Dynamic Optimal Power Flow Method Based on Reinforcement Learning for Offshore Wind Farms Considering Multiple Points of Common Coupling

Yang Fu, *Member*, *IEEE*, Zixu Ren, Shurong Wei, *Member*, *IEEE*, Lingling Huang, *Member*, *IEEE*, Fangxing Li, *Fellow*, *IEEE*, and Yang Liu, *Member*, *IEEE*,

***Abstract*—The widespread adoption of renewable energy sources presents significant challenges for power system dispatching. This paper proposes a dynamic optimal power flow (DOPF) method based on reinforcement learning (RL) to address the dispatching challenges. The proposed method considers a scenario where large-scale offshore wind farms are interconnected and have access to an onshore power grid through multiple points of common coupling (PCCs). First, the operational area model of the offshore power grid at the PCCs is established by combining the prediction results and the transmission capacity limit of the offshore power grid. Built upon this, a dynamic optimization model of the power system and its RL environment are constructed with the consideration of offshore power dispatching constraints. Then, an improved algorithm based on the conditional generative adversarial network (CGAN) and the soft actor-critic (SAC) algorithm is proposed. By analyzing an improved IEEE 118-node example, the proposed method proves to have the advantage of economy over a longer timescale. The resulting strategy satisfies power system operation constraints, effectively addressing the constraint problem of action space of RL, and it has the added benefit of faster solution speeds.**

***Index Terms*—Offshore power grid, optimal scheduling, dynamic optimal power flow (DOPF), reinforcement learning (RL), renewable energy.**

I. Introduction

THE two key concerns in the operation of any power system are enhancing its reliability and improving its economic efficiency. However, with the rapid growth of renewable energy sources and their large-scale integration into the power system, the level of synchronous power sources in the power system has further decreased, posing significant challenges to power system operations.

In 2021, 21.1 GW of offshore wind capacity was installed worldwide, a remarkable surge over previous years [1] that cemented it as a mainstay of coastal power generations. To address the challenges of large-scale offshore wind power integration, many countries are envisioning the concept of offshore power grids. The offshore power grid is formed by interconnecting offshore wind farms through multiple points of common coupling (PCCs) to the onshore power grid [2], as shown in Fig. 1 [3]. The integration of offshore wind power through multiple PCCs can effectively enhance the schedulability of offshore wind power, as demonstrated in [3], [4]. Hence, determining the operational area of these connection points is crucial for power system operation. Unfortunately, as of yet, there is no reliable and fast method for this.

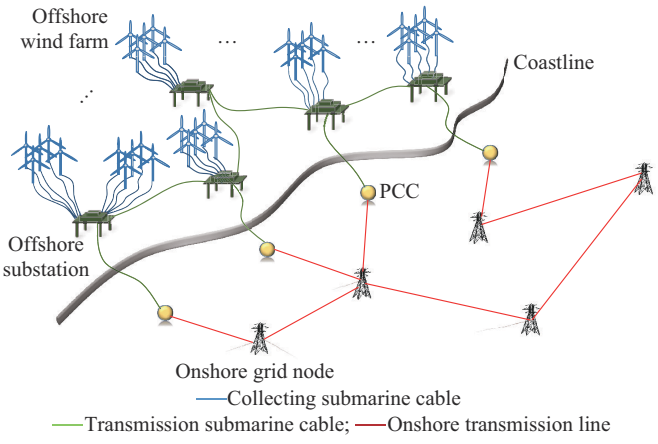

Fig. 1. Offshore power grid.

Manuscript received: October 8, 2023; revised: January 25, 2024; accepted: March 28, 2024. Date of CrossCheck: March 28, 2024. Date of online publication: May 29, 2024.

This work was supported in part by the National Natural Science Foundation of China (No. 52377063), the Shanghai Action Plan for Science, Technology and Innovation (No. 22dz1206100), the Major Natural Science Project of Shanghai Municipal Education Commission (No. 2021-01-07-00-07-E00122), and the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (No. TP2020066).

Y. Fu, S. Wei (corresponding author), L. Huang, and Y. Liu are with the Engineering Research Center of Offshore Wind Technology Ministry of Education (Shanghai University of Electric Power), Shanghai 200090, China (e-mail: mfudong@126. com; wsrmail@163. com; linglinghuang82@126. com; liuyang_ssun@163.com).

Z. Ren is with the Department of Electrical Power Engineering, Shanghai University of Electric Power, Shanghai 200090, China (e-mail: 605211514@qq. com).

F. Li is with the Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN 37996, USA (e-mail: fli6@utk.edu).

DOI: 10.35833/MPCE.2023.000765

The rapid fluctuations in power generation brought about by offshore wind power also result in higher uncertainty in the power system, making it challenging for system dispatchers to achieve more optimized scheduling from a longer-term perspective. The dynamic optimal power flow (DOPF) has proven to be an effective model for addressing such issues [5]. It is an extension of the optimal power flow (OPF) model that can cover multiple time periods [6].

In summary, the concentrated integration of offshore wind power clusters into power system operations faces two major challenges: ① the controllable capacity of offshore wind power clusters needs to be accurately quantified to ensure the safe operation of the offshore transmission network; and ② dispatchers need to make forward-thinking decisions that promote both system security and economic efficiency.

Reinforcement learning (RL) has proven its superior performance in addressing optimal decision-making problems across various types of systems and is widely applied to optimization and scheduling problems in the field of power systems [7], [8]. Traditional RL models face challenges in ensuring the safety of actions. Some recent studies have addressed this issue by introducing security layers after the agent generates actions to enforce safety constraints [8], which may lead to a reduction in computational efficiency. Current models have also not yet addressed the issue of reliable power constraints at the PCCs of offshore power grids.

In light of this, this study introduces a model tailored to the operational area of offshore power grids and integrates it into the conventional DOPF model. Following that, an improved RL algorithm based on the conditional generative adversarial network (CGAN) is proposed. This algorithm enables the construction of a safety-constrained agent to replace the original agent, thereby enhancing solution speed while ensuring system safety.

Currently, the DOPF model lacks a description of constraints related to the concentrated integration of offshore wind power. For quantitatively describing the operational space of offshore power grids, the security region [9] is a commonly used method to describe the operational area of an electrical system and has found widespread application in distribution networks [10]-[12]. Reference [13] introduces an observational method for the security region of distribution networks based on power flow calculations. Reference [14] proposes a security region model that satisfies $N-1$ constraints in smart distribution networks. Reference [15] investigates the convexity and formation principles of distribution system security regions, and suggests that they are internally connected and hole-free.

RL is widely employed in optimal decision-making problems [16]-[18]. RL is also widely applied in the field of electrical engineering, including microgrid energy management [19], [20], electric vehicle charging scheduling [21], [22], and grid optimization and operation [23], [24]. In the field of power systems, the safety of RL is a widely recognized concern. Some commonly used methods in this regard include constraint Markov decision processes [25], Lagrangian methods [26], Lyapunov methods [27], and security layers. Reference [26] proposes an RL-based solution model for DOPF and tackles safety constraint issues using the Lagrangian method. Reference [8] proposes a safe deep deterministic policy gradient (DDPG) algorithm that incorporates a training security layer to correct the action and ensure the constraints of the RL action space.

We can summarize the areas of existing research that need to be supplemented as follows.

First, in [9]-[12], the calculation methods for the operational area mostly involve repetitive power flow calculations to fit the boundaries. As the calculation of multiple power flows is time consuming, it is difficult to adapt to the needs of power system scheduling

Second, while RL-solving strategies that consider safety constraints are proposed by [8], [25]-[27], the Lagrangian method may occasionally produce actions that violate system safety constraints with a low probability. Additionally, security layers can present training challenges and computational burdens during the initial stages of system operation.

This paper investigates a real-time optimization scheduling method tailored to meet the requirements of modern power systems. By combining the power prediction results of offshore wind power clusters with the topology of the offshore power grid interconnection, we establish a model for the operational area of the offshore power grid. The constraints of the offshore power are derived by solving the operational area at the PCCs. A DOPF optimization model for power systems that considers the integration of offshore wind power through multiple PCCs is proposed, and an RL environment for DOPF is constructed. To enhance the safety and efficiency of the solution method, we propose a CGAN-soft actor-critc (SAC) algorithm. The SAC algorithm adopts the CGAN to replace the security layer, which ensures the safety of the action space while considering the exploratory nature of the solution method. This results in a safe and efficient solution to the DOPF problem. The main innovation points of this paper can be summarized as follows.

1) A DOPF model incorporating the operational area of offshore power grids is established: a rapid solution method for the operational area of the offshore power grid is established and incorporated into the DOPF model, thereby reducing the complexity of the optimization model.

2) A CGAN-SAC algorithm is proposed to solve DOPF problems. The new CGAN agent module can directly generate actions that adhere to safety constraints. Furthermore, pretraining of the CGAN has the potential to enhance the solution rate, which has the potential to improve the performance of RL in power systems, particularly when considering safety constraints.

The rest of this paper is organized as follows. Section II introduces the construction of the operational area of the offshore power grid at the PCC. Section III proposes a DOPF model of the power system. Section IV constructs a DOPF method based on improved RL. Section V carries out a case analysis. Section VI concludes this paper.

II. Construction of Operational Area of Offshore Power Grid at PCC

When multiple PCCs are used in the offshore cluster net-

work to integrate the offshore network into the onshore power grid, the offshore power can be dispatched flexibly at each PCC. When integrating offshore power into the power grid, it is essential to consider the uncertainties and schedulability of the offshore power output. Operating boundaries at PCCs are difficult to determine. Considering the operating domain of the offshore power grid at the PCC and forming operational constraints for the offshore power grid can effectively reduce the computational burden of power flow calculations.

A. Constraints of Offshore Power Grid Operation

The following constraints should be considered during the operation of offshore power grids.

1) Output constraint of offshore wind farm

The output of offshore wind farms must not exceed the ideal power generation in the maximum power point tracking (MPPT) mode:

$$P_{Wi,t} \le P_{Wi,t,opt} \tag{1}$$

where $P_{Wi,t,opt}$ is the ideal power generation of the wind farm in MPPT mode at time t; and $P_{Wi,t}$ is the power generation of the wind farm at time t.

2) Transmission line capacity constraint

$$L \le L_{\max} \tag{2}$$

where L is the transmission capacity of the submarine cable; and $L_{\max}$ is the maximum transmission capacity of the submarine cable.

3) Reverse power flow constraint

In order to prevent the reverse power flow from affecting the power grid and to ensure the positive power flow at the PCC, the following constraint is set:

$$L_i > 0 \tag{3}$$

where L_i is the output capacity of PCC i.

4) Node voltage constraint

$$V_{\min} \le V \le V_{\max} \tag{4}$$

where V is the node voltage; and $V_{\min}$ and $V_{\max}$ are the minimum and maximum node voltages, respectively.

5) Power flow constraints

$$P_i = V_i \sum_{j=1}^{n_b} V_j \left(G_{ij} \cos\theta_{ij} + B_{ij} \sin\theta_{ij} \right) \tag{5}$$

$$Q_i = V_i \sum_{j=1}^{n_b} V_j \left(G_{ij} \sin\theta_{ij} - B_{ij} \cos\theta_{ij} \right) \tag{6}$$

where V_i and V_j are the voltages of the nodes i and j, respectively; θ_{ij} is the phase angle difference of voltage between nodes i and j; P_i and Q_i are the active and reactive power of node i, respectively; n_b is the set of nodes connected to node i; and G_{ij} and B_{ij} are the real and imaginary parts of the node admittance matrix, respectively.

B. Operational Area Model of Offshore Power Grid

Based on the constraints in Section II-A, the operational area model of the offshore power grid is established. Three PCCs are selected as observation points and the outputs of offshore wind farms are taken as variables. By doing so, the operational area of the offshore power grid is obtained, which is denoted as Ω_{OFG}, and is expressed as:

$$\Omega_{OFG} = \left\{ P_W | W\left(f_i\right) \le 0 \right\} \quad i = 1, 2, \ldots, c \tag{7}$$

where P_W is the power value of offshore wind farms; $W\left(f_i\right)$ represents the constraint of system operation; and c is the number of constraints.

C. Solution of Operational Area of Offshore Power Grid

Summarizing the model in Section II-A, the operational constraints of offshore power grids can be categorized into three main aspects: ① wind farm output constraint boundaries; ② submarine cable capacity constraint boundaries; and ③ voltage deviation constraint boundaries. Each of these three aspects are addressed separately as follows.

1) Aspect 1: wind farm output constraint boundaries

The operation of offshore power grids should adhere to an active power balance. In power flow calculations, the remaining power is typically handled by the balancing nodes. However, in offshore cluster networks, it is challenging for offshore wind farms to match the regulating capacity of conventional power plants. Therefore, it is essential to consider the adjustment of upper and lower limits for offshore wind farms. Taking WF1 as the balancing node as an example, there exists

$$\underline{P_{WF1}} + \sum_{i=2}^{n_{WF}} P_{WFi} - P_{loss} \le \sum_{i=1}^{n_{PCC}} P_{PCCi} \le \overline{P_{WF1}} + \sum_{i=2}^{n_{WF}} P_{WFi} - P_{loss} \tag{8}$$

where $\underline{P_{WF1}}$ and $\overline{P_{WF1}}$ are the adjustable lower and upper limits of WF1, respectively; P_{loss} is the system power loss; P_{WFi} is the active power of wind farm i; P_{PCCi} is the active power of PCC i; and n_{PCC} and n_{WF} are the numbers of PCCs and wind farms, respectively. Formula (8) represents the power boundary for the sum of power at the PCC.

2) Aspect 2: submarine cable capacity constraint boundaries

For a general electrical network, the branch current matrix $\dot{\boldsymbol{I}}_B$ can be obtained from the branch susceptance and node voltages:

$$\dot{\boldsymbol{I}}_B = \boldsymbol{Y}_B \boldsymbol{A}^{\mathrm{T}} \dot{\boldsymbol{U}}_N \tag{9}$$

where $\boldsymbol{Y}_B$ is the branch susceptance matrix; $\dot{\boldsymbol{U}}_N$ is the node voltage matrix; and $\boldsymbol{A}$ is the node incidence matrix.

The node admittance matrix can be represented as:

$$\boldsymbol{Y}_N \dot{\boldsymbol{U}}_{\boldsymbol{N}} = \dot{\boldsymbol{I}}_N \tag{10}$$

where $\boldsymbol{Y}_N$ is the node susceptance matrix; and $\dot{\boldsymbol{I}}_N$ is the node current matrix.

Thus, the current correlation coefficient matrix $\boldsymbol{C}(\boldsymbol{\lambda})$ can be represented as:

$$\boldsymbol{C}(\boldsymbol{\lambda}) = \boldsymbol{Y}_B \boldsymbol{A}^{\mathrm{T}} \boldsymbol{Y}_N^{-1} \tag{11}$$

Taking a specific line k as an example, the current phasor within the branch is:

$$\dot{I}_{k,B} = \lambda_{k-1} \dot{I}_{1,N} + \ldots + \lambda_{k-i} \dot{I}_{i,N} + \ldots + \lambda_{k-n} \dot{I}_{n,N} \tag{12}$$

where λ_{k-i} $(i = 1, 2, \ldots, n)$ is the current correlation coefficient between branch k and node i; and $\dot{I}_{i,N}$ is the current phasor of node i.

Performing a simple transformation on (12), we can obtain:

$$\frac{\dot{I}_{k,B}^{*}\dot{U}_{k,B}}{\dot{U}_{k,B}}=\frac{\lambda_{k-1}\dot{I}_{1,N}^{*}\dot{U}_{1,N}}{U_{1,N}}+\ldots+\frac{\lambda_{k-i}I_{i,N}^{*}\dot{U}_{i,N}}{\dot{U}_{i,N}}+\ldots+\frac{\lambda_{k-n}I_{n,N}^{*}\dot{U}_{n,N}}{\dot{U}_{n,N}} \tag{13}$$

where $\dot{U}_{k,B}$ is the initial voltage phasor of branch k; and $\dot{U}_{i,N}$ is the voltage phasor of node i.

The transmission power of branch k can be represented as:

$$\begin{aligned}P_{k,B}+\mathrm{j}Q_{k,B}=&\frac{\lambda_{k-1}\left(P_{1,N}+\mathrm{j}Q_{1,N}\right)U_{k,B}\left(\cos\varphi_{k,B}+\mathrm{j}\sin\varphi_{k,B}\right)}{U_{1,N}\left(\cos\varphi_{1,N}+\mathrm{j}\sin\varphi_{1,N}\right)}+\ldots+\\&\frac{\lambda_{k-i}\left(P_{i,N}+\mathrm{j}Q_{i,N}\right)U_{k,B}\left(\cos\varphi_{k,B}+\mathrm{j}\sin\varphi_{k,B}\right)}{U_{i,N}\left(\cos\varphi_{i,N}+\mathrm{j}\sin\varphi_{i,N}\right)}+\ldots+\\&\frac{\lambda_{k-n}\left(P_{n,N}+\mathrm{j}Q_{n,N}\right)U_{k,B}\left(\cos\varphi_{k,B}+\mathrm{j}\sin\varphi_{k,B}\right)}{U_{i,N}\left(\cos\varphi_{n,N}+\mathrm{j}\sin\varphi_{n,N}\right)}\end{aligned} \tag{14}$$

where $P_{k,B}$ and $Q_{k,B}$ are the active and reactive transmission power of branch k, respectively; $P_{i,N}$ and $Q_{i,N}$ are the active and reactive power of node i, respectively; $U_{k,B}$ and $\varphi_{k,B}$ are the magnitude and phase angle of the voltage at the beginning of branch k, respectively; and $U_{i,N}$ and $\varphi_{i,N}$ are the magnitude and phase angle of the voltage at node i, respectively.

For the operational area of the offshore power grid, the current constraints primarily focus on the overstepping of active power. Therefore, in this context, only the active power variables are considered. Taking the real part of (14) and ignoring the reactive power variation of offshore wind farms, we can obtain:

$$\begin{aligned}\Delta P_{k,B}=&\frac{\lambda_{k-1}U_{k,B}}{U_{1,N}}\Delta P_{1,N}\left(\cos\varphi_{k,B}\cos\varphi_{1,N}+\sin\varphi_{k,B}\sin\varphi_{1,N}\right)+\ldots+\\&\frac{\lambda_{k-i}U_{k,B}}{U_{i,N}}\Delta P_{i,N}\left(\cos\varphi_{k,B}\cos\varphi_{i,N}+\sin\varphi_{k,B}\sin\varphi_{i,N}\right)+\ldots+\\&\frac{\lambda_{k-n}U_{k,B}}{U_{n,N}}\Delta P_{n,N}\left(\cos\varphi_{k,B}\cos\varphi_{n,N}+\sin\varphi_{k,B}\sin\varphi_{n,N}\right)\end{aligned} \tag{15}$$

where $\Delta P_{k,B}$ is the incremental power of branch k; and $\Delta P_{i,N}$ is the incremental power of node i.

Based on (15), we can derive the constraints on the power variation at the PCC for branch k. By solving all the branch constraints, the enclosed region represents the safe region that satisfies the transmission power of branch constraints.

3) Aspect 3: voltage deviation constraint boundaries

Following the derivation method presented in aspect 2, the voltage sensitivity for the offshore cluster network can be derived correspondingly.

$$\begin{aligned}\Delta\dot{U}_{k}=&\frac{\mu_{k-1}\Delta P_{1,N}}{U_{1,N}\left(\cos\varphi_{1,N}+\mathrm{j}\sin\varphi_{1,N}\right)}+\ldots+\\&\frac{\mu_{k-i}P_{i,N}}{U_{i,N}\left(\cos\varphi_{i,N}+\mathrm{j}\sin\varphi_{i,N}\right)}+\ldots+\frac{\mu_{k-n}\Delta P_{n,N}}{U_{n,N}\left(\cos\varphi_{n,N}+\mathrm{j}\sin\varphi_{n,N}\right)}\end{aligned} \tag{16}$$

where μ_{k-i} is the current correlation coefficient between node k and node i, and $\boldsymbol{C}(\boldsymbol{\mu})=\boldsymbol{Y}_{N}^{-1}$, $\boldsymbol{\mu}=(\mu_{k-i})$.

III. DOPF Model of Power System

The DOPF model is based on the traditional OPF model with the addition of the unit climbing rate constraint, the rule of grid-connected power for offshore power grids, and the system reserve constraint in different time intervals. Its model is constructed as follows.

A. Objective Function

The DOPF model aims to describe the minimum operating cost of the power system on a specific timescale, including the operating cost of thermal power units, the cost of reserve capacity, and the penalty of wind abandonment:

$$\begin{aligned}\min F=&\sum_{t=0}^{T}\sum_{i=1}^{n_G}\left(a_iP_{G,i,t}^{2}+b_iP_{G,i,t}+c_i\right)+\sum_{t=0}^{T}\sum_{j=1}^{n_{G2}}d_{G,j}^{R}\left(R_{G,j,t}^{+},R_{G,j,t}^{-}\right)+\\&e_{PCC}\sum_{t=0}^{T}\left(\sum_{i=1}^{n_{WF}}\tilde{P}_{WF,i,t}-\sum_{k=1}^{n_{PCC}}P_{PCC,i,t}\right)\end{aligned} \tag{17}$$

where T is the scheduling period; n_G is the number of thermal power plants; a_i, b_i, and c_i are the cost coefficients of synchronous unit i; $P_{G,i,t}$ is the active power of synchronous unit i at time t; n_{G2} is the number of standby thermal power plants provided; $d_{G,j}^{R}$ is the standby cost coefficient of unit j that can provide standby power; $R_{G,j,t}^{+}$ and $R_{G,j,t}^{-}$ are the positive and negative standby power provided by unit j at time t, respectively; e_{PCC} is the penalty coefficient of offshore wind power abandonment; $\tilde{P}_{WF,i,t}$ is the predicted power of wind farm i at time t; and $P_{PCC,i,t}$ is the actual power value at PCC i at time t.

B. Constraint Condition

1) Power flow constraints

Like the offshore power grid, the onshore power grid also needs to meet the power flow constraints as follows:

$$\Delta P_{i,t}=P_{G,i,t}-P_{D,i,t}-V_{i,t}\sum_{j=1}^{n_b}V_{j,t}\left(G_{ij}\cos\theta_{ij,t}+B_{ij}\sin\theta_{ij,t}\right) \tag{18}$$

$$\Delta Q_{i,t}=Q_{G,i,t}-Q_{D,i,t}-V_{i,t}\sum_{j=1}^{n_b}V_{j,t}\left(G_{ij}\sin\theta_{ij,t}-B_{ij}\cos\theta_{ij,t}\right) \tag{19}$$

where $\Delta P_{i,t}$ and $\Delta Q_{i,t}$ are the differences between the transmitted active and reactive power and the load of node i at time t, respectively; $P_{G,i,t}$ and $Q_{G,i,t}$ are the active and reactive power of the power supply plant of node i at time t, respectively; $P_{D,i,t}$ and $Q_{D,i,t}$ are the active and reactive power of the load of node i, respectively; $V_{i,t}$ and $V_{j,t}$ are the voltage amplitudes of node i and node j at time t, respectively; and $\theta_{ij,t}$ is the phase angle difference of the voltage between node i and node j at time t.

2) Thermal power unit output constraints

$$P_{G,i,t}+R_{u,i,t}\le P_{G,i,\max} \tag{20}$$

$$P_{G,i,\min}\le P_{G,i,t}+R_{d,i,t} \tag{21}$$

$$Q_{G,i,\min}\le Q_{G,i,t}\le Q_{G,i,\max} \tag{22}$$

where $R_{u,i,t}$ and $R_{d,i,t}$ are the upper and lower spare capacities provided by unit i at time t, respectively; $P_{G,i,\min}$ and $P_{G,i,\max}$ are the minimum and maximum active power of the thermal

power unit i, respectively; and $Q_{G,i,\min}$ and $Q_{G,i,\max}$ are the minimum and maximum reactive power of the thermal power unit i, respectively. The sum of the current output of the unit and the provided spare capacity should fall within the upper and lower boundaries of the output of the unit.

3) Active power constraint of PCC

$$\left\{P_{PCCi}\right\} \in \Omega \tag{23}$$

where Ω is the operational area obtained from Section II-B.

The power of offshore parallel connection points shall be in the offshore operational area divided in Section II. We consider that offshore wind power provides limited assistance in system voltage support; therefore, the PCCs are treated as PQ nodes. The operational region of the PCC should be within the enclosed region formed by (8), (15), and (16).

4) Unit climbing rate constraints

$$P_{G,i,t} - P_{G,i,t-1} \le R_{\text{up},i}\Delta t \tag{24}$$

$$P_{G,i,t-1} - P_{G,i,t} \le R_{\text{down},i}\Delta t \tag{25}$$

where $R_{\text{up},i}$ and $R_{\text{down},i}$ are the upward and downward adjustment capabilities of unit i, respectively; and Δt is the time interval.

Because this paper focuses on a DOPF problem, the impact of generator start-up and shut-down is not considered.

5) Node voltage constraints

$$V_{i,\min} \le V_{i,t} \le V_{i,\max} \tag{26}$$

$$\theta_{i,\min} \le \theta_{i,t} \le \theta_{i,\max} \tag{27}$$

where $\theta_{i,t}$ is the voltage phase angle of node i at time t; $V_{i,\min}$ and $V_{i,\max}$ are the lower and upper limits of the voltage of node i, respectively; and $\theta_{i,\min}$ and $\theta_{i,\max}$ are the lower and upper limits of voltage phase angle of node i, respectively. In general, phase angles are not constrained in power flow calculation.

6) System reserve constraints

$$\sum_{i=1}^{n_G} R_{u,i,t} + \sum_{i=1}^{n_{WF}} \left(\tilde{P}_{WF,i,t} - P_{WF,i,t}\right) \ge \sum_{i=1}^{n} \varepsilon P_{D,i,t} + (K_G + K_L) f_{\lim}^{-} \tag{28}$$

$$\sum_{i=1}^{n_G} R_{d,i,t} \le \sum_{i=1}^{n} \varepsilon P_{D,i,t} + (K_G + K_L) f_{\lim}^{+} \tag{29}$$

where ε is the reserve factor; K_G and K_L are the unit regulated power of the generator and load, respectively; and $f_{\lim}^{+}$ and $f_{\lim}^{-}$ are the upper and lower limits of system frequency, respectively. When the offshore wind farm is in limited power operation, it can be used for backup to a certain extent.

IV. DOPF Method Based on Improved RL

A. Construction of RL Environment for DOPF

The classic RL framework is shown in Fig. 2. The DOPF problem proposed in this paper is explicitly constructed as follows.

1) State

$$\boldsymbol{s}_t = \left[P_{G,i,t-1} \quad P_{PCC,i,t-1} \quad P_{D,i,t} \quad Q_{D,i,t} \quad \Omega_t\right] \tag{30}$$

where $\boldsymbol{s}_t$ is the system state at time t; and Ω_t is the operational area of the offshore power grid at time t.

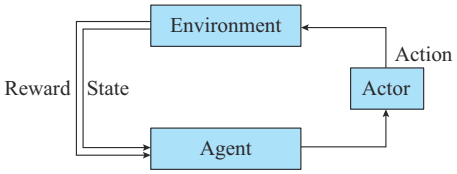

Fig. 2. Classic RL framework.

2) Action space

In the construction of the action space, generators are treated as PV nodes. Their action space includes changes in power output, terminal voltage, and additional upper and lower reserve capacity of the generator. PCCs are treated as PQ nodes. Their action space includes changes in active power.

$$\boldsymbol{a}_t = \left[\Delta P_{G,i,t} \quad V_{G,i,t} \quad \Delta P_{PCC,i,t} \quad R_{u,i,t} \quad R_{d,i,.t}\right] \tag{31}$$

where $\boldsymbol{a}_t$ is the response action taken by the agent at time t; $\Delta P_{G,i,t}$ is the active power change of unit i at time t; $V_{G,i,t}$ is the voltage amplitude of unit i at time t; and $\Delta P_{PCC,i,t}$ is the active power change of PCC i at time t .

3) State transition function

$$\begin{cases} P_{G,i,t} = P_{G,i,t-1} + \Delta P_{G,i,t} \\ P_{PCC,i,t} = P_{PCC,i,t-1} + \Delta P_{PCC,i,t} \end{cases} \tag{32}$$

In the state transition equation, the active power outputs of the generator and the PCC vary continuously. Since the terminal voltage of the generator can change rapidly and is independent of the previous state, it can be directly determined using the current control action.

The load and the operational area of the offshore power grid are not affected by the actions of the agent; their state transitions depend on the actual load and wind speed variations in the power system.

4) Reward

The reward function for the actions of the agent is set to be consistent with the DOPF objective function (17).

B. Structure of CGAN-SAC Algorithm

The SAC algorithm primarily addresses RL problems in discrete and continuous action spaces [28]. Unlike traditional RL, which necessitates a large number of samples for learning, the SAC algorithm is an off-policy algorithm designed for maximum entropy. In this context, entropy $H(X)$ represents the expected amount of information and is utilized to express the uncertainty of random variables:

$$H(X) = -\sum_{x_i \in X} P(x_i) \ln P(x_i) \tag{33}$$

where X is a discrete, randomly distributed system; and $P(x_i)$ is the probability that the system is in state x_i.

The DOPF problem is a multi-step decision-making problem, and an OPF solution exists under the current conditions at every moment. An algorithm is expected to explore all possible action spaces in the current network environment as thoroughly as possible to obtain the optimal strategy over a long timescale.

The objective of the SAC algorithm can be expressed as:

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{(\boldsymbol{s}_t, \boldsymbol{a}_t)\sim\rho_\pi}\left[\sum_t R\left(\boldsymbol{s}_t, \boldsymbol{a}_t\right) + \alpha H\left(\pi\left(\cdot|\boldsymbol{s}_t\right)\right)\right] \quad (34)$$

where $\mathbb{E}_{(\boldsymbol{s}_t, \boldsymbol{a}_t)\sim\rho_\pi}$ is the expectation of the reward from action $\boldsymbol{a}_t$ by strategy ρ_π in state $\boldsymbol{s}_t$; π is the policy, specifically indicating the set of rules that dictate the actions an agent should take when encountering different states; π^* is the optimal policy; R is the reward function; H is the entropy; and α is the weight of entropy used to balance the value of reward and entropy. The SAC algorithm aims to maximize reward expectations while maintaining the maximum action entropy.

Compared with traditional deep RL algorithms such as the deep Q network and DDPG, the SAC algorithm exhibits a superior capability for exploration, enhanced robustness, and faster training speed.

The CGAN-SAC algorithm comprises one actor network and four critic networks. Its structure is illustrated in Fig. 3.

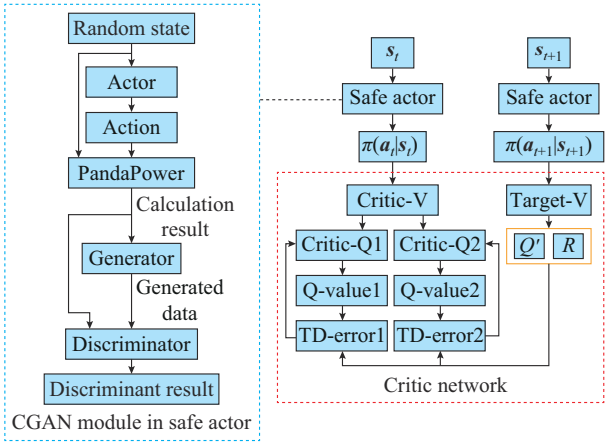

Fig. 3. Structure of CGAN-SAC algorithm.

The actor network represents the action estimation network. The critic-V network outputs the estimated value of the state. The target-V network outputs the estimated value of the next state Q' and reward R. The critic-Q network outputs the estimated value of the action. The purpose of using two identical networks (critic-Q1 and critic-Q2) is to increase the computing speed. TD-error is the temporal difference error, which is defined by the difference between the currently estimated value and the calculated target Q-value.

The CGAN-SAC algorithm can be divided into two major components: the actor network and the critic network. Each will be discussed in detail below.

1) Safe Actor Network

The traditional actor network consists of multiple dense layers. Its principle is to generate the possible actions of the agent through the system state and then input them into the critic network for evaluation. However, for the scheduling decision-making problem of the power system, the action set generated by the traditional actor network may result in power flow calculations that fail to converge or converge to an inoperable solution of the power system. To address this issue, the Lagrangian method is usually used to add system power flow constraints and operational constraints into the return function. However, there remains a slight probability of obtaining an inoperable solution.

The CGAN algorithm [29] is a generative adversarial network algorithm [30] with conditional constraints, which adds conditions to the model through additional information, and sends constraints to the discriminator and generator models, thereby realizing a CGAN. The initial training data are randomly generated, and the power flow is calculated by the simulation tool PandaPower to determine whether the action meets the constraint. The objective function is as follows:

$$\min_G \max_D V\left(D, G\right) = \mathbb{E}_{x\sim p_{data}(x)}\left[\ln D\left(x|y\right)\right] + \mathbb{E}_{z\sim p_z(z)}\left[\ln\left(1 - D\left(G\left(z|y\right)\right)\right)\right] \quad (35)$$

where $V\left(D, G\right)$ is the objective function formed by the combination of D and G; G is the generator function; D is the discriminator function; $\mathbb{E}_{x\sim p_{data}(x)}$ is the expected value for the data x based on its probability distribution $P_{data}(x)$; $\mathbb{E}_{z\sim p_z(z)}$ is the expected value for data z based on its probability distribution $p_z(z)$; and x, y, and z denote the inputs (or outputs) of the network, e.g., $D\left(x|y\right)$ indicates that the input of the discriminator network is x and the output is y.

The discriminator first judges the system actions generated by the generator to determine whether they meet the system power flow and operation constraints, and then inputs them into the critic network in the SAC algorithm. The CGAN module in the safe actor is shown in the blue dashed block of Fig. 3. CGAN module, which is a pre-trained model, does not significantly contribute to an increase in computational time when the actor network in the SAC algorithm is replaced.

2) Critic Network

The critic network is shown in the red dashed block of Fig. 3, which encompasses the critic-V and critic-Q components, along with the policy evaluation and policy improvement processes.

The Q-value and V-value are generated respectively by the critic-Q and critic-V modules. The Q-value focuses on the value of taking a specific action in a given state, while the V-value pertains to the overall value of a state.

From (34), it can be observed that the objective function of the SAC algorithm differs from standard RL in that the SAC algorithm includes an additional term, entropy. Correspondingly, the value function of the SAC algorithm is augmented with entropy. The Q-value function $Q^{\pi}_{soft}(s, a)$ and V-value function $V^{\pi}_{soft}(s)$ are as follows:

$$Q^{\pi}_{soft}(s, a) = \mathbb{E}_{(\boldsymbol{s}_t, \boldsymbol{a}_t)\sim\rho_\pi}\left[\sum_{t=0}^{\infty}\gamma^t R\left(\boldsymbol{s}_t, \boldsymbol{a}_t\right) + \alpha\sum_{t=1}^{\infty}\gamma^t H\left(\pi\left(\cdot|\boldsymbol{s}_t\right)\right)\middle| s_0 = s, a_0 = a\right] \quad (36)$$

$$V^{\pi}_{soft}(s) = \mathbb{E}_{(\boldsymbol{s}_t, \boldsymbol{a}_t)\sim\rho_\pi}\left[\sum_{t=0}^{\infty}\gamma^t R\left(\boldsymbol{s}_t, \boldsymbol{a}_t\right) + \alpha H\left(\pi\left(\cdot|\boldsymbol{s}_t\right)\right)\middle| s_0 = s\right] \quad (37)$$

where γ^t is the discount factor; s_0 and a_0 are the initial states of state s and action a, respectively.

The relationship between the Q-value function and V-val-

ue function can be deduced from (36) and (37):

$$Q_{soft}^{\pi}(s,a)=\mathbb{E}_{(s_t,a_t)\sim\rho_\pi}\left[R(s,a)+\gamma V_{soft}^{\pi}(s')\right] \tag{38}$$

$$V_{soft}^{\pi}(s)=\mathbb{E}_{a\sim\pi}\left[Q_{soft}^{\pi}(s,a)-\alpha\ln\pi(a|s)\right] \tag{39}$$

where $\mathbb{E}_{a\sim\pi}$ is the expected value for action a based on its probability policy π.

Based on (38) and (39), the iterative equations for the Q-value and V-value can be derived.

Policy evaluation and policy improvement are two crucial steps in the policy iteration algorithm of RL. Policy evaluation involves computing the value function under the current policy and providing an estimate of state values. Policy improvement, based on this estimate, enhances the policy by selecting actions with the maximum value function.

As mentioned earlier, the value iteration equations for policy evaluation are given by (38) and (39), and policy improvement is optimized using (40).

$$\pi_{new}=\arg\min_{\pi\in\Pi}D_{KL}\left(\pi(\cdot|\boldsymbol{s}_t)\middle\|\frac{\exp\left(\frac{Q_{soft}^{\pi_{old}}(\boldsymbol{s}_t,\cdot)}{\alpha}\right)}{\exp\left(\frac{V_{soft}^{\pi_{old}}(\boldsymbol{s}_t)}{\alpha}\right)}\right) \tag{40}$$

where Π is the set of policies; π_{new} and π_{old} are the new and old policies, respectively; and D_{KL} is Kullback-Leibler (K-L) divergence.

These two steps are repeated in an alternating manner, gradually converging the policy towards the optimal one. This process ensures that the agent can attain the maximum long-term reward in the environment.

C. Framework of DOPF Solved by CGAN-SAC Algorithm

The flowchart in Fig. 4 illustrates the solving process for the DOPF problem using the CGAN-SAC algorithm.

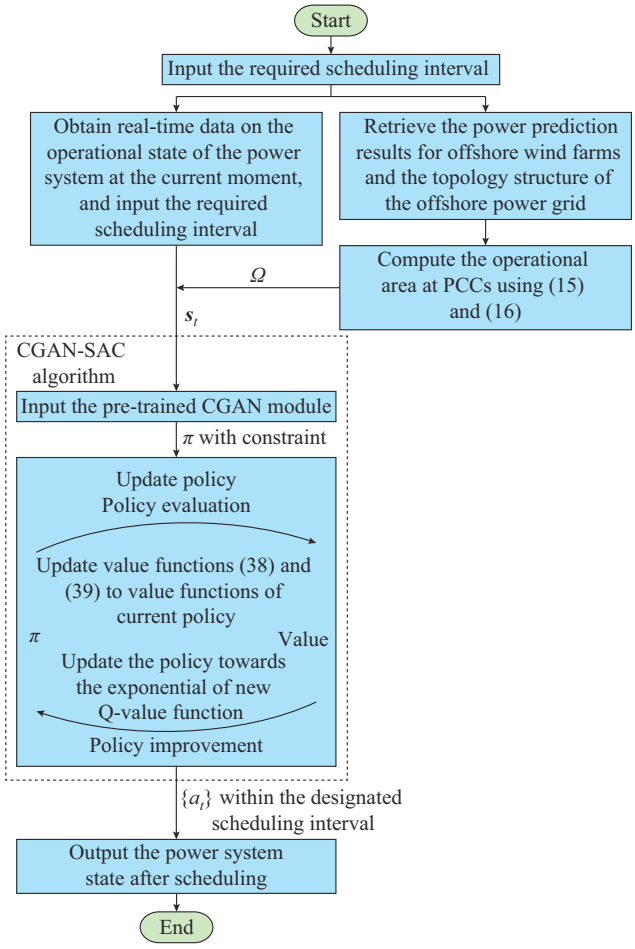

Fig. 4. Flowchart of solving process for DOPF problem using CGAN-SAC algorithm.

V. Example Analysis

A. Case Study

The example adopts the modified IEEE 118-node system, and the access capacity of the offshore wind power is 2090.5 MW, with a penetration rate of 18.6%. The data of the IEEE 118-node system can be found in [31].

The topology of the offshore power grid in the example is shown in Appendix A Fig. A1. Nine offshore wind farms are merged into onshore nodes 63, 81, and 108 through three PCCs. As the wind power prediction method is not the focus of this paper, it will not be repeated here. For the modified IEEE 118-node system, the four-hour period is divided into 16 time intervals, each lasting 15 min, for optimization purposes. The installed capacity of the wind farms and the power prediction results for the 16 time intervals can be found in Appendix A Table AI. The load data of each node are randomly simulated with 5% variance, the climbing rate of each PCC is set to be a maximum of 10 MW per 15 min, and the wind abandonment loss caused by the PCC of the offshore power grid is 1.5 $/MWh.

According to the wind power prediction results, the operational area at the three PCCs of the offshore power grid at a certain moment is shown in Fig. 5, where P_{63}, P_{81}, and P_{108} are the active power of the nodes 63, 81, 108, respectively.

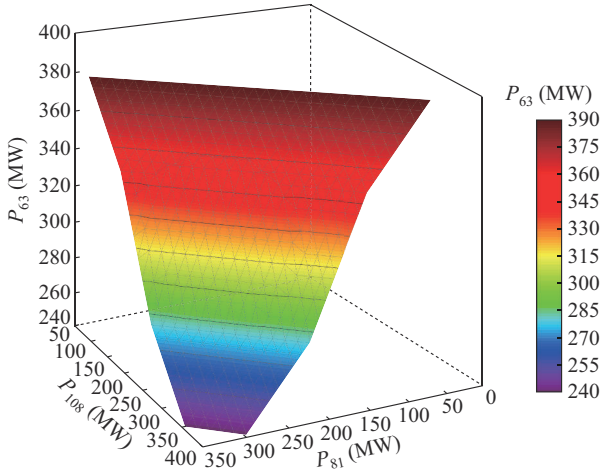

Fig. 5. Operational area at three PCCs of offshore power grid.

The plane enclosed by the middle plane of Fig. 5 and $P_{63}=0$ represents the operational area at the PCC of the offshore power grid under the power prediction result at that mo-

ment. The integration of offshore wind power cluster interconnections into the onshore power grid through multiple PCCs can effectively enhance the offshore power dispatching capability among PCCs. Due to the low inertia and rapid regulation speed of wind turbines, power transfer over a broader range can be quickly achieved.

The CGAN-SAC algorithm is employed to solve the problem across 16 time intervals. Following the verification of the results through power flow calculations, we found that all results are convergent, thus satisfying the power flow constraints. The scheduling results obtained by the proposed model over the 16 time intervals is shown in Fig. 6.

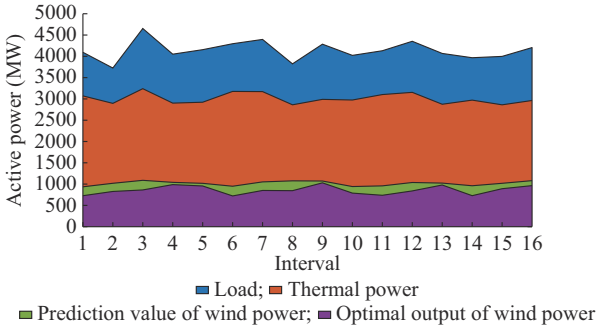

Fig. 6. Scheduling results obtained by proposed model over 16 time intervals.

Due to the large number of generator units in the IEEE 118-node system, individual output details for each unit are not presented separately. Owing to the variations in the power system load, constraints on synchronous unit ramp rates and PCC ramp rates, and the reduction of system standby costs through wind curtailment, the offshore power grid only operates at full capacity during specific periods. Consequently, a power dispatching plan for each offshore wind farm can be developed to achieve optimal operation of the offshore power grid.

Figure 7 depicts the calculation results of node voltage of the modified IEEE 118-node system during a random time interval. All node voltages are observed to be within the allowable range indicated by the red line. Similar conclusions hold for the remaining time intervals; for brevity, they are not presented.

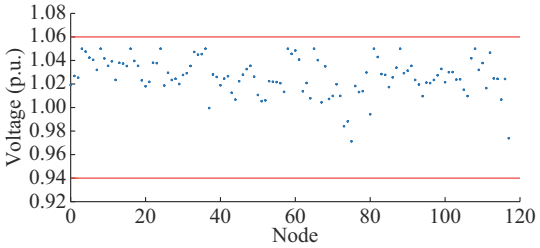

Fig. 7. Calculation results of node voltage of IEEE 118-node system.

To further illustrate the effectiveness of the CGAN-SAC algorithm, the following three similar algorithms are compared.

1) Algorithm 1: an OPF model considering the constraint of unit ramp rate between time intervals is solved by MATPOWER.

2) Algorithm 2: a DOPF model is solved by the commercial software GAMS with CONOPT solver, wherein constraints are addressed through the application of the Lagrangian method.

3) Algorithm 3: a DOPF model described by RL environment is solved using the SAC algorithm, wherein constraints are addressed through the application of the Lagrangian method.

The comparison of power flow results in each branch for the first time interval is illustrated in Fig. 8.

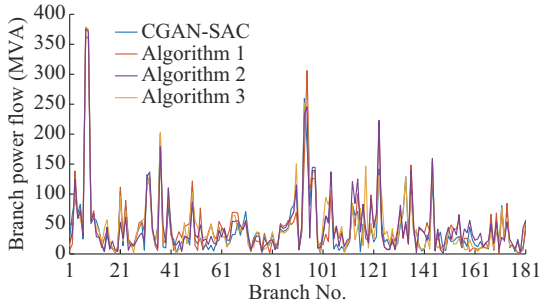

Fig. 8. Comparison results of power flow in each branch.

As observed from Fig. 8, the distribution of power flow in each branch generated by the three algorithms in the first time interval is similar. However, the optimization results of Algorithm 2 and the CGAN-SAC algorithm focus more on the overall revenue of the system over a long timescale, resulting in less economical operation in the early stages of the system compared with that of Algorithm 1. A comparison of the system operating costs of the different algorithms for each time interval is displayed in Fig. 9.

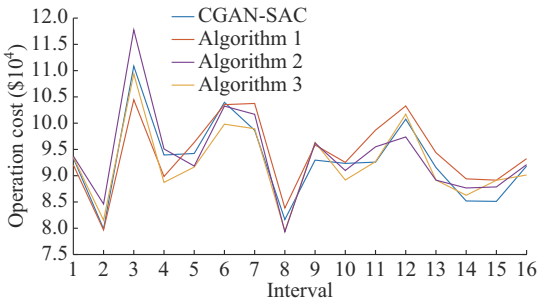

Fig. 9. Comparison of system operating costs with different algorithms in different time intervals.

When comparing the operating costs of continuous time intervals, we found that Algorithm 1 focuses more on optimal scheduling in the current system state, resulting in a more economical solution over the short term. However, since this algorithm does not consider the impact on subsequent stages, its operating costs gradually increase compared with those of Algorithm 2, Algorithm 3, and the CGAN-SAC algorithm, which are based on the DOPF problem and consider the environment composed of the current system state and the load in the next moment.

B. Comparison of Calculation Algorithms

Table I presents a performance comparison of the four algorithms.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS

Algorithm	Operating cost ($)	Constraint	Calculation time (s)
Algorithm 1	1510020	Complete satisfaction	735
Algorithm 2	1504175	The minimal probability of dissatisfaction	77
Algorithm 3	1477614	The minimal probability of dissatisfaction	12
CGAN-SAC	1488897	Complete satisfaction	25

The results of Algorithm 1 satisfy the system constraints but exhibit the longest calculation time and the highest running cost. Algorithm 3 has the lowest average operating cost and the fastest calculation speed; however, there is a small probability that it will not meet the system constraints, mainly in the ramp rate of the balance nodes. Algorithm 2 also has a slight possibility of failing to meet the constraints, with no apparent advantage in operating cost and extended calculation time. The proposed CGAN-SAC algorithm has a slightly inferior calculation speed and average operating cost compared with Algorithm 3, but its results fully satisfy the system constraints. The computational results presented in this paper are obtained based on the AMD Ryzen 7 3700X, NVIDIA GTX1660, and 32 GB RAM platform.

VI. CONCLUSION

The safety of the action policy is crucial in applying RL to solve the DOPF problem. In this paper, by utilizing the CGAN algorithm to reconstruct the action space and enhance the SAC algorithm, a DOPF method for offshore wind farms considering multiple PCCs is proposed. The power system is divided into offshore and onshore subnetworks. The safety of the offshore power grid is ensured by the offshore operational area. The pretrained CGAN algorithm ensures the safety of the onshore system operation. Compared with the traditional SAC algorithm and GAMS commercial solver, the calculation results of the proposed CGAN-SAC algorithm can completely meet system security constraints, without significant differences in the running cost and the solution time. Compared with the OPF problem solved by MATPOWER, the proposed CGAN-SAC algorithm can effectively shorten the running speed and reduce the operating cost by 0.986% on average in all time intervals.

APPENDIX A

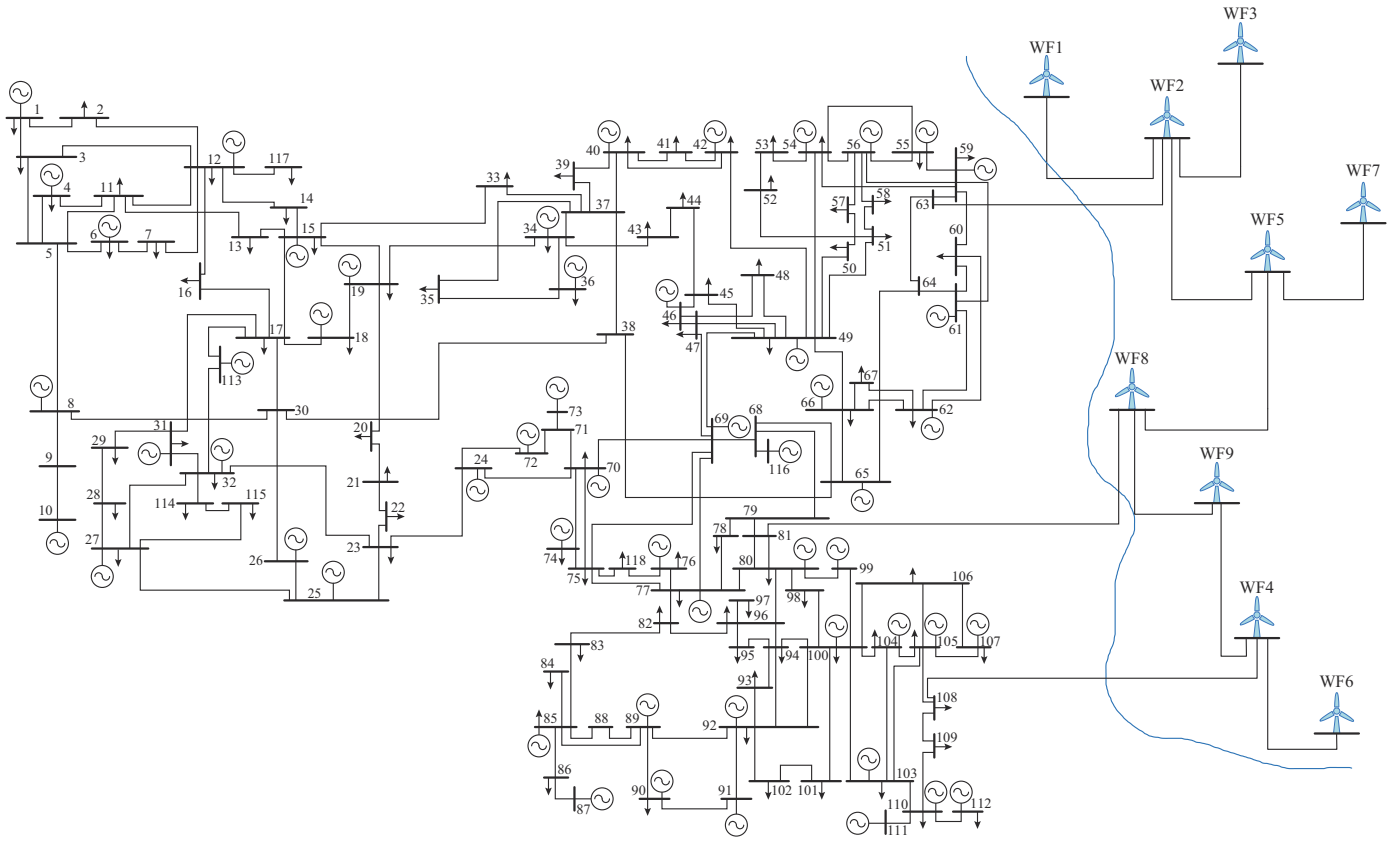

Fig. A1. Modified IEEE 118-node system.

TABLE AI
RATED CAPACITY AND POWER PREDICTION RESULTS

Wind farm No.	Rated capacity (MW)	Power prediction result (MW)															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
WF1	300.0	144.53	160.21	162.39	130.63	148.61	135.68	133.40	111.57	133.03	118.85	141.57	128.97	147.69	134.58	131.46	169.56
WF2	200.0	61.67	106.94	103.56	99.98	84.00	99.71	117.09	107.67	117.12	90.66	119.30	91.40	81.65	56.84	105.76	88.70
WF3	150.0	63.15	26.81	99.82	85.95	73.72	51.10	82.11	59.41	66.91	70.79	84.98	86.37	80.39	72.98	83.17	70.52
WF4	125.0	64.84	82.92	64.93	71.72	64.36	49.22	48.46	77.22	91.51	64.33	20.93	95.64	87.28	51.39	71.66	54.02
WF5	300.0	135.06	151.26	161.25	164.32	124.99	121.14	137.18	171.92	147.03	153.84	119.19	142.22	132.07	145.85	120.22	128.14
WF6	300.0	139.89	163.82	166.66	151.23	153.62	157.66	169.04	139.28	153.47	126.26	86.55	137.48	153.01	137.90	165.21	173.97
WF7	202.0	127.34	100.70	108.46	105.00	98.88	86.00	106.69	125.41	109.22	74.83	106.80	101.27	98.04	85.75	93.12	107.17
WF8	213.5	89.03	94.13	92.09	85.27	130.79	107.44	107.53	117.81	97.43	133.79	124.70	92.33	125.61	130.41	107.28	127.66
WF9	300.0	112.08	133.30	130.49	147.93	140.18	143.11	151.22	167.29	157.68	110.18	155.45	163.18	118.99	146.51	141.80	162.46

REFERENCES

[1] GWEC. (2022, Aug.). Global offshore wind report 2022. [Online]. Available: https://gwec. net/wp-content/uploads/2022/06/GWEC-Global-Offshore-Wind-Report-2022.pdf

[2] M. Koivisto, J. Gea-Bermúdez, and P. Sørensen, "North Sea offshore grid development: combined optimisation of grid and generation investments towards 2050," *IET Renewable Power Generation*, vol. 14, no. 8, pp. 1259-1267, Jun. 2020.

[3] Y. Liu, Y. Fu, L. Huang *et al*., "Optimization of offshore grid planning considering onshore network expansions," *Renewable Energy*, vol. 181, pp. 91-104, Jan. 2022.

[4] J. G. Dedecca and R. A. Hakvoort, "A review of the North Seas offshore grid modeling: current and future research," *Renewable and Sustainable Energy Reviews*, vol. 60, pp. 129-143, Jul. 2016.

[5] K. Xie and Y. Song, "Dynamic optimal power flow by interior point methods," *IET Proceedings: Generation, Transmission, and Distribution*, vol. 148, no. 1, p. 76, Jul. 2001.

[6] S. Gill, I. Kockar, and G. W. Ault, "Dynamic optimal power flow for active distribution networks," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 121-131, Jan. 2014.

[7] Y. Zhou, W. -J. Lee, R. Diao *et al*., "Deep reinforcement learning based real-time AC optimal power flow considering uncertainties," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 5, pp. 1098-1109, Sept. 2022.

[8] Q. Ma and C. Deng, "Simplified deep reinforcement learning based volt-var control of topologically variable power system," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 5, pp. 1396-1404, Sept. 2023.

[9] Y. Yu, "Security region of bulk power system," in *Proceedings of International Conference on Power System Technology*, Kunming, China, Oct. 2002, pp. 13-17.

[10] H. Jia, X. Yu, Y. Yu *et al*., "Power system small signal stability region with time delay," *International Journal of Electrical Power & Energy Systems*, vol. 30, no. 1, pp. 16-22, Jan. 2008.

[11] S. Chen, Q. Chen, Q. Xia *et al*., "Steady-state security assessment method based on distance to security region boundaries," *IET Generation*, *Transmission & Distribution*, vol. 7, no. 3, pp. 288-297, Mar. 2013.

[12] J. Xiao, W. Gu, C. Wang *et al*., "Distribution system security region: definition, model and security assessment," *IET Generation*, *Transmission & Distribution*, vol. 6, no. 10, p. 1029, Oct. 2012.

[13] J. Xiao, L. Zuo, G. Zu *et al*., "Model of distribution system security region based on power flow calculation," *Proceedings of the CSEE*, vol. 37, no. 17, pp. 4941-4949, Sept. 2017.

[14] J. Xiao, W. Guo, G. Zu *et al*., "An observation method based on $N-1$ simulation for distribution system security region," in *Proceedings of 2016 IEEE PES General Meeting*, Boston, USA, Jul. 2016, pp. 1-5.

[15] H. Jiao, X. Lin, J. Xiao *et al*., "Concavity-convexity of distribution system security region – part I: observation results and mechanism," *Applied Energy*, vol. 342, p. 121169, Jul. 2023.

[16] D. Silver, A. Huang, C. J. Maddison *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484-489, Jan. 2016

[17] A. Stooke, J. Achiam, and P. Abbeel. (2020, Jul.). Responsive safety in reinforcement learning by PID Lagrangian methods. [Online]. Available: https://arxiv.org/abs/2007.03964

[18] Y. Du and F. Li, "Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1066-1076, Mar. 2020.

[19] Q. Zhang, K. Dehghanpour, Z. Wang *et al*., "A learning-based power management method for networked microgrids under incomplete information," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1193-1204, Mar. 2020.

[20] F. S. Gorostiza and F. M. Gonzalez-Longatt, "Deep reinforcement learning-based controller for SOC management of multi-electrical energy storage system," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5039-5050, Nov. 2020.

[21] H. Li, Z. Wan, and H. He, "Constrained EV charging scheduling based on safe deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2427-2439, May 2020.

[22] F. Ming, F. Gao, K. Liu *et al*., "Constrained double deep Q-learning network for EVs charging scheduling with renewable energy," in *Proceedings of 2020 IEEE 16th International Conference on Automation Science and Engineering*, Hong Kong, China, Oct. 2020, pp. 636-641.

[23] S. Paudyal, C. A. Canizares, and K. Bhattacharya, "Optimal operation of industrial energy hubs in smart grids," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 684-694, Mar. 2015.

[24] R. E. Martinez, B. E. Z. Pérez, A. E. Reza *et al*., "Optimal planning, design and operation of a regional energy mix using renewable generation – study case: Yucatan Peninsula," *International Journal of Sustainable Energy*, vol. 40, no. 3, pp. 283-309, Mar. 2021.

[25] A. Ray, J. Achiam, and D. Amodei. (2019, May). Benchmarking safe exploration in deep reinforcement learning. [Online]. Available: https://cdn.openai.com/safexp-short.pdf

[26] Z. Yan and Y. Xu, "Real-time optimal power flow: a Lagrangian based deep reinforcement learning approach," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 3270-3273, Jul. 2020.

[27] Y. Chow, O. Nachum, E. Duenez-Guzman *et al*., "A Lyapunov-based approach to safe reinforcement learning," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Montréal, Canada, Dec. 2018, pp. 8103-8112.

[28] T. Haarnoja, A. Zhou, K. Hartikainen *et al.* (2018, Dec.). Soft actor-critic algorithms and applications. [Online]. Available: https://arXiv: 1812.05905

[29] I. Goodfellow, J. Pouget-Abadie, M. Mirza *et al.*, "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Montreal, Canada, Dec. 2014, pp. 2672-2680.

[30] J. Zhao, F. Li, H. Sun *et al.*, "Self-attention generative adversarial network enhanced learning method for resilient defense of networked microgrids against sequential events," *IEEE Transactions on Power Systems*, vol. 38, no. 5, pp. 4369-4380, Sept. 2023.

[31] Illinois Institute of Technology. (2023, Apr.). Index of data. [Online]. Available: http://motor.ece.iit.edu/data/

Yang Fu received the M.S. degree from Southeast University, Nanjing, China, in 1993, and the Ph.D. degree from Shanghai University, Shanghai, China, in 2007. He is currently a Professor at Shanghai University of Electric

Power, Shanghai, China. His research interests include power system analysis and new energy planning and optimization.

Zixu Ren received the B.S. and M.S. degrees in electrical engineering from Shanghai University of Electric Power, Shanghai, China, in 2015 and 2019, respectively. He is currently pursuing the Ph.D. degree in electrical engineering from Shanghai University of Electric Power. His research interests include fault diagnosis and new energy planning and optimization.

Shurong Wei received the B.S. degree from Northeast Electric Power University, Jilin, China, in 2002, the M.S. degree from Hohai University, Nanjing, China, in 2005, and the Ph.D. degree from Shanghai University, Shanghai, China, in 2016. She is currently a Professor at Shanghai University of Electric Power, Shanghai, China. Her research interests include offshore wind farm planning, design, fault diagnosis, and operation maintenance.

Lingling Huang received the B.E. degree from Xi'an Jiaotong University, Xi'an, China, in 2004, the M.S. degree from Zhejiang University, Hangzhou, China, in 2006, and the Ph. D. degree from Shanghai University, Shanghai, China, in 2018. She is currently a Professor with Shanghai University of Electric Power, Shanghai, China. Her research interests include offshore wind farm planning, design, and optimal operation.

Fangxing Li received the B.S.E.E. and M.S.E.E. degrees from Southeast University, Nanjing, China, in 1994 and 1997, respectively, and the Ph.D. degree from Virginia Tech, Blacksburg, USA, in 2001. He is currently the James W. McConnell Professor with the University of Tennessee, Knoxville, USA. His current research interests include electricity market, renewable energy integration, demand response, microgrid, and power system computing.

Yang Liu received the Ph.D. degree from Shanghai University, Shanghai, China, in 2023. Currently, he is a Lecturer with Shanghai University of Electric Power, Shanghai, China. His research interests include offshore wind farm planning, design, and optimal operation.