Energy

AIMS Energy, 12(3): 639–663.
DOI: 10.3934/energy.2024030
Received: 20 March 2024
Revised: 11 May 2024
Accepted: 17 May 2024
Published: 22 May 2024

http://www.aimspress.com/journal/energy

Research article

Energy management of integrated energy system in the park under multiple time scales

Linrong Wang [1], Xiang Feng [1], Ruifen Zhang [1], Zhengran Hou [1], Guilan Wang [2] and Haixiao Zhang [2,*]

[1] Ordos Power Supply Branch of Inner Mongolia Electric Power（Group）Co., Ltd., Inner Mongolia, Ordos, 017004, China

[2] School of Control and Computer Engineering, North China Electric Power University, Baoding 071000, China

* **Correspondence:** Email: 220212221011@ncepu.edu.cn; Tel: +086-18931395665; Fax: +086-18931395665.

Abstract: Considering the problem of time scale differences among subsystems in the integrated energy system of a park, as well as the increasing complexity of the system structure and number of control variables, there may be a deep reinforcement learning (DRL) “curse of dimensionality” problem, which hinders the further improvement of economic benefits and energy utilization efficiency of park-level integrated energy systems (PIES). This article proposes a reinforcement learning optimization algorithm for comprehensive energy PPO (Proximal Policy Optimization) in industrial parks considering multiple time scales for energy management. First, PIES are divided into upper and lower layers, the first containing power and thermal systems, and the second containing gas systems. The upper and lower layers of energy management models are built based on the PPO; then, both layers formulate the energy management schemes of the power, thermal, and gas systems in a long (30 min) and short time scale (6 min). Through confirmatory and comparative experiments, it is shown that the proposed method can not only effectively overcome the curse of dimensionality in DRL algorithms during training but can also develop different energy system management plans for PIES on a differentiated time scale, improving the overall economic benefits of the system and reducing carbon emissions.

Keywords: integrated energy system; deep reinforcement learning; proximal policy optimization; economic performance; multiple time scales

1. Introduction

With the continuous growth of China's energy demand and carbon emissions, environmental issues are becoming increasingly prominent, constraining the development of China's economy and society [1]. Economy and low carbon have become the trends of future energy development [2]. Park-integrated energy systems (PIES) have the characteristics of multi-energy coupling and joint scheduling [3], becoming an important lever for efficient and clean energy utilization and achieving the "dual carbon" goal [4]. PIES couple multiple types of energy, and through complementary operation, the efficient utilization and flexible conversion of various types of energy can be greatly improved [5].

Compared with traditional energy systems, PIES have three types of energy: Electrical, thermal, and gas. These systems have more complex structures and include multiple energy sources and energy conversion equipment [6]. There are multiple uncertainties, as both photovoltaic and wind power generation included in the system have inherent uncertainty [7]. PIES also show flexibility, allowing them to adjust energy supply to meet the system's supply and demand [8]. Additionally, PIES have complex profit-seeking characteristics; the system can use gas turbines and other equipment to convert gas into electricity when the electricity price is high, reducing the consumption of electricity and lowering costs [9]. In order to improve the economic benefits and energy coupling ability of PIES, many scholars have conducted research based on multi-time scale models. In [10], authors developed a multi-time scale hierarchical rolling optimization scheduling model based on the time difference of energy consumption and load of different equipment during the intraday scheduling stage, and adjusted the unit output by perceived load changes. In [11], the system was divided into three time scales, day ahead long-time scale, day ahead predictive control, and real-time scheduling, to perform rolling optimization and reduce operating costs. In [12], an intraday procurement plan was developed based on the bilateral game mechanism of operators, as well as a dynamic scheduling model during the intraday management phase based on the differences in time scales of electricity, heat, and gas energy equipment. In [13], the problem of scheduling time scale differentiation in heterogeneous energy subsystems was solved using a double-layer scheduling time scale. In [14], authors utilized a multi-time scale coordinated optimization method to establish PIES scheduling strategies for three-time scales—day ahead, day in, and real-time—and analyzed the impact of multiple energy storage devices on the economic benefits of the system. From the above, it can be seen that multi-time scale models are beneficial for solving the problem of time scale differences in PIES, but the scheduling decision of the system still mainly relies on an accurate prediction of source load storage. With the increasing variety of energy equipment in PIES, the difficulty of prediction has increased, affecting the optimization of energy management in the system.

In recent years, some scholars have attempted to use deep reinforcement learning (DRL) methods of artificial intelligence to solve the energy management of PIES. In [15], the energy management of generator sets and gas turbines in PIES was optimized based on a deep deterministic strategy gradient algorithm. In [16], the differential evolution deep Q-network (DQN) algorithm was used to improve the overall economic benefits of PIES and the utilization rate of energy storage equipment. Authors in [17] proposed a load scheduling and energy management strategy based on Q-learning algorithm for distributed energy management in microgrids. In [18], a real-time energy management system for microgrids was designed based on the DQN algorithm, achieving the goal of minimizing operating costs. The energy management method based on DRL effectively reduces the dependence on accurate prediction of new energy output and source load storage. However, the increasingly complex structure and increasing variety of energy and equipment in PIES may lead to a "dimensionality disaster" problem. At the same time, reinforcement learning requires agents to execute actions with the same dimensionality [19].

In order to address the time scale differences in PIES and the "curse of dimensionality" of DRL methods, this paper proposes a multi-time scale PPO reinforcement learning optimization algorithm for comprehensive energy management in PIES, which includes three types of energy sources: Electric, heat, and gas. The main contributions of this article are as follows:

(1) To solve the problem of time scale differences among energy subsystems in the PIES, this article divides the PIES into two layers: The upper layer, which includes the power system and the thermal system, and the lower layer, which includes the gas system. The upper and lower layers are coupled and cooperate with each other to meet the energy supply and demand balance in the PIES.

(2) Compared with traditional strategy gradient optimization algorithms, the PPO algorithm in reinforcement learning has the advantages of being insensitive to update step size and not requiring resampling during updates, making it suitable for PIES containing continuous data such as photovoltaic and load. Therefore, this article applies the PPO algorithm to train the upper and lower layers of PIES separately, reducing the difficulty of model training, to set corresponding management time scales for different energy systems in the upper and lower layers, and developing management plans for different energy systems within PIES. Simulation examples show that this method can effectively reduce the operating cost of PIES while meeting the differences in time scales for managing different energy systems.

2. Park-integrated energy system

2.1. *The structure of the park-integrated energy system*

PIES mainly include three types of energy—electrical, heat, and gas—and manage the energy transmission, conversion, and storage to meet energy loads demands, integrating different components: energy supply side, energy conversion side, energy storage side, and energy load side. The system structure is shown in Figure 1.

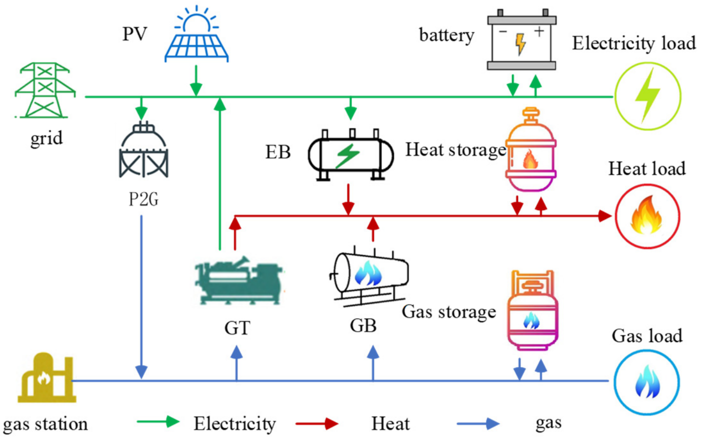


Figure 1. Park-integrated energy system structure.

2.1.1. Energy supply side

The energy supply side provides energy to PIES. The energy supply side mainly includes electricity purchased from the power grid, natural gas purchased from natural gas stations, and new energy equipment for photovoltaic power generation.

(1) Electricity grid

The electricity grid is responsible for providing electricity to PIES, and the constraints for purchasing power from the grid are shown in Eq (1).

$$0 \le P_t^{\mathrm{Ele}} \le P_{\max}^{\mathrm{Ele}} \tag{1}$$

In the equation above, P_t^{Ele} is the power purchased from the external power grid at time *t*; $P_{\max}^{\mathrm{Ele}}$ is the maximum transmission power of the external power grid interconnection line.

(2) Natural gas station

The natural gas station is responsible for providing natural gas to PIES, and the constraint conditions for purchasing gas power are shown in Eq (2).

$$0 \le G_t^{\mathrm{Gas}} \le G_{\max}^{\mathrm{Gas}} \tag{2}$$

G_t^{Gas} is the gas purchasing power of the natural gas station at time *t*; $G_{\max}^{\mathrm{Gas}}$ is the maximum power output of the natural gas station.

2.1.2. Energy conversion side

The energy conversion side includes GT (gas turbine), GB (gas boiler), EB (electric boiler), and P2G (electric to gas) equipment, which are used to convert energy between electricity, natural gas, and heat energy.

(1) Gas turbine

The GT equipment is a device that converts natural gas into electrical and heat energy. The relationship between the consumption of natural gas and the generation of heat and electrical energy by GT equipment at time slot *t* is shown in Eqs (3) and (4), respectively.

$$P_t^{\mathrm{GT}} = G_t^{\mathrm{GT}} \eta^{\mathrm{GT\text{-}E}} \tag{3}$$

$$\mathrm{H}_t^{\mathrm{GT}} = G_t^{\mathrm{GT}} \left(1 - \eta^{\mathrm{GT\text{-}E}} - \mu^{\mathrm{GT-loss}}\right) \tag{4}$$

P_t^{GT} is the power output of the GT equipment; $\mathrm{H}_t^{\mathrm{GT}}$ is the heat generation power of the GT equipment; G_t^{GT} is the gas consumption power of the GT equipment, and $\eta^{\mathrm{GT\text{-}E}}$ is the power generation efficiency of the GT equipment; $\mu^{\mathrm{GT-loss}}$ is the gas loss rate of GT equipment.

The constraint conditions for GT operating power and climbing power are shown in Eqs (5) and (6), respectively.

$$G_{\min}^{\mathrm{GT}} \le G_t^{\mathrm{GT}} \le G_{\max}^{\mathrm{GT}} \tag{5}$$

$$0 \le \left| G_t^{\mathrm{GT}} - G_{t-1}^{\mathrm{GT}} \right| \le \Delta G_{\max}^{\mathrm{GT}} \tag{6}$$

$G_{\min}^{\mathrm{GT}}$ is the minimum operating power of GT; $G_{\max}^{\mathrm{GT}}$ is the maximum operating power of GT; $\Delta G_{\max}^{\mathrm{GT}}$ is the upper limit of GT climbing power.

(2) Gas boiler

The GB is a device that converts natural gas into heat energy. Under time slot *t*, the relationship between the consumption of natural gas and the generation of heat energy by GB devices is shown in Eq (7).

$$\mathrm{H}_t^{\mathrm{GB}} = G_t^{\mathrm{GB}} \eta^{\mathrm{GB}} (1 - \mu^{\mathrm{GB-loss}}) \tag{7}$$

$\mathrm{H}_t^{\mathrm{GB}}$ is the heat generation power of the GB device; G_t^{GB} is the gas consumption power of GB equipment; η^{GB} is the gas-to-heat conversion efficiency of GB equipment; $\mu^{\mathrm{GB-loss}}$ is the gas loss rate of GB equipment.

The constraint conditions for GB operating power and climbing power are shown in Eqs (8) and (9), respectively.

$$G_{\min}^{\mathrm{GB}} \le G_t^{\mathrm{GB}} \le G_{\max}^{\mathrm{GB}} \tag{8}$$

$$0 \le \left| G_t^{\mathrm{GB}} - G_{t-1}^{\mathrm{GB}} \right| \le \Delta G_{\max}^{\mathrm{GB}} \tag{9}$$

$G_{\min}^{\mathrm{GB}}$ is the minimum operating power of GB; $G_{\max}^{\mathrm{GB}}$ is the maximum operating power of GB; $\Delta G_{\max}^{\mathrm{GB}}$ is the upper limit of GB climbing power.

(3) Electric boiler

The electric boiler is a device that converts electrical energy into heat energy. The relationship between the electrical power consumed by the EB device and the generated heat energy at time slot *t* is shown in Eq (10).

$$H_t^{\mathrm{EB}} = P_t^{\mathrm{EB}} \eta^{\mathrm{EB}} \left(1 - \mu^{\mathrm{EB\text{-}loss}}\right) \tag{10}$$

H_t^{EB} is the heat production work of the EB equipment; P_t^{EB} is the power consumption of the EB device; η^{EB} is the electric heating conversion efficiency of the EB equipment; $\mu^{\mathrm{EB\text{-}loss}}$ is the electrical energy loss rate of the EB device.

The constraints on EB operating power and climbing power are shown in Eqs (11) and (12), respectively.

$$P_{\min}^{\mathrm{EB}} \le P_t^{\mathrm{EB}} \le P_{\max}^{\mathrm{EB}} \tag{11}$$

$$0 \le \left| P_t^{\mathrm{EB}} - P_{t-1}^{\mathrm{EB}} \right| \le \Delta P_{\max}^{\mathrm{EB}} \tag{12}$$

$P_{\min}^{\mathrm{EB}}$ is the minimum operating power of EB; $P_{\max}^{\mathrm{EB}}$ is the maximum operating power of EB; $\Delta P_{\max}^{\mathrm{EB}}$ is the upper limit of EB climbing power.

(4) P2G equipment

The P2G is a device that converts electrical energy into natural gas [20]. The P2G device first decomposes water into oxygen and hydrogen through electrolysis, and then reacts to synthesize methane from carbon dioxide and hydrogen [21]. The relationship between the electrical power consumed by the P2G device and the amount of gas generated at time slot *t* is shown in Eq (13).

$$G_t^{\mathrm{P2G}} = P_t^{\mathrm{P2G}} \eta^{\mathrm{P2G}} \left(1 - \mu^{\mathrm{P2G\text{-}loss}}\right) \tag{13}$$

G_t^{P2G} is the gas production power of the P2G equipment; P_t^{P2G} is the power consumption of the P2G device; η^{P2G} is the electrical conversion efficiency of the P2G equipment; $\mu^{\mathrm{P2G\text{-}loss}}$ is the electrical energy loss rate of the P2G device.

The P2G operating power and climbing power constraints are shown in Eqs (14) and (15), respectively.

$$P_{\min}^{\mathrm{P2G}} \le P_t^{\mathrm{P2G}} \le P_{\max}^{\mathrm{P2G}} \tag{14}$$

$$0 \le \left|P_t^{\mathrm{P2G}} - P_{t-1}^{\mathrm{P2G}}\right| \le \Delta P_{\max}^{\mathrm{P2G}} \tag{15}$$

$P_{\min}^{\mathrm{P2G}}$ is the minimum operating power of P2G; $P_{\max}^{\mathrm{P2G}}$ is the maximum operating power of P2G; $\Delta P_{\max}^{\mathrm{P2G}}$ is the upper limit of P2G climbing power.

2.1.3. Energy storage side

A battery is an efficient energy storage element that primarily stores and releases electrical energy through the conversion of electrical and chemical energy. Lithium, sodium sulfur, and lead-acid batteries are currently the most widely used. Although lead-acid batteries have the advantages of low cost and large storage capacity, they have short life cycles and low energy density and result in significant environmental pollution, not being suitable for application in PIES. The technology of sodium sulfur batteries is not yet mature and is not suitable for widespread application at present. Lithium batteries have the characteristics of low self-discharge rate, low energy density, high charging and discharging efficiency, and long battery life cycle [22]. This article will use lithium batteries as storage components for PIES.

Thermal storage devices are divided into sensible thermal energy storage and latent thermal energy storage. Latent thermal energy storage has the advantages of high energy storage density and temperature stability but has the drawbacks of high cost and complex energy storage. Explicit heat storage has the advantages of simplicity, low cost, and long lifespan but low energy storage density, large equipment volume, and unstable temperature [23]. This article considers that PIES have lower requirements for high energy storage density and temperature stability. Therefore, explicit energy storage devices are selected as the thermal storage components of PIES to further reduce the operating cost of PIES.

Natural gas storage technology includes gas tank, underground gas, liquefied natural gas, pipeline gas, and hydrate gas storage, as well as other related storage technologies. In this article, the storage of natural gas is carried out by the widely used gas storage tanks.

The energy storage side includes three types of equipment: batteries, gas storage tanks, and heat storage tanks, which are responsible for storing or releasing electrical energy, gas, and heat, respectively. The mathematical model of energy storage equipment is shown in Eq (16).

$$S_{t+1}^{\mathrm{X}} = \left(1-\mu^{\mathrm{X\text{-}loss}}\right) S_t^{\mathrm{X}} + \left[P_t^{\mathrm{X,ch}} \eta^{\mathrm{X,ch}} \delta_t^{\mathrm{X,ch}} - \left(1-\delta_t^{\mathrm{X,ch}}\right) \frac{P_t^{\mathrm{X,dis}}}{\eta^{\mathrm{X,dis}}} \right] \Delta t \tag{16}$$

X represents the energy category; ES, HS, and GS represent the battery, heat storage tank, and air storage tank, respectively; S_t^{X} and S_{t+1}^{X} represent the energy storage at time slot *t* and at time slot *t* + 1; $\mu^{\mathrm{X\text{-}loss}}$ is the loss coefficient of energy storage device X. $P_t^{\mathrm{X,ch}}$ and $P_t^{\mathrm{X,dis}}$ are the energy storage power and discharge power of energy storage device X at time slot *t*; $\eta^{\mathrm{X,ch}}$ and $\eta^{\mathrm{X,dis}}$ are the energy storage efficiency and release efficiency of energy storage device X; $\delta_t^{\mathrm{X,ch}}$ is a 0–1 variable that represents the energy storage status of the energy storage device X at time slot *t*; Δt is the unit time slot length.

The state constraints, capacity constraints, and energy storage and discharge power constraints of energy storage device X are shown in Eqs (17–19), respectively.

$$\delta_t^{\mathrm{X,ch}} + \left(1-\delta_t^{\mathrm{X,ch}}\right) = 1 \tag{17}$$

$$S_{\min}^{\mathrm{X}} \le S_t^{\mathrm{X}} \le S_{\max}^{\mathrm{X}} \tag{18}$$

$$\begin{gathered} 0 \le P_t^{\mathrm{X,ch/dis}} \le P_{\max}^{\mathrm{X,ch/dis}} \\ P_t^{\mathrm{X,ch/dis}} = \begin{cases} P_t^{\mathrm{X,ch}}, & \delta_t^{\mathrm{X,ch}} = 1 \\ P_t^{\mathrm{X,dis}}, & \delta_t^{\mathrm{X,ch}} = 0 \end{cases} \end{gathered} \tag{19}$$

$S_{\min}^{\mathrm{X}}$ and $S_{\max}^{\mathrm{X}}$ are the lower and upper capacity limits of the energy storage device X; $P_t^{\mathrm{X,ch/dis}}$ is the maximum energy storage or discharge power of the energy storage device X.

2.1.4. Energy load side

The energy load in PIES mainly includes electricity, gas, and heat loads, all of which have the characteristics of temporal uncertainty.

2.2. Objective function of PIES

The goal of PIES energy management is to adjust the output of each unit in the energy system while ensuring the safe operation of the system, so as to minimize the operating cost of the system. The system operating cost includes the cost of purchasing electricity from the power grid C^{Ele}, the cost of purchasing gas from natural gas stations C^{Gas}, the cost of operating energy storage equipment C^{RC}, and the cost of carbon emissions from the system C^{c}. The objective function is shown in Eq (20).

$$F = \min\left(C^{\mathrm{Ele}} + C^{\mathrm{Gas}} + C^{\mathrm{RC}} + C^{\mathrm{c}}\right) \tag{20}$$

The calculation methods for electricity purchase cost and gas purchase cost are shown in Eqs (21) and (22), respectively.

$$C^{\mathrm{Ele}}=\sum_{t=1}^{T}c_t^{\mathrm{Ele}}P_t^{\mathrm{Ele}}\Delta t \tag{21}$$

$$C^{\mathrm{Gas}}=\sum_{t=1}^{T}c_t^{\mathrm{Gas}}\frac{G_{\mathrm{PH}}}{G_{\mathrm{HV}}}G_t^{\mathrm{Gas}}\Delta t \tag{22}$$

C^{Ele} and C^{Gas} are the electricity and gas prices at time slot *t*; P_t^{Ele} and G_t^{Gas} are the electricity and gas purchasing power at time slot *t*; G_{PH} is the equivalent power and heat energy conversion coefficient of gas; G_{HV} is the high calorific value of gas combustion.

The operating cost C^{RC} of energy storage equipment includes the operating cost C^{ES} of the battery, the operating cost C^{HS} of the heat storage tank, and the operating cost C^{GS} of the air storage tank. The representation of each cost is shown in Eqs (23–26).

$$C^{\mathrm{RC}}=C^{\mathrm{ES}}+C^{\mathrm{HS}}+C^{\mathrm{GS}} \tag{23}$$

$$C^{\mathrm{ES}}=\sum_{t=1}^{T}\frac{P_{\mathrm{R}}^{\mathrm{ES}}\Delta t C_{\mathrm{cape}}^{\mathrm{ES}}}{D_{\mathrm{R}}^{\mathrm{ES}}S_{\max}^{\mathrm{ES}}\left(m\left(\Delta E_t^{\mathrm{ES}}\right)^{-n}e^{-q\Delta E_t^{\mathrm{ES}}}\right)} \tag{24}$$

$$C^{\mathrm{HS}}=\sum_{t=1}^{T}c^{\mathrm{HS}}\Delta t\left(1-\delta_t^{\mathrm{HS,ch}}\right)P_t^{\mathrm{HS,ch/dis}} \tag{25}$$

$$C^{\mathrm{GS}}=\sum_{t=1}^{T}c^{\mathrm{GS}}\Delta t\left(1-\delta_t^{\mathrm{GS,ch}}\right)P_t^{\mathrm{GS,ch/dis}} \tag{26}$$

$P_{\mathrm{R}}^{\mathrm{ES}}$ is the rated charging and discharging power of the battery; $C_{\mathrm{cape}}^{\mathrm{ES}}$ is the investment cost of building batteries; $D_{\mathrm{R}}^{\mathrm{ES}}$ is the rated discharge depth of the battery; *m*, *n,* and *q* are the fitting curve parameters for converting the irregular charging and discharging process of the battery into the standard cycle usage times [24]; E_t^{ES} is the state of charge of the battery; C^{HS} and C^{GS} are the unit operating costs for the service life of heat and gas storage tanks.

The carbon emission cost of the system is shown in Eq (27).

$$C^{c}=\sum_{t=1}^{T}c_{co_2}\left(\sigma_{\mathrm{Ele}}P_t^{\mathrm{Ele}}+\sigma_{\mathrm{GT}}P_t^{\mathrm{GT}}-\sigma_{\mathrm{P2G}}P_t^{\mathrm{P2G}}\right)\Delta t \tag{27}$$

σ_{Ele} is the carbon emission coefficient of power grid purchasing; σ_{GT} is the carbon emission coefficient of GT equipment; σ_{P2G} is the carbon absorption coefficient of P2G equipment operation; c_{co_2} is the carbon price.

3. PIES energy management model based on PPO algorithm

3.1. Markov decision process

The Markov decision process is the mathematical foundation for reinforcement learning. The Markov decision process (MDP) consists of (S, A, R, γ) elements, where *S* represents the set of states of the environment, *A* represents the set of actions of the agent, R represents the return function, γ is the discount factor, and $\gamma \in (0,1]$. The state transition process is at time *t*. The intelligent agent selects action a_t to interact with the environment based on the current environmental state s_t, obtains a reward r_t, and enters the next state s_{t+1}. The agent receives a reward for interacting with the environment at each time step until the end of the state. G_t represents the long-term benefits of the intelligent agent, as shown in Eq (28).

$$G_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots + \gamma^{T-t} r_T = \sum_{i=0}^{T-t} \gamma^i r_{t+i} \tag{28}$$

T is the length of the decision sequence.

Using the action value function Q to evaluate the quality of action a in state s and using the state value function V to evaluate the quality of the state, the value of the Q value function can be used to calculate the V value function, as defined in Eqs (29) and (30).

$$Q_\pi(s,a) = E_\pi\left\{G_t \middle| S_t = s, A_t = a\right\} \tag{29}$$

$$V_\pi(s) = \sum_{a \in A} \pi(a|s) Q_\pi(s,a) \tag{30}$$

$\pi(a \mid s)$ represents the probability of executing action a in the current state s, representing the agent's strategy.

3.2. PIES energy management method based on PPO algorithm

Compared with traditional strategy gradient optimization algorithms, the PPO algorithm has the advantages of being insensitive to update step size and not requiring resampling during updates. It is suitable for PIES containing continuous data such as photovoltaic and load and can effectively avoid the curse of dimensionality.

3.2.1. Principles of PPO algorithm

PPO is a benchmark algorithm for reinforcement learning based on the actor critic (AC) framework proposed by OpenAI in 2017 [25]. The AC method includes both value-based and strategy-based learning methods. The AC framework consists of two networks, namely the actor network and the critical network. The actor network, also known as the policy network, is mainly used to generate policy functions. The critical network, also known as value network, is mainly used to evaluate the actions taken by actors in order to improve the strategy function of the actor network. The training flowchart of the PPO algorithm is shown in Figure 2.

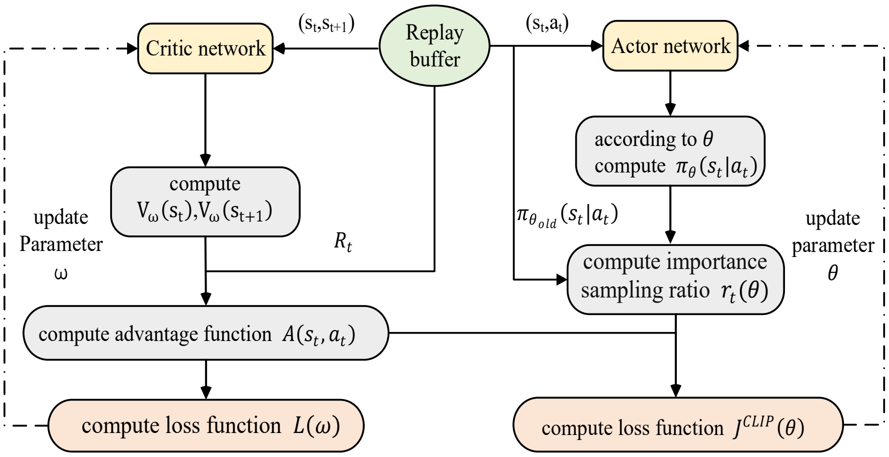

Critic network
(s_t,s_{t+1})
Replay buffer
(s_t,a_t)
Actor network
compute $V_\omega(s_t),V_\omega(s_{t+1})$
according to θ compute $\pi_\theta(s_t|a_t)$
update Parameter ω
R_t
$\pi_{\theta_{old}}(s_t|a_t)$
update parameter θ
compute importance sampling ratio $r_t(\theta)$
compute advantage function $A(s_t,a_t)$
compute loss function $L(\omega)$
compute loss function $J^{CLIP}(\theta)$

Figure 2. PPO algorithm training flow chart.

(1) Actor network training

The actor network updates network parameter θ by optimizing the loss function $J^{CLIP}(\theta)$. The determination of $J^{CLIP}(\theta)$ is shown in Eq (31).

$$J^{CLIP}(\theta)=E_{(s_t,a_t)-\pi(\theta_{old})}\left[\min\left(r_t(\theta)A(s_t,a_t),clip\left(r_t(\theta),1-\varepsilon,1+\varepsilon\right)A(s_t,a_t)\right)\right] \tag{31}$$

$A(s_t,a_t)$ is the dominance function; $r_t(\theta)$ is the importance sampling ratio; θ is the parameter of the actor network; ε is the pruning factor, which is a hyperparameter used to measure the degree of deviation between the new and old strategies. Due to the large update distance between the new and old strategies, the algorithm may become unstable. To avoid this situation, the importance sampling weight is limited to $[1-\varepsilon,1+\varepsilon]$.

The definition of the advantage function in Eq (5) is shown in Eq (32).

$$\begin{aligned}A(s_t,a_t)&=y_t-V_\omega(s_t)\\ y_t&=R_t+\gamma V_\omega(s_{t+1})\end{aligned} \tag{32}$$

$V_\omega(s_t)$ is the output value of the critical network at time *t*; R_t is the reward at time *t*; ω is the network parameter of critical; y_t is the $V_\omega(s_t)$ estimated value for time *t* + 1.

The importance sampling ratio is the ratio of the new strategy distribution function to the old strategy distribution function, as shown in Eq (33).

$$r_t(\theta)=\frac{\pi_\theta(s_t|a_t)}{\pi_{\theta_{old}}(s_t|a_t)} \tag{33}$$

Using the gradient ascent method to update actor network parameters θ, the size of the update equation is shown in Eq (34).

$$\theta \leftarrow \theta + \sigma^{A} \nabla_{\theta} J(\theta) \tag{34}$$

In the equation, σ^{A} is the learning rate of the actor network.

(2) Critical network training

The critical network updates the network parameter ω of critical by optimizing the loss function $L(\omega)$. The definition of $L(\omega)$ is shown in Eq (35).

$$L(\omega) = E\left[y_t - V_{\omega}(s_t)\right]^2 \tag{35}$$

Use the gradient descent method to update the critical network parameter ω, as shown in Eq (36).

$$\omega \leftarrow \omega - \sigma^{C} \nabla_{\omega} L(\omega) \tag{36}$$

In the equation, σ^{C} is the learning rate of the critical network.

3.3. Energy management model based on PPO algorithm

The PPO algorithm is used to solve the PIES energy management model, as shown in Figure 3.

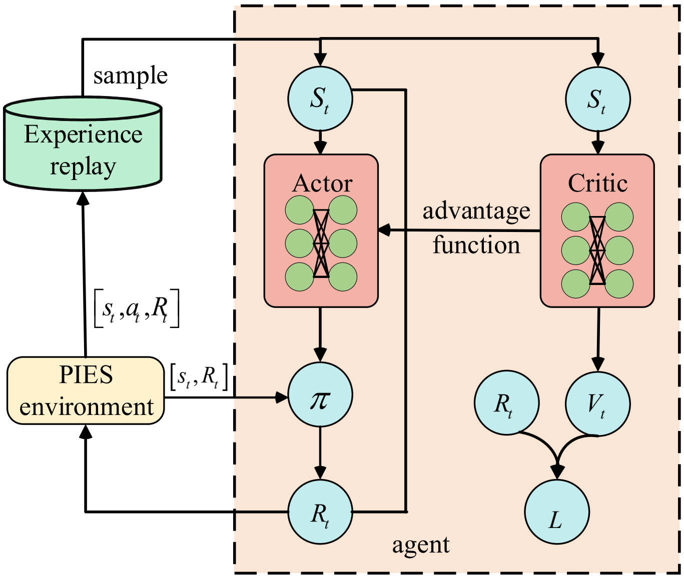


Figure 3. PIES energy management model based on PPO algorithm.

The initial input states of both the critical and the actor network are randomly sampled from the experience pool to obtain state s_t. The advantage of randomly sampling the initial states of each training round of the model from the experience pool is that it can reduce the randomness of the trained model in obtaining PIES energy management schemes. At the same time, the output of the critical network is the V_t value, while the output of the actor network is the action a_t. The agent interacts with the PIES environment according to the time slot and makes action a_t based on the current environment state s_t. The PIES environment returns the reward value R_t to the agent, and the experience pool is used to save the state s_t, action a_t, and reward R_t for each time period. The samples used for updating network weights in intelligent agents are randomly extracted from the experience pool. After offline training of the DRL model based on the PPO algorithm using training data, the model is saved and applied to the energy management of PIES.

4. Optimization of PIES scheduling based on PPO algorithm

The core idea of using the PPO reinforcement learning optimization algorithm for PIES considering multiple time scales is to first construct upper and lower energy management models based on different energy management time scales. The upper energy management model includes the power and heat systems, and the lower energy management model includes the gas system. Then, based on the PPO algorithm, long-term energy management of power and heat systems is achieved at a time scale of 30 min, while short-term energy management of gas systems is achieved at a time scale of 6 min, achieving differentiated energy management.

The upper and lower PPO scheduling models are mutually coupled, and the scheduling plan for the power and heat systems by the upper PPO serves as the environmental state of the lower PPO system. Due to the relatively long update time, the upper PPO scheduling model may experience a shortage of heat or power supply. At this time, the lower PPO scheduling model can be fully utilized to provide electricity and heat supply to the upper heat and power system by utilizing the GT and GB equipment in the lower PPO scheduling model. At the same time, in order to avoid the situation where the upper PPO scheduling excessively relies on the energy supply of the lower PPO scheduling and the system energy state is unstable, the cumulative adjustment of GT and GB operating power in the lower PPO scheduling is used as the reward function penalty term for the upper PPO scheduling. The specific energy management model is shown in Figure 4.

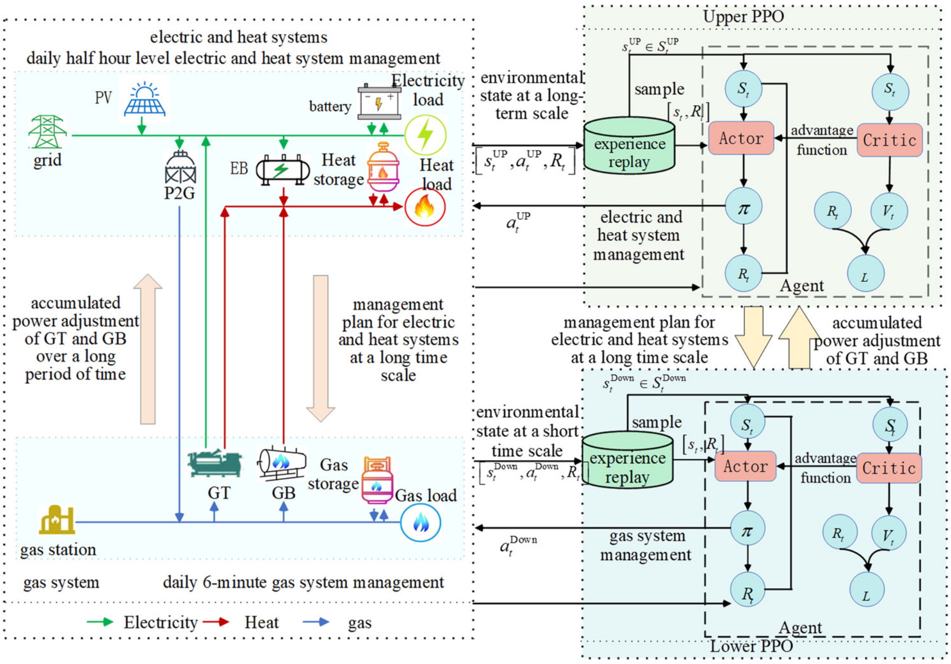

electric and heat systems
daily half hour level electric and heat system management
PV
grid
battery
Electricity load
P2G
EB
Heat storage
Heat load
accumulated power adjustment of GT and GB over a long period of time
management plan for electric and heat systems at a long time scale
GT
GB
Gas storage
Gas load
gas station
gas system
daily 6-minute gas system management
Electricity
Heat
gas
Upper PPO
$s_t^{UP} \in S_t^{UP}$
environmental state at a long-term scale
sample
$[s_t, R_t]$
experience replay
$[s_t^{UP}, a_t^{UP}, R_t]$
S_t
Actor
advantage function
Critic
π
R_t
V_t
L
a_t^{UP}
electric and heat system management
Agent
management plan for electric and heat systems at a long time scale
accumulated power adjustment of GT and GB
$s_t^{Down} \in S_t^{Down}$
environmental state at a short time scale
$[s_t^{Down}, a_t^{Down}, R_t]$
a_t^{Down}
gas system management
Lower PPO

Figure 4. Multi-time scale park comprehensive energy PPO reinforcement learning optimization algorithm.

4.1. Upper PPO long time scale energy management

4.1.1. Upper PPO state space

The state space s_t^{UP} of the upper PPO consists of the observed states of the upper power and heat systems (including photovoltaic power generation), as shown in Eq (37).

$$s_t^{\mathrm{UP}} = \left\{ c_t^{\mathrm{Ele}}, \mathrm{G}_t^{\mathrm{GB}}, \mathrm{G}_t^{\mathrm{GT}}, P_t^{\mathrm{load}}, H_t^{\mathrm{load}}, P_t^{\mathrm{PV}}, \sum_{i=1}^{5} \Delta G_{t+i}^{\mathrm{GB}}, \sum_{i=1}^{5} \Delta G_{t+i}^{\mathrm{GT}}, t \right\} \tag{37}$$

c_t^{Ele} is the electricity purchase price of the power grid; $\mathrm{G}_t^{\mathrm{GB}}$ is the gas consumption power in GB; $\mathrm{G}_t^{\mathrm{GT}}$ is the GT gas consumption power; P_t^{load} is the electrical load; H_t^{load} is the heat load; P_t^{PV} is the photovoltaic power generation; $\sum_{i=1}^{5} \Delta G_{t+i}^{\mathrm{GB}}$ is the gas consumption power of the GB device at a long time scale; $\sum_{i=1}^{5} \Delta G_{t+i}^{\mathrm{GT}}$ is the gas consumption power of the GT equipment; *t* is the time slot.

4.1.2. Improved upper PPO action space

The action space of the upper PPO intelligent agent is shown in Eq (38).

$$a_t^{\mathrm{UP}} = \left\{ P_t^{\mathrm{Ele}}, \mathrm{P}_t^{\mathrm{P2G}}, P_t^{\mathrm{EB}}, P_t^{\mathrm{ES,ch/dis}}, P_t^{\mathrm{HS,ch/dis}} \right\} \tag{38}$$

P_t^{Ele} represents the power purchased from the external power grid; $\mathrm{P}_t^{\mathrm{P2G}}$ is the power consumption of the P2G device; P_t^{EB} is the power consumption of the EB device; $P_t^{\mathrm{ES,ch/dis}}$ and $P_t^{\mathrm{HS,ch/dis}}$ are the storage/discharge power of the battery and the storage/discharge power of the heat storage tank, respectively.

By adding random disturbances in the upper action space to enhance the perception of the environment, the improved upper PPO action space is shown in Eq (39).

$$a_t^{\mathrm{UP}\prime} = \tau a_t^{\mathrm{UP}} + (1-\tau) m_t \tag{39}$$

$a_t^{\mathrm{UP}\prime}$ represents the actual action space; τ represents the proportion of each component in the initial action space; m_t represents an increased random disturbance and the upper limit of $m_t \in [-1,1]$, and τ is 0.9, ensuring that the action space in the later stage of training still has perceptual ability.

4.1.3. Upper PPO reward function

The upper PPO reward function is used to guide the intelligent agent to select actions based on the current state and obtain the maximum cumulative return. The upper PPO reward function consists of three penalty terms. The first penalty term mainly includes the cost of power grid purchase, the operating cost of upper equipment, and the cost of operating power regulation. The significance of considering the cost of equipment operating power regulation is to prevent external power purchase or the fluctuation range of P2G and EB equipment operating power from being too large, causing sharp changes in system load and affecting system stability. The definition of C_t^{U1} is shown in Eq (40).

$$C_t^{\mathrm{U1}} = C_t^{\mathrm{EleRc}} + C_t^{\mathrm{P2GRc}} + C_t^{\mathrm{EBRc}} + \sum_{i=1}^{5} C_{t+i}^{\mathrm{Ele}} + C_t^{\mathrm{ES}} + C_t^{\mathrm{HS}} \tag{40}$$

$\sum_{i=1}^{5} C_{t+i}^{\mathrm{Ele}}$ represents the cost of purchasing electricity from the external power grid over a long period of time; C_t^{ES} is the operating cost of the battery; C_t^{HS} is the operating cost of the heat storage tank; C_t^{EleRc} is the cost of external power purchase and regulation; C_t^{P2GRc} and C_t^{EBRc} are the operating power regulation costs of P2G and EB devices, where C_t^{EleRc}, C_t^{P2GRc}, and C_t^{EBRc} are defined as Eqs (41–43).

$$C_t^{\mathrm{EleRc}} = c_{\mathrm{RC}}^{\mathrm{Ele}} \left| P_t^{\mathrm{Ele}} - P_{t-5}^{\mathrm{Ele}} \right| \tag{41}$$

$$C_t^{\mathrm{P2GRc}} = c_{\mathrm{RC}}^{\mathrm{P2G}} \left| P_t^{\mathrm{P2G}} - P_{t-5}^{\mathrm{P2G}} \right| \tag{42}$$

$$C_t^{\mathrm{EBRc}} = c_{\mathrm{RC}}^{\mathrm{EB}} \left| P_t^{\mathrm{EB}} - P_{t-5}^{\mathrm{EB}} \right| \tag{43}$$

$c_{\mathrm{RC}}^{\mathrm{Ele}}$, $c_{\mathrm{RC}}^{\mathrm{P2G}}$, and $c_{\mathrm{RC}}^{\mathrm{EB}}$ are the prices for purchasing power from the external power grid, P2G equipment, and EB equipment.

The penalty term of the second part of the reward function includes the unbalanced supply and demand costs of electricity and heat energy, as defined in Eq (44).

$$C_t^{\mathrm{U2}} = c^{\mathrm{Enb}} P_t^{\mathrm{Enb}} + c^{\mathrm{Hnb}} H_t^{\mathrm{Hnb}} \tag{44}$$

c^{Enb} and c^{Hnb} are the penalty prices for the imbalance between supply and demand of electricity and heat energy; P_t^{Enb} and H_t^{Hnb} refer to the power with imbalanced supply and demand of electrical and heat energy, where P_t^{Enb} and H_t^{Hnb} are defined as Eqs (45) and (46).

$$P_t^{\mathrm{Enb}} = \left| \left(P_t^{\mathrm{load}} + P_t^{\mathrm{P2G}} + P_t^{\mathrm{EB}} + \delta_t^{\mathrm{ES,ch}} P_t^{\mathrm{ES,ch/dis}} \right) - \left(P_t^{\mathrm{Ele}} + P_t^{\mathrm{PV}} + \left(1 - \delta_t^{\mathrm{ES,ch}}\right) P_t^{\mathrm{ES,ch/dis}} + P_t^{\mathrm{GT}} \right) \right| \tag{45}$$

$$H_t^{\mathrm{Enb}} = \left| \left(H_t^{\mathrm{EB}} + H_t^{\mathrm{GT}} + H_t^{\mathrm{GB}} + \left(1 - \delta_t^{\mathrm{HS,ch}}\right) P_t^{\mathrm{HS,ch/dis}} \right) - \left(H_t^{\mathrm{load}} + \delta_t^{\mathrm{HS,ch}} P_t^{\mathrm{HS,ch/dis}} \right) \right| \tag{46}$$

The penalty term C_t^{U3} of the third part of the reward function includes the cumulative operating power regulation cost of GT equipment and GB equipment. The purpose of considering this regulation cost is to prevent the upper power and heat system from excessively relying on the electricity and heat energy supply of GT and GB equipment in the lower gas system. The definition of C_t^{U3} is shown in Eq (47).

$$C_t^{\mathrm{U3}} = c^{\mathrm{cpa}} \left(\sum_{i=1}^{5} \Delta G_{t+i}^{\mathrm{GB}} + \sum_{i=1}^{5} \Delta G_{t+i}^{\mathrm{GT}} \right) \tag{47}$$

c^{cpa} is the penalty price for the cumulative power adjustment of GT and GB devices.

In summary, the reward function obtained by the upper PPO after executing action a_t^{UP} based on state s_t^{UP} is shown in Eq (48).

$$R_t^{\mathrm{UP}} = -\left[\left(C_t^{\mathrm{U1}} + C_t^{\mathrm{U2}} + C_t^{\mathrm{U3}}\right) + I\left[\left(P_t^{\mathrm{Enb}} + H_t^{\mathrm{Hnb}}\right) < \alpha\right] + r_0\right] \times 0.001 \tag{48}$$

$I\left[\left(P_t^{\mathrm{Enb}} + H_t^{\mathrm{Hnb}}\right) < \alpha\right] + r_0\Big]$ is the indicator function; α is the maximum cumulative electricity and heat energy supply and demand imbalance value; r_0 is a constant (which can change the cumulative return from negative to positive, improving the stability and convergence speed of the model).

4.2. Short time scale energy management of lower PPO

4.2.1. Lower PPO state space

The state space of the lower PPO includes the observed states of the lower gas system and the energy parameters generated by the upper power and heat system, as shown in Eq (49).

$$s_t^{\mathrm{Down}} = \left\{c_t^{\mathrm{Gas}}, P_t^{\mathrm{load}}, H_t^{\mathrm{load}}, G_t^{\mathrm{load}}, t, P_t^{\mathrm{Ele}}, P_t^{\mathrm{P2G}}, P_t^{\mathrm{EB}}, P_t^{\mathrm{ES,ch/dis}}, P_t^{\mathrm{HS,ch/dis}}\right\} \tag{49}$$

c_t^{Gas} is the gas purchase price of the natural gas station; G_t^{load} is the gas load.

4.2.2. Lower PPO state space

The action space of the lower PPO intelligent agent is shown in Eq (50).

$$a_t^{\mathrm{Down}} = \left\{G_t^{\mathrm{Gas}}, G_t^{\mathrm{GT}}, G_t^{\mathrm{GB}}, P_t^{\mathrm{GS,ch/dis}}\right\} \tag{50}$$

G_t^{Gas} represents the power of natural gas station to purchase gas; G_t^{GT} is the gas consumption power of the GT equipment; G_t^{GB} is the gas consumption power of the GB device; $P_t^{\mathrm{GS,ch/dis}}$ is the storage/discharge power of the air storage tank.

Adding random disturbances to the lower action space to enhance the perception of the environment, the improved lower PPO action space $a_t^{\mathrm{Down}\prime}$ is shown in Eq (51).

$$a_t^{\mathrm{Down}\prime} = \tau a_t^{\mathrm{Down}} + (1-\tau) m_t \tag{51}$$

4.2.3. Lower PPO reward function

The lower PPO reward function includes the first part penalty term C_t^{D1} and the second part penalty term C_t^{D2}, which are used to guide the intelligent agent to select actions based on the current state.

The first part of the reward function penalty term is shown in Eq (52).

$$C_t^{\mathrm{D1}} = C_t^{\mathrm{GasRc}} + C_t^{\mathrm{GT}} + C_t^{\mathrm{GB}} + C_t^{\mathrm{Gas}} + C_t^{\mathrm{GS}} \tag{52}$$

C_t^{GasRc} represents the cost of external gas purchase power regulation; C_t^{GT} and C_t^{GB} are the operating power regulation costs of GT equipment and GB equipment, where C_t^{GasRc} , C_t^{GT}, and C_t^{GB} are defined as Eqs (53–55).

$$C_t^{\mathrm{GasRc}} = c_{\mathrm{Rc}}^{\mathrm{Gas}} \left| G_t^{\mathrm{pur}} - G_{t-1}^{\mathrm{pur}} \right| \tag{53}$$

$$C_t^{\mathrm{GT}} = c_{\mathrm{Rc}}^{\mathrm{GT}} \left| G_t^{\mathrm{GT}} - G_{t-1}^{\mathrm{GT}} \right| \tag{54}$$

$$C_t^{\mathrm{GB}} = c_{\mathrm{Rc}}^{\mathrm{GB}} \left| G_t^{\mathrm{GB}} - G_{t-1}^{\mathrm{GB}} \right| \tag{55}$$

The second part of the reward function penalty term is shown in Eq (56):

$$C_t^{\mathrm{D2}} = c^{\mathrm{Gnb}} G_t^{\mathrm{Gnb}} \tag{56}$$

c^{Gnb} is the penalty price for the imbalance between supply and demand in the gas system; G_t^{Gnb} is the power of the imbalance between gas energy supply and demand, where G_t^{Gnb} is defined as shown in Eq (57).

$$G_t^{\mathrm{Gnb}} = \left| \left(G_t^{\mathrm{Gas}} + G_t^{\mathrm{P2G}} + \left(1 - \delta_t^{\mathrm{GS,ch}}\right) P_t^{\mathrm{GS,ch/dis}} \right) - \left(G_t^{\mathrm{load}} + G_t^{\mathrm{GB}} + G_t^{\mathrm{GT}} + \delta_t^{\mathrm{GS,ch}} P_t^{\mathrm{GS,ch/dis}} \right) \right| \tag{57}$$

In summary, the reward function obtained by the lower PPO after executing action a_t^{Down} based on state s_t^{Down} is shown in Eq (58).

$$R_t^{\mathrm{Down}} = -\left[\left(C_t^{\mathrm{D1}} + C_t^{\mathrm{D2}} \right) + I\left[G_t^{\mathrm{Gnb}} < \alpha_0 \right] + r_0 \right] \times 0.001 \tag{58}$$

In the equation, α_0 is the maximum cumulative gas energy supply and demand imbalance value.

5. Example simulation and result analysis

5.1. Example simulation settings

The electricity, heat, and gas loads, and photovoltaic power generation data in this article are sourced from a small domestic park. The structure is shown in Figure 1, and the system equipment parameters and other simulation parameters are shown in Tables 1 and 2 [26,27]. The electricity price of the power system is the time-of-use electricity price as shown in Table 3 [28], and the natural gas unit price is a fixed price of 4.2 yuan/Nm^3 (Normal cubic meter). Most of the formulas in the text are based on references [29]. To verify the superiority of the double-layer PPO algorithm proposed in this article compared to the single-layer PPO algorithm in terms of runtime, it is necessary to control variables and ensure that the CPU and memory settings of the double-layer PPO algorithm and the single-layer PPO algorithm are consistent, both completed by a computer equipped with an I7-7700 CPU and 16 GB RAM.

This experiment was implemented on the TensorFlow platform, with five upper PPO action space control objects and four lower PPO action space control objects. The same actor and critical networks, hidden layer, and activation function are used in the upper and lower levels. The number of hidden layers for both the actor and the critical network is 3, each layer containing 256 neurons; the activation function is ReLU. The network weights are updated using the Adam optimizer.

Table 1. PIES equipment simulation operation parameters.

Equipment	Lower power limit/MW	Upper power limit/MW
Gas turbine	0	1
Gas boiler	0	0.55
Electric boiler	0	0.4
P2G	0	0.5
Battery	0.06	0.24
Heat storage tank	0.04	0.2
Gas storage tank	0.07	0.35

Table 2. PIES remaining simulation parameters.

Parameter	Numeric value	Parameter	Numeric value	Parameter	Numeric value
$\Delta G_{\max}^{\mathrm{GT}}$	0.1 MW	$\Delta G_{\max}^{\mathrm{GB}}$	0.05 MW	$\Delta P_{\max}^{\mathrm{EB}}$	0.1 MW
$\Delta P_{\max}^{\mathrm{P2G}}$	0.04 MW	$\eta^{\mathrm{GT-E}}$	43%	η^{GB}	97%
η^{EB}	93%	η^{P2G}	85%	$\eta^{\mathrm{GT-loss}}$	15%
$\mu^{\mathrm{GB-loss}}$	5%	$\mu^{\mathrm{EB-loss}}$	3%	$\eta^{\mathrm{P2G-loss}}$	20%
$P_{\max}^{\mathrm{ES,ch/dis}}$	0.08 MW	$P_{\max}^{\mathrm{GS,ch/dis}}$	0.08 MW	$P_{\max}^{\mathrm{HS,ch/dis}}$	0.03 MW
$\eta^{\mathrm{ES,ch}}$	90%	$\eta^{\mathrm{HS,ch}}$	80%	$\eta^{\mathrm{GS,ch}}$	90%
$\eta^{\mathrm{ES,dis}}$	110%	$\eta^{\mathrm{HS,dis}}$	115%	$\eta^{\mathrm{GS,dis}}$	110%
$P_{\max}^{\mathrm{Ele}}$	2 MW	$G_{\max}^{\mathrm{Gas}}$	1.5 MW	$c_{\mathrm{RC}}^{\mathrm{Ele}}$	1 yuan
$c_{\mathrm{RC}}^{\mathrm{P2G}}$	1 yuan	$c_{\mathrm{RC}}^{\mathrm{EB}}$	1 yuan	$c_{\mathrm{RC}}^{\mathrm{Gas}}$	1.5 yuan
$c_{\mathrm{RC}}^{\mathrm{GT}}$	1.5 yuan	$c_{\mathrm{RC}}^{\mathrm{GB}}$	15000 yuan	c^{HS}	6 yuan
c^{GS}	6	$c_{\mathrm{cape}}^{\mathrm{ES}}$	45	m	694
n	1.98	q	0.016	D_{R}^{ES}	0.8
σ^{A}	0.001	σ^{C}	0.002	γ	0.9

Table 3. Time-of-use electricity price.

Time span	Time	Electrovalence/(yuan/kwh)
Valley period	23:00–morrow 07:00	0.2
Peacetime period	morrow 07:00–12:00 19:00–23:00	0.6
Peak period	12:00–19:00	1.1

5.2. *Validation experiment and result analysis*

First, under the experimental environment and simulation parameters provided above, the upper and lower PPO models were trained separately, and the convergence characteristics of the upper and lower PPO were obtained, as shown in Figure 5.

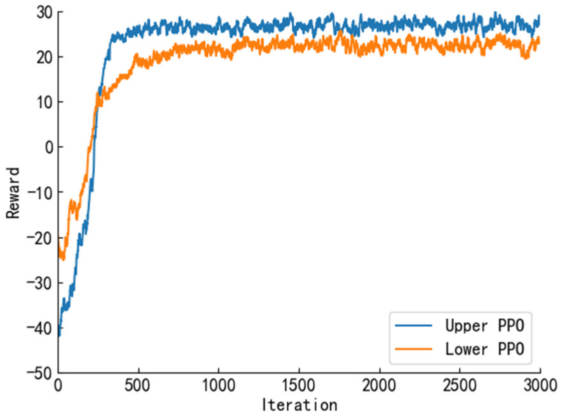
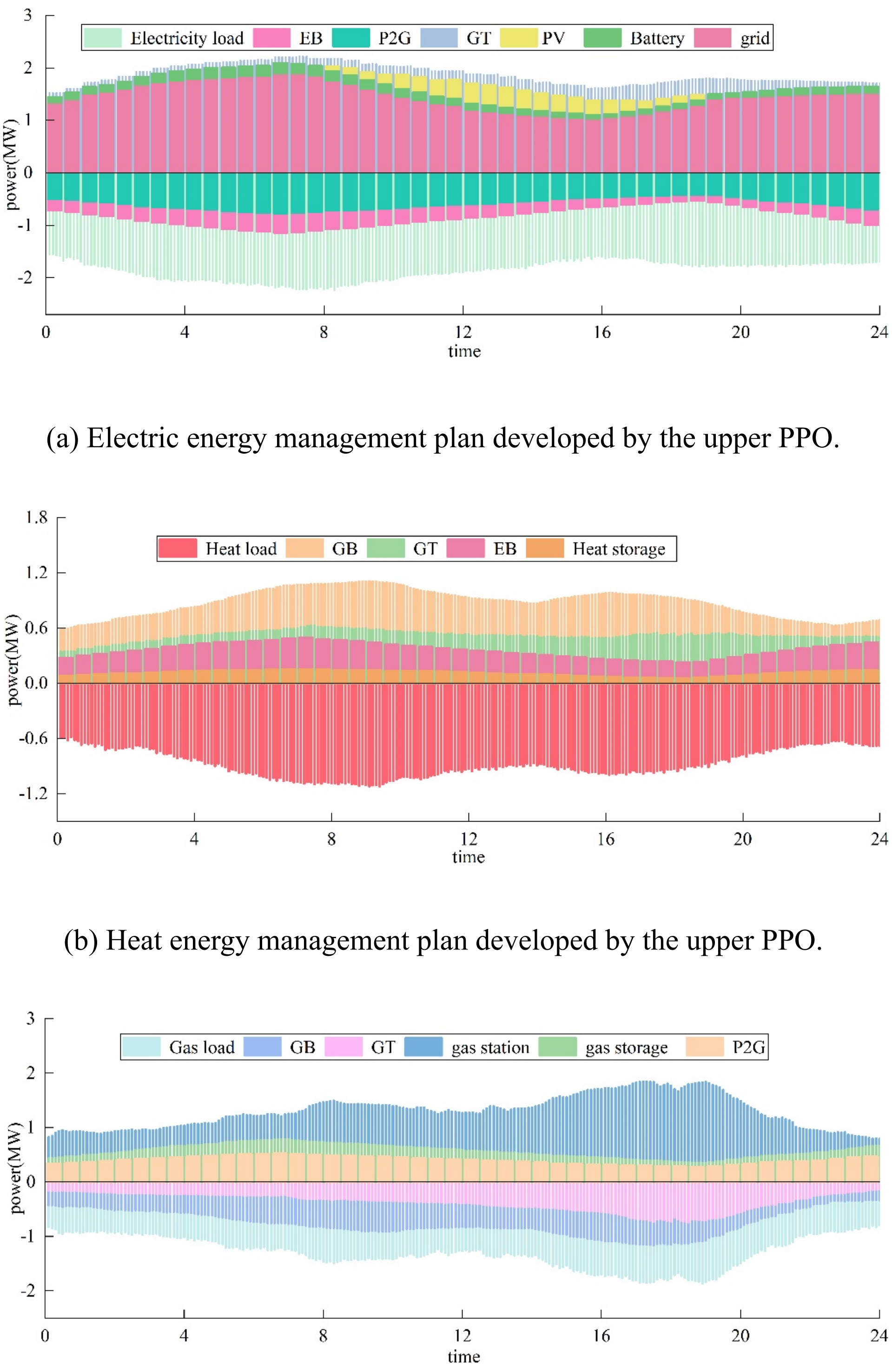


Figure 5. Upper and lower PPO reward function change.

From Figure 5, it can be seen that in the initial stage of training, the decision reward values of the upper and lower agents are relatively small due to their unfamiliarity with the environment. With the continuous interaction between agents and the environment, upper and lower agents continuously accumulate experience to update network weights, and the reward values obtained gradually increase until convergence. The upper PPO converges after approximately 400 rounds of training, while the lower PPO converges after approximately 500 rounds of training. The upper PPO converges faster because the reward function of the upper PPO includes the cumulative adjustment of the operating power of GT and GB devices in the lower PPO. The reward functions of the upper PPO and lower PPO both converge quickly, effectively adjusting the energy purchase, conversion, and storage behavior of the power, heat, and gas systems.

Then, PIES enters normal operation mode, and the upper PPO updates the energy management status of the power and heat systems at a long-term scale of 30 min, while the lower PPO updates the energy management status of the gas system at a short-term scale of 6 min, as shown in Figure 6: The upper and lower parts display the total energy supply and the total energy demand power, respectively.

During the valley electricity price period of the power grid, the lower electricity price drives the increase in external power purchase, causing the GT equipment to operate at lower power, while the operating power of EB and P2G equipment rebounds and energy storage increases. The power system balance is mainly maintained by external power purchase on the grid side, as shown in Figure 6a. On the heating network side, due to the GT equipment adopting the “electricity and heating” mode, the operating power is relatively low, and the supply and demand balance of the heat system is mainly maintained by the GB and EB equipment, as shown in Figure 6b. On the gas network side, the balance of the gas system is mainly maintained by external gas purchasing power and P2G equipment, as shown in Figure 6c.



(a) Electric energy management plan developed by the upper PPO.

(b) Heat energy management plan developed by the upper PPO.

(c) Gas management plan developed by lower PPO.

Figure 6. Management scheme of electric energy, heat energy, and gas output by upper and lower PPO.

From Figure 6a, it can be seen that the grid side maintains a balance between supply and demand of the power system by flexibly adjusting external purchasing power, batteries, GT, P2G, and EB equipment during the normal and peak periods of electricity prices. Due to the increase in electricity prices, external power purchases have decreased, resulting in a corresponding decrease in the operating power of P2G and EB equipment, a decrease in energy storage, and a corresponding increase in the operating power of GT equipment to maintain a balance between supply and demand in the power system. At this point, the power system mainly applies photovoltaic power generation and GT equipment to make up for the supply gap of external purchased power. The photovoltaic output becomes higher during the 10:00–13:00 period, and the output of GT equipment is higher during the 17:00–20:00 period. The decrease in operating power of the P2G and EB equipment reduces the energy supply of the gas system and heat system. The gas network side compensates for the supply gap of P2G equipment by increasing external gas purchasing power, maintaining the balance of the gas system, as shown in Figure 6b. The heating network side mainly uses the GB and GT equipment to fill the supply gap of the EB equipment and maintain the balance of the heat system, as shown in Figure 6c.

5.3. Comparative experiments and result analysis

Comparative experiments were conducted on the proposed PPO reinforcement learning optimization algorithm for PIES considering multiple time scales, PPO algorithm, and traditional methods. The experimental data of the three methods were randomly selected from the test set, with a total scheduling period of 24 h and a time scale of 30 min. The traditional method uses the solving software CPLEX (a mathematical modeling tool that can help solve the optimal or feasible solution in the model). The operating costs of the three methods are shown in Table 4.

Table 4. Daily operation cost of different energy management optimization methods.

Operating cost/yuan	Proposed method	PPO	Traditional method	Effect 1	Effect 2
Maximum value	28735	33594	35386	5.06%	14.46%
Minimum value	22378	30391	32462	6.37%	26.3%
Average value	25576	32358	33490	3.38%	20.9%
Carbon emission	1169	1552	1625	4.50%	24.6%
Training time	5973 s	13685 s	–	–	56.35%

From Table 4, it can be seen that using the algorithm proposed in this article for energy management has the lowest daily average operating cost and carbon emission cost. The former is 73.5% for the PPO algorithm and 69.1% for the traditional method; the carbon emission cost is 75.3% for the PPO algorithm method and 71.9% for the traditional method.

The traditional method relies on an accurate prediction of renewable energy and loads. To solve this problem, this paper adopts the PPO algorithm in reinforcement learning. Reinforcement learning is a model-free method that does not rely on accurate prediction and modeling of source loads and can effectively deal with uncertain energy supply problems such as photovoltaics. Meanwhile, compared to traditional methods, this article divides PIES into upper and lower parts, which can meet the differences in time scales of various energy subsystems. The PPO algorithm reduces operating costs by 3.38% and carbon emissions by 4.50%, compared with traditional methods.

Using a dual-layer PPO management model to partition and manage the same number of control variables can effectively improve the training success rate and convergence speed of the PPO management model, thereby reducing its effective training time. This is because the double-layer PPO overcomes the curse of dimensionality in model training by partitioning control variables. In addition, the single-layer PPO is limited by the system management time scale of 30 min, making it difficult to quickly adjust the three energy sources' supply and demand situations, resulting in the overall economic benefits of PIES being lower than the double-layer PPO algorithm proposed in this article. The simulation results show that the double-layer PPO algorithm reduces operating costs by 20.9% and carbon emissions by 2.5% compared with the single-layer PPO algorithm.

5.4. Analysis of energy loss results

To verify the adaptive ability of the proposed solution to energy loss, the thermal load in the PIES system was incrementally increased, and the dynamic scheduling solution analysis of PIES was conducted again to see if it meets the energy demand of the thermal load in PIES.

(1) The power variation of a gas turbine considering thermal energy loss is shown in Figure 7.

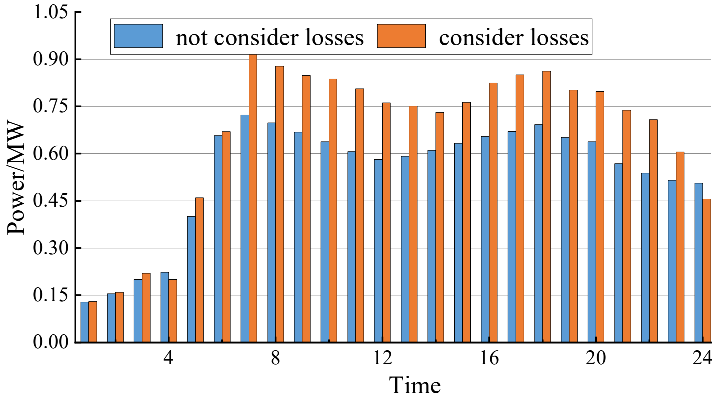


Figure 7. Gas turbine thermal energy loss diagram.

(2) The power variation of a gas boiler considering thermal energy loss is shown in Figure 8.

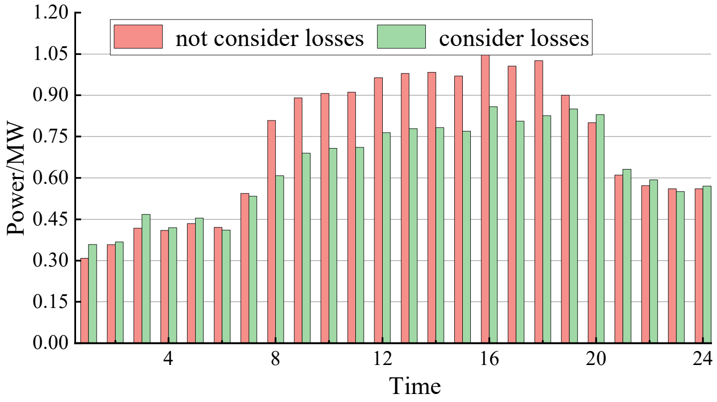


Figure 8. Gas boiler thermal energy loss diagram.

(3) The power variation of an electric boiler considering thermal energy loss is shown in Figure 9.

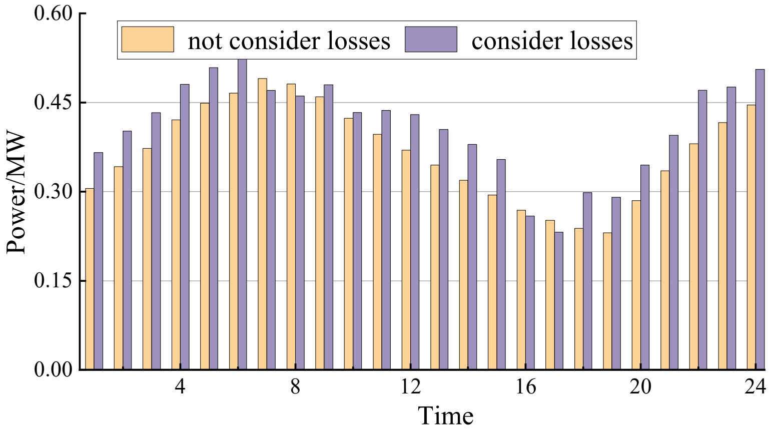


Figure 9. Electric boiler thermal energy loss diagram.

As shown in Figures 7–9, during the valley period of the electricity price, the thermal power output of the electric boiler changes more significantly. During the normal and peak periods of the electricity price, the thermal power output of the gas turbine and gas boiler changes more significantly. This indicates that the gas turbine, gas boiler, and electric boiler proposed in this paper can all adapt to dynamic scheduling decisions and maintain the supply-demand balance of thermal energy in PIES.

6. Conclusions

This article proposes an integrated-energy PPO reinforcement learning optimization algorithm for a park-integrated energy management system that considers multiple time scales to address the uncertainty of photovoltaic output and load changes, as well as the differences in time scales of heterogeneous energy subsystem management. The method divides PIES into two layers, upper and lower, with the upper layer containing power and heat systems (including photovoltaic power generation), and the lower layer containing gas systems. The main conclusions are as follows:

This article uses the PPO algorithm in deep reinforcement learning to establish a PIES energy management model, which can make real-time decisions and effectively respond to the uncertainty of photovoltaic output and load changes.

The different time scales of the upper and lower layers in PIES not only meet the needs of heterogeneous energy subsystems for energy management time scale differences but also timely adjust the output of equipment in each subsystem to meet the energy supply and demand balance in PIES.

Compared with single-layer PPO and traditional energy management methods, the method proposed in this article has advantages in reducing carbon emissions and improving the economic benefits of PIES. The PPO algorithm reduces operating costs by 3.38% and carbon emissions by 4.50% compared with traditional methods. The simulation results show that the double-layer PPO algorithm reduces operating costs by 20.9% and carbon emissions by 2.5% compared with the single-layer PPO algorithm.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The work is supported by Inner Mongolia Electric Power (Group) Co., Ltd Technology Project [Grant No. LX2023-5-11].

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, Linrong Wang; methodology, Haixiao Zhang and Xiang Feng; validation, Ruifen Zhang and Guilan Wang; formal analysis, Haixiao Zhang, Guilan Wang and Zhengran Hou; writing—original draft preparation, Haixiao Zhang; writing—review and editing, Guilan Wang; supervision, Linrong Wang and Guilan Wang. All authors have read and agreed to the published version of the manuscript.

References

1. Feng J, Nan J, Wang C, et al. (2022) Source-load coordinated low-carbon economic dispatch of electric-gas integrated energy system based on carbon emission flow theory. *Energies* 15: 3641–3652. https://doi.org/10.3390/en15103641
2. Bhowmik C, Bhowmik S, Ray A, et al. (2017) Optimal green energy planning for sustainable development: A review. *Renewable Sustainable Energy Rev* 71: 796–813. https://doi.org/10.1016/j.rser.2016.12.105
3. Li P, Wu D, Li Y, et al. (2020) A multi-objective union optimal configuration strategy for multi-microgrid integrated energy system considering bargaining game. *Power Syst Tech* 44: 3680–3690. https://doi.org/10.1016/p.st.20203680
4. Lv J, Zhang S, Cheng H, et al. (2021) Review on district-level integrated energy system planning considering interconnection and interaction. *Pro CSEE* 41: 4001–4021. https://doi.org/10.3390/en20214001
5. Yu X, Xu X, Chen S, et al. (2016) A brief review to integrated energy system and energy internet. *Trans China Electro Society* 31: 1–13. https://doi.org/10.1016/eprint/104480
6. Ding T, Jia W, Shahidehpour M, et al. (2022) Review of optimization methods for energy hub planning, operation, trading, and control. *IEEE Trans Sustainable Energy* 13: 1802–1818. https://doi.org/10.1109/TSTE.2022.3172004
7. Khodadadi A, Abedinzadeh T, Alipour H, et al. (2023) Optimal operation of energy hub systems under resiliency response options. *J Electr Comput Eng* 20: 23–36. https://doi.org/10.1155/2023/2590362
8. Song D, Meng W, Dong M, et al. (2022) A critical survey of integrated energy system: Summaries, methodologies and analysis. *Energy Convers Manage* 266: 58–63. https://doi.org/10.1016/j.enconman.2022.115863
9. Jiang X, Sun C, Cao L, et al. (2022) Semi-decentralized energy routing algorithm for minimum-loss transmission in community energy internet. *Int J Electrical Power Energy Syst* 135: 35–47. https://doi.org/10.1016/j.ijepes.2021.107547

10. Yang M, Cui Y, Huang D, et al. (2022) Multi-time-scale coordinated optimal scheduling of integrated energy system considering frequency out-of-limit interval. *Inter J Elect Power Energy Syst* 141: 68–81. https://doi.org/10.1016/j.ijepes.2022.108268
11. Hu K, Wang B, Cao S, et al. (2022) A novel model predictive control strategy for multi-time scale optimal scheduling of integrated energy system. *Energy Rep* 8: 7420–7433. https://doi.org/10.1016/j.egyr.2022.05.184
12. Li X, Wang W, Wang H (2021) Hybrid time-scale energy optimal scheduling strategy for integrated energy system with bilateral interaction with supply and demand. *Appl Energy* 285: 458–463. https://doi.org/10.1016/j.apenergy.2021.116458
13. Li P, Guo T, Abeysekera M, et al. (2021) Intraday multi-objective hierarchical coordinated operation of a multi-energy system. *Energy* 228: 5–28. https://doi.org/10.1016/j.energy.2021.120528
14. Cheng S, Wang R, Xu J, et al. (2021) Multi-time scale coordinated optimization of an energy hub in the integrated energy system with multi-type energy storage systems. *Sustainable Energy Technol Assess* 47: 327–335. https://doi.org/10.1016/j.seta.2021.101327
15. Zhang B, Hu W, Li J, et al. (2020) Dynamic energy conversion and management strategy for an integrated electricity and natural gas system with renewable energy: Deep reinforcement learning approach. *Energy Convers Manage* 220: 63–75. https://doi.org/10.1016/j.enconman.2020.113063
16. Xu Z, Han G, Liu L, et al. (2021) Multi-energy scheduling of an industrial integrated energy system by reinforcement learning-based differential evolution. *IEEE Trans Green Commun Netw* 5: 1077–1090. https://doi.org/10.1109/TGCN.2021.3061789
17. Foruzan E, Soh LK, Asgarpoor S (2018) Reinforcement learning approach for optimal distributed energy management in a microgrid. *IEEE Trans Power Syst* 33: 5749–5758. https://doi.org/10.1109/TPWRS.2018.2823641
18. Gorostiza FS, Gonzalez-Longatt FM (2020) Deep reinforcement learning-based controller for SOC management of multi-electrical energy storage system. *IEEE Trans Smart Grid* 11: 5039–5050. https://doi.org/10.1109/TSG.2020.2996274
19. Zhang X, Liu Y, Duan J, et al. (2021) DDPG-based multi-agent framework for SVC tuning in urban power grid with renewable energy resources. *IEEE Trans Power Syst* 36: 5465–5475. https://doi.org/10.1109/TPWRS.2021.3081159
20. Zhu X, Yang J, Liu Y, et al. (2019) Optimal scheduling method for a regional integrated energy system considering joint virtual energy storage. *IEEE Access* 7: 138260–138272. https://doi.org/10.1109/ACCESS.2020.3046743
21. Li Y, Zhang F, Li Y, et al. (2021) An improved two-stage robust optimization model for CCHP-P2G microgrid system considering multi-energy operation under wind power outputs uncertainties. *Energy* 223: 48–60. https://doi.org/10.1016/j.energy.2021.120048
22. Fotopoulou M, Pediaditis P, Skopetou N, et al. (2024) A Review of the Energy Storage Systems of Non-Interconnected European Islands. *Sustainability* 16: 1572. https://doi.org/10.3390/su16041572
23. Rious V, Perez Y (2014) Review of supporting scheme for island power system storage. *Renewable Sustainable Energy Rev* 29: 754–765. https://doi.org/10.1016/j.rser.2013.08.015
24. Guo M, Mu Y, Jia H, et al. (2021) Electric/thermal hybrid energy storage planning for park-level integrated energy systems with second-life battery utilization. *Adva Appl Energy* 4: 64–75. https://doi.org/10.1016/j.adapen.2021.100064

25. Li Z, Zhang F, Liang J, et al. (2015) Optimization on microgrid with combined heat and power system. *Proc CSEE* 35: 3569–3576. https://doi.org/10.13334/j.0258-8013.pcsee.2015.14.011
26. Zhou S, Hu Z, Gu W, et al. (2020) Combined heat and power system intelligent economic dispatch: A deep reinforcement learning approach. *Inter J Electrical Power Energy Syst* 120: 106016. https://doi.org/10.1016/j.ijepes.2020.106016
27. Yang HZ, Li ML, Jiang ZY, et al. (2020) Multi-time scale optimal scheduling of regional integrated energy systems considering integrated demand response. *IEEE Access* 8: 5080–5090. https://doi.org/10.1109/ACCESS.2019.2963463
28. Yang T, Zhao L, Liu Y, et al. (2021) Dynamic economic scheduling of integrated energy systems based on deep reinforcement learning. *Power Syst Autom* 45: 39–47. https://doi.org/10.7500/AEPS20200405004
29. Dong J, Wang HX, Zhou XR, et al. (2023) Low carbon economic dispatch of electricity gas heat integrated energy system considering comprehensive demand response. *J North China Electr Power Univ, Nat Sci Ed* 50: 81–90. https://doi.org/10.3969/j.ISSN.1007-2691.2023.03.08