-MOPSO, which is different to conventional PSO.

4.2. Reinforcement learning-based multi-objective particle swarm optimization algorithm

As an important research field in artificial intelligence, reinforcement learning has been widely used to solve optimization problems [31]. The main concepts in reinforcement learning are an agent, an environment, actions,

states and rewards. By taking different actions and getting rewards, the agent learns an optimal strategy in an unknown environment. The flowchart of RL-MOPSO is given in Figure 1. In RL-MOPSO, particles act in a similar way to the reinforcement learning agents. The environment is defined by the solution space of DD-MOUC. The states represent the position of each particle. Actions are characterized when a particle moves from one state to another. Rewards are decided based on the performance of each particle. The framework of RL-MOPSO is shown in Figure 2.

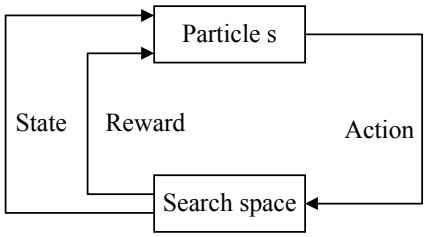


Figure 2: Framework of the proposed RL-MOPSO

4.2.1. Actions

In RL-MOPSO, a particle update its state by taking actions. In this research, four actions with different features are designed as shown in Figure 3.

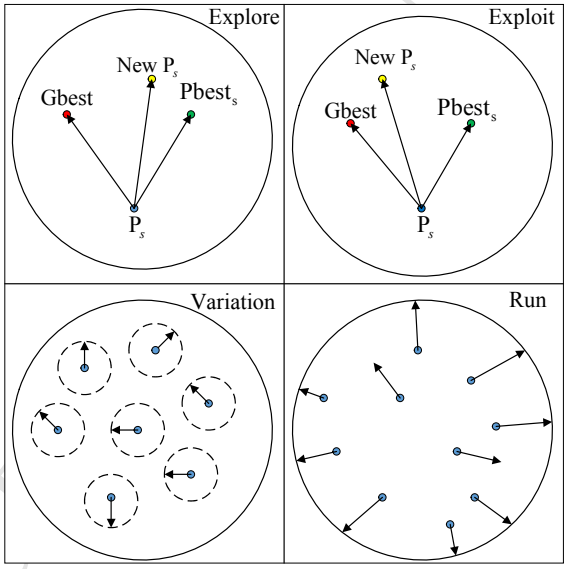


Figure 3: Four actions in RL-MOPSO

Explore: In the early times of iterations, particles are expected to explore in the search space to approximate global optimal solutions. Therefore, action Explore is designed to endow particles with strong capabilities of global exploration. According to the features of PSO parameters, the inertia weight $w_{explore}$ is set as a high value and $c1_{explore}$ is larger than $c2_{explore}$. The velocity of particle s is updated as below:

$$v_s = w_{explore} * v_s + c1_{explore} * Rand * (Pbest_s - P_s) + c2_{explore} * Rand * (Gbest - P_s). \quad (26)$$

Exploit: When particles are converging, they are expected to exploit in the local search space to obtain more accurate optimal solutions. Thus, action Exploit is designed to endow particles with strong capabilities of local exploit. Contrary to the Explore operation, a low $w_{exploit}$ value is utilized and $c1_{exploit}$ is smaller than $c2_{exploit}$. The

velocity of particle s is updated as below:

$$v_s = w_{exploit} * v_s + c1_{exploit} * Rand * (Pbest_s - P_s) + c2_{exploit} * Rand * (Gbest - P_s). \quad (27)$$

Variation: During the evolvement, particles are easy to get trapped into local optimization. To mitigate the above problem, the Variation action is designed to slow down the local convergence procedure by producing random disturbances. In the Variation operation, a random value α is first generated in interval $(0, 1)$, then α is compared with a predefined threshold α_L. When $\alpha > \alpha_L$, the u_i^t, p_i^t and h of particle s are updated as follows:

$$u_i^t = 1 - u_i^t, \quad (28)$$

$$p_i^t = p_i^t + random(-1, 1) * p_L, \quad (29)$$

$$h = h + random(-1, 1) * h_L, \quad (30)$$

where p_L and h_L are a predefined maximal disturbance value of the power generation and bandwidth.

Run: To further mitigate the local convergence problem, the Run action is designed to help the particles run out of local optimization and restart in a new position. If the archive of the optimal solutions cannot be updated for a predefined iterations, the particles in RL-MOPSO are considered to be trapped in the local optimization. Then the speed of particle s is modified by the escape speed v_{es}. If v_{es} is set as a sufficiently large value, the particle will be able to run out of the local optimization. The speed and position of particle s are updated as follows:

$$v_s = v_s + v_{es}, \quad (31)$$

$$P_s = P_s + v_s. \quad (32)$$

4.2.2. Reward function

Particles in RL-MOPSO interact with the environment by getting rewards. For particle s, positive rewards are given to well-preformed actions and negative rewards are given to actions with poor performance. Thus, the following strategy is proposed to determine the reward after executing an action:

Scenario I: If the current solution is dominated by the new one after action a was executed, the reward to a is set to 1.

Scenario II: If the current solution dominates the new one after executing action a, the reward to a is set to -1.

Scenario III: If the current solution and the new solution cannot dominate each other after action a was executed, the reward to a is set to 0.

4.2.3. Action selection strategy

As mentioned before, actions determine the position update of each particle, which have an significant influence on the evolvement of particles. Therefore, it is important to select a proper action. In this research, actions are selected according to the rewards they received. Considering that different actions may have the same reward value, different priorities are set to the actions as below:

Explore>Exploit>Variation>Run

For particle s, the following strategy is utilized to select a proper action.

Algorithm 1 Action selection strategy

Input: Particle s
Output: The proper action

set threshold L;
for particle s **do**
  execute the four actions;
  record the reward of each action;
  if the rewards of all actions are -1 **then**
    Calculate the iterations that the archive have not been updated as l;
    if $l < L$ **then**
      return Explore.
    else
      return Run.
    end if
  else
    Among the four actions, select the actions with the highest reward as set A;
    return the action with the highest priority in set A.
  end if
end for

4.3. Detailed steps of the proposed algorithm

Based on the above improvements, the detailed steps of RL-MOPSO is summarized as follows:

[Step 1]. Particle initialization: The particles are first initialized.

Step 1.1. Each particle contains a real value h which represents the bandwidth of NPKDM. The initial value of h is generated randomly.

Step 1.2. Each particle contains three matrices: a $T \times I$ binary matrix (BM) which denotes the on/off state of units over the scheduling horizon, a $T \times I$ real value matrix (RM) which denotes the corresponding power generation amount and a $T \times I_w$ real value matrix (WM) which denotes the wind power output of each wind turbine, that is:

$$\left[\begin{array}{cccc|cccc|ccc} u_1^1 & u_2^1 & \cdots & u_I^1 & p_1^1 & p_2^1 & \cdots & p_I^1 & w_1^1 & w_2^1 & \cdots & w_{I_w}^1 \\ u_1^2 & u_2^2 & \cdots & u_I^2 & p_1^2 & p_2^2 & \cdots & p_I^2 & w_1^2 & w_2^2 & \cdots & w_{I_w}^2 \\ \cdots & \cdots & u_i^t & \cdots & \cdots & \cdots & p_i^t & \cdots & \cdots & \cdots & w_i^t & \cdots \\ u_1^T & u_2^T & \cdots & u_I^T & p_1^T & p_2^T & \cdots & p_I^T & w_1^T & w_2^T & \cdots & w_{I_w}^T \end{array}\right] \tag{33}$$

$$BM \qquad\qquad RM \qquad\qquad WM$$

The initial value of u_i^t, p_i^t and w_i^t are generated randomly, where u_i^t equals 0 or 1, p_i^t and w_i^t are limited by the generation constraints.

[Step 2]. χ^2 test: The bandwidth value initialized above is revised to satisfy the χ^2 test constraint.

Step 2.1. Calculate the critical value $\chi^2_{1-\alpha,k-1}$ according to the χ^2 distribution with freedom degree $k-1$.

Step 2.2. Monte Carlo simulation: Suppose that $Z_{i_w}^t$ and Z_j^t indicate the data samples of $p_{i_w}^t$ and ξ_j^t. The maximal iteration times is set as M and ψ^t represents $\sum_{j\in J}\xi_j^t - \sum_{i_w\in I_w} p_{i_w}^t$.

Step 2.2.1. Set $m = 1$.

Step 2.2.2. Generate $Z_{i_w}^t$ and Z_j^t randomly for each wind turbine and load point.

Step 2.2.3. Calculate the mth data sample of ψ^t as:

$$\psi_m^t = \sum_{j\in J} Z_j^t - \sum_{i\in I} Z_i^t. \tag{34}$$

Step 2.2.4. If $m < M$, $m = m + 1$, then go to Step 2.2.2.

Step 2.2.5. Calculate the probability density function of ψ^t by NPKDM.

Step 2.2.6. Finally, $\mathrm{VaR}_{1-\beta}$ is calculated based on Equation (12).

Step 2.3. Calculate the value of χ^2 according to Equation (22).

Step 2.4. If $\chi^2 \geq \chi^2_{1-\alpha,k-1}$, generate a new bandwidth value randomly and then go to Step 2.2. Otherwise, stop the algorithm.

[Step 3]. Priority list: Make the priority list (PL) of all the units according to the ascending order of the average fuel cost ($/MW) when the units are at their maximal output. The priority coefficient of unit i is calculated as follows:

$$\lambda_i = \frac{G(p_i^{max})}{p_i^{max}} = \frac{a_i}{p_i^{max}} + b_i + c_i p_i^{max}. \tag{35}$$

[Step 4]. BM adjustment: The BM initialized above is revised to satisfy constraints (16) and (21).

Step 4.1. The total capacity of all committed units at time period t is examined as:

$$\triangle p^t = \sum_{i=1}^{I} u_i^t p_i^{max} + \sum_{i_w=1}^{I_w} p_{i_w}^{rated} - \xi_U^t. \tag{36}$$

If $\triangle p^t < 0$, an offline unit with the lowest priority coefficient is committed until $\triangle p^t \geq 0$.

Step 4.2. A heuristic method is utilized to remove any violation on the minimum on/off time constraints.

Step 4.2.1. For each unit i, compute its on/off time at each time period:

$$T_{i,on}^t = \begin{cases} T_{i,on}^{t-1} + 1 & if\ u_i^t = 1 \\ 0 & if\ u_i^t = 0, \end{cases} \tag{37}$$

$$T_{i,off}^t = \begin{cases} T_{i,off}^{t-1} + 1 & if\ u_i^t = 0 \\ 0 & if\ u_i^t = 1. \end{cases} \tag{38}$$

Step 4.2.2. Set t=1 (From period 1).

Step 4.2.3. Set i=1 (From the first unit).

Step 4.2.4. If $u_i^t = 0$, $u_i^{t-1} = 1$, and $T_{i,on}^{t-1} < T_{i,up}$, u_i^t is set as 1.

Step 4.2.5. If $u_i^t = 0$, $u_i^{t-1} = 1$, $t + T_{i,down} - 1 \leq T$, and $T_{i,off}^{t+T_{i,down}-1} < T_{i,down}$, u_i^t is set as 1.

Step 4.2.6. If $u_i^t = 0$, $u_i^{t-1} = 1$, $t + T_{i,down} - 1 > T$ and $\sum_{y=t}^{T} u_i^t > 0$, y is any integer, u_i^t is set as 1.

Step 4.2.7. Use Equations (37) and (38) to update $T_{i,on}^t$ and $T_{i,off}^t$.

Step 4.2.8. If $i < I$, $i = i + 1$, then go to Step 4.2.4.

Step 4.2.9. If $t < T$, $t = t + 1$, then go to Step 4.2.3. Otherwise, stop the algorithm.

Step 4.3. The above steps may result in redundant generation capacity by committing excessive off-line units. The following method is used to check whether some units can be de-committed without against constraints (16) and (21) based on the descending order of priority list (DPL).

Step 4.3.1. Set $t = 1$.

Step 4.3.2. Make the DPL of the committed units based on the priority coefficient. Select the first unit in DPL as D_1.

Step 4.3.3. Compute the excessive capacity as:

$$\triangle p^t = \sum_{i=1}^{I} u_i^t p_i^{max} + \sum_{i_w=1}^{I_w} p_{i_w}^{rated} - \xi_U^t. \tag{39}$$

Step 4.3.4. If $\triangle p^t < p_{D_1}^{max}$, D_1 is deleted from DPL.

If $\triangle p^t \geq p_{D_1}^{max}$ and decommit unit D_1 will violate the minimum on/off time constraints, D_1 is deleted from DPL.

If $\triangle p^t < p_{D_1}^{max}$ and decommit unit D_1 will not go against the minimum on/off time constraints, decommit D_1, remove D_1 from DPL and update the on/off periods of all the units.

Step 4.3.5. If DPL is not empty, go to Step 4.3.4.

Step 4.3.6. If $t < T$, $t = t + 1$, go to Step 4.3.3. Otherwise, stop the algorithm and record the BM matrix.

[Step 5]. RM adjustment: The RM matrix is generated randomly and should be revised to fulfill the flexible power demand balance.

Step 5.1. First, the corresponding elements of BM and RM are multiplied, then combine the results with WM to obtain a comprehensive matrix (CM).

$$CM = \left[\begin{array}{ccc|ccc} u_1^1 p_1^1 & \cdots & u_I^1 p_I^1 & w_1^1 & \cdots & w_{I_w}^1 \\ u_1^2 p_1^2 & \cdots & u_I^2 p_I^2 & w_1^2 & \cdots & w_{I_w}^1 \\ \cdots & u_i^t p_i^t & \cdots & \cdots & w_{i_w}^t & \cdots \\ u_1^T p_1^T & \cdots & u_I^T p_I^T & w_1^T & \cdots & w_{I_w}^T \end{array}\right] \tag{40}$$

Step 5.2. Set $t = 1$.

Step 5.3. Calculate the summation of each line in CM, if $\xi_L^t \le \sum_{i\in I, i_w\in I_w}(u_i^t p_i^t + w_{i_w}^t) \le \xi_U^t$, go to Step 5.4. Otherwise, if $\sum_{i\in I, i_w\in I_w}(u_i^t p_i^t + w_{i_w}^t) \le \xi_L^t$, the following approach is applied to revise the above matrix.

Step 5.3.1. Calculate:

$$\triangle p^t = \xi_L^t - \sum_{i\in I, i_w\in I_w}(u_i^t p_i^t + w_{i_w}^t). \tag{41}$$

Step 5.3.2. Select the first element in PL as H_1, if $u_{H_1}^t = 0$, H_1 is removed. Otherwise, calculate

$$\triangle p_{H_1}^t = p_{H_1}^{max} - p_{H_1}^t. \tag{42}$$

Step 5.3.3. If $\triangle p_{H_1}^t \ge \triangle p^t$, then $p_{H_1}^t = p_{H_1}^t + \triangle p^t$, go to Step 5.4. Otherwise, $p_{H_1}^t = p_{H_1}^t + Rand \cdot (p_{H_1}^{max} - p_{H_1}^t)$, where *Rand* is a random value in (0,1). Then H_1 is removed from PL and go to Step 5.3.2.

Similar approach can be used to repair the violation of $\sum_{i=1}^{I} u_i^t p_i^t + w_{i_w}^t \ge \xi_U^t$, in which DPL instead of PL is used.

Step 5.4. If $t < T$, $t = t + 1$, go to Step 5.3. Otherwise, stop the algorithm.

[Step 6]. Ramp rate constraints: The RM matrix is adjusted to satisfy the ramp rate constraints.

Step 6.1. Calculate the output variations of all the units in RM:

$$\triangle p_i^t = u_i^t(p_i^t - p_i^{t-1}). \tag{43}$$

Step 6.2. Set $t = 2$ (the unit initial states are considered when $t = 1$).

Step 6.3. Set $i = 1$.

Step 6.4. If $-RR_i \le \triangle p_i^t \le RR_i$, $i = i + 1$, go to Step 6.5. Otherwise, if $\triangle p_i^t < -RR_i$, calculate

$$\begin{aligned} \triangle D_i &= -RR_i - \triangle p_i^t, \\ p_i^t &= p_i^{t-1} - RR_i. \end{aligned} \tag{44}$$

According to DPL, $\triangle D_i$ is dispatched to specified units subject to the ramp rate and generation constraints. Similar approach can be applied when $\triangle p_i^t > RR_i$, where PL instead of DPL is used.

Step 6.5. If $i \le I$, go to Step 6.4.

Step 6.6. If $t < T$, $t = t + 1$, go to Step 6.3. Otherwise, stop the algorithm.

[Step 7]. Transmission line constraint: The RM matrix is revised to satisfy the transmission line constraint.

Step 7.1. Calculate the power flow in each transmission line as follows:

$$f_k^t = \sum_{n\in N}(\sum p_i^t h_{kn} + \sum p_{i_w}^{rated} h_{kn} - \sum \xi_j^{te} h_{kn}). \tag{45}$$

Step 7.2. Set $t = 1$.

Step 7.3. Set $k = 1$.

Step 7.4. If $-C_k < f_k^t < C_k$, $k = k + 1$, go to Step 7.5. Otherwise, if $f_k^t > C_k$, calculate

$$\triangle C_k = C_k - f_k^t. \tag{46}$$

Step 7.4.1. Make the priority list of all the buses according to the ascending order of h_{kn}.

Step 7.4.2. According to the PL, $\triangle C_k$ is dispatched to specified buses subject to the ramp rate, generation and transmission line constraints. Similar approach can be utilized when $f_k^t < -C_k$.

Step 7.5. If $k < K$, $k = k + 1$, go to Step 7.4.

Step 7.6. If $t < T$, $t = t + 1$, go to Step 7.3. Otherwise, stop the algorithm.

[Step 8]. Fitness values: The obtained matrix CM can be taken as a feasible solution. Then calculate the fitness value of operation costs and reliability according to Equations (8) and (13).

[Step 9]. Iterations.

Step 9.1. First, the proposed action selection strategy is used to determine a proper action for each particle.

Step 9.2. The multi-objective mechanism introduced in [36] is applied to determine the Gbest and Pbest of each iteration, whereafter the velocity, position and bandwidth of each particle are updated. It is noted that BM and RM are two different types of matrices and the bandwidth is a real value. Therefore, three different types of settings are applied to RL-MOPSO parameters such as v_{max} and v_{min}.

[Step 10]. Record the Pareto-optimal solutions after running RL-MOPSO for predefined iterations.

5. Experimental analysis

In this section, a modified IEEE 118 bus test system is employed to exemplify the effectiveness of the proposed model and algorithm. The system is consisted of 118 buses, 54 units, 3 wind farms and 186 transmission lines. The exact knowledge is given in [32] and the shift factors are obtained by MATPOWER [33]. The wind power output data is based on the statistics from National Renewable Energy Laboratory (NREL) [34] and the load data comes from the statistics of Slovakia Eastern Power Company [37]. The emission cost coefficients are assumed to be equal to 3% of the related generation cost coefficients, and the dispatch period is considered as 24 hours. Then all the following experiments were implemented with C++ code on an Intel i7-7700 3.6GHz CPU personal computer, while the parameters of RL-MOPSO are given in Table 1.

Table 1: Parameter setting of RL-MOPSO

Parameter	Value	Parameter	Value	Parameter	Value
Particle number	60	Max velocity (BM/RM/h)	1.5/20/10	$c1_{explore}$ (BM/RM/h)	2.5/2.5/2.5
Iteration times	800	Min velocity (BM/RM/h)	-1.5/-20/-10	$c1_{explore}$ (BM/RM/h)	0.5/0.5/0.5
$c1_{exploit}$ (BM/RM/h)	0.5/0.5/0.5	$c2_{exploit}$ (BM/RM/h)	2.5/2.5/2.5	$w_{explore}$ (BM/RM/h)	1.5/1.5/1.5
$w_{exploit}$ (BM/RM/h)	0.5/0.5/0.5	$1-\beta$	0.95	System lead time	4
Threshold L	10	Threshold α_L	0.9	Max disturbance (h_L/p_L)	10/10

5.1. Performance of the prosed model and algorithm

First, the performance of DD-MOUC and RL-MOPSO is illustrated by solving the above UC problem. The final Pareto-optimal solutions are shown in Figure 4. In this study, the system reliability is evaluated by VaR and a smaller VaR represents a higher system reliability. And, VaR with negative values means that the supply system is capable of satisfying the demand with its remain capabilities while accidents happen.

Figure 4 shows that, for the Pareto-optimal solutions, the smaller VaR is, the higher system cost will be. This phenomenon is consistent with the reality: an SO should increase the total capability of the supply system to mitigate the negative influences of hybrid uncertainties. Thus, some units with high average costs will be turned on to ensure sufficient generation capacity when accidents happen. Meanwhile, the generation amount of some units will be decimated to provide more spinning reserve. As a result, an extra expense on the start up of more units and a larger amount of spinning reserve will be paid to improve system reliability. According to the above analysis, it can be concluded that the conflict between reliability and cost is reflected in the proposed model.

Then four different β values are set to exemplify the impact of confidence levels on the obtained UC solutions. Figure 5 shows that confidence level has an significant impact on the experimental results as different confidence levels result in different Pareto-optimal fronts. The reason is that different extreme scenarios are included in the model when confidence levels change. It is also shown in Figure 5 that the Pareto-optimal front with a smaller $1-\beta$ value dominates that with a higher one. This phenomenon is reasonable since a higher confidence level means the consideration of more extreme scenarios. When a higher confidence level is set, the reliability for a certain UC solution will decrease

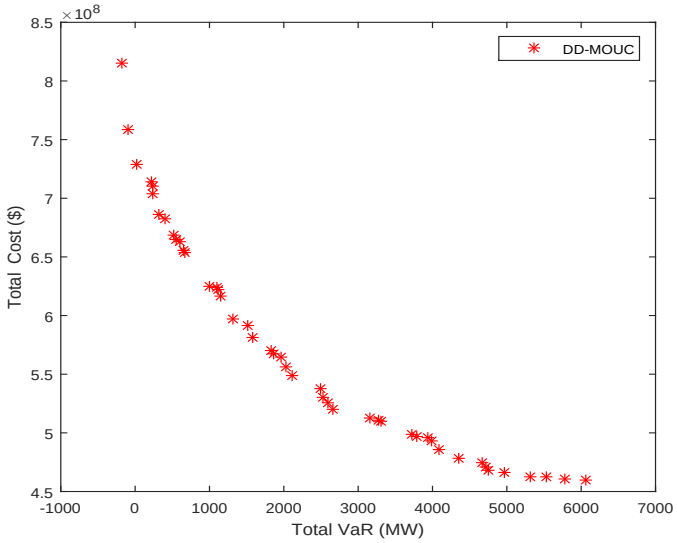


Figure 4: Performance of the proposed DD-MOUC model

as more serious accidents are involved in the model. Based on the above analysis, a smaller β value indicates a more rigorous control on system reliability. And, SOs can adjust the system conservativeness by setting different confidence levels of VaR.

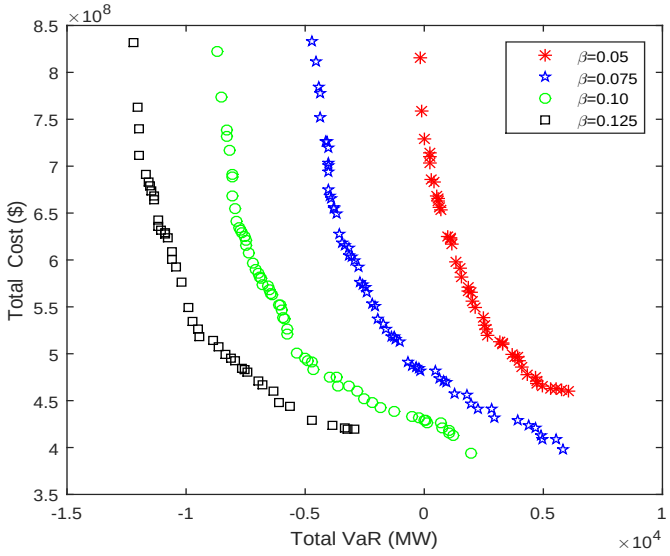


Figure 5: Pareto-optimal fronts of the proposed DD-MOUC model with different confidence levels

5.2. Effect of the proposed bandwidth selection strategy

In this portion, the performance of the proposed bandwidth selection strategy is discussed by comparing with two existing strategies, as follows:

Strategy 1: The bandwidth is decided by the dfittool toolbox in MATLAB [35].

Strategy 2: The bandwidth is determined by the method proposed in [12].

Figure 6 shows that the final Pareto-optimal front with the proposed strategy dominates those with the other two strategies. The reason is that, the correlation between uncertainty representation and UC decision is considered in this study. Specially, the bandwidth optimization is performed by considering the economy and reliability aspects of UC in the proposed bandwidth selection strategy. During the evolvement of particles, bandwidths with lower costs and higher reliability solutions are assigned to a larger fitness value. Therefore, the hybrid distribution of wind power and future load is properly adjusted to get more economical and reliable UC solutions. By contrast, the two existing strategies decided the optimal bandwidth before UC optimization by maximizing estimating accuracy. Thus, the correlation between bandwidth selection and UC optimization is ignored.

Based on the above analysis, it can be concluded that the correlation between uncertainty representation and UC optimization is reflected in this study, and the proposed bandwidth selection strategy can improve UC scheduling from the aspects of economy and reliability.

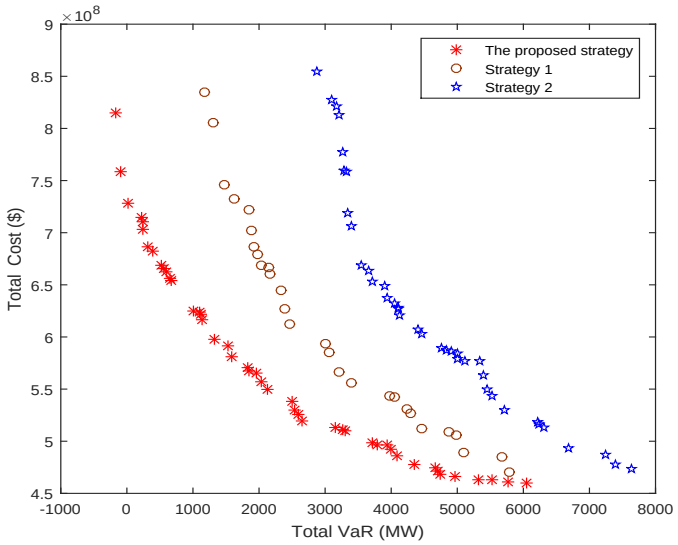


Figure 6: Comparison of Pareto-optimal fronts with different strategies

5.3. Impact of transmission line constraint

In this portion, the impact of TLC on the UC optimization is discussed. It is obviously shown in Figure 7 that the Pareto-optimal front of DD-MOUC with TLC is dominated by the one without considering TLC. The reason is that, units with low average costs are committed to generate as much as possible when TLC is not considered, which may cause transmission lines congestion. Therefore, when TLC is considered, the generation amounts of some cheap units will be reduced and some expensive units will be turned on to satisfy system load, which increases operation costs. According to the experimental result, it can be concluded that the proposed DD-MOUC is more practical than the models without considering TLC.

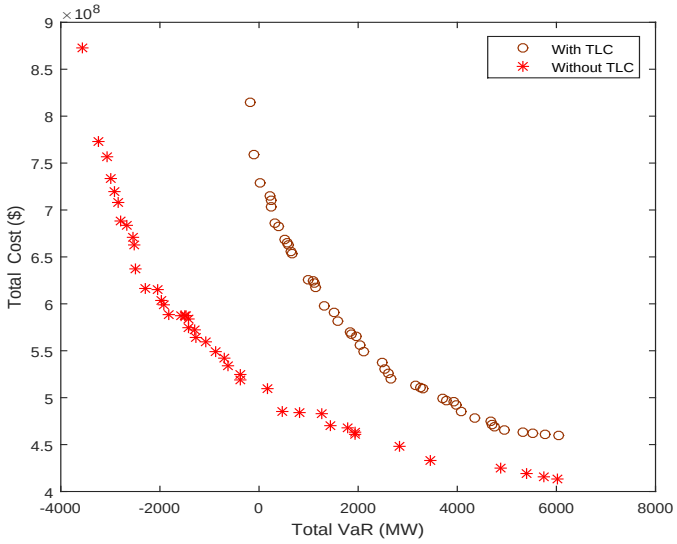


Figure 7: Different Pareto-optimal fronts whether TLC is considered

5.4. *Effect of Algorithm*

In this portion, the effectiveness of the proposed RL-MOPSO is exemplified by comparing with existing DT-MOPSO [36] and TV-MOPSO [26]. The results are given in Figure 8. It is shown that the Pareto-optimal solutions obtained by TV-MOPSO are with low cost and high VaR, which illustrates that the particles in TV-MOPSO are trapped into local optima. And, the Pareto-optimal front obtained by RL-MOPSO dominates those of DT-MOPSO. Therefore, it can be concluded that RL-MOPSO has better global exploration capability than TV-MOPSO and stronger local exploit capability than DT-MOPSO.

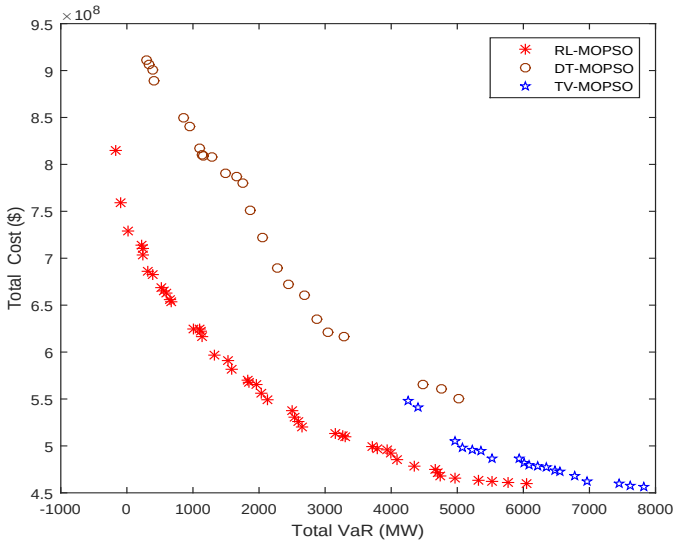


Figure 8: Comparison of Pareto-optimal fronts obtained by different algorithms

However, it is difficult to evaluate the solution set as there are different metrics for an algorithm such as diversity and convergence. To quantify the different performance of each algorithm, the popular-used metric hypervolume [38]

is utilized as a measurement of the Pareto-optimal sets. To obtain the above metric, a point outside of the solution set is first set as the reference point among the objective space. Then the hypercube volume of the reference and solutions is calculated as hypervolume. And, the hypervolumes of the Pareto-optimal solutions obtained by RL-MOPSO, DT-MOPSO and TV-MOPSO are 2.44×10^{12}, 1.68×10^{12} and 0.943×10^{12} while the reference point is set as $(8 \times 10^3, 9.5 \times 10^8)$, which proves that RL-MOPSO obtains the best performance among the three algorithms.

The above results illustrate that the reinforcement learning mechanism improves the performance of MOPSO. At the beginning, particles search in the solution space to approximate the global optimization. After a number of iterations, they can carry out local exploit to get more accurate optimal solutions. To slow down the local convergence procedure, they are designed to search in the neighborhood space with a predefined probability $1 - \alpha_L$. When convergence happens, they can run out of the local optimization and restart in a new position. More importantly, the optimal action of each particle is decided by the feedback from the environment. By contrast, the existing two algorithms cannot provide effective strategies to mitigate the local convergence problem.

According to the above facts, it can be concluded that the overall performance of the proposed RL-MOPSO is better than existing ones, and it can be used to realize a convergent and diversified Pareto-optimal set.

5.5. Comparisons with stochastic and robust unit commitment

Finally, the effectiveness of the VaR-based reliability measurement (β is set as 0.075) is proved by comparing with existing approaches.

Figure 9 shows that the Pareto-optimal front of SUC dominates that of the DD-MOUC. However, these solutions ignore system robustness since SUC optimizes the expected performance of UC solutions under manifold scenarios. Thus, the above solutions may not perform well when scenarios with high load burden and low wind generation happen. The proposed DD-MOUC evaluates the system reliability by the largest energy not served under a predefined confidence level. As UC schedules are optimized under a predefined worst-case scenario, the obtained solutions are robust when accidents happen. Therefore, it can be concluded that DD-MOUC can obtain more robust solutions than SUC.

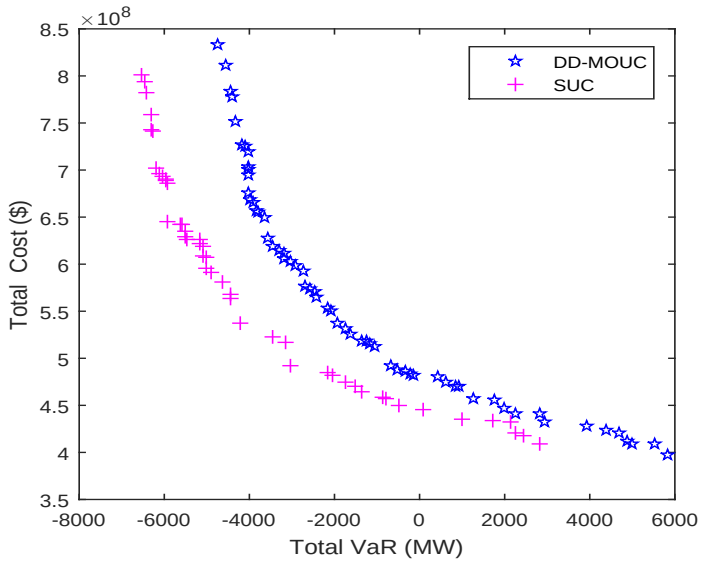


Figure 9: Pareto-optimal fronts of DD-MOUC and SUC

Figure 10 illustrates the difference between the solutions of DD-MOUC and RUC, where the uncertainty representation is degenerated from random variables to interval values in RUC. Obviously, the Pareto-optimal solutions of DD-MOUC are better than those of RUC (higher system reliability with the same cost, and vice versa). The main

reason is that uncertainties in DD-MOUC are represented by probability density functions which provide more information than interval values. Therefore, the solutions obtained by DD-MOUC are less conservative than those of RUC.

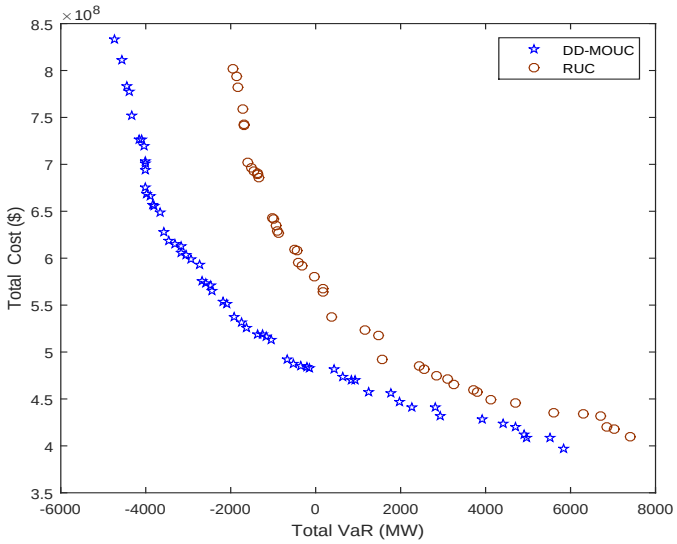


Figure 10: Pareto-optimal fronts of DD-MOUC and RUC

6. Conclusion and future work

In this study, a data-driven approach for the unit commitment optimization is proposed for system operators to handle uncertainties that exist in both demand and supply sides. Considering the difficulty of uncertainty representation, a non-parameter kernel density method is performed to represent the wind power and future load uncertainties, which makes no assumption on the distribution of uncertainties. Then a novel bandwidth selection strategy for the above method is proposed by capturing the correlation between uncertainty representation and unit commitment optimization. To solve the complicated model, a reinforcement learning-based multi-objective particle swarm optimization algorithm is proposed, where Monte Carlo simulation is applied to realize the calculation among different distributions. Finally, a modified IEEE standard test system is applied to justify the effectiveness of this research. Experimental results illustrate that the proposed model is able to achieve different trade-offs between reliability and economy by adjusting the confidence level. Besides, the strategy comparisons prove that the proposed bandwidth selection strategy serves for unit commitment optimization effectively. Furthermore, the proposed model improves the robustness of stochastic unit commitment and mitigates the conservativeness of robust unit commitment. Moreover, the reinforcement learning-based multi-objective particle swarm algorithm is proved to be able to get diversified and convergent Pareto-optimal front. In general, the proposed model and algorithm are useful and meaningful for unit commitment optimization.

Nevertheless, there are still much room for future discussion and improvement. From the modeling perspective, the integration of storage systems will be considered. From the algorithm perspective, the parallel cell coordinate system to evaluate the environment can be investigated. Our future work will be focused on the above points.

7. Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 61603176), the Natural Science Foundation of Jiangsu Province (Grant No. BK20160632), and the Fundamental Research Funds for the Central Universities.

Reference

[1] Zhao C, Wang J, Watson JP, Guan Y. Multi-stage robust unit commitment considering wind and demand response uncertainties. IEEE Trans Power Syst 2013;28(3):2708-17.
[2] Knyazkin V, Canizares CA, Soder LH. On the parameter estimation and modeling of aggregate power system loads. IEEE Trans Power Syst 2004;19(2):1023-31.
[3] Mota LTM, Mota AA. Load modeling at electric power distribution substations using dynamic load parameters estimation. Int J Elec Power 2004;26(10):805-11.
[4] Seguro JV, Lambert TW. Modern estimation of the parameters of the weibull wind speed distribution for wind energy analysis. J Wind Eng Aerod 2000;85(1):75-84.
[5] Dorvlo ASS. Estimating wind speed distribution. Energ Convers Manage 2002;43(17):2311-18.
[6] Trivedi A, Srinivasan D, Sharma D, Singh C. Evolutionary Multi-Objective Day-Ahead Thermal Generation Scheduling in Uncertain Environment. IEEE Trans Power Syst 2013;28(2):1345-54.
[7] Bagheri A, Wang J, Zhao C. Data-driven stochastic transmission expansion planning. IEEE Trans Power Syst 2017;32(5):3461-70.
[8] Weng Y, Negi R, Faloutsos C, Ilić MD. Robust data-driven state estimation for smart grid. IEEE Trans Smart Grid 2017;8(4):1956-67.
[9] Zhao C, Guan Y. Data-Driven Stochastic Unit Commitment for Integrating Wind Generation. IEEE Trans Power Syst 2016;31(4):2587-96.
[10] Blanco I, Morales JM. An Efficient Robust Solution to the Two-Stage Stochastic Unit Commitment Problem. IEEE Trans Power Syst 2017;32(6):4477-88.
[11] Xu X, Yan Z, Xu S. Estimating wind speed probability distribution by diffusion-based kernel density method. Elect Power Syst Res 2015;121:28-37.
[12] Qin Z, Li W, Xiong W. Estimating wind speed probability distribution using kernel density method. Elect Power Syst Res 2011;81(12):2139-46.
[13] Zheng QP, Wang J, Liu AL. Stochastic Optimization for Unit Commitment-A Review. IEEE Trans Power Syst 2015;30(4):1913-24.
[14] Aghaei J, Nikoobakht A, Siano P. Exploring the reliability effects on the short term AC security-constrained unit commitment: A stochastic evaluation. Energy 2016;114:1016-32.
[15] Quan H, Srinivasan D, Khosravi A. Integration of renewable generation uncertainties into stochastic unit commitment considering reserve and risk: A comparative study. Energy 2016;103:735-45.
[16] Schulze T, Mckinnon K. The value of stochastic programming in day-ahead and intra-day generation unit commitment. Energy 2016;101:592-605.
[17] Fattahi S, Ashraphijuo M, Lavaei J. Conic Relaxation of the Unit Commitment Problem. Energy 2017;134:1079-95.
[18] An Y, Zeng B. Exploring the Modeling Capacity of Two-Stage Robust Optimization: Variants of Robust Unit Commitment Model. IEEE Trans Power Syst 2015;30(1):109-22.
[19] Lee C, Liu C, Mehrotra S, Shahidehpour M. Modeling Transmission Line Constraints in Two-Stage Robust Unit Commitment Problem. IEEE Trans Power Syst 2014;29(3):1221-31.
[20] Jiang R, Wang J, Guan Y. Robust Unit Commitment With Wind Power and Pumped Storage Hydro. IEEE Trans Power Syst 2012;27(2):800-10.
[21] Jabr RA. Robust self-scheduling under price uncertainty using conditional value-at-risk. IEEE Trans Power Syst 2005;20(4):1852-58.
[22] Zhu Q, Lin Q, Chen W, Wong WC, Coello Coello CA, Li J. An external archive-guided multiobjective particle swarm optimization algorithm. IEEE Trans Cybernetics 2017;47(9):2794-808.
[23] Hu W, Yen GG. Adaptive multiobjective particle swarm optimization based on parallel cell coordinate system. IEEE Trans Evolut Comput 2015;19(1):1-18.
[24] Wang B, Li Y, Watada J. Supply Reliability and Generation Cost Analysis Due to Load Forecast Uncertainty in Unit Commitment Problems. IEEE Trans Power Syst 2013;28(3):2242-52.
[25] Heo JS, Lee KY, Garduno-Ramirez R. Multiobjective control of power plants using particle swarm optimization techniques. IEEE Trans Energy Convers 2016;21(2):552-61.
[26] Tripathi PK, Bandyopadhyay S, Pal SK. Multi-objective particle swarm optimization with time variant inertia and acceleration coefficients. Inform Sciences 2007;177(22):5033-49.
[27] Wang B, Wang S, Zhou X, Watada J. Two-Stage Multi-Objective Unit Commitment Optimization Under Hybrid Uncertainties. IEEE Trans Power Syst 2016;31(3):2266-77.
[28] Samma H, Lim CP, Saleh JM. A new reinforcement learning-based memetic particle swarm optimizer. Appl Soft Comput 2016;43:276-97.
[29] Duffie D, Pan J. An overview of value-at-risk. J. Derivarives 1997;4(3):7-49.
[30] Kennedy J, Eberhart R. Particle swarm optimization. IEEE Int Conf Neural Networks 1995;1942-1948.
[31] Li J, Kiumarsi B, Chai T, Lewis FL, Fan J. Off-Policy Reinforcement Learning: Optimal Operational Control for Two-Time-Scale Industrial Processes. IEEE Trans Cybernetics 2017;47(12):4547-58.
[32] Zhao C, Guan Y. Unified Stochastic and Robust Unit Commitment. IEEE Trans Power Syst 2013;28(3):3353-61.
[33] Zimmerman RD, Murillo-Sanchez CE, Thomas RJ. MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education. IEEE Trans Power Syst 2011;26(1):12-19.
[34] Tewari S, Geyer CJ, Mohan N. A statistical model for wind power forecast error and its application to the estimation of penalties in liberalized markets. IEEE Trans Power Syst 2011;26(4):2031-39.
[35] Santhosh D, Srinivas VV. Bivariate frequency analysis of floods using a diffusion based kernel density estimator. Water Resour Res 2013;49(12):8328-43.
[36] Wang B, Wang S, Zhou X, Watada J. Multi-objective unit commitment with wind penetration and emission concerns under stochastic and fuzzy uncertainties. Energy 2016;111:18-31.
[37] Available: http://neuron.tuke.sk/competition2/.

[38] Emmerich M, Beume N, Naujoks B, An EMO Algorithm Using the Hypervolume Measure as Selection Criterion. Int Conf Evolutionary Multi-Criterion Optimization 2005;3410:67-76.

Highlights

Non-parameter kernel density method is used to represent hybrid uncertainties.

A novel bandwidth selection strategy is proposed to adjust uncertainty distribution.

A data-driven multi-objective unit commitment model is proposed.

A reinforcement learning-based particle swarm optimization algorithm is developed.

The model and algorithm are effective to solve unit commitment under uncertainties.