Transfer-based Approximate Dynamic Programming for Rolling Security-constrained Unit Commitment with Uncertainties

Jianquan Zhu, *Senior Member*, *IEEE*, Kai Zeng, Jiajun Chen, Wenmeng Zhao, Wenhao Liu, and Wenkai Zhu

***Abstract*—This paper studies the rolling security-constrained unit commitment (RSCUC) problem with AC power flow and uncertainties. For this NP-hard problem, it is modeled as a Markov decision process, which is then solved by a transfer-based approximate dynamic programming (TADP) algorithm proposed in this paper. Different from traditional approximate dynamic programming (ADP) algorithms, TADP can obtain the commitment states of most units in advance through a decision transfer technique, thus reducing the action space of TADP significantly. Moreover, compared with traditional ADP algorithms, which require to determine the commitment state of each unit, TADP only needs determine the unit with the smallest on-state probability among all on-state units, thus further reducing the action space. The proposed algorithm can also prevent the iterative update of value functions and the reliance on rolling forecast information, which makes more sense in the rolling decision-making process of RSCUC. Finally, numerical simulations are carried out on a modified IEEE 39-bus system and a real 2778-bus system to demonstrate the effectiveness of the proposed algorithm.**

***Index Terms*—Rolling security-constrained unit commitment, approximate dynamic programming, decision transfer, probability-based decision priority criterion, uncertainty.**

I. INTRODUCTION

The security-constrained unit commitment (SCUC) problem, which is used to determine the optimal operation schedule of units, plays an essential role in the secure and economic operation of power systems. With the rapid development of renewable energy, the day-ahead unit commitment (UC) schedule may become unsuitable for actual application due to the strong uncertainties. In this context, the rolling SCUC (RSCUC), which allows the UC schedule to be flexibly adjusted to accommodate these uncertainties, has received more and more attentions nowadays [1]–[3]. In some independent system operators (ISOs) (e.g., CAISO [4], [5], ERCOT [6], PJM [7], and MISO [8]), RSCUC has been successfully applied to ensure the security and economy of the power systems.

Mathematically, the SCUC problem is a mixed integer nonlinear programming (MINLP) problem. The common solution algorithms can be classified into heuristic algorithms and mathematical algorithms. Heuristic algorithms such as priority list [9] and extend priority list [10] can quickly obtain the UC schedule by establishing the priority list of units. However, these algorithms can only consider the simple operation constraints of power systems. As substitutes, meta heuristic algorithms (e.g., genetic algorithm [11], particle swarm optimization [12], immune algorithm [13], and simulated annealing algorithm [14]) are proposed to take more detailed operation constraints into account. These algorithms iteratively seek the optimal solution through some strategies inspired by nature or human social behaviors. However, although meta heuristic algorithms are accomplished in handling the optimization problems with complex constraints, their computational burden rises exponentially with the scale of problems. This limits their application in the large-scale RSCUC problem, which requires high decision efficiency.

Mathematical optimization algorithms, such as branch-and-bound (BB) [15], [16], Lagrangian relaxation (LR) [17]–[20], Benders decomposition (BD) [21]–[23], and outer approximation (OA) [24], [25], have also been used to solve the SCUC problem. Among them, BB and LR are only applicable to the simplified SCUC problem, in which the AC power flow is replaced by the DC power flow or is neglected directly. BB relaxes the SCUC problem into a sequence of linear pro-

Received: November 12, 2023
Accepted: May 4, 2024
Published Online: September 1, 2024
Jianquan Zhu (corresponding author), Kai Zeng, Jiajun Chen, Wenhao Liu, and Wenkai Zhu are with the School of Electric Power Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: zhujianquan@scut.edu.cn; 1026708117@qq.com; epchenjiajun@mail.scut.edu.cn; 867086614@qq.com; epuzwki@foxmail.com).
Wenmeng Zhao is with the Electric Power Research Institute, China Southern Power Grid, Guangzhou 510700, China (e-mail: zhaowm@csg.cn).
DOI: 10.23919/PCMP.2023.000140

gramming problems along the branch and bound tree, and then iteratively solves them to narrow the gap between the upper and lower bounds of the objective. Similarly, LR decouples the SCUC problem into a number of subproblems, whose solutions are coordinated according to Lagrangian multipliers. Although these algorithms are stable and efficient for the simplified SCUC problem, the neglect of AC power flow may make the decisions unfeasible in applications. Different from BB and LR, BD and OA decouple the original SCUC problem into two subproblems, i.e., the mixed integer linear programming (MILP) subproblem for handling the UC schedule and the nonlinear programming (NLP) subproblem for handling the optimal power flow. Then, the subproblems are solved iteratively to generate a series of non-increasing upper bounds and non-decreasing lower bounds. The difference of these two algorithms lies in the outer approximation process, in which BD and OA employ the dual information and primal information, respectively. However, both BD and OA cannot meet the computing time requirement of RSCUC due to its nonlinearity and nonconvexity.

In recent years, machine learning (ML) has also been applied to the SCUC problem, and the existing algorithms can be divided into two categories. The first category of algorithms employs ML to detect and remove the redundant and inactive constraints, and thus simplify the SCUC problem [26]–[28]. However, this category of algorithms can only reduce the number of constraints rather than the number of decisions. When the system scale is large, it is still difficult for them to handle so many decisions. The second category of algorithms utilizes ML to directly determine the UC schedules according to the relationship between the system states and decisions [29]–[31]. However, they can only provide the commitment states of units, while cannot distinguish whether the commitment states are reliable or unreliable. If those unreliable commitment states are applied, the accuracy of UC schedules will be deteriorated accordingly.

In order to cope with the uncertainties, it is necessary to adjust the day-ahead UC schedule through RSCUC. In [32], the hourly RSCUC problem is performed to adjust the UC schedule according to the rolling updated forecast information of load and wind power. In [33] and [34], the hourly RSCUC problem is further developed into sub-hourly ones to utilize more accurate forecast information. Model predictive control (MPC) is a commonly used optimization framework, which repeatedly solves the RSCUC problem according to the rolling forecast information [35]. In [36] and [37], the MPC-based RSCUC approach has been successfully applied to a simplified 12-bus power system of New York. However, when a large-scale system is considered, the computing time will become unacceptable for RSCUC. Furthermore, MPC deeply relies on the rolling forecast information, and if the rolling forecast information is unreliable or unavailable, MPC will become invalid accordingly.

To facilitate the rolling decision-making, the RSCUC problem can be formulated as a Markov decision process (MDP), which can then be solved successively. The existing algorithms to tackle the MDP problem include reinforcement learning (RL)-based algorithms [38], [39], and approximated dynamic programming (ADP)-based algorithms [40], [41]. RL handles the MDP problem through continuous interactions with the environment, and thus does not rely on the model information. However, it is difficult for RL to satisfy the complex constraints of the problem [42], which makes it unsuitable for RSCUC with strict security constraints. By contrast, ADP deals with the MDP problem through solving Bellman's equation, in which the complex constraints can be taken into sufficient consideration. Therefore, ADP is able to satisfy the strict security constraints of RSCUC. Nevertheless, ADP still faces challenges when dealing with the RSCUC problem. On the one hand, due to the discreteness of RSCUC, ADP needs to search for the UC schedule in a large action space. On the other hand, the approximate value functions require to be updated iteratively [43], [44]. Both features make ADP time-consuming in the rolling decision-making process of RSCUC. To fill these gaps, a transfer-based ADP (TADP) algorithm is proposed in this paper, and the main contributions are listed as follows:

1) A deep neural network (DNN)-based decision transfer (DT) technique is proposed for the first time to the best of our knowledge, to reduce the action space of ADP. Through transferring the decision experience using DNN, the commitment states of most units can be determined in advance, which speeds up the algorithm greatly.

2) A probability-based decision priority (PDP) criterion is also proposed for the first time to further reduce the action space of ADP. Different from traditional ADP algorithms, which require to determine the commitment state of each unit, the proposed algorithm only needs determine the unit with the smallest on-state probability among all on-state units. Consequently, both the action space and computing time can be further reduced.

3) A TADP-based rolling optimization framework is proposed for the RSCUC problem. By training the approximate value functions offline, the UC schedule can be adjusted period by period to adapt to the uncertainties flexibly, while avoiding the iterative update of value functions and the reliance on rolling forecast information, thus making the decision-making process more convenient and efficient.

The remainder of this paper is structured as follows. The RSCUC model description is presented in Section II, while the TADP algorithm is proposed in Section III. In Section IV, the numerical results are presented to illustrate the effectiveness of the proposed TADP algorithm. Finally, conclusions are drawn in Section V.

II. RSCUC PROBLEM FORMULATION

The RSCUC problem is formulated in this section under the MDP framework. State variables, decision variables, exogenous information, transition functions, and objective function are the fundamental components of MDP, which are described below.

A. State Variables

The state variable $\boldsymbol{S}_t$ is used to reflect the current operation condition of the system, given as:

$$\boldsymbol{S}_t=\left\{\boldsymbol{T}_t^{\mathrm{G,on}},\boldsymbol{T}_t^{\mathrm{G,off}},\boldsymbol{I}_{t-\Delta t}^{\mathrm{G}},\boldsymbol{P}_{t-\Delta t}^{\mathrm{G}},\boldsymbol{P}_t^{\mathrm{WT}},\boldsymbol{P}_t^{\mathrm{load}},\boldsymbol{Q}_t^{\mathrm{load}},t\right\},\forall t\in\mathcal{T} \tag{1}$$

where $\mathcal{T}$ is the set of periods; $\boldsymbol{T}_t^{\mathrm{G,on}}$ and $\boldsymbol{T}_t^{\mathrm{G,off}}$ represent the on and off time of units at period t; $\boldsymbol{I}_{t-\Delta t}^{\mathrm{G}}$ represents the commitment states of units at period $(t-\Delta t)$; $\boldsymbol{P}_{t-\Delta t}^{\mathrm{G}}$ represents the active power generations of units at period $(t-\Delta t)$; $\boldsymbol{P}_t^{\mathrm{WT}}$ represents the active power generations of wind turbines at period t; $\boldsymbol{P}_t^{\mathrm{load}}$ and $\boldsymbol{Q}_t^{\mathrm{load}}$ represent the active and reactive loads at period t.

B. Decision Variables

The decision variable $\boldsymbol{a}_t$ is the controllable factor in RSCUC. Since the main target of RSCUC is to determine the commitment states of units, they are selected as the decision variable, i.e.:

$$\boldsymbol{a}_t=\left\{\boldsymbol{I}_t^{\mathrm{G}}\right\} \tag{2}$$

Note that in addition to the commitment states $\boldsymbol{I}_t^{\mathrm{G}}$, the outputs of units also need to be determined in RSCUC. Since these variables can be obtained by solving the optimal power flow problem once the commitment states have been determined, they are not included in the decision variables.

To ensure the security of the system operation, the decision-making should be subjected to the following constraints [45]:

$$V_{i,t}\sum_{j\in\mathcal{N}}V_{j,t}(G_{ij}\cos\delta_{ij,t}+B_{ij}\sin\delta_{ij,t})=P_{i,t}^{\mathrm{G}}+P_{i,t}^{\mathrm{WT}}-P_{i,t}^{\mathrm{load}},\quad\forall i\in\mathcal{N},\forall t\in\mathcal{T} \tag{3}$$

$$V_{i,t}\sum_{j\in\mathcal{N}}V_{j,t}(G_{ij}\sin\delta_{ij,t}-B_{ij}\cos\delta_{ij,t})=Q_{i,t}^{\mathrm{G}}-Q_{i,t}^{\mathrm{load}}\quad\forall i\in\mathcal{N},\forall t\in\mathcal{T} \tag{4}$$

$$V_i^{\min}\leqslant V_{i,t}\leqslant V_i^{\max},\forall i\in\mathcal{N},\forall t\in\mathcal{T} \tag{5}$$

$$P_{ij,t}^{\mathrm{L}}=V_{i,t}^2G_{ij}-V_{i,t}V_{j,t}(G_{ij}\cos\delta_{ij,t}+B_{ij}\sin\delta_{ij,t}),\quad\forall i,j\in\mathcal{N},\forall t\in\mathcal{T} \tag{6}$$

$$Q_{ij,t}^{\mathrm{L}}=-V_{i,t}^2B_{ij}+V_{i,t}V_{j,t}(B_{ij}\cos\delta_{ij,t}-G_{ij}\sin\delta_{ij,t}),\quad\forall i,j\in\mathcal{N},\forall t\in\mathcal{T} \tag{7}$$

$$(P_{ij,t}^{\mathrm{L}})^2+(Q_{ij,t}^{\mathrm{L}})^2\leqslant(S_{ij,\max}^{\mathrm{L}})^2,\forall i,j\in\mathcal{N},\forall t\in\mathcal{T} \tag{8}$$

$$\min\left\{\sum_{i\in\mathcal{G}}I_{i,t}^{\mathrm{G}}P_{i,\max}^{\mathrm{G}}-\sum_{i\in\mathcal{N}}P_{i,t}^{\mathrm{load}},\sum_{i\in\mathcal{G}}R_i^{\mathrm{up}}\right\}\geqslant R_t^{+},\forall t\in\mathcal{T} \tag{9}$$

$$\min\left\{\sum_{i\in\mathcal{N}}P_{i,t}^{\mathrm{load}}-\sum_{i\in\mathcal{G}}I_{i,t}^{\mathrm{G}}P_{i,\min}^{\mathrm{G}},\sum_{i\in\mathcal{G}}R_i^{\mathrm{dn}}\right\}\geqslant R_t^{-},\forall t\in\mathcal{T} \tag{10}$$

$$I_{i,t}^{\mathrm{G}}P_{i,\min}^{\mathrm{G}}\leqslant P_{i,t}^{\mathrm{G}}\leqslant I_{i,t}^{\mathrm{G}}P_{i,\max}^{\mathrm{G}},\forall i\in\mathcal{G},\forall t\in\mathcal{T} \tag{11}$$

$$I_{i,t}^{\mathrm{G}}Q_{i,\min}^{\mathrm{G}}\leqslant Q_{i,t}^{\mathrm{G}}\leqslant I_{i,t}^{\mathrm{G}}Q_{i,\max}^{\mathrm{G}},\forall i\in\mathcal{G},\forall t\in\mathcal{T} \tag{12}$$

$$P_{i,t}^{\mathrm{G}}-P_{i,t-\Delta t}^{\mathrm{G}}\leqslant I_{i,t}^{\mathrm{G}}R_i^{\mathrm{up}}+(I_{i,t}^{\mathrm{G}}-I_{i,t-\Delta t}^{\mathrm{G}})P_{i,\max}^{\mathrm{G}}+(1-I_{i,t}^{\mathrm{G}})P_{i,\min}^{\mathrm{G}},\quad\forall i\in\mathcal{G},\forall t\in\mathcal{T} \tag{13}$$

$$P_{i,t-\Delta t}^{\mathrm{G}}-P_{i,t}^{\mathrm{G}}\leqslant I_{i,t-\Delta t}^{\mathrm{G}}R_i^{\mathrm{dn}}+(I_{i,t-\Delta t}^{\mathrm{G}}-I_{i,t}^{\mathrm{G}})P_{i,\max}^{\mathrm{G}}+(1-I_{i,t-\Delta t}^{\mathrm{G}})P_{i,\min}^{\mathrm{G}},\quad\forall i\in\mathcal{G},\forall t\in\mathcal{T} \tag{14}$$

$$(T_{i,t}^{\mathrm{G,on}}-T_{i,\min}^{\mathrm{G,on}})(I_{i,t-\Delta t}^{\mathrm{G}}-I_{i,t}^{\mathrm{G}})\geqslant 0,\forall i\in\mathcal{G},\forall t\in\mathcal{T} \tag{15}$$

$$(T_{i,t}^{\mathrm{G,off}}-T_{i,\min}^{\mathrm{G,off}})(I_{i,t}^{\mathrm{G}}-I_{i,t-\Delta t}^{\mathrm{G}})\geqslant 0,\forall i\in\mathcal{G},\forall t\in\mathcal{T} \tag{16}$$

where $\mathcal{N}$ is the set of nodes; $\mathcal{G}$ is the set of nodes with units; G_{ij} and B_{ij} are the real and imaginary parts of the ith row and the jth column of the nodal admittance matrix; $V_{i,t}$ is the amplitude of the voltage at node i; $\delta_{i,t}$ is the phase angle of the voltage at node i; $P_{i,t}^{\mathrm{load}}$ and $Q_{i,t}^{\mathrm{load}}$ are the active and reactive loads at node i and period t; $P_{i,t}^{\mathrm{G}}$ and $Q_{i,t}^{\mathrm{G}}$ are the active and reactive power generations of the unit at node i and period t; $P_{i,t}^{\mathrm{WT}}$ is the active power generation of the wind turbine at node i and period t; $V_i^{\min}$ and $V_i^{\max}$ are the lower and upper limits of the voltage magnitude at node i; $P_{ij,t}^{\mathrm{L}}$ and $Q_{ij,t}^{\mathrm{L}}$ are the active and reactive power of the branch between nodes i and j at period t; $S_{ij,\max}^{\mathrm{L}}$ is the apparent power limit of the branch between nodes i and j; R_i^{up} and R_i^{dn} are the ramp-up and ramp-down rate limits of the unit at node i; $I_{i,t}^{\mathrm{G}}\in\{0,1\}$ is the commitment state of the unit at node i and period t; $P_{i,\min}^{\mathrm{G}}$ and $P_{i,\max}^{\mathrm{G}}$ are the lower and upper active power limits of the unit at node i; $Q_{i,\min}^{\mathrm{G}}$ and $Q_{i,\max}^{\mathrm{G}}$ are the lower and upper reactive power limits of the unit at node i; $T_{i,t}^{\mathrm{G,on}}$ and $T_{i,t}^{\mathrm{G,off}}$ are the on and off time of the unit at node i and period t; $T_{i,\min}^{\mathrm{G,on}}$ and $T_{i,\min}^{\mathrm{G,off}}$ are the minimum on and off time limits of the unit at node i.

Equations (3) and (4) are the AC power flow constraints, while (5) is the voltage amplitude constraint. Equations (6)–(8) are the branch power flow constraints, whereas (9) and (10) are the system spinning reserve constraints. Equations (11) and (12) represent the output limits of units, equations (13) and (14) represent the ramp rate limits of units, while (15) and (16) represent the minimum on and off time limits of units.

C. Exogenous Information

The exogenous information $\boldsymbol{W}_t$, which represents the stochastic factor between periods $(t-\Delta t)$ and t, is given as:

$$\boldsymbol{W}_t=\left\{\overline{\boldsymbol{P}}_t^{\mathrm{load}},\overline{\boldsymbol{Q}}_t^{\mathrm{load}},\overline{\boldsymbol{P}}_t^{\mathrm{WT}}\right\} \tag{17}$$

where $\overline{\boldsymbol{P}}_t^{\text{load}}$ is the change between $\boldsymbol{P}_t^{\text{load}}$ and its forecast value $\boldsymbol{P}_{t-\Delta t}^{\text{load}}$; while $\overline{\boldsymbol{Q}}_t^{\text{load}}$ and $\overline{\boldsymbol{P}}_t^{\text{WT}}$ have similar meanings.

D. Transition Functions

With the given decision $\boldsymbol{a}_t$ and exogenous information $\boldsymbol{W}_{t+\Delta t}$, the current state $\boldsymbol{S}_t$ can be mapped to the next state $\boldsymbol{S}_{t+\Delta t}$ through the transition functions as:

$$\boldsymbol{S}_{t+\Delta t} = T(\boldsymbol{S}_t, \boldsymbol{a}_t, \boldsymbol{W}_{t+\Delta t}) \tag{18}$$

Specifically, the transition functions can be described as:

$$P_{i,t+\Delta t}^{\text{load}} = P_{i,t}^{\text{load}} + \overline{P}_{i,t+\Delta t}^{\text{load}}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \tag{19}$$

$$Q_{i,t+\Delta t}^{\text{load}} = Q_{i,t}^{\text{load}} + \overline{Q}_{i,t+\Delta t}^{\text{load}}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \tag{20}$$

$$P_{i,t+\Delta t}^{\text{WT}} = P_{i,t}^{\text{WT}} + \overline{P}_{i,t+\Delta t}^{\text{WT}}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \tag{21}$$

$$T_{i,t+\Delta t}^{\text{G,on}} = I_{i,t}^{\text{G}}(T_{i,t}^{\text{G,on}} + 1), \forall i \in \mathcal{G}, \forall t \in \mathcal{T} \tag{22}$$

$$T_{g,t+\Delta t}^{\text{G,off}} = (1 - I_{i,t}^{\text{G}})(T_{i,t}^{\text{G,off}} + 1), \forall i \in \mathcal{G}, \forall t \in \mathcal{T} \tag{23}$$

where $\overline{P}_{i,t+\Delta t}^{\text{load}}$, $\overline{Q}_{i,t+\Delta t}^{\text{load}}$ and $\overline{P}_{i,t+\Delta t}^{\text{WT}}$ are the components of $\overline{\boldsymbol{P}}_{t+\Delta t}^{\text{load}}$, $\overline{\boldsymbol{Q}}_{t+\Delta t}^{\text{load}}$ and $\overline{\boldsymbol{P}}_{t+\Delta t}^{\text{WT}}$, respectively.

E. Objective Function

The objective of RSCUC is to minimize the expectation of operation cost during the whole horizon, i.e.:

$$\min E\left[\sum_{t\in\mathcal{T}} C_t(\boldsymbol{S}_t, \boldsymbol{a}_t)\right] \tag{24}$$

where $C_t(\boldsymbol{S}_t, \boldsymbol{a}_t)$ is the operation cost of units at period t, which can be calculated by:

$$C_t(\boldsymbol{S}_t, \boldsymbol{a}_t) = \sum_{i\in\mathcal{G}} \begin{pmatrix} I_{i,t}^{\text{G}}(a_i^{\text{G}}(P_{i,t}^{\text{G}})^2 + b_i^{\text{G}} P_{i,t}^{\text{G}} + c_i^{\text{G}})\Delta t + \\ C_i^{\text{SU}} I_{i,t}^{\text{G}}(1 - I_{i,t-\Delta t}^{\text{G}}) + C_i^{\text{SD}} I_{i,t-\Delta t}^{\text{G}}(1 - I_{i,t}^{\text{G}}) \end{pmatrix} \tag{25}$$

where a_i^{G}, b_i^{G} and c_i^{G} are the cost coefficients of the unit at node i; C_i^{SU} and C_i^{SD} are the start-up and shut-down costs of the unit at node i.

III. Proposed TADP Algorithm

In theory, the above RSCUC model can be solved by the traditional ADP algorithm. However, with the increase of the system scale, its computational efficiency drops sharply, and thus is difficult to satisfy the rolling decision-making requirement of RSCUC. To overcome this problem, a TADP algorithm is proposed in this section.

A. Basic ADP Theory

According to Bellman's equation [46], equation (24) can be divided into T nested subproblems as follows:

$$\begin{aligned} V_t(\boldsymbol{S}_t) = \min_{a_t}\left\{E\left[\sum_{\tau=t}^{T} C_\tau(\boldsymbol{S}_\tau, \boldsymbol{a}_\tau)\right]\right\} = \\ \min_{a_t}\left\{C_t(\boldsymbol{S}_t, \boldsymbol{a}_t) + E\left[V_{t+\Delta t}(\boldsymbol{S}_{t+\Delta t})\right]\right\} \end{aligned} \tag{26}$$

where $V_t(\cdot)$ is the value function, which represents the operation cost of the subsequent periods.

According to [47], when the post-decision state is introduced, equation (26) can be rewritten as:

$$V_{t-\Delta t}^a(\boldsymbol{S}_{t-\Delta t}^a) = E\left[\min_{a_t}\left\{C_t(\boldsymbol{S}_{t-\Delta t}^a, \boldsymbol{a}_t, \boldsymbol{W}_t) + V_t^a(\boldsymbol{S}_t^a)\right\}\right] \tag{27}$$

where $\boldsymbol{S}_t^a$ is the post-decision state, which represents the state after $\boldsymbol{a}_t$ but before $\boldsymbol{W}_{t+\Delta t}$; $V_t^a(\boldsymbol{S}_t^a)$ is the value function corresponding to $\boldsymbol{S}_t^a$.

Considering $\boldsymbol{W}_t$ is a high-dimension stochastic vector, it is challenging to directly calculate the expectation in (27). To address this issue, it is approximated through the Monte Carlo (MC) method as:

$$\overline{V}_{t-\Delta t}^a(\boldsymbol{S}_{t-\Delta t}^a) = \frac{1}{N_{\text{s}}}\sum_{n=1}^{N_{\text{s}}}\min_{a_t}\left\{C_t(\boldsymbol{S}_{t-\Delta t}^a, \boldsymbol{a}_t, \boldsymbol{W}_t^n) + \overline{V}_t^a(\boldsymbol{S}_t^a)\right\} \tag{28}$$

where $\overline{V}_t^a(\boldsymbol{S}_t^a)$ is the MC estimate of the value function; N_{s} is the number of sampled scenarios; and $\boldsymbol{W}_t^n$ is the exogenous information in the nth sampled scenario.

Similar to [40], the value table is employed to approximate the value function in (28). According to [48], the iterative update process can be described by:

$$\hat{v}_t^n = \min_{a_t^n}\left\{C_t(\boldsymbol{S}_{t-\Delta t}^{a,n}, \boldsymbol{a}_t, \boldsymbol{W}_t^n) + \overline{V}_t^{a,n-1}(\boldsymbol{S}_t^{a,n})\right\} \tag{29}$$

$$\overline{V}_{t-\Delta t}^{a,n}(\boldsymbol{S}_{t-\Delta t}^{a,n}) = (1-\alpha^n)\overline{V}_{t-\Delta t}^{a,n-1}(\boldsymbol{S}_{t-\Delta t}^{a,n}) + \alpha^n \hat{v}_t^n \tag{30}$$

where $\hat{v}_t^n$ is the MC estimate value with $\boldsymbol{S}_{t-\Delta t}^{a,n}$ and $\boldsymbol{W}_t^n$ in the nth iteration; $\overline{V}_t^{a,n-1}(\cdot)$ is the value function obtained in the (n–1)th iteration; and $\alpha^n \in (0,1)$ is the learning rate of ADP.

When ADP is applied to the large-scale RSCUC problem, it will face two challenges:

1) ADP needs search in a large action space, which grows exponentially with the number of units.

2) ADP needs update the value functions iteratively, which further increases the computational burden.

To deal with these two challenges, three techniques (i.e., DT technique, PDP criterion, and value function transfer (VFT) technique) are proposed below.

B. DT Technique

In power systems, the decision-making always contains certain regularity. For example, for the days with similar system states, the UC schedules are almost identical in most cases. This motivates us to employ the historical experience to determine some units' commitment states in advance, such that the action space of ADP can be reduced accordingly.

Mathematically, the optimization of commitment states can be described as a classification problem, which can be formulated by decision tree [49], support vector machine (SVM) [50], DNN [31], etc. Considering the excellent learning and generalization ability of DNN, it is adopted in this paper.

First, the input and output of DNN are defined as the state variables $\boldsymbol{S}_t$ and the on-state probability of each unit, as shown in (1) and (31) respectively.

$$\boldsymbol{o}_t = \left[p(I_{1,t}=1 \mid \boldsymbol{S}_t;\boldsymbol{\omega}), \cdots, p(I_{N_G,t}=1 \mid \boldsymbol{S}_t;\boldsymbol{\omega}) \right] \tag{31}$$

where $\boldsymbol{o}_t$ represents the output of DNN; $\boldsymbol{\omega}$ represents the weight parameters of DNN; and $p(I_{j,t}=1 \mid \boldsymbol{S}_t;\boldsymbol{\omega})$ represents the posterior on-state probability of unit j at period t with $\boldsymbol{S}_t$; N_{G} represents the number of the units.

Theoretically, the posterior on-state probability in (31) can be formulated as a logistic regression function as:

$$p(I_{j,t}=1 \mid \boldsymbol{S}_t;\boldsymbol{\omega}) = \frac{1}{1+\exp[-h(\boldsymbol{S}_t;\boldsymbol{\omega})]} \tag{32}$$

where $h(\boldsymbol{S}_t;\boldsymbol{\omega})$ can be described by any differentiable functions, and the hidden layers of DNN is used in this paper.

When the dataset $\left\{(\boldsymbol{S}_t^{(i)}, \boldsymbol{I}_t^{(i)})\right\}_{i=1}^{N_{\mathrm{d}}}$ is given, the weight parameter $\boldsymbol{\omega}$ in (32) can be obtained by minimizing the log-likelihood function as:

$$l(\boldsymbol{\omega}) = -\sum_{i=1}^{N_{\mathrm{d}}}\sum_{j=1}^{N_{\mathrm{G}}} \left\{ \begin{array}{l} I_{j,t}^{(i)} \log p(I_{j,t}^{(i)}=1 \mid \boldsymbol{S}_t^{(i)};\boldsymbol{\omega}) + \\ (1-I_{j,t}^{(i)}) \log\left[1-p(I_{j,t}^{(i)}=1 \mid \boldsymbol{S}_t^{(i)};\boldsymbol{\omega})\right] \end{array} \right\} \tag{33}$$

where N_{d} is the number of training data.

Once DNN has been trained, it can be used to calculate the on-state and off-state probabilities of each unit according to (31) and (34), respectively.

$$\left[p(I_{1,t}=0 \mid \boldsymbol{S}_t;\boldsymbol{\omega}), \cdots, p(I_{N_G,t}=0 \mid \boldsymbol{S}_t;\boldsymbol{\omega}) \right] = \boldsymbol{1} - \boldsymbol{o}_t \tag{34}$$

where $p(I_{j,t}=0 \mid \boldsymbol{S}_t;\boldsymbol{\omega})$ is the posterior off-state probability of unit j at period t with $\boldsymbol{S}_t$.

Then, the commitment states of each unit can be determined according to the probabilities as:

$$I_{j,t} = \begin{cases} 1, & \text{if } p(I_{j,t}=1 \mid \boldsymbol{S}_t;\boldsymbol{\omega}) \geqslant \beta \\ 0, & \text{if } p(I_{j,t}=0 \mid \boldsymbol{S}_t;\boldsymbol{\omega}) \geqslant \beta \end{cases} \tag{35}$$

where β is the confidence level.

Through (35), the unreliable commitment states of units can be detected, and then redetermined by ADP. The procedure of the DT technique is outlined in Table I.

TABLE I
ALGORITHM 1: DT TECHNIQUE

Algorithm 1: DT technique
1: Input the operation dataset $\left\{(\boldsymbol{S}_t^{(i)}, \boldsymbol{I}_t^{(i)})\right\}_{i=1}^{N_d}$.
2: Initialize the weight parameter $\boldsymbol{\omega}$ of DNN.
3: Solve (33) to obtain the optimal $\boldsymbol{\omega}$.
4: Output the on-state and off-state probabilities of each unit according to (31) and (34), respectively.
5: Determine commitment states of units according to (35).

C. PDP Criterion

Although some units' commitment states can be determined in advance through the DT technique, the action space of ADP may still be huge for large-scale systems. Taking a 220-unit system as an example, assume that the commitment states of 200 units have been determined by the DT technique, ADP still needs to search 2^{20} decisions at each period. To address this issue, a PDP criterion is proposed.

The PDP criterion is proposed based on the assumption that the unit with a higher on-state probability should have a priority of on state. Here, a four-unit example is given to illustrate its principle in Fig. 1. First, the units are arranged in descending order according to their on-state probabilities, which generates a decision priority sequence $(I_{(1)}, I_{(2)}, I_{(3)}, I_{(4)})$. Then, we can decide which units should be on in the reverse order. For example, if $I_{(4)}$ is set to 1 in Case 1, $I_{(3)}$, $I_{(2)}$ and $I_{(1)}$ should also be set to 1, since they have higher on-state probabilities. Similarly, we can form other four alternative UC schedules corresponding to Cases 2–5, respectively.

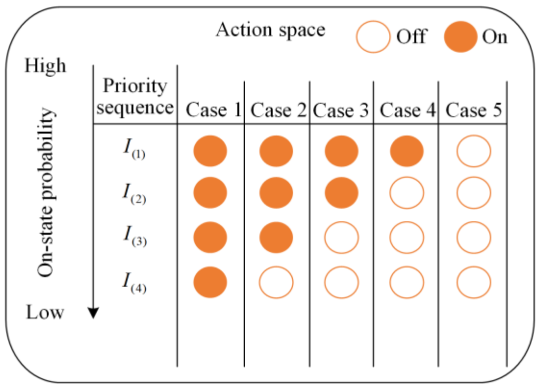

Fig. 1. Principle of the PDP criterion.

It can be found that compared with the traditional UC decision-making mode, which searches for the commitment state of each unit, we now only need to search for the unit with the lowest on-state probability among all on-state units. In this way, the action space can be reduced greatly (e.g., for N_u units, the action space is reduced from 2^{N_u} to N_u+1).

The procedure of the PDP criterion is outlined in Table II.

TABLE II
ALGORITHM 2: PDP CRITERION

Algorithm 2: PDP criterion
1: Input the on-state probabilities of units.
2: Generate the decision priority sequence $(I_{(1)}, \cdots, I_{(N_u)})$ according to their on-state probabilities.
3: **for** $k=0$ to N_u **do**:
4: Set the commitment states of the first k units to be on, and set the states of other units to be off.
5: Add the decision to the action space.
6: **end for**

D. VFT Technique

As mentioned above, ADP needs update the value functions iteratively, which is difficult to satisfy the rolling decision-making requirement of RSCUC. To tackle this problem, a VFT technique is employed in this paper.

The VFT technique includes two stages: the off-line training and on-line application of value functions. In the off-line training stage, value functions are iteratively updated using the historical data. When sufficient historical data are used, the decision experience can be embedded into the value functions [29]. Then, in the on-line application stage, these experienced value functions can be applied directly to get satisfactory solutions, even though the forecast information is not obtained.

E. Procedure of TADP

When the DT technique, PDP criterion and VFT technique are embedded into ADP, a novel ADP algorithm, named TADP, is formed. The procedure is summarized in Table III, and the corresponding flow charts are presented in [51].

TABLE III
ALGORITHM 3: TADP

Algorithm 3: TADP

Stage I: Off-line learning

Input the pre-trained DNN and the history scenario set $\mathcal{W}$; initialize the value functions.
for $n=1$ to $N_{\mathcal{W}}$ **do**:
  Extract the nth scenario.
  for $t=1$ to T **do**:
    Generate the action space $\mathcal{A}_t$ by algorithm 1 and algorithm 2.
    for $\boldsymbol{a}_t$ in $\mathcal{A}_t$
      Solve the optimal power problem by IPOPT, and then obtain $C_t(\boldsymbol{S}_t, \boldsymbol{a}_t)$.
    end for
    Obtain the optimal decision $\boldsymbol{a}_t^*$ according to (29), and update the value functions according to (30).
    Obtain $\boldsymbol{S}_{t+1}$ according to (19)–(23).
  end for
end for

Stage II: On-line application

Input the well-trained DNN and value functions.
for $t=1$ to T **do**:
  Receive the exogenous information $\boldsymbol{W}_t$.
  Generate the action space $\mathcal{A}_t$ by Algorithm 1 and Algorithm 2.
  for $\boldsymbol{a}_t$ in $\mathcal{A}_t$
    Solve the optimal power problem by IPOPT, and then obtain $C_t(\boldsymbol{S}_t, \boldsymbol{a}_t)$.
  end for
  Obtain the optimal decision $\boldsymbol{a}_t^*$ according to (29).
  Obtain $\boldsymbol{S}_{t+1}$ according to (19)–(23).
end for

IV. NUMERICAL ANALYSIS

In this section, the effectiveness of the proposed algorithm is evaluated on a modified IEEE 39-bus system and a real 2778-bus system. The optimization horizon and time interval are set to 24 hours and 1 hour, respectively. All case studies are run using Matlab and GAMS on a computer with i7-9700 CPU@3.0GHz and 16 GB RAM. The deep neural network is trained using Matlab.

A. Modified IEEE 39-bus System

The modified IEEE-39 system includes 10 thermal units, 2 WTs and 46 branches. The WTs are located at buses 20 and 39. The forecast information of WTs and loads is presented in Fig. 2, and the other parameters can be found in [52].

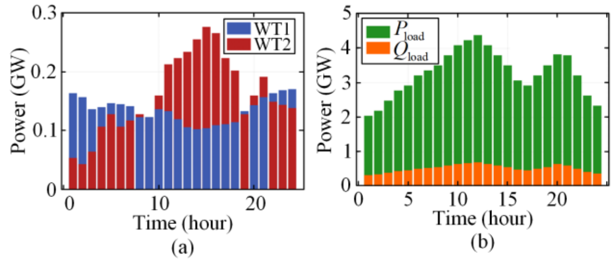

Fig. 2. Forecast information. (a) Outputs of WTs. (b) Loads.

1) Test on the DT Technique

In order to obtain the training data of DNN, similar to [53], [54], it assumes that the forecast errors of the active load, reactive load and wind power all obey Gaussian distribution, and the variance is set to 10% of the forecast value. Then, 1500 scenarios (i.e., 1000 training scenarios and 500 testing scenarios) are generated using MC, as shown in Fig. 3. Each scenario represents the observations of uncertainties for a day, and the solutions of SCUC problem obtained by the MINLP solver (i.e., SBB in this paper) in each scenario are regarded as the operation data for training and testing. The maximal iteration number of DNN training is set to 320. The training process of DNN is presented in Fig. 4, and the identification results are shown in Table IV.

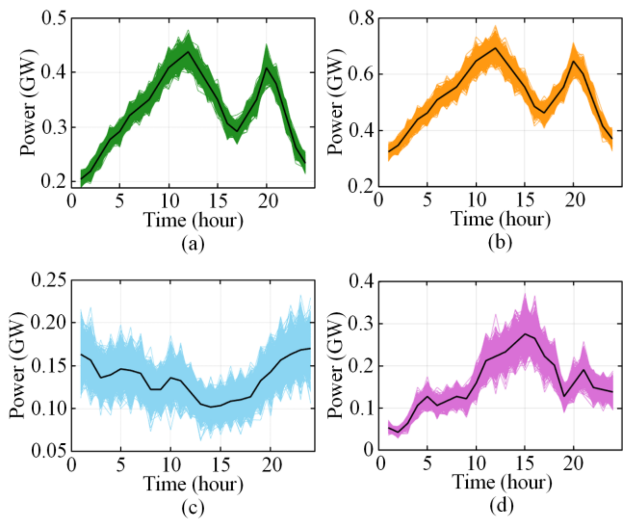

Fig. 3. Sampled scenarios. (a) Active load. (b) Reactive load. (c) Output of WT1. (d) Output of WT2.

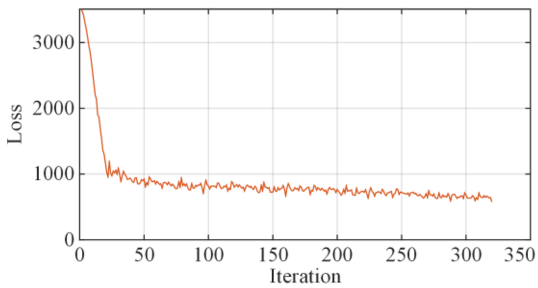

Fig. 4. Training process of DNN.

TABLE IV
IDENTIFICATION RESULTS ($\beta = 0.5$)

Commitment states	Correct	Incorrect	Total	Accuracy (%)
On	101479	3347	104826	96.81
Off	14352	822	15174	94.58

As shown in Table IV, the accuracies of the on-state and off-state results are 96.81% and 94.58%, respectively. This indicates that most of commitment states can be accurately identified by DNN when β is set to 0.5. To further discover the distribution of the misidentified commitment states, their probabilities are presented in Table V. Observe that the misidentified commitment units are mainly those with the probability lower than 0.90. Therefore, if the confidence level β is set to a larger one, these misidentified commitment states will be excluded from the identification results. This is consistent with the principle of the proposed algorithm, which tends to determine the commitment states of most units rather than all the units through DNN. For the units with the on-state or off-state probabilities lower than β, we can redetermine their commitment states through ADP accurately.

TABLE V
PROBABILITY DISTRIBUTION OF MISIDENTIFIED COMMITMENT STATES

Misidentified commitment states	$\beta \in [0,0.90]$	$\beta \in (0.90,1]$
On	2895 (86.50%)	452 (13.5%)
Off	741 (90.15%)	81 (9.85%)

To discover an appropriate β, it is set to different values, and then ADP is executed to redetermine those unidentified commitment states in the forecast scenario. The results are compared with the ones provided by ADP (i.e., all commitment states are determined by ADP), which are regarded as benchmarks in this test. The relative error e_d is used to evaluate decision precision as (36). The results are shown in Table VI.

$$e_{\mathrm{d}} = \frac{1}{TN_{\mathrm{G}}} \sum_{t=1}^{T} \sum_{g=1}^{N_{\mathrm{G}}} \left| I_{g,t}^{\mathrm{pro}} - I_{g,t}^{\mathrm{ben}} \right| \times 100\% \tag{36}$$

where $I_{g,t}^{\mathrm{pro}}$ and $I_{g,t}^{\mathrm{ben}}$ are the commitment states of the proposed algorithm and the benchmark algorithm, respectively.

TABLE VI
RESULTS WITH DIFFERENT CONFIDENCE LEVELS

Algorithm	β	e_{d} (%)	Action space	Time (s)
Benchmarked			23168	2880.63
Proposed	0.8	1.25	94	12.19
	0.9	0.42	186	24.49
	0.95	0.42	186	24.66
	0.99	0.00	244	32.37
	0.995	0.00	1208	159.71

It can be found in Table VI that the bigger the β is, the higher decision precision of the proposed algorithm has. This is because more commitment states with low on-state or off-state probabilities can be redetermined by ADP when β is set to a larger value. However, the action space of ADP also grows with the increase of β, reducing the computational efficiency of the algorithm accordingly. Taking both the accuracy and computational efficiency into consideration, β is set to 0.99 in the subsequent tests.

2) Test on the PDP Criterion

In this paper, a PDP criterion is proposed to further reduce the action space of ADP using the probability information provided by DNN. To validate its effectiveness, the aforementioned algorithm (i.e., ADP with the DT technique) is applied to determine the UC schedule with and without the PDP criterion, respectively. The results are shown in Table VII.

TABLE VII
RESULTS WITH AND WITHOUT THE PDP CRITERION

β	PDP	e_{d} (%)	Action space	Time (s)
0.9	Without	0.42	186	24.49
	With	0.42	94	12.42
0.95	Without	0.42	186	24.66
	With	0.42	94	12.56
0.99	Without	0.00	244	32.37
	With	0.00	102	13.64
0.995	Without	0.00	1208	159.71
	With	0.00	144	19.24

It can be found in Table VII that, in all confidence levels, the accuracies of the proposed algorithm with and without the PDP criterion are almost identical. However, in terms of the efficiency, both the action space and the computational time can be reduced evidently when the PDP criterion is adopted, which demonstrates the effectiveness of the proposed PDP criterion.

3) Test on the VFT Technique

In this paper, the value functions are trained offline with a VFT technique before being applied to the rolling decision-making process. To evaluate its performance, the proposed algorithm with and without the VFT technique is carried out to solve the RSCUC problem in the testing scenarios. And the results are presented in Table VIII.

TABLE VIII
RESULTS WITH AND WITHOUT THE VFT TECHNIQUE

VFT	Average cost ($)	Average iteration	Average time (s)
Without	1 578 082	182	71.85
With	1 589 378		13.23

Observe that when the VFT technique is used, the computational efficiency is greatly improved, while the average cost is almost identical. Additionally, the proposed algorithm is independent of the forecast information in the rolling decision process, which is favorable in application, especially when the rolling forecast information is inaccurate or unavailable.

4) Comparison of Different Algorithms

In this case, the proposed algorithm is compared with the myopic policy and the MPC algorithm. For myopic policy, it does not require the rolling forecast information, because it only considers the information of the current period. For MPC, the optimization horizons are set to 4 and 6 to provide more results for comparison. Similar to [55]–[57], at the initial period, the commitment states of units are all set to 1, while the minimum on time limits is satisfied. The optimization error in each scenario and the average error are calculated as follows:

$$e_n = \frac{F_n - F_n^*}{F_n^*} \times 100\% \tag{37}$$

$$\bar{e} = \frac{1}{N_{\text{test}}} \sum_{n=1}^{N_{\text{test}}} e_n \tag{38}$$

where e_n is the optimization error in the nth scenario; F_n and F_n^* are the objective costs in the nth scenario obtained by the compared algorithm and the benchmark algorithm, respectively; N_{test} is the number of testing scenarios; and $\bar{e}$ is the average error.

In Fig. 5 and Table IX, the optimization results for the 500 testing scenarios are presented. It can be found that the proposed algorithm outperforms the myopic policy and the MPC algorithm in the following two aspects:

1) Regarding the computational accuracy, it can be observed that the proposed algorithm has smaller optimization errors than the myopic policy and MPC. This is because the value functions enable the proposed algorithm to make rolling decisions from a global perspective, while the comparison algorithms (i.e., myopic policy and MPC) only consider one or several periods in the rolling decision-making process.

2) Regarding the computational efficiency, it can also be found that the calculation time of the proposed algorithm is close to that of the myopic policy, but is much less than that of MPC. Because of the DT technique and the PDP criterion, the action space of ADP is reduced significantly. Meanwhile, the VFT technique helps the proposed algorithm avoid the iterative update of value functions. Thus, the computational efficiency of the proposed algorithm can be greatly improved during the rolling decision-making process. Note that the speed advantage of the proposed method will become more obvious with the increase of the system scale, as will be demonstrated in the last test.

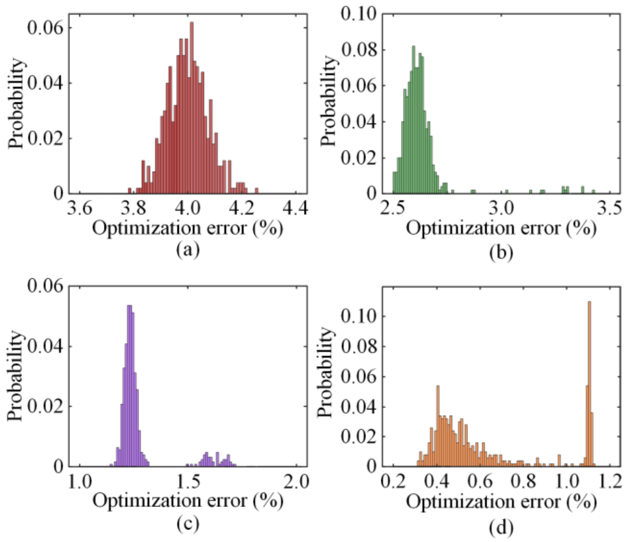

Fig. 5. Probability distribution of optimization errors of various algorithms. (a) Myopic policy. (b) MPC-4. (c) MPC-6. (d) The proposed algorithm.

TABLE IX
OPTIMIZATION RESULTS OF DIFFERENT ALGORITHMS

Algorithm	Average cost ($)	Average error (%)	Average time (s)
Myopic	1 642 838	4.10	13.02
MPC-4	1 620 165	2.67	66.01
MPC-6	1 598 608	1.30	120.06
Proposed	1 589 378	0.72	13.23

In Fig. 6, the UC schedules in a randomly selected scenario are presented as examples to further compare the performance of various algorithms.

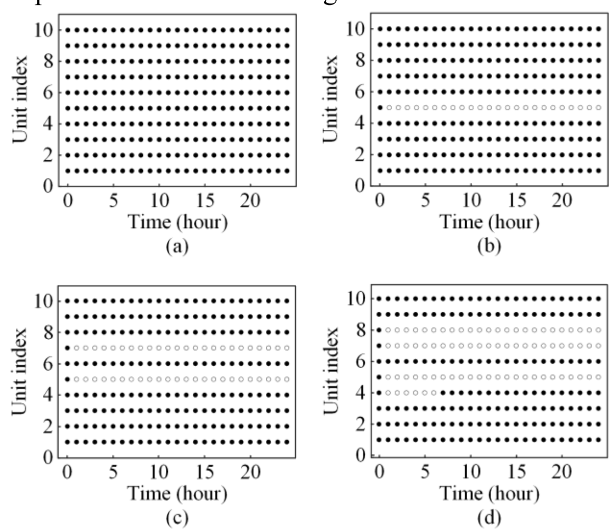

Fig. 6. UC schedules of various algorithms in a randomly selected scenario. (a) Myopic policy. (b) MPC-4. (c) MPC-6. (d) The proposed algorithm.

It can be found that when the myopic policy is adopted, the commitment states of all units do not change throughout the entire dispatch horizon. The reason is that the myopic policy only focuses on the situation of the current period, and the change of commitment states of units will bring extra start up and shut down cost, which is higher than what it will reduce at the current period. When MPC-4 and MPC-6 are employed, the commitment states of one and two units are adjusted since four and six periods are taken into consideration, respectively. As for the proposed algorithm, the commitment states of the four units are adjusted because it makes decisions from a global perspective. This indicates that the more periods considered in the rolling decision-making process, the more commitment states of units may be adjusted to enhance the operational economy.

5) Test on the Adaptability to Various Random Environments

In the proposed algorithm, both the DT and VFT techniques rely on the scenario information. Thus, the performance of the proposed algorithm may be affected if the random distributions of testing scenarios vary from those of the training scenarios. To test the adaptability of the proposed algorithm to this situation, the random distributions of the testing scenarios are set as follows.

Case1:

$$\boldsymbol{\varepsilon}_t^{\text{Pload}} \sim N(0,0.12^2), \boldsymbol{\varepsilon}_t^{\text{Qload}} \sim N(0,0.12^2), \boldsymbol{\varepsilon}_t^{\text{WT}} \sim N(0,0.17^2)$$

Case2:

$$\boldsymbol{\varepsilon}_t^{\text{Pload}} \sim N(0,0.15^2), \boldsymbol{\varepsilon}_t^{\text{Qload}} \sim N(0,0.15^2), \boldsymbol{\varepsilon}_t^{\text{WT}} \sim N(0,0.20^2)$$

Case3:

$$\boldsymbol{\varepsilon}_t^{\text{Pload}} \sim N(0,0.17^2), \boldsymbol{\varepsilon}_t^{\text{Qload}} \sim N(0,0.17^2), \boldsymbol{\varepsilon}_t^{\text{WT}} \sim N(0,0.22^2)$$

Case4:

$$\boldsymbol{\varepsilon}_t^{\text{Pload}} \sim N(0,0.20^2), \boldsymbol{\varepsilon}_t^{\text{Qload}} \sim N(0,0.20^2), \boldsymbol{\varepsilon}_t^{\text{WT}} \sim N(0,0.25^2)$$

Five hundred sets of fresh testing scenarios are generated using the MC method. Other parameter settings of the proposed algorithm remain unchanged, and the optimization errors are shown in Table X. It can be seen that the maximal optimization error will increase slowly with the increase of the stochastic forecast error. However, in all four cases, the maximal optimization errors do not exceed 1.5%, which indicates the adaptivity of the proposed algorithm to different random environments.

TABLE X
OPTIMIZATION RESULTS WITH DIFFERENT FORECAST ERROR DISTRIBUTIONS

Case	Average error (%)	Maximal error (%)	Minimum error (%)
1	0.52	1.11	0.35
2	0.59	1.28	0.32
3	0.66	1.39	0.30
4	0.71	1.45	0.30

B. A Real 2778-bus System

To demonstrate the scalability of the proposed algorithm in large scale systems, a real 2778-bus system, which contains 2778 buses, 3035 branches, 225 units and 10 WTs, is selected for testing. In order to form the proposed algorithm, 600 sets of operation data from 2016 to 2017 are employed to train DNN in the DT technique and the value functions in the VFT technique. Then, 300 sets of operation data in 2018 are utilized to test its performance. The proposed algorithm is compared with the myopic policy and MPC again, and the results are given in Table XI.

TABLE XI
OPTIMIZATION RESULTS OF DIFFERENT ALGORITHMS IN THE 2778-BUS SYSTEM

Algorithm	Average cost ($)	Time (s)
Proposed	3.3259×10^8	174
Myopic	×	×
MPC-4	×	×
MPC-6	×	×

Note: '×' indicates the algorithm fail to find solutions within 20 hours.

It can be found that the myopic policy and MPC fail to solve the RSCUC problem of the large scale power system within 20 hours due to the discreteness and nonlinearity. However, thanks to the DT technique, PDP criterion and VFT technique, the proposed algorithm can still obtain solutions within 3 minutes, which is sufficient to meet the rolling decision-making requirement of RSCUC.

V. CONCLUSION

This paper proposes a TADP algorithm for the RSCUC problem with AC power flow and uncertainties. The main conclusions are as follows:

1) The proposed TADP algorithm is accurate, since its optimization results are almost identical to those obtained by the benchmark algorithm.

2) The proposed TADP algorithm is efficient, as the DT technique, PDP criterion and VFT technique improve the computational efficiency to a great extent.

3) The proposed TADP algorithm is independent of the forecast information, so it has high accuracy even when the forecast information is not considered.

4) Compared with other rolling optimization algorithms (i.e., myopic policy and MPC), the proposed TADP algorithm can provide better solutions. In addition, the proposed algorithm has excellent scalability for large-scale RSCUC problems, while other algorithms fail to provide solutions within 20 hours.

ACKNOWLEDGMENT

Not applicable.

AUTHORS' CONTRIBUTIONS

Jianquan Zhu: guidance at all stage and improving the manuscript. Kai Zeng and Jiajun Chen: the idea, mathematical and practical design, case study and writing the paper. Wenmeng Zhao, Wenhao Liu, and Wenkai Zhu: supervision, reviewing and improving the manuscript. All authors read and approved the final manuscript.

FUNDING

This work is supported in part by the State Key Laboratory of HVDC (No. SKLHVDC-2021-KF-09), and in part by the National Natural Science Foundation of China (No. 51977081).

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

DECLARATIONS

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

AUTHORS' INFORMATION

Jianquan Zhu received the B.S., M.S. and Ph.D. degrees in electrical engineering from Fuzhou University, Guangxi University, and Tsinghua University, in 2005, 2008, and 2012, respectively. He is currently an associate professor in the School of Electric Power Engineering, South China University of Technology, Guangzhou, China. His research interests are modeling and optimization of power systems.

Kai Zeng received the B.S. degree in electrical engineering from South China University of Technology, Guangzhou, China, where he is currently pursuing the M.S. degree. His research interests focus on the application of reinforcement learning in power systems.

Jiajun Chen received the B.S. degree in electrical engineering from South China University of Technology, Guangzhou, China, where he is currently pursuing the Ph.D. degree. His research interests focus on applying approximate dynamic programming and optimization techniques in power systems.

Wenmeng Zhao received the B.S. and Ph.D. degrees in electrical engineering from Zhengzhou University and South China University of Technology, in 2011 and 2016, respectively. He is currently with the Electric Power Research Institute, China Southern Power Grid, Guangzhou. His research interests include electricity market and economic dispatch of power systems.

Wenhao Liu received the B.S. degree in electrical engineering from South China University of Technology, Guangzhou, China, where he is currently pursuing the M.S. degree. His research interests focus on the decentralized optimization of power systems and electricity market.

Wenkai Zhu received the B.S. degree in electrical engineering from South China University of Technology, Guangzhou, China, where he is currently pursuing the M.S. degree. His research interests focus on the distributed and robust optimization of power systems.

REFERENCES

[1] S. Cordova, H. Rudnick, and A. Lorca *et al.*, "An efficient forecasting-optimization scheme for the intraday unit commitment process under significant wind and solar power," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1899-1909, Oct. 2018.

[2] J. Warrington, C. Hohl, and P. J. Goulart *et al.*, "Rolling unit commitment and dispatch with multi-stage recourse policies for heterogeneous devices," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 187-197, Feb. 2015.

[3] M. Kazemi, P. Siano, and D. Sarno *et al.*, "Evaluating the impact of sub-hourly unit commitment method on spinning reserve in presence of intermittent generators," *Energy*, vol. 113, no. 1, pp. 338-354, Oct. 2016.

[4] Y. V. Makarov, C. Loutan, and J. Ma *et al.*, "Operational impacts of wind generation on California power systems," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 1039-1050, Mar. 2009.

[5] Y. V. Makarov, P. V. Etingov, and J. Ma *et al.*, "Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 4, pp. 433-442, Oct. 2011.

[6] H. Hui, C. N. Yu, and R. Surendran *et al.*, "Wind generation scheduling and coordination in ERCOT nodal market," in *2012 IEEE Power & Energy Society General Meeting*, San Diego, USA, Jul. 2012, pp. 1-8.

[7] A. L. Ott, "Development of smart dispatch tools in the PJM market," in *2011 IEEE Power & Energy Society General Meeting*, Detroit, USA, Jul. 2011, pp. 1-3.

[8] Y. Chen, Q. Wang, and X. Wang *et al.*, "Applying robust optimization to MISO look-ahead commitment," in *2014 IEEE Power & Energy Society General Meeting*, National Harbor, USA, Jul. 2014, pp. 1-5.

[9] R. C. Johnson, H. H. Happ, and W. J. Wright, "Large scale hydro-thermal unit commitment-method and results," *IEEE Transactions on Power Apparatus and Systems*, vol. 90, no. 3, pp. 1373-1384, May 1971.

[10] T. Senjyu, K. Shimabukuro, and K. Uezato *et al.*, "A fast technique for unit commitment problem by extended priority list," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 882-888, May 2003.

[11] J. Qin, H. Li, and Y. Sun *et al.*, "Optimal method of minimum start-up mode for conventional units considering new energy limit grid-connection strength constraint," *Power System Protection and Control*, vol. 51, no. 16, pp. 127-135, Aug. 2023. (in Chinese)

[12] W. Tong, D. Liu, and Z. Hu *et al*., "Hybridizing genetic algorithm with grey prediction evolution algorithm for solving unit commitment problem," *Applied Intelligence*, vol. 53, no. 1, pp. 19922-19939, Sept. 2023.

[13] M. T. Tsai, H. J. Gow, and W. M. Lin, "Hybrid Taguchi-immune algorithm for the thermal unit commitment," *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 4, pp. 1062-1069, May 2011.

[14] M. Haberg, "Fundamentals and recent developments in stochastic unit commitment," *International Journal of Electrical Power & Energy Systems*, vol. 109, no. 1, pp. 38-48, Jul. 2019.

[15] C. L. Chen and S. C. Wang, "Branch-and-bound scheduling for thermal generating units," *IEEE Transactions on Energy Conversion*, vol. 8, no. 2, pp. 184-189, Jun. 1993.

[16] M. Shafie-Khah, M. P. Moghaddam, and M. K. Sheikh-El-Eslami, "Unified solution of a non-convex SCUC problem using combination of modified branch-and-bound method with quadratic programming," *Energy Conversion and Management*, vol. 52, no. 12, pp. 3425-3432, Dec. 2011.

[17] F. Feng, P. Zhang, and M. A. Braginand *et al*., "Novel resolution of unit commitment problems through quantum surrogate Lagrangian relaxation," *IEEE Transactions on Power Systems*, vol. 38, no. 3, pp. 2460-2471, May 2023.

[18] M. R. Scuzziato, E. C. Finardi, and A. Frangioni, "Solving stochastic hydrothermal unit commitment with a new primal recovery technique based on Lagrangian solutions," *International Journal of Electrical Power & Energy Systems*, vol. 127, May 2021.

[19] N. Raghunathan, M. A. Bragin, and B. Yan, *et al*., "Exploiting soft constraints within decomposition and coordination methods for sub-hourly unit commitment," *International Journal of Electrical Power & Energy Systems*, vol. 139, Jul. 2022.

[20] M. Zhou, J. Zhai, and G. Li *et al*., "Distributed dispatch approach for bulk ac/dc hybrid systems with high wind power penetration," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3325-3336, May 2018.

[21] Y. Zhang, Q. Guo, and Y. Zhou *et al*., "Frequency-constrained unit commitment for power systems with high renewable energy penetration," *International Journal of Electrical Power & Energy Systems*, vol. 153, Nov. 2023.

[22] J. Guo, J. Zhang, and T. Zhao *et al*., "Stochastic unit commitment for power systems with offshore wind farms towards frequency resiliency," *IET Renewable Power Generation*, vol. 28, no. 7, pp. 1218-1229, May 2024.

[23] H. Xiong, Z. Chen, and X. Zhang *et al*., "Robust scheduling with temporal decomposition of integrated electrical-heating system based on dynamic programming formulation," *IEEE Transactions on Industry Applications*, vol. 59, no. 4, pp. 5087-5100, Aug. 2023.

[24] D. Han, J. Jian, and L. Yang, "Outer approximation and outer-inner approximation approaches for unit commitment problem," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 505-513, Mar. 2014.

[25] A. Castillo, S. Member, and C. Laird *et al*., "The unit commitment problem with ac optimal power flow constraints," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4853-4866, Nov. 2016.

[26] S. Pineda, J. M. Morales, and A. Jiménez-Cordero, "Data-driven screening of network constraints for unit commitment," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3695-3705, Sept. 2020.

[27] Á. Porras, S. Pineda, and J. M. Morales *et al*., "Cost-driven screening of network constraints for the unit commitment problem," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 42-51, Jan. 2023.

[28] F. Mohammadi, M. Sahraei-Ardakani, and D. N. Trakas *et al*., "Machine learning assisted stochastic unit commitment during hurricanes with predictable line outages," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5131-5142, Nov. 2021.

[29] X. Tang, X. Bai, and Z. Weng *et al*., "Graph convolutional network-based security-constrained unit commitment leveraging power grid topology in learning," *Energy Reports*, vol. 9, no. 1, pp. 3544-3552, Dec. 2023.

[30] T. Wu, Y. A. Zhang, and S. Wang, "Deep learning to optimize: Security-constrained unit commitment with uncertain wind power generation and BESSs," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 1, pp. 231-240, Jan. 2022.

[31] J. Wu, P. B. Luh, and Y. Chen *et al*., "Synergistic integration of machine learning and mathematical optimization for unit commitment," *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 391-401, Jan. 2024.

[32] A. Tuohy, P. Meibom, and E. Denny *et al*., "Unit commitment for systems with significant wind penetration," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 592-601, May 2009.

[33] M. Costley, M. J. Feizollahi, and S. Ahmed *et al*., "A rolling-horizon unit commitment framework with flexible periodicity," *International Journal of Electrical Power & Energy Systems*, vol. 90, no. 1, pp. 280-291, Sept. 2017.

[34] E. A. Bakirtzis, P. N. Biskas, and D. P. Labridis *et al*., "Multiple time resolution unit commitment for short-term operations scheduling under high renewable penetration," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 149-159, Jan. 2014.

[35] Y. Huo, F. Bouffard, and G. Joós, "Integrating learning and explicit model predictive control for unit commitment in microgrids," *Applied Energy*, vol. 306, Jan. 2022.

[36] L. Xie and M. D. Ilic, "Model predictive dispatch in electric energy systems with intermittent resources," in *2008 IEEE International Conference on Systems, Man and Cybernetics*, Singapore, Singapore, Oct. 2008, pp. 42-47.

[37] L. Xie and M. D. Ilic, "Model predictive economic/environmental dispatch of power systems with intermittent resources," in *2009 IEEE Power & Energy Society General Meeting*, Calgary, Canada, Jul. 2009, pp. 1-6.

[38] C. Hu, Z. Cai and Y. Zhang *et al*., "A soft actor-critic deep reinforcement learning method for multi-timescale coordinated operation of microgrids," *Protection and Control of Modern Power Systems*, vol. 7, no. 3, pp. 1-10, Jul. 2022.

[39] H. Liu, W. Wu, and Y. Wang, "Bi-level off-policy reinforcement learning for two-timescale Volt/VAR control in active distribution networks," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 385-395, Jan. 2023.

[40] H. Shuai, J. Fang, and X. Ai *et al*., "Optimal real-time operation strategy for micro-grid: an ADP-based stochastic nonlinear optimization approach," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 931-942, Apr. 2019.

[41] Z. Chen, L. Wu, and M. Shahidehpour, "Effective load carrying capability evaluation of renewable energy via stochastic long-term hourly based SCUC," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 188-197, Jan. 2015.

[42] D. Qiu, Y. Wang, and W. Hua *et al*., "Reinforcement learning for electric vehicle applications in power systems: a critical review," *Renewable & Sustainable Energy Review*, vol. 173, Mar. 2023.

[43] H. Ye, J. Zhu, and J. Chen *et al*., "Quadratic Taylor expansion-based approximate dynamic programming for fully decentralized AC-OPF of multi-area power systems," *IEEE Transactions on Power Systems*, vol. 38, no. 5, pp. 4940-4949, Sept. 2023.

[44] W. Zhang, C. Shao, and B. Hu *et al*., "Proactive security-constrained unit commitment against typhoon disasters: an approximate dynamic programming approach," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 7076-7087, May 2023.

[45] A. Castillo, S. Member, and C. Laird *et al*., "The unit commitment problem with AC optimal power flow constraints," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4853-4866, Nov. 2016.

[46] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34-37, Jul. 1966.

[47] W. Powell, "Learning value function approximations," in *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, Hoboken, USA: John Wiley & Sons, 2007, pp. 337-382.

[48] D. R. Jiang and W. B. Powell, "Optimal hour-ahead bidding in the real-time electricity market with battery storage using approximate dynamic programming," *Informs Journal on Computing*, vol. 27, no. 3, pp. 525-543, Aug. 2015.

[49] R. B. Damala, R. K. Patnaik, and A. R. Dash, "A simple decision tree-based disturbance monitoring system for VSC-based HVDC transmission link integrating a DFIG wind farm," *Protection and Control of Modern Power Systems*, vol. 7, no. 2, pp. 1-19, Apr. 2022.

[50] Z Liu, Y Cui, and W. Li, "A classification method for complex power quality disturbances using EEMD and rank wavelet SVM," *IEEE Transactions on Smart Grid*, vol. 4, no. 6, pp. 1678-1685, Jul. 2015.

[51] Flow charts of the TADP algorithm, [Online]. Available: https://github.com/ZK-scut/Flowchart

[52] Parameters of the IEEE 39-bus system, [Online]. Available: https://github.com/ZK-scut/IEEE-39-bus-test-data

[53] T. Ding, Q. Yang, and X. Liu *et al*., "Duality-free decomposition based data-driven stochastic security-constrained unit commitment," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 1, pp. 82-93, Jan. 2019.

[54] S. Gao, C. Xiang, and M. Yu *et al*., "Online optimal power scheduling of a microgrid via imitation learning," *IEEE Transactions on Smart Grid*, vol. 13, no. 2, pp. 861-876, Mar. 2023.

[55] Y. Bai, H. Zhong, and Q. Xia *et al*., "A decomposition method for network-constrained unit commitment with AC power flow constraints," *Energy*, vol. 88, no. 1, pp. 595-603, Aug. 2015.

[56] M. Zhou, S. Xia, and G. Li *et al*., "Interval optimization combined with point estimate method for stochastic security-constrained unit commitment," *International Journal of Electrical Power & Energy Systems*, vol. 63, no. 1, pp. 276-284, Dec. 2014.

[57] A. Khodaei and M. Shahidehpour, "Transmission switching in security-constrained unit commitment," *IEEE Transactions on Power Systems*, vol. 25, no. 4, pp. 1937-1945, Nov. 2010.