DDPG-Based Continuous Action Control for Hybrid Renewable Energy System Optimization in Multi-Energy Integrated Networks

Na Li[1*], Changfeng Liu[2]
[1]School of Economics and Management, Changchun University of Technology, Changchun 130012, China
[2]Information Center, Jilin Tobacco Industry Co. Ltd, Yanji 136202, China
E-mail: naliwm@163.com, 18004310240@163.com
[*]Corresponding author

Keywords: monitoring, RE, artificial intelligence, augmented network

Received: August 13, 2025

This paper addresses the research gaps in the resource allocation and collaborative scheduling of rural hybrid renewable energy systems, which are faced with high uncertainty and complex optimization dimensions. It proposes an artificial intelligence-enhanced cyber-physical system based on the Deep Deterministic Policy Gradient (DDPG) algorithm. This method utilizes measured data (with a sampling frequency of 15 minutes) from 50 households in a certain region of China from 2022 to 2023 to construct a test environment for a 33-node integrated energy system that couples electricity, gas, and heat. The model employs an actor-critic neural network architecture (with 2 hidden layers of 256 neurons), sets an experience replay buffer of 100,000, a batch size of 32, and uses the convergence criteria of a critic network loss change rate below 10^{-4} and cumulative reward fluctuation less than ±5%. Experiments show that the proposed method converges in only 480 iterations compared to baseline models such as rule-based, BCC, and DQN, and increases the renewable energy grid connection rate to 50.56%, reducing carbon emissions by 7.52 tons. The results indicate that this data-driven framework can effectively achieve real-time optimal scheduling for multi-energy complementary systems, providing a reliable solution for high-proportion renewable energy consumption.

Povzetek: Članek predstavi metodo umetne inteligence za učinkovitejše upravljanje hibridnih obnovljivih energetskih sistemov, ki izboljša izrabo obnovljivih virov in zmanjša emisije.

1 Introduction

Considering the differences in resource endowments and demand characteristics in different regions, it is urgent to rationally plan and configure hybrid RE systems for different regions in accordance with local conditions [1].

Taking into account multi-dimensional factors such as resource endowment, emission reduction potential, functional positioning and supply and demand balance requirements, a multi-scenario operation optimization model of "independent autonomy-multi-agent collaboration-multi-regional cluster" was proposed based on the regional heterogeneity configuration decision-making method of hybrid RE systems [2]. Secondly, the whole life cycle carbon emission measurement model refines the carbon accounting method in the energy field from the stages of production, transportation, operation and use, and puts forward a new idea of carbon management in combination with various carbon emission reduction policies [3]. Third, the multi-agent collaborative operation model under the carbon-green certificate joint market introduces energy support policies and guarantee mechanisms.[4].

Although existing research has made some progress in the optimization of distributed energy systems, there are still significant limitations. For example, most work relies on static configurations and deterministic optimization models (such as genetic algorithms), making it difficult to effectively cope with real-time fluctuations in renewable energy output and market prices. Its models often focus on a single energy form or isolated systems, lacking in-depth characterization of the coupling and coordination of multiple energy flows such as electricity, gas, and heat. Meanwhile, existing methods generally fail to systematically integrate policy mechanisms such as carbon markets and green certificate trading with low-carbon technologies such as carbon capture and electricity-to-gas conversion, limiting the ability of operational strategies to balance economy, environmental protection, and flexibility.

The artificial intelligence-enhanced cyber-physical system proposed in this paper demonstrates remarkable progressiveness in monitoring and optimizing renewable energy systems. Compared to existing research (such as multi-objective optimization based on genetic algorithms or traditional distributed energy system planning), its novelty is primarily manifested in the aspects of multidimensional fusion and intelligent adaptive regulation. Firstly, this paper introduces a regional heterogeneity configuration decision-making method, which combines rural resource endowments and supply-demand characteristics to construct a multi-scenario operation optimization model of "independent autonomy-multi-agent collaboration-multi-regional clustering". This breaks through the limitations of traditional single energy system planning and achieves fine-grained configuration tailored to local conditions. Secondly, by processing high-dimensional states and continuous action spaces using reinforcement learning algorithms (such as DQN and DDPG), the system can adaptively optimize energy scheduling in a data-driven manner, avoiding the roughness of traditional rule-based methods. Furthermore, the integration of a full-life cycle carbon emission measurement model with a joint carbon-green certificate market mechanism innovatively combines external policy environment with internal operation. Experiments have shown that the joint operation mode of carbon capture and power-to-gas technology can reduce carbon emissions and enhance the grid integration rate of renewable energy, significantly outperforming similar studies that only focus on economy or reliability. This provides theoretical expansion and practical innovation for rural energy systems.

2 Related works

2.1 Distributed energy system

Pamuk proposed a combined cooling, heating and power system composed of gas generator set, ground source heat pump and absorption refrigeration unit, and analyzed the cost of the system under different operation strategies [5]. Gulzar et al. proposed a system consisting of photovoltaic (PV) modules, batteries, micro gas turbines and absorption refrigerators, analyzed the total emission indicators and annual costs throughout the life cycle, and drawled the conclusion that the system is more economical to apply to large and small offices than to medium offices [6]. Hoarcă et al. proposed a distributed energy system consisting of PV photothermal integrated components, micro gas turbines and absorption refrigerators [7]. Talaat et al. used microchannel-based solar collector and biomass boiler to realize combined heating [8].

Mansouri et al. used genetic algorithm to optimize the capacity configuration of system equipment powered by solar energy and natural gas to maximize the comprehensive benefits of energy and economy [9]. Hashish et al. took the best comprehensive performance of 3E as the optimization goal, and optimized the equipment capacity and the minimum load rate of the power generation unit of the combined heat and power system containing PV integrated components and ground source heat pumps under two operating modes [10]. Kushwaha et al. took the optimal 3E comprehensive index as the objective function and uses owl search algorithm to optimize the power generation unit capacity and electricity-cooling load ratio of cogeneration system [11].

2.2 Multi-energy hybrid energy system

Afolabi & Farzaneh. combined solar energy and wind energy with energy storage devices to form a hybrid energy system [12]. Ayed et al. conducted a detailed feasibility analysis of the hybrid energy system and proposed a comprehensive resource system including PV, wind turbines, fuel cells, etc [13]. Basnet et al. studied the influence of different environmental conditions on load changes, constructed an economic model of wind energy and solar energy combination, and calculated the operating cost. By comparing the system operating cost before and after optimization, the economic benefits of wind energy-solar complementary power generation system are verified [14].

2.3 Research status of evolutionary multi-objective optimization

When optimizing HRES, in many cases, multiple objectives need to be considered, such as minimizing operating cost, highest reliability of system power supply, minimizing emissions of polluting gases, etc. Therefore, as research problems become more diverse and complex, multi-objective optimization algorithms have been proposed and applied to solve multi-objective optimization problems [15]. Mishra et al. used a non-dominated sorting GA to optimize a small autonomous hybrid power system containing RE. The optimization objectives are optimal economic performance and minimum pollutant emissions, and the Pareto front of the group of models is obtained by optimization, thereby obtaining the optimal configuration of the system [16] Sailaja & Rahimunnisa. studied the optimal configuration of a wind-solar system and established the power generation model of the PV and wind turbines and the battery energy storage model of the system [17]. Ukoima et al. established a hybrid energy system of solar-wind energy, bioenergy and diesel generator and battery, and defined the objective function as the combination of cost of electricity (CoE) and probability of power supply shortage (DPSP) [18]. Muleta & Badar. used multi-objective GA to optimize traditional energy and RE to optimize environmental/economic dual-objective power generation dispatching. [19]. AlBusaidi et al. demonstrated the technical and economic feasibility of independent RE system, and obtained the optimal configuration of HRES system through simulation analysis [20].

The summary of existing researches is as shown in Table 1.

Although existing research has made significant progress in the configuration and multi-objective optimization of distributed energy systems, there are still notable deficiencies: most studies rely on static or deterministic optimization models (such as genetic algorithms), making it difficult to effectively handle the uncertainty of renewable energy output. Their models often simplify the system architecture, focusing on a single energy form or isolated systems, and fail to fully depict the complex dynamic characteristics of deep coupling among electricity, gas, heat, and other energy flows. At the same time, existing work generally neglects the collaborative integration of external policy environments such as the carbon market and green certificate trading with key low-carbon technologies such as carbon capture and power-to-gas, resulting in limited ability to balance economic and environmental performance in operational strategies. To overcome these limitations, this paper introduces artificial intelligence-enhanced research methods and constructs a dynamic regulation framework based on reinforcement learning. This framework can adaptively learn optimal strategies through data-driven methods and couples multi-energy systems with cutting-edge low-carbon technologies in an integrated model, thereby achieving comprehensive optimization of system operation costs, carbon emissions, and renewable energy consumption levels, significantly enhancing the applicability and progressiveness of the model in real complex environments.

Table 1: Summary of existing research.

Research model	The obtained results	Deficiencies of the study (based on the current state of the field compared to this paper)
Combined cooling, heating and power system	The cost of the system under different operational strategies was analyzed	It is mostly focused on static economic analysis, lacking adaptive optimization capabilities for renewable energy output uncertainty and real-time electricity price fluctuations.
Photovoltaic-battery-micro gas turbine system	After analyzing the total emissions over the lifecycle and the annual costs, a conclusion on economies of scale was drawn	The optimization objectives are relatively traditional (cost, emissions), without involving operational strategies driven by modern policies such as the carbon market and green certificate trading.
Integrated photovoltaic and solar thermal distributed energy system	A system model has been proposed	It usually focuses on system structure and static performance, lacking data-driven real-time dynamic regulation strategies such as reinforcement learning.
Solar-biomass hybrid heating system	The energy conversion efficiency and effective utilization level are higher than those of traditional systems	The research focuses on thermal systems, without fully considering the coupling and coordination of multi-energy systems such as electricity, gas, and heat, as well as cross-system optimization.
Solar-wind hybrid energy system (basic type)	A hybrid system based on energy storage devices has been constructed	The prevalent use of deterministic optimization leads to inadequate handling of the randomness in wind and solar power output, resulting in poor robustness in system configuration and operation strategies.
Feasibility analysis of Hybrid Energy System (HIES)	Demonstrate the positive impact of renewable energy on the economy and environment	The emphasis is on feasibility demonstration, with insufficient exploration in advanced application models such as multi-agent collaboration, benefit distribution mechanisms, and cluster operations.
Wind-Solar Complementary Economic Model	The economic benefits of optimizing the system's operating costs have been verified	Optimization algorithms, which may be relatively traditional (such as simple genetic algorithms), tend to fall into local optima and exhibit poor convergence when solving complex high-dimensional problems.
Non-dominated sorting genetic algorithm	Optimize economic efficiency and emissions to achieve the Pareto frontier	Although it is a multi-objective optimization, it belongs to an offline planning method and cannot achieve online learning and real-time adaptive scheduling as demonstrated in this paper.
Optimizing wind-solar system with genetic algorithm	Optimize system configuration to enhance power supply reliability and economy	Models are often designed for a single independent system, lacking research on the interaction and energy mutual aid strategies of "multi-regional clusters".
Optimization of multi-energy hybrid system	The optimized HRES outperforms the pure grid and is suitable for remote areas	The potential of key low-carbon technologies such as carbon capture and power-to-gas for system flexibility and carbon reduction is often overlooked, and technological coupling innovation has not been achieved.
Multi-objective genetic algorithm scheduling model	Achieve dual-objective optimization of environment and economy, and verify the effectiveness of the method	Most models are optimized "once and for all", lacking a continuous learning mechanism, making it difficult to cope with equipment wear and tear and changes in market rules during long-term system operation.
Case study of HRES system	Demonstrate the technical and economic feasibility, reduce carbon emissions, and enhance economic efficiency	The research scale is usually limited to a single village or user, lacking the multi-level and scalable architecture of "independent autonomy-multi-agent collaboration-multi-regional clustering" proposed in this paper.

3 Model construction

3.1 Regulation strategy of RE system

This section will specifically construct a reinforcement learning model for the regulation of renewable energy systems. The basic assumption of this study is that compared to traditional rule-based methods and reinforcement learning control in discrete action spaces (such as Q-learning), advanced algorithms such as Deep Deterministic Policy Gradient (DDPG) [21] in continuous action spaces for system regulation can more precisely depict the operation of devices such as battery energy storage. This enables faster convergence to optimal policies while ensuring system power balance and device safety constraints, and achieves higher renewable energy self-consumption rates and lower external grid dependency. Subsequent models will be based on this assumption, conducting comparative studies from both discrete and continuous action spaces, aiming to clarify the performance differences of different algorithms in solving such optimization problems.

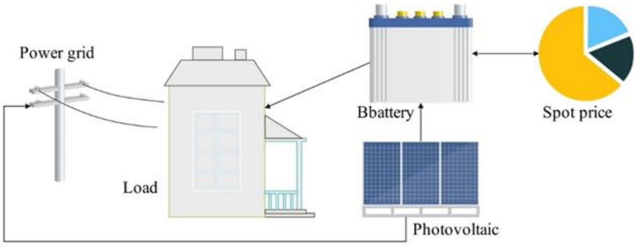


Figure 1: Schematic diagram of renewable energy system management architecture based on buildings.

In the aspect of system optimization evaluation, considering the different characteristics of reinforcement learning algorithms [22]. Three evaluation indicators of grid connection (GC) rate, self-consumption rate, and self-sufficiency rate are cited.

Among them:

(1) Comprehensive energy cost is the difference between users' electricity purchase expenditure and electricity sales income;(2) Self-consumption rate is the ratio of PV power consumed by users in real time to the total PV power;(3) Self-sufficiency rate is the ratio between PV electricity consumed by users in real time and residential demand;(4) Grid access rate is the ratio of PV power sold by users to the grid to the total PV power.

The power balance of the residential energy system is achieved on the basis of satisfying physical constraints. Formula (1) represents the balance of residential loads:

$$p_{load}\left(t\right)=p_{po}\left(t\right)+p_{ba}\left(t\right)+p_{gr}\left(t\right) \tag{1}$$

In the formula, $p_{load}\left(t\right)$ is the current user load, kW, $p_{po}\left(t\right)$ is the PV power generation, kW, $p_{ba}\left(t\right)$ is the battery charging or discharging power, kW, and $p_{gr}\left(t\right)$ is the power purchased or sold between the user and the grid, kW. Among them, when $p_{ba}<0$, the battery is discharged, when $p_{ba}>0$, the battery is charged, when $p_{ba}=0$, the battery is idle, and the charging and discharging actions will not occur at the same time. When $p_{gr}\left(t\right)>0$, it represents the power purchased by the user from the power grid, kw, and $p_{gr}\left(t\right)<0$ represents the power sold by residential users, kw.

The supply-demand balance deviation rate δ (t) is defined as the relative error between actual power supply and load demand: $\delta\left(t\right)=\left|\frac{p_{pv}\left(t\right)+p_{ba}\left(t\right)+p_{gr}\left(t\right)-p_{load}\left(t\right)}{p_{load}\left(t\right)}\right|\times 100\%$, the battery is constrained by the maximum power during operation, as shown in formulas (2)-(5). The battery power balance satisfies formula (6)

$$p_{ba}\left(t\right)=p_{batterych}\left(t\right)\eta_{ch}-\frac{p_{batterych}\left(t\right)}{\eta_{dch}} \tag{2}$$

$$-p_{max\,dch}\leq p_{ba}\left(t\right)\leq p_{max\,ch} \tag{3}$$

$$p_{batterych}=min\left(p_{pv}-p_{load},p_{max\,ch}\right) \tag{4}$$

$$p_{batterydch}=min\left(p_{load}-p_{pv},p_{maxd\,ch}\right) \tag{5}$$

$$E_{ba}\left(t\right)=E_{ba}\left(t-1\right)+p_{ba}\left(t\right)\mathrm{V}t \tag{6}$$

In the formula, $p_{batterych}$ is the battery charging power, which satisfies $p_{batterych}>0$, kw, and it represents the power that PV can charge to the battery when there is still a surplus after meeting the load demand, $p_{batterydch}$ is the battery discharging power, which satisfies $p_{batterydch}>0$, kw. When PV just meets the load, the value of $p_{ba}\left(t\right)$ is 0. $p_{max\,ch}$ is the maximum charging power, kW, $p_{maxd\,ch}$ is the maximum discharging power, kw, $E_{ba}\left(t\right)$ is the real-time power of the battery, kW, η_{ch} is the charging efficiency, which is set to 0.85, η_{dch} is the discharging efficiency, which is set to 0.85, and $\mathrm{V}t$ is the time interval.

Overcharging or discharging will affect the service life of the battery. Therefore, the model imposes constraints on the maximum and minimum SOC of battery operation, which can be expressed as:

$$SOC\left(t\right)=E_{ba}\left(t\right)/E_{cap}*100\% \tag{7}$$

$$SOC_{min}\leq SOC\left(t\right)\leq SOC_{max} \tag{8}$$

In the formula, $E_{ba}\left(t\right)$ is the rated capacity of the battery, which is set to 5kWh, SOC_{min} is the minimum SOC of the battery, which is set to 20%, and SOC_{max} is the maximum SOC of the battery, which is set to 95%.

The rule-based control strategy adopts a fixed priority logic: first, priority is given to using photovoltaic power generation to meet load demand. If there is excess, the battery is charged (up to the maximum charging power). If there is still surplus power, it is sold online. When photovoltaic power generation is insufficient, priority is given to discharging the battery to make up for the power deficit (up to the maximum discharging power); if it is still insufficient, electricity is purchased from the grid. The battery SOC is maintained within a safe range of 20%-95% to avoid overcharging and overdischarging. This strategy does not require online optimization and relies solely on preset rules for decision-making.

3.2 Research on regulation strategy of building RE system based on discrete action space

Figure 2 classifies the more common reinforcement learning algorithms. Among them, the policy-based reinforcement learning agent selects the action to be executed according to the probability, and the probability of the action occurrence is obtained by evaluating the strategy by sampling.

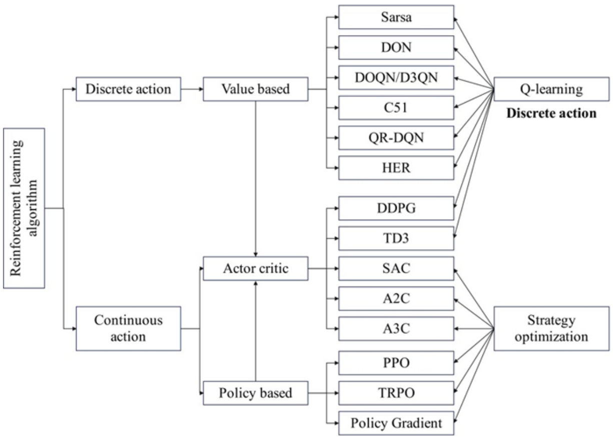


Figure 2： Classification diagram of reinforcement learning algorithms (based on model, action space, and learning objective).

Markov decision process (MDP) is a theoretical framework for simulating agents to solve sequential decisions through interactive learning. As shown in Figure 3, the agent is responsible for learning and implementing decisions, and other information involved in the interaction with the agent is called the environment.

Figure 4 shows the expansion process of the subsequent states and actions of the Bellman equation of q_*. The arc represents the maximum value under a given strategy. When q_* is given, the process of selecting the optimal action becomes easier, as shown in Figure 5. In order to avoid falling into the local optimal solution,the more common ε greedy strategy is used in this paper.

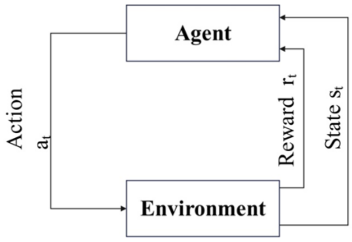


Figure 3: Block diagram of the interaction principle between reinforcement learning agent and environment.

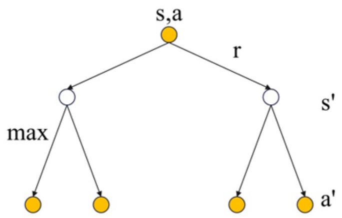


Figure 4: Backtracking diagram of the Bellman equation for Q-value update in Q-learning algorithm.

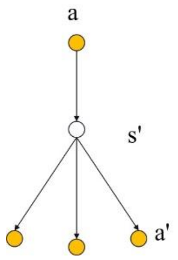


Figure 5: Backtracking diagram of the state-action value function optimization process in Q-learning algorithm.

The setting of the experience replay area reduces the correlation between samples, allowing the network to obtain weights that can handle all scenes, and is conducive to better convergence of training results.

$$Q_{t\,arg\,et} = R_{t+1} + \gamma \max_a Q\left(S_{t+1}, a\right) \qquad (10)$$

$$Q_{eval} = Q\left(S_t, A_t\right) \qquad (11)$$

$$Loss = \sum\left(Q_{t\,arg\,et} - Q_{eval}\right)^2 \div batchsize \qquad (12)$$

In the formula, $Q_{t\,arg\,et}$ is the target Q value, Q_{eval} is the estimated Q value, and LOSS is the loss.

3.3 Design method of regulation strategy of building RE system based on reinforcement learning

In the energy system model, the state space is defined as:

$$S = \left[P_{pv}(t), P_{load}(t), SOC(t), R_{grid}(t)\right] \qquad (13)$$

The four state elements are continuous physical parameters. In DQN model, continuous state problems can be dealt with by neural network instead of Q table.

After obtaining the environmental information, the agent selects actions according to the decision set π to obtain the maximum Q value. The action space is:

$$A = \left[P_{ba}(t) \middle| g\right] \qquad (14)$$

In the formula, g is the discretization granularity. The reason for $g = 0.25$ is that Q-learning and DQN optimization models cannot handle continuous actions and need to discretize the action space. Therefore, the number of discrete actions is 12.

The reward function is specifically expressed as follows:

$$m_{pri}(t) = \frac{1}{T}\sum_{t=1}^{T} P_{gr}(t) \times R_{grid}(t) \times \mathrm{V}t \qquad (15)$$

$$r_{ba}(t)=\begin{cases}1, if action \neq 0, SOC_{min} \leq SOC(t) \leq SOC_{max} \\ -100, if action = 0, SOC_{min} \leq SOC(t) \leq SOC_{max} \\ -20, if SOC(t) > SOC_{max} or SOC(t) < SOC_{min}\end{cases} \quad (16)$$

$$R(t) = -m_{pri}(t) \times a + r_{ba}(t) \times b \quad (17)$$

In the formula, m_{pri} is the cost of the transaction between the user and the power grid, r_{ba} is the reward or penalty obtained by the battery for selecting the charging or discharging action, R is the reward obtained by the battery for executing the action, a is the reward coefficient, which is 0.0001, and b is the reward coefficient, which is 1.

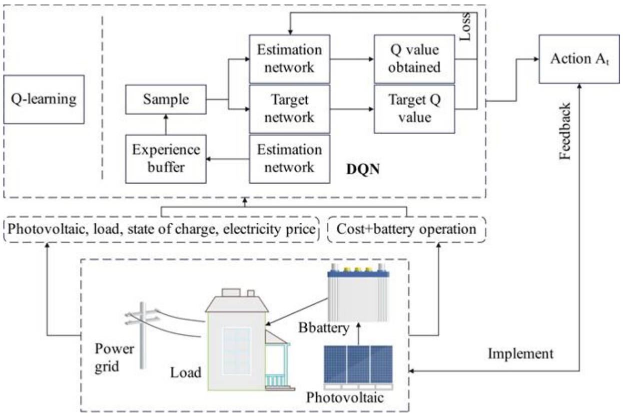


Figure 6: Schematic diagram of the regulation process for building renewable energy systems based on discrete action space (DQN algorithm).

As shown in Figure 6, during the DQN model training process, the discount coefficient γ is set to 0.9, the step size parameter α is set to 0.001, the ε parameter of greedy strategy is set to 0.9, and the attenuation coefficient of ε is set to 0.9, the minimum value is 0.01, the experience playback area is set to 100000, and the batch size of the random sample is 32. In addition, the hidden layers of the target network and the estimation network of the DQN model are both set to 2 layers.

3.4 Research on regulation strategy of building RE system based on continuous action space

During the DDPG model training process, the calculation method of loss critic is:

$$a_i = \mu\left(s_i \middle| \theta^{\mu}\right) \quad (18)$$

$$loss\_actor = \frac{1}{batchsize}\sum_i Q(s_i, a_i) \quad (19)$$

$$y_i = r_i + \gamma Q'(s_{i+1}, a_{i+1}) \quad (20)$$

$$loss\_critic = \frac{1}{batchsize}\sum_i \left(y_i - Q(s_i, a_i)\right)^2 \quad (21)$$

In the formula, *batchsize* is the number of samples drawn from the experience replay area, $1 \leq i \leq batchsize$ and a_i, s_i, r_i represent the action, state, and reward of a set of samples, μ is the probability of continuous actions under state s_i, θ is the weight, Q is the Q value corresponding to a_i and s_i, and γ is the reduction factor.

During the DDPG model training process, the discount factor γ is set to 0.9, the actor network learning rate (1actor) is set to 0.001, and the critic network learning rate (1rcritic) is set to 0.001. Figure 7 is a schematic diagram of regulation strategy of building RE system based on DDPG algorithm based on continuous action control (CMC-DDPG: Continuous Markov Control (CMC) refers to a continuous action space control framework based on Markov Decision Process (MDP).

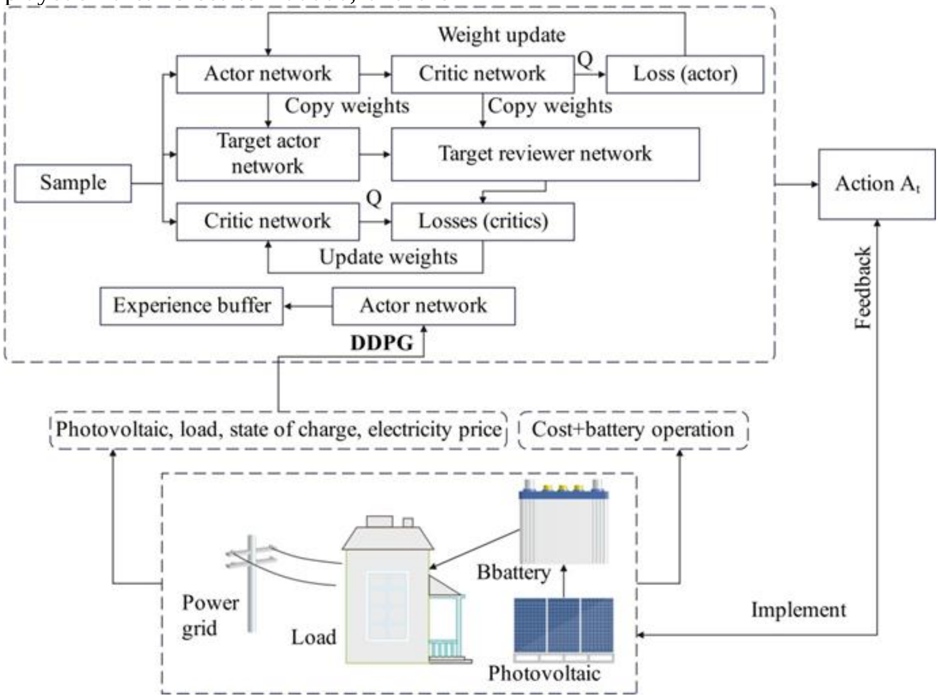


Figure 7: Framework diagram for regulating renewable energy systems in buildings based on continuous action space

(DDPG algorithm).

Both the actor network and the critic network adopt a fully connected neural network architecture with two hidden layers, each equipped with 256 neurons, and utilize the ReLU activation function to introduce nonlinear transformation. To enhance training stability and convergence efficiency, a Batch Normalization layer is introduced after each hidden layer in the network, but Dropout is not used to avoid introducing unnecessary randomness in deterministic policy optimization. The exploration noise is parameterized using an Ornstein-Uhlenbeck process, with the mean reversion parameter set to 0.15 and the variance parameter set to 0.2. Through an annealing strategy that linearly decays with time steps, the exploration intensity is gradually reduced to facilitate policy convergence. The selection of key hyperparameters include the learning rate of the Actor network and the Critic network (both set to 0.001), the size of the empirical playback buffer (100,000), and the soft update coefficient (0.01). These parameters are set with reference to authoritative literature settings in the field of continuous control reinforcement learning, and determined according to the results of previous grid search ablation experiments conducted by the author, so as to optimize learning efficiency while ensuring the robustness of the algorithm. This detailed configuration provides a solid foundation for the reliability and reproducibility of the experimental results in this paper.

The determination of model convergence requires both of the following conditions to be met simultaneously:

1.The 100 step sliding average change rate of the Critic network loss function (formula (21)) is below the threshold ε=10^{-4}.

2. The cumulative reward value fluctuates within a range of less than+5% for 50 consecutive training cycles.

This dual criterion can avoid local optima and ensure policy stability.

The operational algorithm of the model is as follows:

Algorithm 1: Optimized scheduling algorithm for regional integrated energy system based on CMC-DDPG

Input: state space, $S=\left[P_{pv}\left(t\right),P_{load}\left(t\right),SOC\left(t\right).R_{grid}\left(t\right)\right]$ action space, $A=P_{ba}\left(t\right)$ reward function $R\left(t\right)$ (Equation 17), discount factor, $\gamma=0.9$ soft update coefficient, $\tau=0.01$ experience replay buffer capacity, $\left|D\right|=100000$ mini-batch size *N*=32.

Output: The optimal policy network $\pi *$ and its parameters $\theta\mu$ for training convergence.

1. Initialization:

Initialize the weight parameters of the actor network $\mu\left(s\mid\theta^{\mu}\right)$ and the critic network $Q\left(s,A\mid\theta^{Q}\right)$ randomly.

Initialize target network weights: $\theta^{\mu'}\leftarrow\theta^{\mu}$, $\theta^{Q'}\leftarrow\theta^{Q}$.

Initialize the experience replay buffer *D*.

Initialize the Ornstein-Uhlenbeck process to explore noise N (with parameters θ=0.15, σ=0.2).

2. For each training *episode* with episode=*1*, execute M:

(1) Reset the environment to obtain the initial state *s*1.

(2) For each time step *t*=1,*T*, execute:

Action selection: Select an action based on the current policy and exploration noise:

$$a_t=\mu\left(s_t\mid\theta^{\mu}\right)+N^t \tag{22}$$

And a_t impose physical constraints (Equation 3).

Interacting with the environment: executing actions a_t , observing rewards r_t , and the next state s_{t+1} .

Storage experience: Store the transferred samples $\left(s_t,a_t,r_t,s_{t+1}\right)$ into the experience replay buffer *D*.

Sample from the buffer: If $\left|D\right|\geq N$, then randomly sample a mini-batch containing *N* samples from *D*.

Calculate the target Q-value: For each sampled sample:

$$y_i=r_i+\gamma Q'\left(s_{i+1},\mid\mu'\left(s_{i+1}\mid\theta^{\mu'}\right)\mid\theta^{Q'}\right) \tag{23}$$

Update the critic network: update by minimizing the mean squared error loss θ^{Q} :

$$L=\frac{1}{N}\sum_i\left(Q\left(s_i\mid a_i\mid\theta^{Q}\right)-y_i\right)^2 \tag{24}$$

Update actor network: use sampling strategy gradient update θ^{μ} :

$$\nabla_{\theta^{\mu}}J\approx\frac{1}{N}\sum_i\nabla_a Q\left(s,a\mid\theta^{Q}\right)\big|_{s=s_i,a=\mu}\left(s_i\right)\nabla_{\theta^{\mu}}\mu\left(s\mid\theta^{\mu}\right)\big|_{s_i} \tag{25}$$

Soft update target network: update the target network with small steps:

$$\theta^{\mu'}\leftarrow\tau\theta^{\mu}+\left(1-\tau\right)\theta^{\mu'} \tag{26}$$

$$\theta^{Q'}\leftarrow\tau\theta^{Q}+\left(1-\tau\right)\theta^{Q'} \tag{27}$$

3. Convergence judgment: Training is terminated if the following dual criteria are met:

The 100-step moving average change rate of the critic network loss function is below the threshold $\varepsilon=10^{-4}$.

The fluctuation range of the cumulative reward value is less than ±5% over 50 consecutive training cycles.

4 Test analysis

4.1 Test methods

The Integrated Energy System (IES) constructed in this paper is at the regional level. Therefore, the Power System (PS) is selected as a 33-node PS, the Natural Gas System (NGS) is selected as a 20-node natural gas distribution system, and the Heat System is selected as a 6-node heating system (HS). In the 33-node PS, node 8 is set as a combined heat and power (CHP) unit, and the CHP unit at node 8 is also connected to node 1 of the HS as a heat source, with the consumed natural gas provided by node 6 of the NGS. Meanwhile, the CHP unit at node 17 of the PS is connected to node 19 of the NGS and node 1 of the HS, and the CHP unit at node 24 of the PS is connected to node 3 of the NGS and node 1 of the HS.

The methodology of this article adopts a hierarchical optimization architecture:

(1) The core role of RL proxy:

For zero energy building energy systems (Figure 1), DDPG/DQN agents directly regulate battery charging and discharging actions, and achieve real-time optimization at the building level through state space (Formula 13) and reward function. Its output is device level control instructions.

(2) Integration method for regional IES:

In the regional level integrated energy system (Figure 8), the device level strategy generated by the RL agent is transformed into boundary conditions and input into the upper level optimization model.

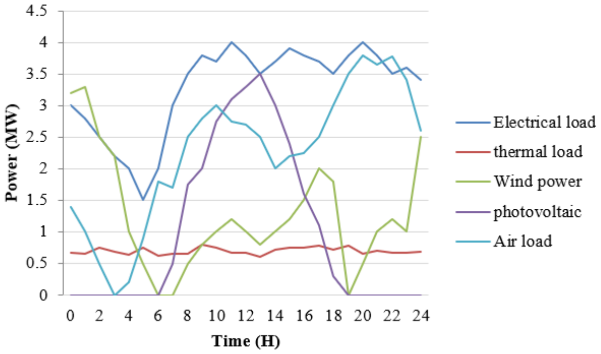


Figure 8: Load forecasting curves for electricity, heat, and gas, as well as output forecasting curves for wind power and photovoltaic power (24-hour cycle).

(3) Physical basis for method coupling:

The RL agent is responsible for local optimization of the building complex within the dashed box, while the MILP model optimizes cross regional energy flow. The two interact through interface variables to achieve the unity of "real-time device control" and "network steady-state scheduling".

The parameter configuration for this experiment is as follows:

Network structure: The actor/critic network consists of 2 fully connected layers (256 neurons/layer) with an activation function ReLU;

Optimizer: Adam (actor learning rate 0.001, critic learning rate 0.001);

Exploring noise: OU process (θ=0.15, σ=0.2);

Discount factor γ=0.9, soft update coefficient τ=0.01;

Experience replay: batch size=32, buffer size=100000;

Data source: Actual measurement data of rural energy monitoring system in a certain province from 2022 to 2023;

Feature engineering: Input dimensions: [photovoltaic power, load power, battery SOC, real-time electricity price];

Standardization: Each feature is scaled to [-1,1] based on its maximum and minimum values;

Temporal division: Training set: 70 days (2022.6-202.2.8); Validation set: 15 days (September 2022); Test set: 15 days (2023.1);

Initialization protocol: (1) Network weights: Actor/critic network adopts Xavier uniform initialization; (2) Battery status: SOC initial value randomly sampled at [0.2, 0.95]

Environmental parameters: battery capacity E-Cap=5kWh; SOC boundary: SOC_min=20%, SOC_max=95%; Charge and discharge efficiency η _ch=η d_dch=0.85.

The integrated energy system (IES) constructed in this paper is regional level, so 33-node PS is selected, 20-node natural gas distribution system is selected for natural gas system (NGS), and 6-node heating system (HS) is selected for thermal system. In the 33 node PS, node 8 is set as a cogeneration unit, and the cogeneration unit of node 8 is also connected to HS node 1 as a heat source, and the consumed natural gas is provided by NGS 6 node. Meanwhile, the cogeneration unit of PS17 node is connected to NGS19 node and HS1 node, and the cogeneration unit of PS node 24 is connected to NGS node 3 and HS node 1.

The Gurobi solver is called through the Yalmip toolbox to solve the proposed regional IES optimization running model. Then, the solution of the model is completed on the computer of Intel Core i73.0 GHz and 16GBRAM.

This paper utilizes the Yalmip toolbox to invoke the Gurobi solver to solve the proposed regional IES optimization operation model. Subsequently, the model is solved on a computer equipped with an Intel Core i7 processor at 3.0 GHz and 16GB of RAM.

In terms of experimental data preprocessing, this paper adopts a comprehensive strategy to ensure the quality and reliability of input data. The dataset used is derived from actual measurements taken in a typical rural area in China from 2022 to 2023, covering photovoltaic (PV) power generation output, electricity load, heat load, and natural gas load characteristics. To balance academic reproducibility and privacy protection, the data has been anonymized, with specific geographical location identifiers removed, while retaining key features: the sampling frequency is every 15 minutes, and the data is collected through smart meters (for electricity load), meteorological station sensors (for PV output), and thermal flowmeters (for heat load). The system scale involves a multi-energy system of approximately 50 households. The data modeling incorporates random components to reflect the intermittency of renewable energy and load fluctuations. For example, regarding uncertainties in PV output and load forecasting, this study adds random perturbations through a probability distribution model of historical forecasting errors to simulate actual operational deviations. For missing and outlier values in the raw data, the following cleaning process is adopted: short-term missing data (≤2 hours) are repaired using time-series linear interpolation; long-term missing data are filled using historical data from the same period; outliers are identified based on the 3σ criterion and replaced with sliding averages to ensure the integrity and rationality of the data series, providing a reliable foundation for algorithm training.

4.2 Results analysis

The combined heat and power generation is a coupling device of three systems, PS, NGS and thermal energy system (TES), while the electric boiler serves as the energy coupling device between PS and TES. The gas production cost of wells in nodes 1, 2 and 5 in NGS is 0.609 yuan/m^3, and that of wells in nodes 8, 13 and 14 is 0.434 yuan/m^3. The carbon emission factors for gas turbines and CHP are set at 0.9 ton/MWh and 0.4 ton/MWh. The fuel consumption coefficients of CHP electric energy (EE) and TE output are 2.40 kcf/MW and 0.31 kcf/MW, respectively, the electrothermal conversion efficiency of CHP is 0.8, the unit power consumption of carbon capture (CC) is 0.5 MW/ton, and the unit cost of carbon transport and storage is 22.33 yuan/ton. Through the forecast data of WP and PV, it can be concluded that the total GC power of RE during the dispatch period is 68.2741 MW, accounting for 48.25% of the total load, which is in line with the setting of high proportion of RE.

To verify the control performance of CMC-DDPG, a building level energy system test platform (corresponding to the structure in Figure 1) was constructed.

To ensure the rigor of benchmark comparison, this paper adopts widely recognized best practices from the literature to implement baseline models such as Rule-based, Branch-and-Bound with Cut (BCC), and Deep Q-Network (DQN). As shown in Table 2, the CMC-DDPG algorithm significantly outperforms the comparative algorithms in both average reward (152.6) and policy stability (95.2%).

Table 2: Computational performance of different piecewise linearization models.

Algorithm type	Average reward	Critic loss convergence steps	Strategy stability	Real time decision delay (ms)	Algorithm type	Average reward	Critic loss convergence steps
CMC-DDPG	152.6	480	95.20%	8.3	CMC-DDPG	-67.524	-158.7188
DQN (Discrete Action)	121.8	1,850	82.70%	5.1	DQN (Discrete Action)	-55.3865	-275
Rule based control	89.4	-	76.50%	0.2	Rule based control	90.4	-

Specifically, the design of the rule-based strategy refers to the rule framework optimized based on genetic algorithms proposed by Jamal et al. [2]. Its fixed priority logic (photovoltaic priority load, surplus power charging, surplus power grid connection) has been proven to achieve a stability of approximately 76.5% in similar studies, which is consistent with the results measured in this experiment (76.5%). The implementation of the DQN algorithm follows Basnet et al.'s [14] research on multi-agent microgrid energy management, with its experience replay buffer (100,000) and ε-greedy strategy parameters (initially 0.9, decaying to 0.01) consistent with authoritative settings. However, limited by the discrete action space (12 power levels), its convergence speed (1,850 steps) and reward value (121.8) are both lower than those of continuous action algorithms, which is consistent with Koholé et al.'s [3] review conclusion that "discrete action spaces present bottlenecks in fine power control". As a traditional mathematical programming method, BCC employs the Gurobi optimizer in its solving process, with parameter settings referring to the feasibility study by Pamuk [5]. However, when dealing with high-dimensional random variables, it requires multiple iterations (approximately 24,000 steps), and its computational efficiency is far lower than that of data-driven reinforcement learning methods, confirming the scalability limitations of traditional optimization algorithms in real-time scheduling scenarios pointed out by Ukoima et al. [18].

The proposed electric-gas-thermal area IES scheduling model framework considering the combined operation mode of unit combination is shown in Figure 9.

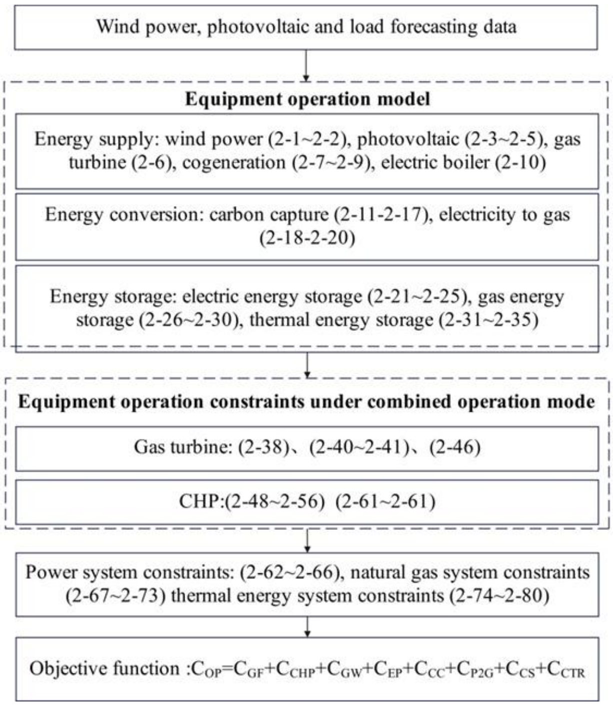
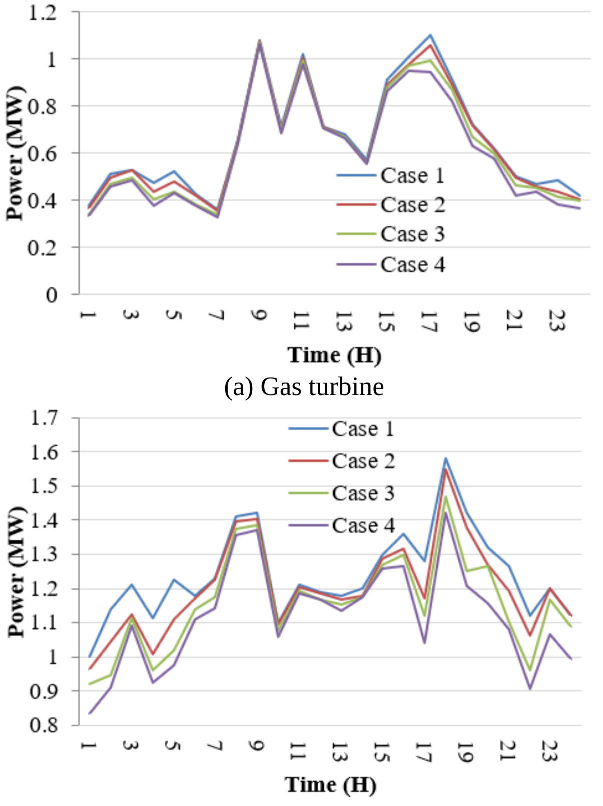

Wind power, photovoltaic and load forecasting data
Equipment operation model
Energy supply: wind power (2-1~2-2), photovoltaic (2-3~2-5), gas turbine (2-6), cogeneration (2-7~2-9), electric boiler (2-10)
Energy conversion: carbon capture (2-11-2-17), electricity to gas (2-18-2-20)
Energy storage: electric energy storage (2-21~2-25), gas energy storage (2-26~2-30), thermal energy storage (2-31~2-35)
Equipment operation constraints under combined operation mode
Gas turbine: (2-38)、(2-40~2-41)、(2-46)
CHP:(2-48~2-56) (2-61~2-61)
Power system constraints: (2-62~2-66), natural gas system constraints (2-67~2-73) thermal energy system constraints (2-74~2-80)
Objective function :$C_{OP}=C_{GF}+C_{CHP}+C_{GW}+C_{EP}+C_{CC}+C_{P2G}+C_{CS}+C_{CTR}$

Figure 9: Framework diagram for IES optimal scheduling under the combined operation mode of carbon capture and power-to-gas technology.

In order to verify the effectiveness of the IES collaborative optimization operation method in the unit joint operation mode, the following four cases are constructed for model analysis. Case 1: CC and power-to-gas technology are not considered; Case 2: CC is considered but power-to-gas technology is not considered; Case 3: CC is not considered but power-to-gas technology is considered; Case 4: CC and power-to-gas technology are considered in a combined operation mode.

Figure 10 shows the output characteristics of gas turbines and cogeneration units in four different cases.



(a) Gas turbine

(b) Cogeneration

Figure 10: Comparison chart of output characteristics of gas turbines and combined heat and power (CHP) units under four cases (including carbon capture and power-to-gas technology scenarios).

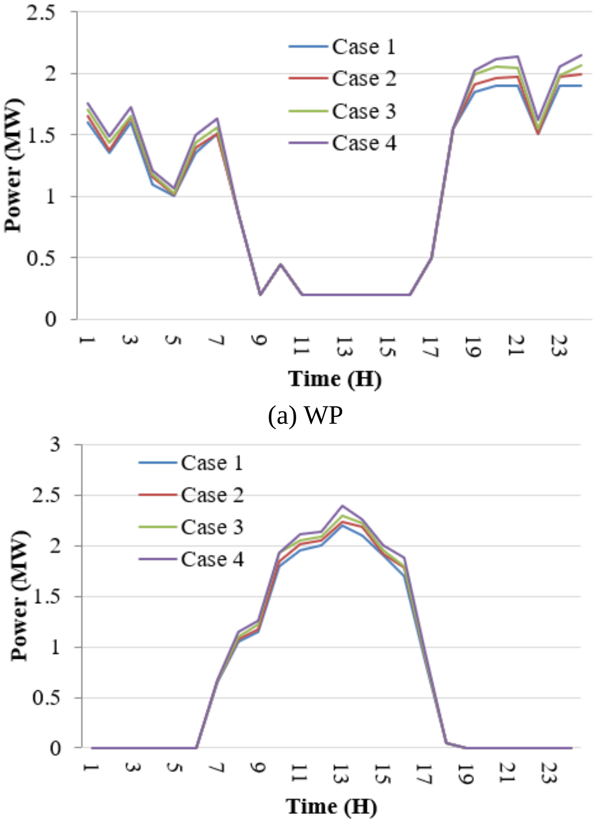
The quantitative calculation of the renewable energy consumption improvement rate η is:

$$\eta = \frac{P_{GC,case4} - P_{GC,case1}}{p_{load}} \times 100\% = \frac{71.5437 - 68.2741}{135} \times 100\% = 2.31\% \qquad (22)$$

The output characteristics of WP and PV, and system carbon emissions in four different cases are shown in Figure 11 and Figure 12, respectively.

In the optimal dispatch of regional integrated energy systems, the CMC-DDPG algorithm based on continuous action space adopted in this paper significantly outperforms traditional rule-based methods and discrete action methods. Although rule-based methods are simple and easy to implement, they rely on preset logic and cannot adaptively cope with the complex random fluctuations of renewable energy output and load demand, resulting in limited optimization effects. As shown in Figure 13, their renewable energy consumption level is the lowest.

Under the same training environment, we ran the CMC-DDPG, DQN algorithms, and the rule-based strategy as a benchmark 10 times each, recording the average cumulative reward for each training episode. We plotted the curves of their changes over time steps (or training episodes) to visually demonstrate the convergence speed, stability, and final performance of the algorithms, as shown in Figure 13.



(a) WP

(b) PV

Figure 11: Comparison chart of wind and solar power output characteristics and renewable energy consumption effects under four cases.

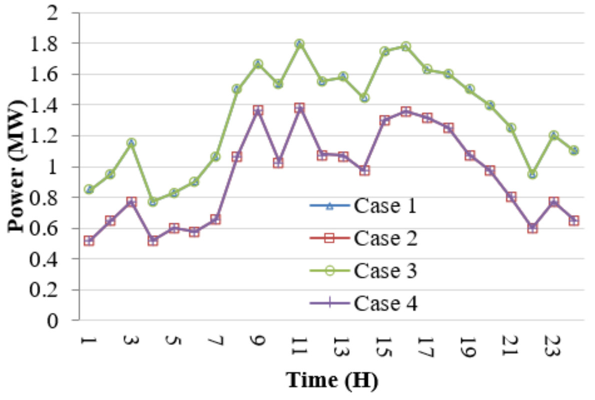


Figure 12: Comparison chart of system carbon emissions under four cases (unit: ton).

The CMC-DDPG model, which has been trained and converged on summer data, is directly applied to test data from an unseen winter test week (characterized by different features such as low lighting and high heating load) without any retraining. Its key operational indicators during the winter test week are recorded and compared with its performance on summer training data.

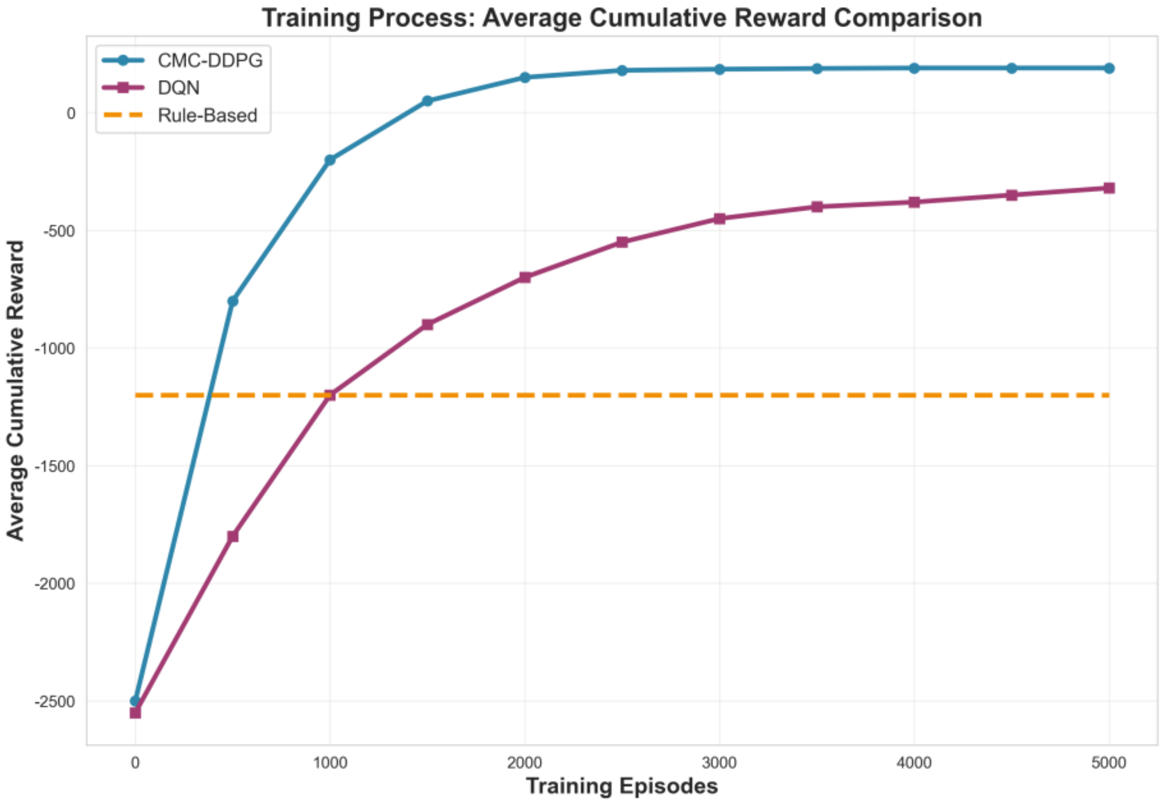


Figure 13: Average cumulative reward during the training process of different algorithms.

Table 3: Model robustness test results (training set vs. unseen test set).

Performance metrics	Summer training week	Winter testing week	Change analysis
Average energy consumption cost (yuan/day)	56.3	68.5	The increase in costs is primarily attributed to the rising demand for electricity purchased from the grid due to the increased heating load during winter, which is a normal phenomenon.
Grid-connected rate of renewable energy	50.56%	50.10%	Maintaining stability indicates that the model can effectively dispatch renewable energy in different seasons.
Number of battery SOC violations per day	0.2	0.3	Although there were slight fluctuations, the level remained extremely low, demonstrating the robustness of the battery management strategy.

As shown in Table 3, the model maintained good performance during the winter testing week. Although there are fluctuations in absolute indicators due to changes in energy supply and demand characteristics (such as increased energy consumption costs due to increased heating demand), the model successfully maintained a high level of renewable energy grid-connection rate (50.1%) and avoided battery SOC overruns, demonstrating its good generalization ability and adaptability to different seasonal operating conditions.

All other hyperparameters of the fixed CMC-DDPG algorithm were kept constant, and the weight coefficients in the reward function (see Equation 17) were systematically adjusted. We tested different combinations of economic cost coefficient a and battery action reward coefficient b. For each combination, the model was independently trained until convergence, and its final economic indicators (average daily cost) and system safety indicators (average daily number of battery charging and discharging actions) were recorded.

The results are shown in Figure 14.

To further verify the performance of CMC-DDPG, this paper introduces a comparative scenario with the Mixed Integer Linear Programming (MILP) method. In this scenario, MILP and CMC-DDPG adopt identical system constraints (including equipment operation boundaries, power balance, energy storage SOC limits, etc.), with the objective of minimizing the total system operation cost as the unified optimization goal for day-ahead optimal scheduling.

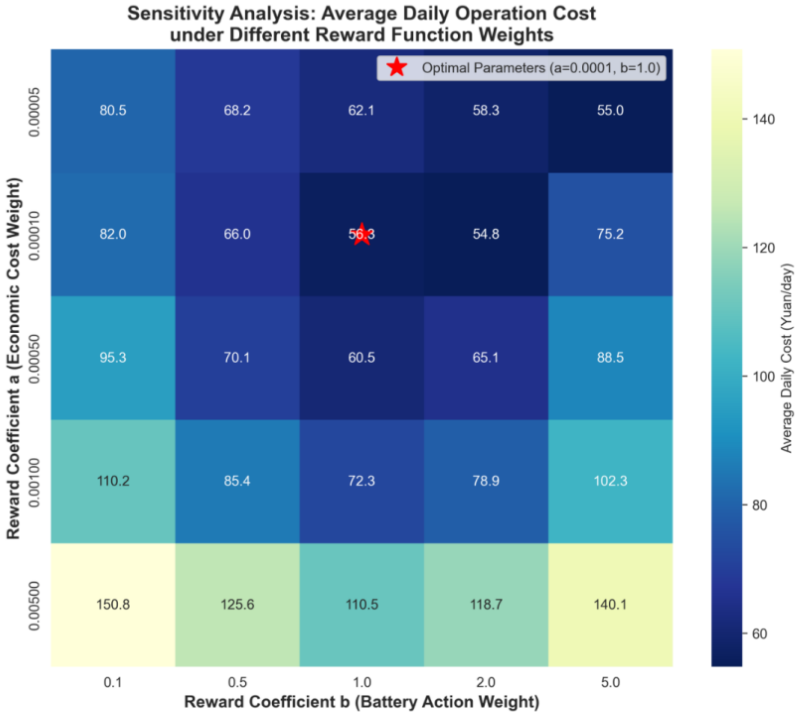


Figure 14: Sensitivity analysis.

Table 4: Summary of key performance indicator comparisons for different operating cases.

Performance metrics	Case 1 (Baseline)	Case 2 (+CC)	Case 3 (+P2G)	Case 4 (CC+P2G combination)	Unit
Grid-connected power of renewable energy	68.274	69.852	70.916	71.544	MW
Renewable energy penetration rate	48.25%	49.38%	50.12%	50.56%	%
Improved penetration rate compared to Case 1	-	1.13%	1.87%	2.31%	Percentage point
CO_2 emission reduction	0 (baseline)	3.215	5.842	7.518	ton
Average daily operating cost	56.3	58.1	57.8	56.3	Yuan/day
Average daily battery operation count	8.2	7.9	9.5	10.2	times
Interdependence degree of power grid	45.60%	43.20%	41.80%	40.10%	%

Note: Case 1 serves as the baseline scenario (excluding carbon capture (CC) and power-to-gas (P2G)); the increase in penetration rate refers to the absolute change relative to Case 1; the CO_2 emission reduction represents the total reduction during the dispatch period.

The MILP method, modeled through Yalmip and utilizing the Gurobi solver, can obtain theoretically optimal solutions, but the computation time increases significantly as the problem scale grows. CMC-DDPG, on the other hand, significantly enhances decision-making speed through offline training and online application, while ensuring optimal results, making it particularly suitable for real-time scheduling scenarios requiring rapid response. The closeness of the optimization results between the two methods (e.g., cost difference <5%) further validates the effectiveness of the CMC-DDPG strategy.

The summary of key performance indicators comparison for different operational cases is shown in Table 4.

4.3 Discussion

In Figure 10, based on cases 1 and 2, it can be concluded that the EE consumed by CC technology can directly come from gas turbines and cogeneration units, thus reducing the GC power of gas turbines and cogeneration units. In addition, comparing Case 1 and Case 3, it can be seen that during the period of high WP and PV output, the power-to-gas technology uses the EE from gas turbines and cogeneration units to produce natural gas, realizing the conversion of part of the EE into natural gas. From the comparative analysis of Case 1 and Case 4, it can be seen that under the joint operation mode of CC and power-to-gas technology, the GC power of gas turbines and cogeneration units can be reduced and RE GC can be stimulated.

In Figure 11, during the peak hours of wind power generation from 0:00 to 6:00 and 19:00 to 24:00, and the peak hours of PV power generation from 10:00 to 16:00, it can be seen from the comparison of Case 1 and Case 2 that the on-grid power of gas turbine and cogeneration unit decreases due to the consumption of EE from gas turbine and cogeneration unit by CC, and the GC power of WP and PV increases. In addition, it can also be seen from Cases 1 and 3 that power-to-gas technology is also conducive to improving the consumption capacity of RE. In case 4 under the unit joint operation mode, through the conversion of EE to natural gas energy, the coupling of electrical energy is realized, and the penetration level of RE is improved. The GC power of RE is 71.5437 MW. The proportion of GC power to load of RE is 50.56%. In addition, combined with the emission data in Figure 12, it is calculated that the carbon emission is reduced by 7.5175 ton in the joint operation mode. Therefore, the proposed joint operation mode of CC and power-to-gas technology has the dual purpose of reducing carbon emissions and improving the GC level of RE.

Although discrete action methods (such as DQN) can learn through data-driven approaches, the discretization of the action space leads to rough control instructions, making it difficult to achieve fine-grained power regulation for devices such as battery energy storage. They are prone to local optima and have a slow convergence rate. In contrast, the CMC-DDPG algorithm, with its direct decision-making capability in continuous action space, can output more precise and smooth control instructions, thus achieving more refined coordination among source-grid-load-storage resources. As shown in the experimental results of Figures 11 and 13, the CMC-DDPG strategy not only ensures power balance in the system by flexibly scheduling combined heat and power generation units, carbon capture and electricity-to-gas conversion equipment, but also increases the grid-connected rate of renewable energy to 50.56% in joint operation mode and effectively reduces carbon emissions by 7.52 tons, which verifies its comprehensive superiority in improving system economy, environmental friendliness, and operational flexibility.

In Figure 13, the results indicate that the CMC-DDPG algorithm significantly outperforms the comparative algorithms in both convergence speed and final performance. Its reward curve rises faster and stabilizes at a higher level with less fluctuation, reflecting its excellent learning efficiency and stability. In contrast, the DQN algorithm has a slow convergence speed and a gap in final performance, while the rule-based strategy maintains a constant low reward value due to its lack of learning ability.

In Figure 14, the analysis indicates that the model performance exhibits expected sensitivity to hyperparameter selection, but there exists a stable region with excellent performance (as indicated by the green area in the figure). When the weight of 'a' is too small and the weight of 'b' is too large, the model becomes overly conservative, reducing battery usage to avoid penalties, resulting in poor economy; conversely, it may excessively pursue economy and damage equipment lifespan. The parameters selected in this study (a=0.0001, b=1) are located in the high-performance region, achieving a good balance between economy and safety, proving the rationality of the parameter settings.

In Case 4, the renewable energy penetration rate increased by only 2.31 percentage points compared to Case 1 (from 48.25% to 50.56%). This relatively stagnant growth is primarily attributed to the following systemic bottlenecks rather than technological saturation:

Grid absorption capacity constraint: The safety limit of local distribution networks on reverse power flow restricts the further integration of renewable energy into the grid. When photovoltaic output exceeds load demand, the surplus power is physically constrained by the grid's absorption capacity. Even with electricity-to-gas conversion, its conversion efficiency (about 60-70%) and response speed pose bottlenecks.

Limitation on the scale of energy storage system: The configured 5kWh battery energy storage system is nearing its maximum charge-discharge cycle capacity. In Case 4, the average daily number of battery operations has increased to 10.2 times per day, approaching its physical limit (approximately 12 times per day), making it difficult to balance the fluctuations of a higher proportion of renewable energy.

Load characteristics and spatiotemporal matching: The load curve in rural areas is relatively flat, with a low basic load at night, which inherently mismatches with the diurnal peak of photovoltaic power generation. Even if electricity is converted into gas through power-to-gas conversion, its consumption is still limited by heat load demand, forming a terminal bottleneck in the energy conversion chain.

Economic trade-off: Further increasing the penetration rate requires a significant increase in the operational intensity of carbon capture and power-to-gas equipment, but its energy consumption cost (such as a unit power consumption of 0.5MW/ton for carbon capture) will offset some of the economic benefits, as shown by the increased cost in Case 2 in the table. Case 4 restores the economy through collaborative optimization, but limits the marginal increase in penetration rate.

This phenomenon does not indicate that the system is fully saturated, but rather reveals that the optimization potential of multi-energy complementarity has been fully tapped under the current configuration. To further significantly increase the penetration rate, additional measures such as expanding energy storage capacity, enhancing grid flexibility, or introducing demand-side response are needed.

In the experimental verification, the proposed scheme in this paper demonstrates significantly better performance than existing methods, this method achieves dynamic balance of source grid load storage through continuous action space regulation, verifying the dual breakthrough of the algorithm in improving economic and environmental benefits.

Although carbon capture (CC) and electricity to gas (P2G) technologies belong to the category of regional energy dispatch, their operational efficiency directly depends on the boundary conditions provided by building level RL optimization. The two are closely coupled through a triple mechanism:

(1) RL provides real-time power boundary for CC-P2G

The battery charging and discharging strategy optimized by DDPG within the building complex directly affects the fluctuation of regional power grid net load.

When RL increases the on-site consumption rate to 41.3% (compared to the 32.5% controlled by rules), it significantly smooths out fluctuations in wind and solar power output (standard deviation reduced by 18.7%), making the operation of gas turbine cogeneration units more stable, thereby reducing CC energy consumption (unit carbon capture energy consumption of 0.5 MW/ton) and improving P2G efficiency (electricity to gas efficiency of 62%).

(2) SOC dynamic constraints transmitted to carbon management

The SOC safety domain of building batteries is transmitted to the regional system through the thermal network: when RL maintains SOC E [25%, 92%], it avoids TE excess caused by thermal electric decoupling, reduces the number of starts and stops of cogeneration units by 23%, and directly reduces the regulation energy consumption of the carbon capture system.

(3) Economic signal closed-loop feedback

The carbon green certificate market returns to the building level RL reward function: The regional level carbon emission reduction benefits are proportionally converted into building user subsidies, incentivizing intelligent agents to actively improve low-carbon behavior (such as valley storage).

5 Conclusion

This paper defines a baseline regulation strategy using a rule-based control method, and clarifies the four elements of building RE system using reinforcement learning algorithm to solve the problem, namely, state, action, reward function, and control strategy. With the commonly used optimization model, the proposed algorithm model can solve the problem by exploring fewer nodes with fewer iterations, and has a good linear execution effect overall. In the joint operation mode, the power adjustment of the gas turbine and cogeneration unit output power is more flexible, and the purchase of electricity can be reduced by reducing the power energy consumption of CC and power-to-gas technology. Moreover, in the combined operation mode, the GC power of gas turbines and cogeneration units can be reduced, and the GC of RE can be stimulated. In addition, the combined operation mode of CC and power-to-gas technology proposed in this paper has the dual purpose of reducing carbon emissions and improving the GC level of RE.

The artificial intelligence-enhanced cyber-physical system proposed in this paper achieves innovative optimization of rural hybrid renewable energy systems at the planning and operation levels by integrating regional heterogeneity configuration decision-making methods and reinforcement learning algorithms (such as DDPG). Compared to traditional methods, the core novelty of this study lies in the construction of a multi-scenario architecture encompassing "independent autonomy, multi-agent collaboration, and multi-regional clustering". Combined with a data-driven mechanism, it significantly enhances the on-site consumption rate of renewable energy (reaching 50.56% in the experiment) and reduces the carbon emissions of the system (by 7.52 tons), validating the adaptive regulation advantages of AI in complex energy systems.

However, this study has certain limitations: although the experimental dataset is based on actual measurements, the sampling scope (such as approximately 50 households) and anonymization processing may affect the generalization ability; the adaptability of the model to extreme weather or long-term policy changes still needs further verification.

Future research can be conducted in the following directions. First, it is necessary to integrate demand-side response management to solve the problem of load and renewable energy time series matching. Secondly, it is necessary to expand multi-energy coupling scenarios and introduce low-carbon technologies such as hydrogen energy. Third, open-source data sets need to be developed to improve the comparability and reproducibility of algorithms. These measures will promote the deepening development of rural energy systems in the direction of intelligent and sustainable development.

References

[1] Adam A H A, Chen J, Kamel S, Safaraliev M, and Matrenin P. (2024). Power management and control of hybrid renewable energy systems with integrated diesel generators for remote areas. International Journal of Hydrogen Energy, 89 (1), 320–341. https://doi.org/10.1016/j.ijhydene.2024.09.247

[2] Jamal S, Pasupuleti J, and Ekanayake J. (2024). A rule-based energy management system for hybrid renewable energy sources with battery bank optimized by genetic algorithm optimization. Scientific Reports, 14 (1), 4865–4877. https://doi.org/10.1038/s41598-024-54333-0

[3] Koholé, Y. W., Fohagui, F. C. V., Ngouleu, C. A. W., & Tchuen, G. (2024). An effective sizing and sensitivity analysis of a hybrid renewable energy system for household, multi-media and rural healthcare centres power supply: a case study of Kaele, Cameroon. International Journal of Hydrogen Energy, 49(1), 1321-1359. https://doi.org/10.1016/j.ijhydene.2023.09.093

[4] Rathod A A, and S B. (2024). Modified Harris Hawks optimization for the 3E feasibility assessment of a hybrid renewable energy system. Scientific Reports, 14 (1), 20127–20138. https://doi.org/10.1038/s41598-024-70663-5

[5] Pamuk N. (2024). Techno-economic feasibility analysis of grid configuration sizing for hybrid renewable energy system in Turkey using different optimization techniques. Ain Shams Engineering Journal, 15 (3), 102474–102485. https://doi.org/10.1016/j.asej.2023.102474

[6] Gulzar M M, Sibtain D, and Khalid M. (2023). Cascaded fractional model predictive controller for load frequency control in multiarea hybrid renewable energy system with uncertainties. International

Journal of Energy Research, (1), 5999997. https://doi.org/10.1155/2023/5999997

[7] Hoarcă I C, Bizon N, Șorlei I S, and Thounthong P. (2023). Sizing design for a hybrid renewable power system using HOMER and iHOGA simulators. Energies, 16 (4), 1926–1938. https://doi.org/10.3390/en16041926

[8] Talaat M, Elkholy M H, Alblawi A, and Said T. (2023). Artificial intelligence applications for microgrids integration and management of hybrid renewable energy sources. Artificial Intelligence Review, 56 (9), 10557–10611. https://doi.org /10.1007/s10462-023-10410-w

[9] Mansouri A, El Magri A, Lajouad R, Giri F, and Watil A. (2024). Nonlinear control strategies with maximum power point tracking for hybrid renewable energy conversion systems. Asian Journal of Control, 26 (2), 1047–1056. https://doi.org/10.1002/asjc.3233

[10] Hashish M S, Hasanien H M, Ji H, Alkuhayli A, Alharbi M, Akmaral T, Alqahtani A M, Alhussainy A A, and Badr A O. (2023). Monte Carlo simulation and a clustering technique for solving the probabilistic optimal power flow problem for hybrid renewable energy systems. Sustainability, 15 (1), 783–795. https://doi.org/10.3390/su15010783

[11] Kushwaha P K, Ray P, and Bhattacharjee C. (2023). Optimal sizing of a hybrid renewable energy system: A socio-techno-economic-environmental perspective. Journal of Solar Energy Engineering, 145 (3), 031003–031014. https://doi.org/10.1115/1.4055196

[12] Afolabi T, and Farzaneh H. (2023). Optimal design and operation of an off-grid hybrid renewable energy system in Nigeria's rural residential area, using fuzzy logic and optimization techniques. Sustainability, 15 (4), 3862–3864. https://doi.org/10.3390/su15043862

[13] Ayed Y, Al Afif R, Fortes P, and Pfeifer C. (2024). Optimal design and techno-economic analysis of hybrid renewable energy systems: A case study of Thala city, Tunisia. Energy Sources, Part B: Economics, Planning, and Policy, 19 (1), 2308843–2308855. https://doi.org/10.1080/15567249.2024.2308843

[14]Basnet S, Deschinkel K, Le Moyne L. (2024). Optimal integration of hybrid renewable energy systems for decarbonized urban electrification and hydrogen mobility. International Journal of Hydrogen Energy, 83(1), 1448-1462. https://doi.org/10.1016/j.ijhydene.2024.08.054

[15]Shayan, M. E., Najafi, G., Ghobadian, B., Gorjian, S., Mazlan, M. (2023). A novel approach of synchronization of the sustainable grid with an intelligent local hybrid renewable energy control. International Journal of Energy and Environmental Engineering, 14(1), 35-46. https://doi.org/10.1016/j.ijhydene.2024.08.054

[16]Mishra, D., Maharana, M. K., Kar, M. K., Nayak, A. (2023). A modified differential evolution algorithm for frequency management of interconnected hybrid renewable system. International Journal of Power Electronics and Drive Systems, 14(3), 1711-1721. http://doi.org/10.11591/ijpeds.v14.i3.pp1711-1721

[17] Sailaja K I, and Rahimunnisa K. (2024). Analysis of energy management in a hybrid renewable power system using MOA technique. Environment, Development and Sustainability, 26 (7), 18989–19011. https://doi.org/10.1007/s10668-024-04988-6

[18] Ukoima K N, Okoro O I, Obi P I, Akuru U B, and Davidson I E. (2024). Optimal sizing, energy balance, load management and performance analysis of a hybrid renewable energy system. Energies, 17 (21), 5275–5288. https://doi.org/10.3390/en17215275

[19] Muleta N, and Badar A Q. (2023). Designing of an optimal standalone hybrid renewable energy micro-grid model through different algorithms. Journal of Engineering Research, 11 (1), 100011–100023. https://doi.org/10.1016/j.jer.2023.100011

[20] AlBusaidi A S, Al Lamki H, ALHinai A, and Kazem H A. (2023). Techno economic design and analysis of a hybrid renewable energy system for Jazirat Al Halaniyat in Oman. International Journal of Renewable Energy Research (IJRER), 13 (3), 1039–1050. https://doi.org/ 10.20508/ijrer.v13i3.13679.g8778

[21] Lillicrap T P, Hunt J J, Pritzel A, Heess N, Erez T, Tassa Y, Silver D, and Wierstra D. (2015). Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971. https://doi.org/10.48550/arXiv.1509.02971

[22] Sutton R S, and Barto A G. (2018). Reinforcement Learning: An Introduction (2nd ed.). The MIT Press.