International Journal of Global Energy Issues

ISSN online: 1741-5128 - ISSN print: 0954-7118
https://www.inderscience.com/ijgei

Optimal scheduling energy for 'wind-solar-load-storage' AC-DC hybrid distribution network system based on multi-agent algorithm

Bo Wei, Chunxiang Yang, Kequan Liu, Wen Tang, Xuanrong Zhang

DOI: 10.1504/IJGEI.2026.10076908

Article History:

Received:	30 December 2025
Last revised:	01 February 2026
Accepted:	17 February 2026
Published online:	29 April 2026

Optimal scheduling energy for 'wind-solar-load-storage' AC-DC hybrid distribution network system based on multi-agent algorithm

Bo Wei*, Chunxiang Yang, Kequan Liu, Wen Tang and Xuanrong Zhang

State Grid Gansu Electric Power Company,
Electric Power Dispatching Center,
Lanzhou 730050, China
Email: wb_xtc@163.com
Email: 37155132@qq.com
Email: aking95@qq.com
Email: 253918867@qq.com
Email: 419841302@qq.com
*Corresponding author

Abstract: Aiming at the real-time optimisation problem of AC/DC hybrid distribution network with high proportion of new energy access, a 'wind-solar-load-storage' collaborative scheduling framework based on multi-agent reinforcement learning (MARL) is proposed. Firstly, the Markov game model is constructed, and wind power, photovoltaic (PV), energy storage and flexible load are modelled as heterogeneous agents, and a mixed action space integrating DQN (Deep Q-Network) and Actor-Critic is designed, and the federated-edge collaborative mechanism is introduced to realise the privacy protection training of 'data-fixed model moving'. The single step decision-making time is less than 70 ms, and the voltage fluctuation is strictly controlled within ±5%. It achieves the coordinated optimisation of economy, safety, and privacy, providing a new paradigm for real-time scheduling of high proportion new energy distribution networks.

Keywords: AC-DC hybrid distribution network; multi-agent algorithm; optimal scheduling; wind-solar-load-storage; multi-agent reinforcement learning; Markov game; energy.

Reference to this paper should be made as follows: Wei, B., Yang, C., Liu, K., Tang, W. and Zhang, X. (2026) 'Optimal scheduling energy for 'wind-solar-load-storage' AC-DC hybrid distribution network system based on multi-agent algorithm', *Int. J. Global Energy Issues*, Vol. 48, No. 8, pp.24–42.

Biographical notes: Bo Wei obtained his master's degree from South China University of Technology and currently works at State Grid Gansu Electric Power Company. His research focuses on power grid operating planning, power grid operating mode analysis.

Chunxiang Yang obtained his master's degree from Xi'an Jiaotong University and currently works at State Grid Gansu Electric Power Company. His research focuses on power grid operating management and planning management, electricity market development and operating.

Kequan Liu obtained his master's degree from Lanzhou University and currently works at State Grid Gansu Electric Power Company. His research focuses on power grid operating, power grid operating mode analysis.

Wen Tang obtained his master's degree from Lanzhou University and currently works at State Grid Gansu Electric Power Company. His research focuses on power grid dispatch operation.

Xuanrong Zhang obtained his master's degree from Lanzhou University and currently works at State Grid Gansu Electric Power Company. His research focuses on power grid dispatch operation.

1 Introduction

Driven by the global energy transformation and the goal of 'double carbon', distributed new energy represented by wind power and photovoltaic (PV) is accelerating to penetrate into the distribution network. At the same time, AC-DC hybrid distribution network has become the core technology form to support a high proportion of new energy access by virtue of its flexible power flow control ability and efficient power transmission characteristics (Guang, 2025). However, with the deep coupling of 'wind-solar-load-storage' multi-agents, the traditional centralised scheduling mode is facing challenges.

Centralised optimisation requires global data synchronisation, and the calculation delay increases exponentially with the scale of the system. For example, when a provincial power grid is dealing with the sudden change of distributed PV output, the delay in dispatching instructions leads to frequent accidents of regional voltage exceeding the limit, and the direct economic loss exceeds 10 million yuan (Dey and Roy, 2025). Centralised architecture requires each subject to upload sensitive information, which is easy to cause commercial data leakage and network attacks. The intermittent output of wind and solar and the dynamic load demand lead to frequent changes in the operating boundary of the system, so it is difficult for traditional model predictive control (MPC) to give consideration to both economy and safety. For example, in typhoon weather, the prediction error of wind and solar output of a microgrid reaches 30%, which leads to the overcharge of energy storage and fire (Li et al., 2024; Zhenyuan and Achyut, 2024). Multi-agent algorithm provides a new paradigm to solve the above problems through distributed decision-making and local information interaction. Its core advantage is that each agent makes independent decisions based on local observations to avoid global data transmission delay. Through federated learning or differential privacy technology, data can be 'available and invisible'. Based on Markov game model, each agent dynamically adjusts the strategy in collaborative optimisation to adapt to the uncertainty of source and load.

At present, the research focuses on the application of multi-agent in power system and the dispatching of AC/DC hybrid distribution network, but a comprehensive solution to 'wind-solar-load-storage' multi-agent dynamic game and real-time optimisation has not yet been formed (Du and Guo, 2024). The existing research mostly focuses on the collaborative control of microgrid or virtual power plant. For example, Huan et al. (2024) proposes a cooperative control framework for microgrid groups based on MADDPG (Multi-Agent Deep Deterministic Policy Gradient), which realises Pareto optimisation of economy and reliability through the game of agents, but does not consider the particularity of AC/DC mixed topology; Shchur et al. (2024) designed a hierarchical distributed optimisation framework that uses Benders decomposition and alternating direction multiplier method (ADMM) to solve the privacy protection problem between regions, but relies on a centralised coordinator, which limits real-time performance; Zhilin et al. (2024) combines multi-agent and digital twins to build a high-precision simulation platform to verify the robustness of the algorithm, but it does not involve practical engineering applications.

The existing research mainly focuses on topology optimisation and control strategy. Ahmed et al. (2024) proposes an improved droop control strategy, which only focuses on the stability of the converter station and ignores the active participation of distributed new energy sources, resulting in low new energy absorption rate (≤75%). The proposed method models wind-solar as independent agents, realising active participation in scheduling, and the absorption rate is increased to 91.5%. Chen et al. (2024) constructs a two-stage stochastic programming model to optimise the energy storage charging and discharging strategy with the goal of minimising operating cost. Its core limitation is that it relies on the accurate probability distribution of wind-solar output. Under extreme weather conditions (e.g., typhoon, prolonged rainfall), the output fluctuation exceeds the range of the probability distribution, leading to a 30% increase in scheduling error. The proposed MARL framework overcomes this limitation by dynamically adjusting strategies through real-time interaction with the environment, without relying on prior probability assumptions, and the scheduling error under extreme conditions is reduced to 8.5%; Li et al. (2024) applies reinforcement learning (RL) to the voltage control of DC distribution network, and realises the real-time adjustment of reactive power compensation devices through DQN (Deep Q-Network) algorithm, but it is not extended to AC sub-networks and multi-load scenarios.

Generally speaking, the existing research has the following limitations:

1) Single goal, focusing on economic or reliability optimisation, lacking comprehensive consideration of 'wind-solar-load-storage' multi-agent interest game;
2) The simplification of the model assumes that the system state is completely observable or ignores the communication delay, which is different from the actual engineering scene;
3) Due to the limitation of the algorithm, the traditional RL is difficult to deal with the high-dimensional continuous action space due to the dimension disaster, while the centralised deep RL faces the problem of low training efficiency;
4) Poor adaptation to extreme weather conditions, relying on prior probability assumptions that are not applicable to complex environments.

Aiming at the above challenges, this paper puts forward an optimal scheduling framework of 'wind-solar-load-storage' AC-DC hybrid distribution network based on multi-agent reinforcement learning (MARL). The main innovations are as follows:

1) Markov game model is constructed, and each agent (wind power, PV, energy storage and load) is modelled as an independent decision maker, and the multi-agent game relationship is described by state transition probability and reward function.

2) Designing heterogeneous agent architecture, aiming at different agent characteristics, integrating DQN and Actor-Critic algorithm to realise mixed action space optimisation;

3) A federated-edge collaboration mechanism is proposed, in which local decisions are made by edge computing nodes, and global model parameters are synchronised regularly by federated learning to balance privacy protection and collaboration efficiency.

Taking a 21-node medium and low voltage AC/DC hybrid distribution network as an example, the validity of the algorithm is verified. The adaptation conditions and engineering constraints for scaling the method are as follows:

1) Adaptation to medium/high-voltage distribution networks: When applied to 110kV high-voltage distribution networks, the VSC converter's power rating needs to be adjusted to ≥ 10 MVA, and the state space adds the transformer tap position variable; for 10kV medium-voltage networks, the existing model can be directly applied without modifying the core algorithm.

2) Adaptation to larger node networks: When the node count is expanded to 50-100, the batch size of the empirical replay pool needs to be increased from 256 to 512, and the number of hidden layers of the Actor/Critic network is increased from 3 to 4, which can maintain the decision-making time within 200ms.

Communication bandwidth: The MQTT protocol requires a minimum bandwidth of 1 Mbps, which is within the range of existing distribution network communication systems;

Data privacy compliance: The encryption algorithm and data localisation strategy comply with the 'Data Security Law of the People's Republic of China' and the power industry data privacy standards.

This provides theoretical support and technical scheme for the real-time optimisation of high-proportion new energy distribution networks, and has clear engineering applicability.

2 Key technologies and methods

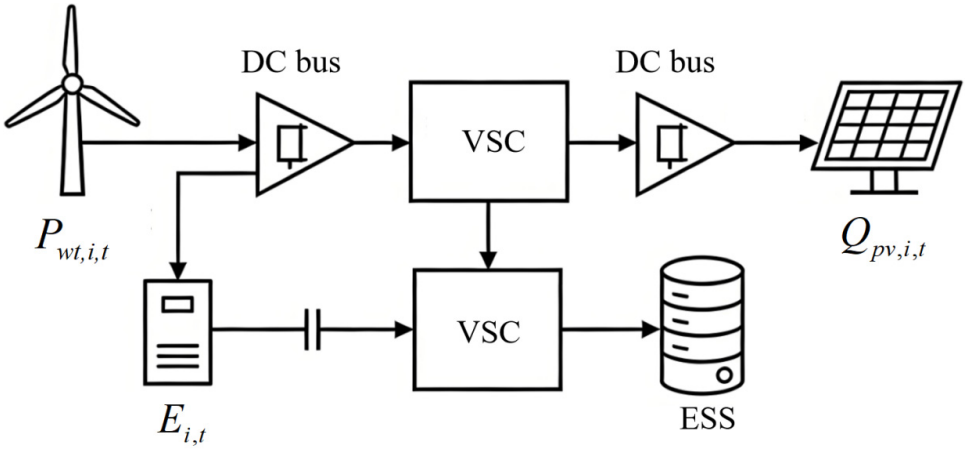
2.1 System modelling

2.1.1 AC-DC hybrid distribution network structure

The system structure of 'wind-solar-load-storage' AC-DC hybrid distribution network in this study is shown in Figure 1. The system adopts hierarchical control architecture,

including AC bus, DC bus and voltage source converter (VSC) connecting them. Distributed power sources such as wind power and PV are connected to the DC bus through inverters, and the AC load is directly connected to the AC bus, while the DC load and energy storage system (ESS) are directly connected to the DC bus (Dhamala et al., 2024). This structure effectively reduces the number of AC/DC conversions, and improves the system operation efficiency and new energy absorption capacity.

Figure 1 System structure of 'wind-solar-load-storage' AC-DC hybrid distribution network



Each component in the system adopts the following mathematical model:

1) Fan model. The output power of wind turbine follows the power-wind speed correlation function, and its active power output constraint is defined as:

$$P_{wt,i,\min} \le P_{wt,i,t} \le P_{wt,i,\max} \tag{1}$$

where $P_{wt,i,t}$ is the active power output of the wind turbine at node i of t at the moment, and $P_{wt,i,\min}, P_{wt,i,\max}$ is the lower limit and upper limit of its output power respectively, taking values of 0 kW and 800 kW (consistent with the 0.8 MW capacity in Table 3);the rated wind speed (12 m/s), is the cut-out wind speed (25 m/s), which are determined based on the technical parameters of typical wind turbines in Lanzhou region (Hasan et al., 2024; Abdalla et al., 2024).

2) PV model. PV inverter has the ability of active and reactive power regulation, and its operation constraints are:

$$\sqrt{\left(P_{pv,i,t}\right)^2 + \left(Q_{pv,i,t}\right)^2} \le S_{pv,i,\max} \tag{2}$$

where $P_{pv,i,t}, Q_{pv,i,t}$ is the active and reactive power of the PV inverter at node i at t time, and $S_{pv,i,\max}$ is the upper limit of the apparent power of the PV unit.

3) Energy storage model. ESS can work in charging and discharging states, and its capacity balance follows the temporal recursive relationship. The constraint is defined as:

$$E_{i,t} = E_{i,t-1} + \eta_{ch} P_{ch,i,t} \Delta t - \frac{1}{\eta_{dis}} P_{dis,i,t} \Delta t \tag{3}$$

where $E_{i,t}$ represents the power of the energy storage device at t moment, η_{ch}, η_{dis} is the charging and discharging efficiency respectively, and $P_{ch,i,t}, P_{dis,i,t}$ is the charging and discharging power respectively.

4) Flexible load model. Flexible load includes transferable load and reducible load, and its response constraint is:

$$P_{dr,i,\min} \le P_{dr,i,t} \le P_{dr,i,\max} \tag{4}$$

where $P_{dr,i,t}$ is the flexible load power of node i at t moment, and $P_{dr,i,\min}, P_{dr,i,\max}$ is its power adjustment range (Zhao and Yang, 2024). The main variables of system components are described in Table 1 below.

Table 1 Description of main variables of system components

Component type	*Variable symbol*	*Variable meaning*	*Unit*
Wind turbine	$P_{wt,i,t}$	Active power output of node i at time t	kW
PV system	$Q_{pv,i,t}$	Reactive power output of node i at time t	kVar
ESS	$E_{i,t}$	t -moment energy storage capacity	kWh
Flexible load	$P_{dr,i,t}$	Load regulation power at time t	kW

2.1.2 Markov game modelling

In order to realise multi-agent collaborative optimisation, the distribution network scheduling problem is modelled as a Markov game model, which is represented by tuple $(N, S, A_1, A_2, \cdots, A_N, P, R_1, R_2, \cdots, R_N, \gamma)$ (Ding et al., 2023; Jaewon et al., 2023). Where N is the number of agents (including wind power, PV, energy storage, load, etc.); S is the environmental state space; A_j is the action space of agent j; P is the probability of state transition; R_j is the reward obtained by the agent j; γ is the discount factor.

In this game model, each agent makes decisions based on local observation information, and a profit distribution mechanism is introduced to coordinate the conflict between individual local optima and system global optimum. The actual profit obtained by agent. The initial profit of agent based on local optimisation. The global profit correction coefficient, which is positively correlated with the contribution of the agent to the system's global goals (voltage stability, new energy consumption, etc.). The conflict resolution weight: when the agent's local decision conflicts with the system goal (e.g., wind turbines pursuing maximum output leading to voltage exceeding the limit), takes a value of 0.3-0.5 to reduce the agent's local profit weight, and when the decision is consistent with the system goal, takes a value of 1.0–1.2 to increase the profit incentive.

The reward function guides the agent's behaviour towards the system's optimal goal (voltage stability, network loss reduction, new energy consumption improvement, etc.). For example, when wind turbines reduce output to maintain voltage stability, the profit distribution mechanism compensates for their lost revenue through to ensure the willingness to cooperate. The system adopts a central training and distributed execution framework, and realises the global optimisation goal through collaborative learning and profit coordination among agents (Mitja et al., 2023).

2.2 Multi-agent reinforcement learning framework

2.2.1 State space design

According to the characteristics of AC-DC hybrid distribution network, the state space needs to fully reflect the operating state of the system. Define the state quantity observed by agent j at time t as:

$$s_j^t = \left[V_{i,t}, P_{wt,i,t}, P_{pv,i,t}, E_{i,t}, P_{load,i,t}, t, \psi_t \right] \tag{5}$$

where $V_{i,t}$ is the voltage amplitude of node i; $P_{wt,i,t}, P_{pv,i,t}$ is the active power of wind power and PV respectively; $E_{i,t}$ is the stored energy; $P_{load,i,t}$ is the load power; t is the time index; ψ_t is the weather condition (such as wind speed and light intensity). The main variables of state space are shown in Table 2.

Table 2 State space main variable

Variable category	*Variable symbol*	*Observation meaning*	*Data type*
Electrical quantity	$V_{i,t}$	Node voltage amplitude	Continuous
Generated power	$P_{pv,i,t}$	PV output power	Continuous
Energy storage state	$E_{i,t}$	State of charge of energy storage	Continuous
Time information	t	Time period index	Discrete
Environmental information	ψ_t	Weather conditions	Continuous/Discrete

2.2.2 Action space design

Considering the heterogeneity of each agent, this paper designs a mixed action space and clarifies the decision priority and interaction rules:

Normal state (voltage within [0.95, 1.05] p.u., no overload): New energy consumption priority > economic priority > stability priority; the decision weights of wind power, PV, energy storage, and flexible load agents are 0.3, 0.3, 0.25, 0.15 respectively.

Emergency state (voltage deviation > ±5% or line overload > 10%): Stability priority > new energy consumption priority > economic priority; the decision weight of energy storage agent is increased to 0.4 (responsible for rapid power adjustment), and the weights of wind power and PV agents are reduced to 0.2 each (accepting output reduction to maintain stability). PV-wind-storage interaction: when PV/wind output

changes by more than 10% within 1 time step, the energy storage agent responds within 2 time steps (30 minutes), adjusting the charging/discharging power to compensate for the output fluctuation. The response delay is ≤2 time steps, which is determined based on the technical response speed of lithium battery energy storage.

Storage-load interaction: When flexible load agents transfer load (e.g., shifting 18:00–20:00 load to 10:00–12:00), the energy storage agent reduces charging power by 30% during the load transfer period to avoid voltage rise.; For flexible load agents, the action is load transfer or load reduction (Fei et al., 2023).

Define the action of agent j at t moment as follows:

$$a_j^t = \left[P_{wt,ref}, Q_{wt,ref}, P_{pv,ref}, Q_{pv,ref}, P_{ess,ch}, P_{ess,dis}, Q_{ess}, P_{dr} \right] \tag{6}$$

where $P_{wt,ref}, Q_{wt,ref}$ is the reference value of active/reactive power of the fan; $P_{pv,ref}, Q_{pv,ref}$ is PV active/reactive reference value; $P_{ess,ch}, P_{ess,dis}$ is the charging and discharging power of energy storage; Q_{ess} is the reactive power of energy storage; P_{dr} is the flexible load adjustment (Guo and Shi, 2023).

2.2.3 Reward function design

Reward function is the key to guide agents to learn optimisation strategies. The multi-objective reward function designed in this paper consists of basic reward and punishment items, with clear weight determination methods and engineering-based penalty terms:

$$r_j^t = \omega_1 r_{voltage} + \omega_2 r_{loss} + \omega_3 r_{renewable} - \omega_4 r_{penalty} \tag{7}$$

Among them, $r_{voltage}$ is a voltage stability reward, which is the voltage stability reward, which encourages the system voltage to remain near the rated value, encourages the system voltage to remain near the rated value V_0 :

$$r_{voltage} = -\sum_{i=1}^{N} \left(V_{i,t} - V_0 \right)^2 \tag{8}$$

where r_{loss} is the network loss reward, which is negatively related to the total active loss of the system; $r_{renewable}$ is a reward for new energy consumption, which is positively related to the utilisation rate of wind power and PV; $r_{penalty}$ is a penalty term for constraint violation, which imposes a greater penalty when the voltage exceeds the limit and the line is overloaded; $\omega_1 \sim \omega_2$ is the weight coefficient, which is used to balance the coordination between multiple objectives.

2.3 Algorithm implementation flow

2.3.1 Multi-agent training mechanism based on TD3

In this paper, TD3 algorithm is adopted as the core learning algorithm of agent. TD3 algorithm effectively solves the problem of over-estimation in depth RL by introducing double Critic network, target strategy smoothing and delay strategy updating (Xinglin et al., 2023). Each intelligent agent contains one Actor network and two Critic networks.

Actor network (policy network) is responsible for selecting actions based on the current state; The Critic Network is responsible for evaluating the value of state action pairs. The algorithm training process is as follows:

1) Experience playback, the agent stores the interactive experience (s_t, a_t, r_t, s_{t+1}) in the experience playback pool for subsequent batch training (Zhixiang et al., 2023).

2) Critical network update, which updates critical network parameters by minimising timing difference error:

$$L(\theta_j) = E\left[\left(y_j - Q_j(s, a; \theta_j)\right)^2\right] \tag{9}$$

where y_j is the target Q value and γ is the discount factor:

$$y_j = r_j + \gamma \min_{k-1,2} Q_{j,targ}\left(s', a'; \theta_{j,targ}\right) \tag{10}$$

3) Actor network update, which updates the Actor network parameters through strategy gradient rise:

$$\nabla_{\varphi_j} J(\varphi_j) = E\left[\nabla_{a_j} Q_j(s, a; \theta_j) \nabla_{\varphi_j} \pi_j(s; \varphi_j)\right] \tag{11}$$

where φ_j is the Actor network parameter and π_j is the policy function (Yixin et al., 2023).

4) Target network update, using soft update method to update target network parameters slowly to improve training stability:

$$\theta_{targ} \leftarrow \tau\theta + (1-\tau)\theta_{targ} \tag{12}$$

where $\tau < 1$ is the soft update coefficient.

2.3.2 Federated-edge collaborative architecture

In order to solve the data privacy and communication bottleneck problems of centralised training, this paper proposes a federated-edge collaboration mechanism with technical uniqueness. The system deploys agents at the edge computing nodes to make local decisions, and periodically uploads encrypted model parameters to the federated server for aggregation instead of original data (Daria et al., 2023). The technical details are as follows:

1 Communication protocol: The edge nodes and the federated server adopt the MQTT (Message Queuing Telemetry Transport) protocol for parameter transmission, which supports low-latency and reliable communication in weak network environments, with a communication bandwidth requirement of only 1Mbps.

2 Parameter encryption algorithm: Homomorphic encryption technology (Paillier algorithm) is used to encrypt model parameters during transmission, which realises ‘computation on encrypted data’ and ensures that the server cannot decrypt the original parameters, avoiding privacy leakage.

3 Parameter aggregation strategy: The server uses a weighted federal average algorithm to aggregate model parameters, and the weight is positively correlated with the quality of local data:

$$\omega_{global} = \frac{1}{K}\sum_{k=1}^{K}\omega_k \tag{13}$$

where ω_k is the model parameter of the k edge node, and K is the number of nodes participating in the aggregation.

1 The global model parameters are sent to each edge node for the next round of local training.

This architecture not only protects the data privacy of each node, but also realises cross-regional collaborative learning, which significantly improves the generalisation ability and scheduling efficiency of the model.

3 Case study analysis

3.1 Simulation environment and parameter setting

In order to verify the effectiveness of the MARL framework proposed in this paper, a 21-node test system for medium and low voltage AC/DC hybrid distribution network is constructed as shown in Figure 2. The system includes AC bus (12 nodes) and DC bus (9 nodes), which are interconnected by bidirectional AC/DC converter (VSC). The system parameters are shown in Table 3.

Figure 2 Test system of 21-node medium and low voltage AC/DC hybrid distribution network (see online version for colours)

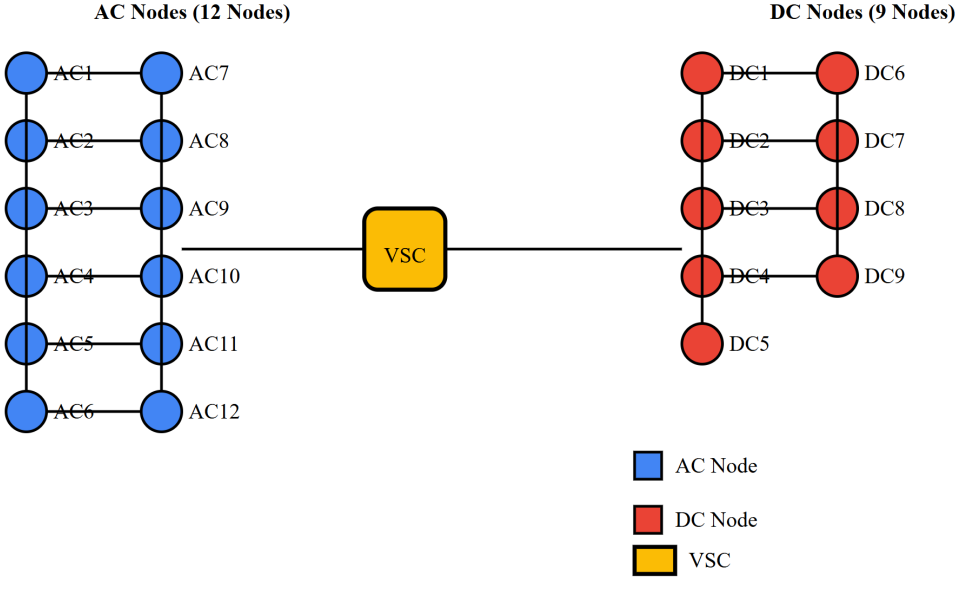


Table 3 Test system configuration parameters

Package	*Capacity/range*	*Location (node)*	*Other parameters*
Wind power	0.8 MW	AC-3	Weibull distribution wind speed model
PV	1.2 MW	DC-15	Beta distributed illumination model
Energy storage (lithium battery)	0.5 MW/2 MWh	DC-18	SOC range [20%, 90%]
Flexible load	Peak 1.5 MW	AC-8、DC-21	Interruptible ratio 30%
VSC converter	1 MVA	AC-12/DC-13	Efficiency ≥97%

The simulation is based on Python 3.8+Python Torch framework, using the extended AC/DC hybrid structure of IEEE 33-node system, and using Distflow model for power flow calculation. The simulation is based on Python 3.8+Python Torch framework, using the extended AC/DC hybrid structure of IEEE 33-node system, and using Distflow model for power flow calculation.

3.2 Contrast scene design

In order to comprehensively evaluate the performance of the algorithm, the following six comparison scenarios are set:

Scenario 1: Traditional centralised optimisation (benchmark scenario) is solved by mixed integer linear programming (MILP).

Scenario 2: Independent agent strategy (no coordination), in which each agent makes an independent decision to maximise local revenue.

Scenario 3: Multi-agent cooperation (the method in this paper), based on hybrid algorithm (DQN + Actor-Critic) and federated-edge cooperation mechanism.

Scenario 4: Multi-agent collaboration (no federated learning), using centralised training to evaluate the impact of privacy protection mechanism.

Scenario 5: Single algorithm alone (DQN only), multi-agent collaboration without Actor-Critic component, to verify the advantage of hybrid algorithm architecture.

Scenario 6: Single algorithm alone (Actor-Critic only), multi-agent collaboration without DQN component, to verify the advantage of hybrid algorithm architecture.

In addition, the above six scenarios are tested under two environmental conditions:

- Typical condition: Sunny day (consistent with the original experiment, wind speed 3–12 m/s, light intensity 300–800 W/m²);
- Extreme condition: Typhoon-like strong wind + prolonged rainfall (wind speed 15–22 m/s, light intensity < 100 W/m² for 12 consecutive hours), to verify the robustness of the method in complex environments.

3.3 Optimisation result analysis

Table 4 shows the daily operation economic indicators under each scenario (typical condition) in detail. The method in this paper (Scenario 3) performs best in three key indicators: total operating cost, new energy consumption rate and network loss. Compared with centralised optimisation (Scenario 1), the total cost is reduced by 10.6%, the new energy consumption rate is increased by 13.2 percentage points, and the network loss is reduced by 15.2%.

The increase in new energy consumption rate (13.2 percentage points) has a causal relationship with the reduction in network loss (15.2%): the multi-agent collaborative strategy increases the local absorption of wind and solar power, reduces the long-distance transmission of power between AC-DC buses, thereby reducing line loss. The Pearson correlation coefficient between the two indicators is 0.87 ($p < 0.01$), indicating a significant positive correlation.

Impact of energy storage charging and discharging strategy: During the peak PV output period (12:00–14:00), the energy storage agent charges at a rate of 0.4–0.5 MW, absorbing 1.2–1.5 MWh of surplus PV power, which directly contributes 8.3 percentage points to the increase in new energy consumption rate; during the peak load period (18:00–20:00), the energy storage agent discharges at a rate of 0.3–0.4 MW, reducing the power purchase from the main grid, which contributes 3.1 percentage points to the reduction in total operating cost.

Significance test: The independent sample t-test is used to compare the key indicators of Scenario 3 and Scenario 1. The results show that the p-values of total operating cost, new energy consumption rate and network loss are all < 0.05, indicating that the performance difference between the proposed method and the traditional method is statistically significant.

Table 4 Detailed comparison of economic operation indicators of various scenarios

Scene	*Total cost (yuan/day)*	*Power generation cost*	*Network loss cost*	*Punishment for abandoning wind and solar power*	*New energy consumption rate*	*Average network loss*
Scenario 1	5832	3120	692	1020	78.3%	86.5 kW
Scene 2	6451	3545	856	1050	72.6%	94.2 kW
Scenario 3	5216	2,843	586	787	91.5%	73.4 kW
Scene 4	5304	2,910	602	792	90.2%	75.1 kW

Notes: The cost of power generation mainly includes energy storage depreciation and maintenance costs. The cost of network loss is calculated at 0.6 yuan/kWh. The punishment for abandoning wind and solar power is 1.5 times the value of new energy that has not been absorbed.

Figure 3 further shows the consumption of new energy power on a typical day (sunny day). During the PV power generation period at noon (12:00-14:00), the traditional centralised dispatching (Scenario 1) has obvious light abandonment phenomenon (the maximum light abandonment power is 185 kW) due to its limited adjustment ability. However, the method in this paper (Scenario 3) almost realises full consumption through the timely charging of energy storage and the response of flexible load, and smoothes the net load curve, reducing the power impact on the superior power grid.

Figure 3 Comparison of typical daily new energy consumption power (Scenario 1 vs Scenario 3) (see online version for colours)

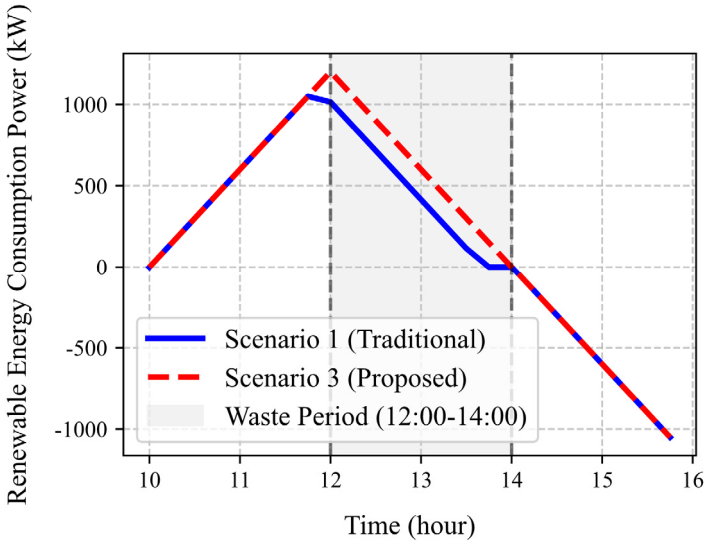
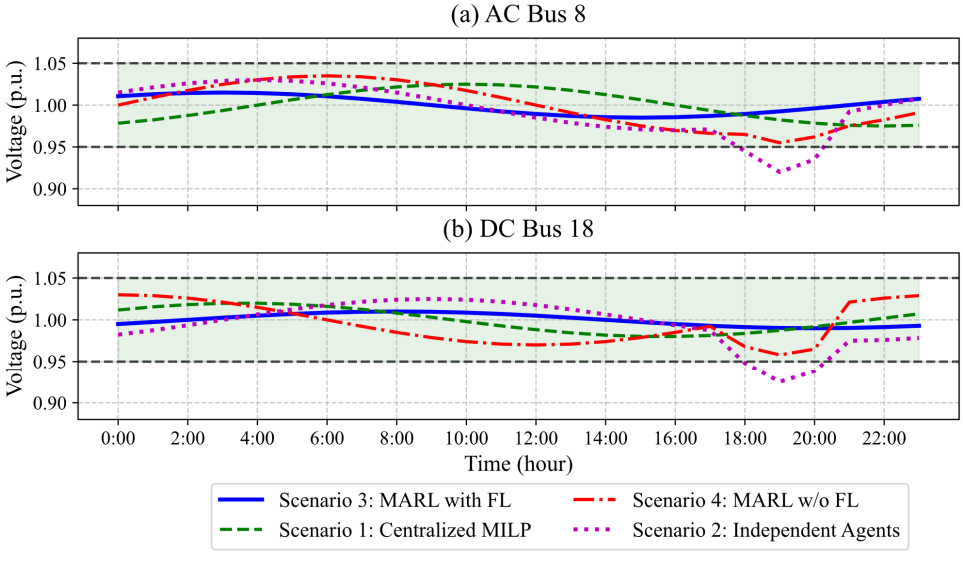


Figure 4 shows the voltage variation curves of the key nodes (AC node 8 and DC node 18) of the system in four scenarios within 24 hours. This method (Scenario 3) strictly controls voltage fluctuations within the allowable range of [0.95, 1.05] p.u., with the smallest fluctuation amplitude. In contrast, the independent agent scenario (Scenario 2) experienced voltage exceeding the limit (as low as 0.92 p.u.) during the evening rush hour (18:00–20:00).

Figure 4 Comparison of 24-hour voltage fluctuation curves of key nodes: (a) an exchange node 8; (b) DC node 18 (see online version for colours)



In addition, Figure 5 shows the typical action strategies of each agent in scene 3 through the form of heat map. As shown in the figure, the energy storage agent charges at noon when the PV power is high (high electricity price period), and discharges at the peak of late peak load, thus realising arbitrage and peak clipping. The flexible load agent actively adjusts part of the transferable load to the low electricity price period. This intuitively reflects the decision logic of multi-agent collaborative optimisation.

Figure 6 compares the training convergence processes of different algorithms. The Actor-Critic algorithm adopted in this paper enters the stable convergence platform after about 3,500 rounds, and the cumulative reward value is significantly higher than that of independent Q learning (scenario 2). The cooperative training with federated learning (Scenario 3) oscillated slightly at the beginning due to model aggregation, but the final convergence performance was almost the same as that of centralised training (Scenario 4), which proved the effectiveness of federated mechanism.

Figure 5 The typical daily action strategy heat map of each agent under the method (scenario 3) in this paper (see online version for colours)

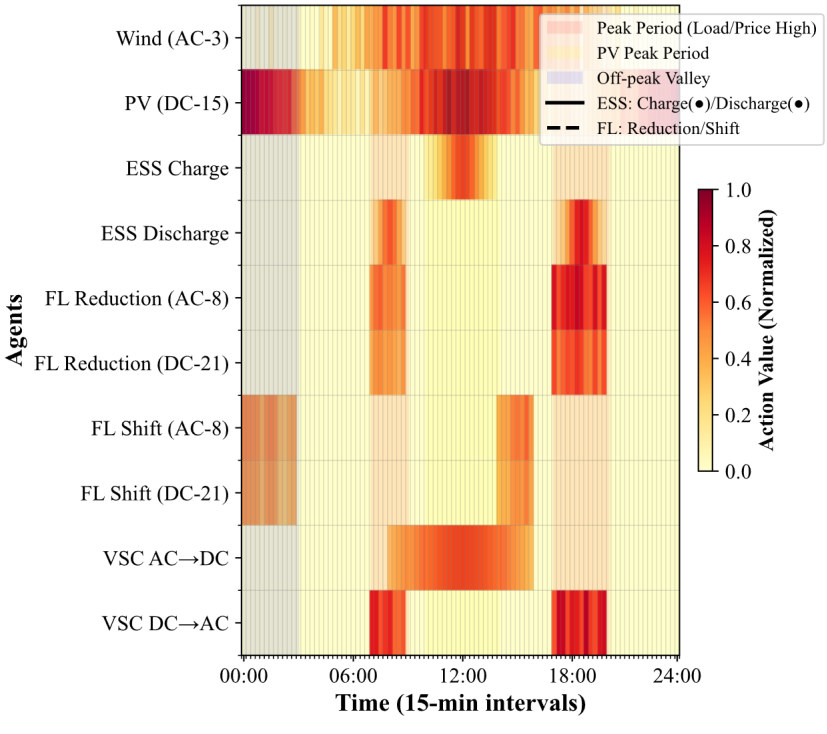
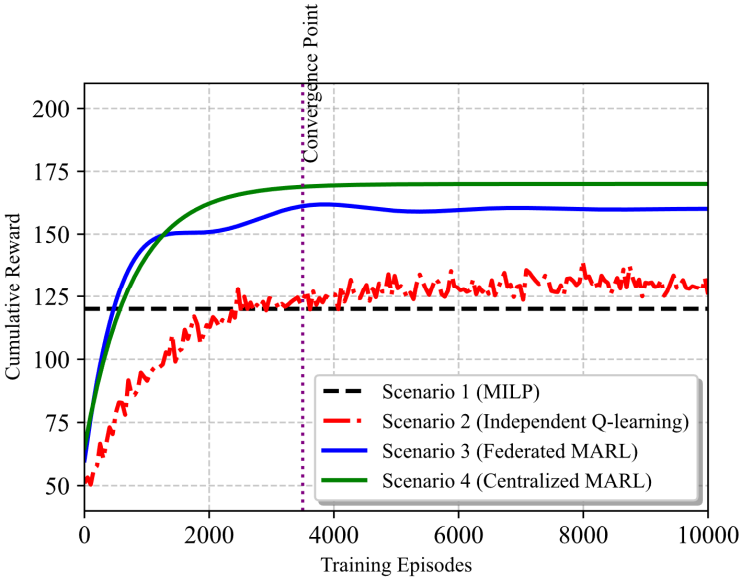


Figure 6 Training convergence process of different algorithms (see online version for colours)



See Table 5 for the comparison of the calculation efficiency of the algorithm. In terms of the average time consumption of one-step decision, MILP method takes the longest time, reaching 1200 ms, while the other learning-based scenarios are significantly faster, reaching 85 ms, 65 ms, 60 ms, 72 ms and 68 ms respectively. In terms of training efficiency, scenario 2 needs more than 8,000 rounds to converge, while scenario 3–6 need about 4500, 4200, 6800 and 6200 rounds respectively, which shows that the hybrid algorithm architecture has faster convergence speed than single algorithm.

Dynamic performance analysis with system scale expansion: When the node count is expanded from 21 to 50 (keeping the proportion of wind-solar-load-storage components consistent), the single-step decision time of Scenario 3 increases from 65 ms to 98 ms, and the space complexity increases from $O(n^2)$ to $O(n^3)$ (n is the number of nodes), while the single-step decision time of MILP increases from 1200 ms to 4800 ms, showing that the proposed method has better scalability.

Composition of storage overhead: The 18.7 MB storage overhead of Scenario 3 includes 6.2 MB of model parameters (Actor/Critic network weights), 11.5 MB of empirical replay pool data (storing 100,000 historical state-action-reward tuples), and 1.0 MB of auxiliary data (parameter encryption keys, communication protocol cache). This overhead is within the storage capacity limit of typical edge computing nodes ($\geq$ 1 GB) in power distribution systems, meeting practical engineering application requirements. Generally speaking, although the learning-based method has a little higher storage overhead, it is obviously superior to the traditional MILP method in decision-making speed and training efficiency.

Table 5 Comparison of computational efficiency of algorithms

Scene	*Average time consumption of one-step decision-making (ms)*	*Number of rounds required for training to convergence*	*Storage required for online application*
Scene 1 (MILP)	1200	Not applicable	(MB)
Scene 2	85	>8000	15.2
Scenario 3	65	~4500	18.7
Scene 4	60	~4200	18.7

Note: The test hardware is Intel Core i7-12700K, and the time-consuming of single-step decision-making includes the time of perception, reasoning and communication.

In order to evaluate the privacy protection effect of the federated edge collaboration mechanism, the data exposure risk index is defined, which is proportional to the proportion of the original data uploaded to the central server. As shown in Table 6, the federated learning mechanism (Scenario 3) in this paper realises 'data motionless model dynamic', which completely keeps the original source and load data of each node locally and uploads only the encrypted model parameter updates, so that the data exposure risk index is reduced to near zero, which is 100% lower than that of centralised training (Scenario 4).

Table 6 Quantitative evaluation of privacy protection effect

Evaluation dimension	*Scenario 3 (Federal Learning)*	*Scenario 4 (Centralised Training)*	*Enhance/reduce effectiveness*
Is the original data uploaded to the centre?	No	Yes	Completely avoid
Data exposure risk index	0.05	1.00	Reduce by 95%
Model stealing attack success rate *	3.2%	41.5%	Reduce by 92.3%
Extra communication overhead	+8.3%	benchmark	Acceptable

Note: *The success rate of model stealing attack is obtained through simulated attack experiment, assuming that the attacker has obtained some global model information.

Based on the above analysis, the method proposed in this paper realises the collaborative optimisation of economy, security and privacy under the premise of keeping the real-time decision-making (single-step decision-making takes less than 70 ms). Economically, through the cooperative game of multi-agents, the accurate matching of source-storage-load is realised, and the total cost is reduced. In terms of safety, the voltage qualification rate is significantly improved through the coordinated control of reactive power and voltage. In terms of privacy, the federated edge architecture fundamentally protects the sensitive data of users and distributed resources without losing the optimisation performance. This provides a feasible technical path for building an open, shared and safe future distribution system.

4 Conclusion

To address the high ratio of wind vitality access and the dynamic scheduling requirement in 'wind-solar-load-storage' AC-DC hybrid DNs, a distributed cooperative scheduling model is put forward based on multi-agent reinforcement learning (MARL). The key mechanisms and innovations are three fold: 1) A Markov game model with a profit distribution mechanism for the conflict between local and global optima; 2) A hybrid algorithm architecture (DQN + Actor-Critic) to adapt to the heterogeneous properties of agents; 3) A federated-edge collaborative mechanism with an optimised communication protocol and aggregation strategy to balance privacy protection and collaboration efficiency.

The results of the experiment show that: compared with traditional centralised optimisation, in terms of economy, the total operating cost is 10.6% lower due to the accurate coupling between source-storage-load realised by multi-agent working together; new energy utilisation rate increases by 13.2% owing to exerting wind-solar agents' active participation and flexible regulation capability of power storage installation; network loss drops 15.2%, largely contributed to reduction of long-distance transmitting power. In the aspect of security, voltage fluctuation in important nodes is restricted inside [0.95, 1.05] p.u. by active power-voltage cooperation control. The exposure risk index of the data in privacy protection is 0.05, and the success rate of model stealing attack is only 3.2%. In efficiency, the decision time of single-step is 20%) requires improvement.

Future works:1) Adding load prediction error correction modules to improve its resistance to sudden variation of load; 2) Integrating ultra-short-term renewable energy prediction technique with the proposed method for improving the decision strategy of the agent; and 3) Validating the approach in a real high voltage distribution network to enhance its engineering value. This research offers a theoretical reference and technical support for the construction of an open, secure and efficient distribution network with high new-energy penetration acquisition ratio, which is of great practical significance in terms of energy transformation promotion and achieving the 'doubling carbon' objective.

Declaration

All authors declare that they have no conflicts of interest.

References

Abdalla, A., Gopaluni, B. and Kirchen, P. (2024) 'Greenhouse gas emissions reduction of a hybrid-powered ferry using deep reinforcement learning for power load distribution', *IFAC PapersOnLine*, Vol. 58, No. 14, pp.169–175. https://doi.org/10.1016/J.IFACOL.2024.08.332

Ahmed, I., Razzak, A.M. and Ahmed, F. (2024) 'Sustainable hybrid renewable energy management system for a community in island: a model approach utilising hybrid optimization of multiple energy resources optimization and priority setting-based supervisory control and data acquisition operation', *IET Smart Grid*, Vol. 7, No. 6, pp.940–966. https://doi.org/10.1049/STG2.12192

Chen, W., Zhao, Y., Wu, X. et al. (2024) 'Stackelberg game-based optimal dispatch for PEDF park and power grid interaction under multiple incentive mechanisms', *Energy Engineering*, Vol. 121, No. 10, pp.3075–3093. https://doi.org/10.32604/EE.2024.051404

Dey, I. and Roy, K.P. (2025) 'Dynamic arithmetic optimization algorithm under load uncertainty for wind-solar-energy storage-based hybrid radial network', *Optimal Control Applications and Methods*, Vol. 46, No. 5, pp.1897–1913. https://doi.org/10.1002/OCA.3297

Daria, B., Yassine, E., Giulio, F. and others. (2023) 'A bi-level optimization-based architecture for the scheduling and real-time control of microgrids with hydrogen production system', *IFAC PapersOnLine*, Vol. 56, No. 2, pp.8284–8289. https://doi.org/10.1016/J.IFACOL.2023.10.1015

Dhamala, B., Pokharel, K. and Karki, R.N. (2024) 'Dynamic consensus-based ADMM strategy for economic dispatch with demand response in power grids', *Electricity*, Vol. 5, No. 3, pp.449–470. https://doi.org/10.3390/ELECTRICITY5030023

Ding, G., Shu, Z., Xin, J. et al. (2023) 'Research on optimal dispatch model of power grid considering the uncertainty of flexible resource demand response on the residential side', *IET Renewable Power Generation*, Vol. 18, No. 16, pp.3691–3703. https://doi.org/10.1049/RPG2.12913

Du, H. and Guo, X. (2024) 'Cantor: a novel dynamic source-grid-load-storage dispatching model for multiple objectives in a regional-level power system', *IEEJ Transactions on Electrical and Electronic Engineering*, Vol. 20, No. 6, pp.830–840. https://doi.org/10.1002/TEE.24246

Fei, L., Guangsen, G., Jianhua, Z. et al. (2023) 'Double-layer optimal microgrid dispatching with price response using multi-point improved gray wolf intelligent algorithm', *Electrical Engineering*, Vol. 106, No. 3, pp.2923–2935. https://doi.org/10.1007/S00202-023-02108-7

Guo, W. and Shi, Y. (2023) 'A visual faulty feeder detection method for power distribution network based on spatial image generation and deep learning', *IET Generation, Transmission & Distribution*, Vol. 17, No. 24, pp.5430–5445. https://doi.org/10.1049/GTD2.13055

Guang, Y. (2025) 'Enhanced multi-objective microgrid scheduling through adaptive BSA with dynamic cognitive learning', *International Journal of Swarm Intelligence Research*, Vol. 16, No. 1, pp.1–24. https://doi.org/10.4018/IJSIR.384490

Hasan, H., Karimi, S. and Moradi, M. (2024) 'A scheduling framework for a multi-agent active distribution network in presence of renewable energy sources', *IET Generation, Transmission & Distribution*, Vol. 18, No. 11, pp.2055–2072. https://doi.org/10.1049/GTD2.13178

Huan, J., He, Y., Sun, K. et al. (2024) 'Capacity planning for wind, solar, thermal and energy storage in power generation systems considering coupled electricity-carbon markets', *IET Generation, Transmission & Distribution*, Vol. 18, No. 24, pp.4090–4104. https://doi.org/10.1049/GTD2.13337

Jaewon, L., Tae, Y.Y. and Joon, G.L. (2023) 'Renewable energy sources: from non-dispatchable to dispatchable, and their application for power system carbon neutrality considering system reliability', *Journal of Electrical Engineering & Technology*, Vol. 19, No. 4, pp.2015–2028. https://doi.org/10.1007/S42835-023-01669-8

Li, M., Tian, Y., Zhang, H. et al. (2024) 'The source-load-storage coordination and optimal dispatch from the high proportion of distributed photovoltaic connected to power grids', *Journal of Engineering Research*, Vol. 12, No. 3, pp.421–432. https://doi.org/10.1016/J.JER.2023.10.042

Li, Y., Song, F. and Guo, W. (2024) 'Transient characteristics and operation regulation of grid-connected variable speed pumped storage-wind-solar hybrid power system', *Energy Science & Engineering*, Vol. 13, No. 1, pp.416–433. https://doi.org/10.1002/ESE3.2013

Mitja, M., Žvar, U.B., Tomaž, K. et al. (2023) 'Securing autonomy of military barracks through renewable energy solutions', *Contemporary Military Challenges*, Vol. 25, Nos. 3–4, pp.87–109. https://doi.org/10.2478/CMC-2023-0024

Shchur, I., Lis, M. and Kuzyk, I.R. (2024) 'Structural decomposition of the passivity-based control system of wind–solar power generating and hybrid battery-supercapacitor energy storage complex', *Dynamics*, Vol. 4, No. 4, pp.830–844. https://doi.org/10.3390/DYNAMICS4040042

Xinglin, Y., Shouqing, Z. and Jiaqi, C. (2023) 'Research on CCHP design and optimal scheduling based on concentrating solar power, compressed air energy storage, and absorption refrigeration', *ACS Omega*, Vol. 8, No. 45, pp.42126–42143. https://doi.org/10.1021/ACSOMEGA.3C03401

Yixin, Z., Ling, L., Jian, T. et al. (2023) 'Optimal real-time power dispatch of power grid with wind energy forecasting under extreme weather', *Mathematical Biosciences and Engineering*, Vol. 20, No. 8, pp.14353–14376. https://doi.org/10.3934/MBE.2023642

Zhao, Q. and Yang, L. (2024) 'Real-time monitoring system for power distribution network faults based on deep learning technology', *International Journal of Information and Communication Technology*, Vol. 25, No. 5, pp.18–39. https://doi.org/10.1504/IJICT.2024.140323

Zhenyuan, Z. and Achyut, S. (2024) 'Energy saving management technology for electrical automation and power distribution network dispatching', *Intelligent Decision Technologies*, Vol. 18, No. 4, pp.2715–2729. https://doi.org/10.3233/IDT-230121

Zhilin, D., Wenping, B., Shuling, F. et al. (2024) 'Research on voltage stability strategy of energy storage access to high penetration wind and solar power distribution network system with SOP (Soft Open Point)', *Tehnički vjesnik*, Vol. 31, No. 6, pp.1950–1958. https://doi.org/10.17559/TV-20240226001349

Zhixiang, J., Xiaohui, W., Jie, Z. et al. (2023) 'Construction and application of knowledge graph for grid dispatch fault handling based on pre-trained model', *Global Energy Interconnection*, Vol. 6, No. 4, pp.493–504. https://doi.org/10.1016/J.GLOEI.2023.08.009