ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Intelligent body-based power macromodeling in the process of achieving dual-carbon goals in key industries

Zeqi Zhang[1,*], *Qingge Ji*[1] *and Enhong Wu*[1]

[1] Sun Yat-sen University, Guangzhou, Guangdong, 510275, China

SUMMARY: *In this paper, an integrated dispatch planning model including carbon trading cost is designed with multiple types of constraints, backed by the power provider model and taking into account the uncertainty within the power system. Subsequently, relying on the Actor - Critic framework in reinforcement learning, the dynamic optimal dispatch model of energy in key industries is constructed, and the Deep Q-Network (DQN) algorithm is employed to fine - tune and train the model's parameters, aiming to acquire the most optimal dispatch strategy. The outcomes of the simulation indicate that the model put forward in this paper can effectively exploit the carbon emission reduction potential of key industries within the framework of the dual - carbon goal, and there are also large differences in the power supply planning of key enterprises under different carbon trading costs. Therefore, By leveraging the intelligent physical data of the large - scale electric power model, it is possible to comprehensively explore the energy consumption of crucial industries within the power grid. so as to help the key industries to better formulate the dual-carbon target and decision-making, and to enhance the green development level of the key industries.*

KEYWORDS: *electricity supplier model; carbon transaction cost; reinforcement learning; Actor-Critic framework; DQN algorithm; key industries; dual-carbon targets*

1 Introduction

Over the past few years, nations have placed greater emphasis on eco - friendly and low - carbon progress. More than 130 countries have established the bold objective of attaining carbon neutrality around the mid - 21st century, and the global acceleration of green and low-carbon transformation will inject new green momentum into the economic recovery in the post-epidemic era [1-3]. China, on the other hand, emphasizes that carbon dioxide emissions strive to peak by 2030 and work towards carbon neutrality by 2060 [4, 5]. Realizing the vision of carbon peaking and carbon neutrality on schedule will become one of the main lines of China's economic and social development in the coming decades. As the basic industry of national economic operation, the power system is not only an important hub for energy transition, but also a core force to promote the realization of low-carbon and intelligent in buildings, transportation, and industry [6, 7].

In the wake of the advancement of artificial intelligence, particularly the utilization of cognitive capabilities and large - scale models, it effectively promotes the change of power system [8, 9]. The key parameters of big models are big data, high parameters, and strong arithmetic, i.e., high-parameter models trained by strong arithmetic with the support of big data [10, 11]. Big models are able to learn richer and finer feature representations and perform well

*ZZq_111025@163.com
https://doi.org/10.65102/is2026608
Published: 30 April, 2026

in tasks such as data prediction, classification and generation [12]. An intelligent body is an “autonomous unit” with the ability to perceive, make decisions, and act in different environments, following goals, responding to changes in the environment, and continuously adapting its strategies [13, 14]. In the wave of AI technology iteration sweeping the world and the depth of digital transformation, the combination of intelligent bodies and large models as the core carrier to achieve human-machine collaboration and automated decision-making, is reshaping the development pattern of various industries [15]. For the power system, the deep integration of the two has given rise to a new pattern of intelligent power big model, which can accelerate the power system to low-carbon and digital. Simultaneously, during the process of attaining the dual - carbon goal in key industries, it is also possible to provide impetus for structural optimization and improve the carbon management capabilities.

Confronted with the present uncertainty regarding the power system's load and the inability of crucial industries to attain the dual - carbon goal, a comprehensive optimal scheduling model that takes load uncertainty and carbon trading into account has been established, and a dynamic optimization strategy based on the Actor-Critic framework of reinforcement learning has been designed for iterative solving of the model. In order to analyze the role of power dispatch in key industries for emission reduction, the sensitivity of carbon trading cost under different scenarios and the impact of carbon trading price on the expansion of key enterprises' units in three levels, namely, high, normal and low, are analyzed. This study provides over a new direction for the dual-carbon target decision-making of key industries, which can effectively promote the greening of key enterprises.

2 System model

Guided by the dual-carbon goal, the question at hand is how to select energy - saving and emission - reducing technologies to achieve low - carbon emission reduction in key industries that are characterized by high energy consumption and high emissions, has become an important topic for research in this field. With the support of electric power big model, comprehensive planning of electricity consumption of key industries is carried out to realize the optimal scheduling of grid resources, which can meet the production of key industries while reducing the cost of carbon trading, and promote the key industries to effectively realize the dual-carbon goal.

2.1 Electricity supplier model

2.1.1 Traditional energy suppliers

For any supplier J, we use l_h^j to denote the total amount of electricity produced by the supplier at moment h. In general, we can assume that the power supplier in a moment of power production costs with the increase in power gradually increase and increase the magnitude of the gradual decrease in the cost of the power supplier's function is monotonically increasing and strictly convex, quadratic function can be a very good representation of the law of change of the cost of power production, the formula is expressed as follows:

$$C_h^j\left(l_h^j\right)=a_h^j\left(l_h^j\right)^2+b_h^j l_h^j+c_h^j \tag{1}$$

where C_h^j is the cost of electricity production of electricity supplier j at moment h, $a_h^j>0$, $b_h^j\geq 0$, $c_h^j\geq 0$ are constant coefficients.

The electricity production of the supplier cannot exceed the capacity. The capacity of power generation is determined by the limitations of the power generation infrastructure, which includes the limitations of heat generation, the limitations of the number of power generation equipments and the upper limit of power generation of the equipments, etc. We denote the capacity of power generation by the power supplier. Let us represent the power - generation capacity of the power provider by the symbol $l_{\max}^{j}$. The limitations on the power generation of the power provider can be presented in the following manner:

$$l_h^j \le l_{\max}^j \tag{2}$$

We consider that different power suppliers supply different amounts of electricity to each user, and use $s_{i,h}^j$ to denote the amount of electricity that power supplier j gives to sell to user i at moment h, which can be expressed as:

$$s_{i,h}^j = \min\left(l_{i,h}^j, D_{i,h}^j\right) \tag{3}$$

where $l_{i,h}^j$ and $D_{i,h}^j$ denote, respectively, the amount of electricity produced by the electricity supplier j at moment h planned for the customer i and the amount of electricity demanded by the customer i at moment h for the electricity supplier j.

In general, the benefit of the electricity supplier is the profit obtained from all the electricity sold minus its cost of electricity production, and using $p_{i,h}^j$ to denote the unit price of electricity set by the electricity supplier j to the user i, the welfare function of the electricity supplier j at moment h, f_h^j, can be presented in the following manner:

$$f_h^j\left(p_{i,h}^j, s_{i,h}^j\right) = \sum_{i=1}^{I} p_{i,h}^j s_{i,h}^j - C_h^j\left(l_h^j\right) \tag{4}$$

The first term of the equation represents the sum of the profits made by the electricity supplier j from selling electricity to all customers, and the latter term is the electricity production cost of the supplier.

2.1.2 New energy suppliers

The power supply decision of the new energy supplier is similar to that of the traditional energy supplier, with the difference that the new energy generator has upper and lower limits and there is a mechanism to recover the excess PV power from the customers.

Let the power generation of the new energy supplier in each time period be S_r^k, then $S_{r,\min}^k \le S_r^k \le S_{r,\max}^k$. Considering that there are losses in the process of power generation by the new energy supplier, the loss rate is set to be κ_r and $0 < \kappa_r < 1$. When the actual power generation is S_r^k, the production loss is $I_r^k = \kappa_r S_r^k$, and the actual power that can be sold is:

$$L_r^k = S_r^k - I_r^k = \left(1 - \kappa_r\right) S_r^k \tag{5}$$

First,a discussion will take place regarding the combination of the electricity storage of the new - energy power provider and the electricity retrieved from the customer end. Let the total

purchased power of the new energy supplier to acquire new energy electricity from all the users in the k time period at the end of the $k-1$ moment be Q_r^k, i.e:

$$Q_r^k = \sum_{n=1}^{N} q_{n,r}^k \tag{6}$$

Dividing Q_r^k into two parts, the minimum subscription $Q_{r,\min}^k$ and the adjustable subscription Q_{ru}^k, then:

$$Q_r^k = Q_{r,\min}^k + Q_{ru}^k \tag{7}$$

$$Q_{ru}^k = \sum_{n=1}^{N} q_{n,du,r}^k \tag{8}$$

The expense associated with energy storage for providers of new energy sources is:

$$C_r\left(R_r^k\right) = \rho_r R_r^{k2} + \varphi_r \tag{9}$$

where $\rho_r > 0$, φ_r is a constant, and R_r^k denotes the amount of storage at the beginning of the k time period.

The amount of electricity recovered from the customer side by the new energy supplier in the k th time period is:

$$Q_{pv}^k = \sum_{n=1}^{N} q_{n,pv,r}^k \tag{10}$$

At this time, it can be divided into two cases, $R_r^k + Q_{pv}^k \le Q_{r,\min}^k$ and $R_r^k + Q_{pv}^k > Q_{r,\min}^k$, and notate that $L_{r,\max}^k$ denotes the maximum amount of electricity sold by new energy power suppliers in the k th time period.

When $R_r^k + Q_{pv}^k \le Q_{r,\min}^k$, the total storage capacity of the new energy power provider and recovered power is less than the customer's minimum ordering power, then its minimum power sale is:

$$L_{r,\min}^k = Q_{r,\min}^k - R_r^k - Q_{pv}^k \tag{11}$$

L_{ru}^k is the fraction of user-adjustable subscriptions Q_{ru}^k that can or cannot be satisfied, then there are:

$$L_r^k = L_{r,\min}^k + L_{ru}^k \tag{12}$$

The range of values is $L_{ru}^k \in \left[0, L_{r,\max}^k - L_{r,\min}^k\right]$.

Correspondingly, the real electricity production of the new - energy provider is in order $S_r^k = L_r^k / \left(1-\kappa_r\right)$, $S_{r,\min}^k = L_{r,\min}^k / \left(1-\kappa_r\right)$, $S_{ru}^k = L_{ru}^k / \left(1-\kappa_r\right)$, then:

$$S_r^k = S_{r,\min}^k + S_{ru}^k \tag{13}$$

The cost of generating $S_{r,\min}^k$, S_{ru}^k sequentially by the new energy supplier is $C_{r,\min}^k$, C_{ru}^k, which can be expressed as:

$$C_{r,\min}^k\left(L_{r,\min}^k\right) = dL_{r,\min}^k + e \tag{14}$$

$$C_{ru}^k\left(L_{r,\min}^k, L_{ru}^k\right) = \int_{L_{r,\min}^k}^{L_{r,\min}^k + L_{ru}^k} dC_{r,\min}^k - C_{r,\min}^k\left(L_{r,\min}^k\right) \tag{15}$$

where $d > 0, e \geq 0$, is a constant.

At this time the new energy supplier's revenue is:

$$\begin{aligned} W_{sr}^k = {} & p_r^k\left(R_r^k + Q_{pv}^k\right) - C_{sr}\left(R_r^k\right) + p_r^k L_{r,\min}^k - C_{r,s}^k \\ & - C_{r,\min}^k\left(L_{r,\min}^k\right) + p_r^k L_{ru}^k \\ & - C_{ru}^k\left(L_{ru}^k\right) - p_{pv}^k Q_{pv}^k + \sum_{n=1}^{N} C_{DSM}\left(q_{n,r}^k, Q_{n,r}^k\right) \\ & - \sum_{n=1}^{N} R_{DSM}\left(q_{n,r}^k, Q_{n,r}^k\right) + \left[\Delta W_{sr}^{k-1}\right]^{++} \end{aligned} \tag{16}$$

where $C_{r,s}^k$ is the start-up cost of the new energy supplier and is constant. $\left[\Delta W_{sr}^{k-1}\right]^{++}$ is the loss due to the deviation of the purchased power from the generated power in the $k-1$th time period, and defines ΔW_{sr}^{k-1} as the difference between the welfare value W_{sr}^{k-1} in the $k-1$th time period and that of the negotiation phase $W_{sr,d}^{k-1}$ is the difference between:

$$\left[\Delta W_{sr}^{k-1}\right]^{++} = \begin{cases} 0, & \Delta W_{sr}^{k-1} \geq 0 \\ W_{sr}^{k-1} - W_{sr,d}^{k-1}, & \Delta W_{sr}^{k-1} < 0 \end{cases} \tag{17}$$

If $\left[\Delta W_{sr}^{k-1}\right]^{++}$ is positive, the cost of the k th time period is not included; if it is negative, the cost of the k th time period needs to be included. The reserves for the next time period are updated as $R_r^{k+1} = 0$.

When $R_r^k + Q_{pv}^k > Q_{r,\min}^k$, the comparison of the sum of the new energy supplier's stored and recovered electricity with the amount of electricity subscribed by the customer will occur in two cases: $Q_r^k > R_r^k + Q_{pv}^k > Q_{r,\min}^k$, at which time the interval of the amount of electricity supplied by the new energy supplier that can be supplied to the customer L_r^k is is $\left[0, L_{r,\max}^k\right]$; $R_r^k + Q_{pv}^k \geq Q_r^k$, at this time the new energy power supplier does not need to start the production equipment.

When $Q_r^k > R_r^k + Q_{pv}^k > Q_{r,\min}^k$, the new energy power supplier's revenue is:

$$W_{sr}^{k} = p_r^k \left(R_r^k + Q_{pv}^k\right) - C_{sr}\left(R_r^k\right) + p_r^k L_r^k - \left[L_r^k\right]^+ \cdot C_{r,s}^k - p_{pv}^k Q_{pv}^k + \sum_{n=1}^{N} C_{DSM}\left(q_{n,r}^k, Q_{n,r}^k\right) - \sum_{n=1}^{N} R_{DSM}\left(q_{n,r}^k, Q_{n,r}^k\right) + \left[\Delta W_{pv}^{k-1}\right]^{++} \tag{18}$$

where $\left[L_r^k\right]^+ = \begin{cases} 0, & L_r^k = 0 \\ 1, & L_r^k \neq 0 \end{cases}$. The next moment of storage is updated to $R_r^{k+1} = 0$.

When $R_r^k + Q_{pv}^k \geq Q_r^k$, the new energy supplier's revenue is:

$$W_{sr}^{k} = p_r^k Q_r^k - C_{sr}\left(R_r^k\right) - p_{pv}^k Q_{pv}^k + \sum_{n=1}^{N} C_{DSM}\left(q_{n,r}^k, Q_{n,r}^k\right) - \sum_{n=1}^{N} R_{DSM}\left(q_{n,r}^k, Q_{n,r}^k\right) + \left[\Delta W_{pv}^{k-1}\right]^{++} \tag{19}$$

The reserves in the next moment are updated as $R_r^{k+1} = R_r^k + Q_{pv}^k - Q_r^k$.

2.1.3 Models for carbon trading

The core of the carbon trading system involves the buying and selling of carbon emissions or the rights to emit carbon. The intention is that by enabling the trading of these carbon emission rights, it will be an effective mechanism to encourage key industries to reduce carbon emissions. Power market environmental protection management department based on the different key industries of electricity consumption and other factors on their allocation of carbon emission quotas, it is anticipated that by putting into effect the carbon trading mechanism and imposing restrictions on the power - generating capacity of conventional generating units, key industries will be motivated to voluntarily shift towards the utilization of cleaner energy for power generation. This shift can lead to a reduction in carbon emissions, thereby lowering environmental costs, and can even be sold through carbon trading surplus carbon emission allowances to obtain income. Evidently, the carbon trading system can, to a certain degree, prompt major industries to proactively invest in the utilization of clean energy. Moreover, it can encourage them to continuously enhance low - carbon power generation technology. This, in turn, helps major industries to consistently boost their profits.

Carbon emission quotas, that is, The environmental protection division of the power market distributes the permitted carbon emission quotas to major enterprises. This distribution is determined by considering various factors, including the installed capacity of these key enterprises. In this paper, the carbon emission quotas obtained by each key enterprise are calculated as follows:

$$R = \sum_{i \in \Omega} \lambda Q_i \tag{20}$$

where R denotes the carbon emission quota allocated to a key enterprise, Q_i denotes the actual power generation of a unit in the key enterprise in the trading process, λ denotes the carbon emission quota allocation rate of the unit, and Ω denotes the serial number of all the units of the key enterprise that are actually involved in the power generation.

C_c^{nt} is the carbon trading cost of the key enterprise n in year t, i.e.:

$$C_c^{nt} = P_t^c \left(M_{nt} - R_{nt} \right) \tag{21}$$

Among them, P_t^c represents the carbon trading price set for the electricity market in the t th year, while M_{nt} and R_{nt} respectively denote the carbon emission volume of the key enterprise n in the t th year and the allocated carbon emission quota.

The annual carbon emission volume is as follows:

$$M = \sum_{i \in \Omega_n} \delta_i G_{i,c} + \sum_{j \in \Omega_n} \varepsilon_j G_{j,q} \tag{22}$$

where δ_i is the carbon emission intensity factor of thermal power unit i and ε_j is the carbon emission intensity factor of gas unit j.

Through the carbon trading system, it becomes evident that when the actual carbon emissions of a major industry surpass the quota of permitted carbon emissions allotted by the energy sector, the major industry is required to acquire supplementary carbon emission allowances from the market. On the contrary, the main sector has the opportunity to sell the surplus carbon allowances to make a profit.

2.2 Integrated Dispatch Planning Model

2.2.1 Objective function

For an existing power system, considering its load growth, the demand of load growth can be met by reasonable access to distributed power sources without changing the network structure. Stochastic Chance Constrained Planning (SCCP) is a class of stochastic planning whose distinguishing feature is that the stochastic constraints hold at least at a certain confidence level. It is essentially a principle adopted by the decision maker to consider the possibility that the decision made may not satisfy the constraints when unfavorable circumstances occur, i.e., the decision made is allowed to fail to satisfy the constraints to a certain extent, and it is only required that the probability of the decision to make the constraints hold is not less than a certain confidence level.

A stochastic chance constrained planning model is usually represented as:

$$\begin{cases} \min \quad \overline{f} \\ \text{s.t.} \quad p_r \left\{ f(x,\xi) \geq \overline{f} \right\} \geq \beta \\ \qquad p_r \left\{ g_i(x,\xi) \leq 0, j = 1,2,\cdots,v \right\} \geq \alpha \end{cases} \tag{23}$$

Among them, x and ξ are the decision and random vectors respectively, $p_r\{\cdot\}$ represents the probability of an event, α and β are the confidence levels pre-determined by the decision-maker, and $\overline{f}$ is the maximum value of the objective function $f(x,\xi)$ when the confidence level is at least β.

When taking into account the attainment of the dual - carbon goal for the key industries within the large - scale electric power model, this paper establishes the objective function from the perspectives of electricity procurement cost, power loss cost, and operational cost, and environmental benefit from the uncertainty of distributed energy. Namely:

$$\min C = C_{loss} + C_{DG} + C_{DSM} - C_b - C_e \tag{24}$$

where C_{loss} denotes the annual system network loss cost, C_{DG} denotes the distributed power source investment cost and annual operation cost, C_{DSM} is the interruptible load compensation cost, C_b is the purchased power cost, and C_e denotes the environmental benefit.

(1) System network loss cost

This portion of the cost results from the active network loss within the system, namely:

$$C_{loss} = C_{ps} \times \sum_{i=1}^{k} \left(P_{loss\_i} \times \tau_{\max\_i} \right) \tag{25}$$

Here, C_{ps} represents the per - unit selling price of electrical energy, k is the total number of branches in the distribution system, P_{loss_i} is the active network loss power of the i th branch, and $\tau_{\max\_i}$ is the annual maximum load loss hours of the i th branch.

(2) Distributed power investment cost and annual operating cost

$$C_{DG} = \sum_{i=1}^{n_{DG}} \left(\frac{a(1+a)^m}{(1+a)^m - 1} \times r_i \times P_{DG\_i} + W_{DG\_i} \right) \tag{26}$$

Here, n_{DG} represents the number of distributed power sources, a is the discount rate, m is the service life of the distributed power sources, r_i is the cost of installing a unit capacity of distributed power source at node i, P_{DG_i} is the capacity of the distributed power source connected at node i, and W_{DG_i} is the annual operation and maintenance cost of the distributed power source connected at node i.

(3) Interruptible load compensation cost

$$C_{DSM} = \sum_{i=1}^{n_{DSM}} P_{DSM\_i} \times T_{DSM\_i} \times \left(C_{ps} + C_{pi} \right) \tag{27}$$

where n_{DSM} is the number of interruptible load users, P_{DSM_i} and T_{DSM_i} are the fulfillment interruptible load as well as the interruption time of the i th interruptible user, respectively, and the unit compensation expense of interruptible load is represented by C_{pi} .

(4) Savings in power purchase cost

$$C_b = \left(\sum_{i=1}^{n_{DG}} P_{DG\_i} \times T_{DG\_i} + \sum_{i=1}^{n_{DSM}} P_{DSM\_i} \times T_{DSM\_i} \right) \times C_{pb} \tag{28}$$

Here T_{DG_i} represents the yearly operating hours of the i - th decentralized power supply, and C_{pb} denotes the unit grid - connected electricity tariff.

(5) Advantages for the environment

$$C_e = \left(\sum_{i=1}^{n_{DG}} P_{DG\_i} \times T_{DG\_i} + \sum_{i=1}^{n_{DSM}} P_{DSM\_i} \times T_{DSM\_i} \right) \times C_{pe} \tag{29}$$

where C_{pe} is the environmental cost per unit of electricity supplied by a conventional thermal power source.

2.2.2 Constraints

(1) AC current constraints

$$\begin{cases} P_{n,t} = U_{n,t} \sum_{j=1}^{N} U_{j,t} \left[G_{nj} \cos\left(\delta_{n,t} - \delta_{j,t} \right) + B_{nj} \sin\left(\delta_{n,t} - \delta_{j,t} \right) \right] \\ Q_{n,t} = U_{n,t} \sum_{j=1}^{N} U_{j,t} \left[G_{nj} \sin\left(\delta_{n,t} - \delta_{j,t} \right) - B_{nj} \cos\left(\delta_{n,t} - \delta_{j,t} \right) \right] \end{cases} \tag{30}$$

In the formula, $P_{n,t}$ represents the net active power injected by node n at time t, $Q_{n,t}$ represents the net reactive power injected by node n at time t, $U_{n,t}$ represents the voltage amplitude of node n at time t, $U_{j,t}$ represents the voltage amplitude of node j at time t, $\delta_{n,t}$ represents the voltage phase angle of node n at time t, and $\delta_{j,t}$ represents the voltage phase angle of node j at time t. G_{nj} is the real part of the element in the nth row and j th column of the node admittance matrix, B_{nj} is the imaginary part of the element in the nth row and j th column of the node admittance matrix, and N is the total number of distribution network nodes.

(2) Active power balance constraints

$$\begin{aligned} & P_t^{grid} + \sum_{n=1}^{N} P_{n,t}^{wind} + \sum_{n=1}^{N} P_{n,t}^{solar} \\ & = P_t^{loss} + \sum_{n=1}^{N} \left[D_{n,t} + \sum_{i=1}^{I} A_{n,i} \cdot \left(DR_{i,t}^{+} - DR_{i,t}^{-} \right) \right] \end{aligned} \tag{31}$$

where $P_{n,t}^{win}$ is the power value of distributed wind power at node n that receives distribution grid dispatch at moment t, $P_{n,t}^{solar}$ is the power value of distributed photovoltaic at node n that receives distribution grid dispatch at moment t, P_t^{loss} is the system's t-moment active network loss, and $D_{n,t}$ is the baseline load of node n at moment t. $A_{n,i}$ is a matrix characterizing the relationship between user i and node n, and the values within the matrix are 0 and 1. $DR_{i,t}^{+}$ is the up-regulated response of the up-regulated Demand Response (DR) resource of the ith user at the t-moment? and $DR_{i,t}^{-}$ is the down-regulated response of the down-regulated DR resource of the ith user at the t-moment.

(3) Node Voltage Constraints

$$U_n^{\min} \le U_{n,t} \le U_n^{\max} \tag{32}$$

For this constraint, $U_n^{\max}$ refers to the maximum allowable voltage magnitude at node n, and $U_n^{\min}$ corresponds to the minimum allowable voltage magnitude at node n.

(4) Line transmission active power constraints

$$-P_l^{\max} \le P_{l,t} \le P_l^{\max} \tag{33}$$

where $P_{l,t}$ is the active power of line l at the t moment, and $P_l^{\max}$ is the maximum active transmission power of line l.

(5) Node active power upper limit constraint

$$-P_n^{\max} \le P_{n,t} \le P_n^{\max} \tag{34}$$

where $P_n^{\max}$ is the upper limit of active power at node n.

(6)Restrictions on the power of the connection line between the distribution network and the primary network

$$P_t^{grid} \ge 0 \tag{35}$$

(7) Constraints on the upper and lower bounds of distributed wind power and photovoltaic power output

$$\begin{cases} 0 \le P_{i,t}^{wind} \le cap_i^{wind} \cdot output_i^{wind} \\ 0 \le P_{i,t}^{solar} \le cap_i^{solar} \cdot output_i^{solar} \end{cases} \tag{36}$$

where $output_i^{wind}$ is the upper limit of distributed wind power output (standardized value), which is affected by wind speed, and $output_i^{solar}$ is the upper limit of distributed photovoltaic output (standardized value), which is affected by sunshine intensity. cap_i^{wind} is the installed capacity of distributed wind power on node i, and cap_i^{solar} is the installed capacity of distributed photovoltaic on node i.

3 Algorithm design

The proposal of the dual-carbon target not only demonstrates China's determination to actively respond to climate change and strive to realize green and low-carbon development, but also reflects China's image as a great power that adheres to the trend of global development and assumes international responsibility for global climate governance. As a major polluter, key industries need to be comprehensively analyzed to ensure that they can truly achieve the dual-carbon goal.

3.1 Enhanced learning techniques

3.1.1 Definition of Enhanced Learning

Reinforcement Learning (RL), a machine learning approach, aims to either maximize the rewards or minimize the penalties an agent (or an intelligent entity) gets from its surrounding

environment through interaction. It is a learning method that operates on a trial - and - error basis. The agent learns to select actions by considering its conduct within the environment and the rewards or penalties it obtains, all with the objective of achieving the highest long - term rewards.

A Markov Decision Process (MDP) serves as a mathematical construct employed to depict a reinforcement learning issue. It encompasses a collection of states, a group of feasible actions, a state transition probability function, and a reward function.Within an MDP, an intelligent agent influences the state it occupies by carrying out diverse actions and gains a corresponding reward. The objective is to discover a policy that maximizes the overall reward acquired by the agent.The key notions of MDP involve:

(1) State. A state that the system may be in, an abstraction that describes the object of interest in the problem.

(2) Action. Optional actions that the intelligent body can perform in each state.

(3) State transfer probability function. Considering the present state and the action being taken, the function that depicts the probability of state transfer is described which states the intelligent body may transfer to after executing the action, and with what probability for each state.

(4) Reward function. In each state, the reward or cost that the intelligent body obtains.

(5) Strategy. The action taken by the intelligent body in each state can be seen as a mapping relationship.

Denote by $V^{\pi}(s)$ the expected return generated by employing the strategy π in state s. Since this process is related to the state, It is termed the state value function because:

$$V^{\pi}(s) = E_{\pi}\left(R_t \left| s_t = s\right.\right) \tag{37}$$

The equation presented above depicts the aggregate reward value that can be obtained by adopting the strategy π in the state s_t at the moment t, and R_t denotes the reward generated at the moment t as:

$$R_t = r_{t+1} + r_{t+2} + r_{t+3} + r_{t+k} \tag{38}$$

The above equation represents the sum of the rewards that the intelligence receives from the environment in time period k. According to Markovianity, it can be seen that the reward at moment $t+k$ is minimal for the here and now, and the impact decreases exponentially, so a discount factor is introduced to discount the reward at other moments as:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{39}$$

Thus, Eq. (37) can again be transformed into the following form:

$$V^{\pi}(s) = E_{\pi}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \left| s_t = s\right.\right] \tag{40}$$

Eq. (40) continues to be derived and can be obtained:

$$\begin{aligned} V^{\pi}(s) &= E_{\pi}\left(r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots \middle| s_t = s\right) \\ &= E_{\pi}\left(r_{t+1} + \gamma\left(r_{t+2} + \gamma r_{t+3} + \cdots\right) \middle| s_t = s\right) \\ &= E_{\pi}\left(r_{t+1} + \gamma V^{\pi}\left(s_{t+1}\right) \middle| s_t = s\right) \end{aligned} \tag{41}$$

Further Eq. (41) is transformed into the basic form of Bellman's equation, viz:

$$\begin{aligned} V^{\pi}(s) &= \sum_a \pi(a|s) \sum_{s'} p(s'|s,a)\left[R(s'|s,a) + \gamma V^{\pi}(s')\right] \\ &= \sum_a \pi(a|s) Q_{\pi}(s,a) \end{aligned} \tag{42}$$

Denote by $Q^{\pi}(s,a)$ the expected payoff generated by performing action a in state s according to policy π. Since this process is related to both the state and the action, it is called the state-action value function and is defined as:

$$Q^{\pi}(s,a) = E_{\pi}\left[R_t \middle| s_t = s, a_t = a\right] \tag{43}$$

Similar to the transformation in state-valued functions, Eq. (43) can be converted to:

$$Q^{\pi}(s,a) = E_{\pi}\left(\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \middle| s_t = s, a_t = a\right) \tag{44}$$

Similarly, continuing the derivation of Eq. (44) yields:

$$Q^{\pi}(s,a) = E_{\pi}\left(r_{t+1} + \gamma Q^{\pi}\left(s_{t+1}, a_{t+1}\right) \middle| s_t = s, a_t = a\right) \tag{45}$$

Further Eq. (45) is transformed into the basic form of Bellman's equation:

$$\begin{aligned} Q^{\pi}(s,a) &= \sum_{s'} p(s'|s,a)\left[R(s'|s,a) + \gamma \sum_{a'} \pi(a'|s') Q^{\pi}(s',a')\right] \\ &= \sum_{s'} p(s'|s,a)\left[R + \gamma V^{\pi}(s')\right] \end{aligned} \tag{46}$$

In reinforcement learning there exists at least one optimal policy, defined as π^*. The goal of reinforcement learning is to maximize the return value as much as possible, which implies that an optimal policy is required, and then the value function must be maximized as well. Therefore according to equation (42), the optimal state value function $V^*(s)$ is defined as:

$$V^*(s) = \max_a Q^*(s,a) = \max_a \sum_{s'} p(s'|s,a)\left[R + \gamma V^*(s')\right] \tag{47}$$

In state s, for different actions $a \in A$, the calculation is performed according to Eq. (42),the optimal action a is the one that corresponds to the maximum value of the action. (policy).

Similarly, according to equation (46), the optimal state action value function $Q^*(s,a)$ is defined as:

$$Q^*(s,a)=\sum_{s'} p(s'|s,a)\left[R+\gamma \max_{a'} Q^*(s',a')\right] \tag{48}$$

$$\pi^*(s)=\arg\max_{a\in A} Q^*(s,a) \tag{49}$$

From Eq. (48), it is clear that the action corresponding to the maximum value in $Q^*(s,a)$ is the optimal action, and does not require a clear environmental model p.

3.1.2 Actor-Critic Framework

The Actor - Critic framework is a reinforcement learning algorithm employed to tackle decision - making issues within continuous action spaces. It integrates policy gradient and value function estimation techniques. This algorithm comprises two distinct types of neural networks: the Actor network and the Critic network.The Actor network, functioning as a policy network, is responsible for choosing the appropriate action within a continuous action space. It empowers the Agent to pick the optimal action in the current state, thereby maximizing the long - term cumulative reward. To put it in a more relatable way, the Actor network can be likened to an athlete. Just as an athlete constantly refines their performance during a competition to secure a higher score from the referee, the Actor network consistently improves its action - selection process.On the other hand, the Critic network is a value network. It permits single - step updates and is used to evaluate actions. Similar to a referee in a sports event, its main function is to estimate the expected value of the reward that can be obtained by performing an action in the current state. This estimated value then serves as a guide for updating the Actor network.Value function estimation methods play a crucial role. They enable intelligent agents to gain a better understanding of the environmental dynamics and the probability of future rewards. This enhanced understanding ultimately leads to more informed and effective decision - making.

Actor observes the current state s, controls Agent to make action a, Critic generates q value based on s and a, and outputs the value to Actor.Actor calculates approximate policy gradient through s, q and a, and updates the parameters by policy ascent, and this update will get higher and higher $\bar{q}$. However, higher q does not mean more accurate future prediction, because the value network is randomly initialized in the initial state, and Critic's accuracy is improved by the reward r of environmental feedback.

A Markov decision process is used to create a recursive expression for the state value function V and the action value function Q. This recursive expression is known as the Bellman equation. The Bellman equation assesses the worth of each state or action, thereby helping the intelligence to make more optimal decisions. Subsequently:

$$Q^{\pi_\theta}(s,a)=r_{ss'}^{a}+\gamma\sum_{s'} P_{ss'}V^{\pi_\theta}(s')=E\left[r+\gamma V^{\pi_\theta}(s')\right] \tag{50}$$

$$V^{\pi_\theta}(s)=\sum_{a\in A}\pi_\theta(a|s)Q^{\pi_\theta}(s,a)=E\left[Q^{\pi_\theta}(s,a)\right] \tag{51}$$

where $r_{ss'}^{a}$ denotes the immediate payoff for state s to take action a to transfer to new state s' and $P_{ss'}$ denotes the state transfer probability.Based on equation (50), the formula for updating the strategy gradient can be presented as:

$$\nabla_\theta J(\theta)=E_\theta\left[\left(r_t+\gamma V^{\pi_\theta}(s_{t+1})-V^{\pi_\theta}(s_t)\right)\nabla_\theta \log \pi_\theta(s_t,a_t)\right] \tag{52}$$

Typically, the value functions of strategies and actions remain unknown. To address this, we can employ two distinct neural networks to estimate these two functions respectively. Specifically, we can utilize the Actor - Critic approach. This approach integrates value learning and strategy learning, enabling the simultaneous learning of the two networks. Subsequently, the parameters of these networks are updated via gradient descent, which is denoted as:

$$\theta_{t+1} \to \theta_t - \alpha \nabla_\theta \log \pi_\theta \left(s_t, a_t \right) V_t \tag{53}$$

where α denotes the learning rate.

3.2 Dynamic Optimized Scheduling Algorithm

3.2.1 Dynamic Optimized Scheduling Strategy

Within this research paper, the Actor - Critic model of reinforcement learning is employed as a support to construct a dynamic optimal scheduling algorithm for energy in key industries, which approximates the strategy through the Actor network and estimates the dominance function through the Critic network, and then evaluates and improves the current scheduling strategy.

(1) Actor network construction

In terms of Actor network construction, the importance sampling technique is adopted, This objective is achieved by employing two Actor networks that have the same architecture but distinct parameters. Specifically, the sampling Actor network is employed to engage in continuous interaction with the environment to acquire sampling data. Subsequently, this sampling data is repurposed to train and update the policy Actor network. The sampling Actor network shares an identical structure with the strategy Actor network. At regular intervals, the parameters of the sampling Actor network are updated by applying weighted values from the parameters of the strategy Actor network in the following manner:

$$\theta' \leftarrow \alpha\theta + \left(1-\alpha\right)\theta' \tag{54}$$

where α is the weighting coefficient.

The truncated function method is used to replace the KL scattering constraint, to truncate the objective function, and to achieve the purpose of slowing down the fluctuation of the strategy and making the direction of the strategy update show a stable improvement by selecting the lesser value between the truncated objective function and the non - truncated objective function. Namely:

$$\begin{aligned} L\left(\pi'_\theta\right) = E_{\pi_\theta}\Bigg[& \min\left(\frac{\pi'_\theta\left(a_t \middle| s_t\right)}{\pi_\theta\left(a_t \middle| s_t\right)} A_{\pi_\theta}\left(s_t, a_t\right) \right. \\ & \left. clip\left(\frac{\pi'_\theta\left(a_t \middle| s_t\right)}{\pi_\theta\left(a_t \middle| s_t\right)}, 1-\varepsilon, 1+\varepsilon \right) A_{\pi_\theta}\left(s_t, a_t\right) \right) \end{aligned} \tag{55}$$

where ε denotes the truncation factor.

After adding constraints for the policy update, the policy Actor network uses Adam's stochastic gradient ascent algorithm to maximize the objective function to update the policy, i.e.:

$$\theta_{k+1}=\arg\max_{\theta}\frac{1}{|D_k|T}\sum_{\tau\in D_k}\sum_{t=0}^{T}\begin{matrix}\min\left(\dfrac{\pi'_\theta(a_t|s_t)}{\pi_\theta(a_t|s_t)}A_{\pi_\theta}(s_t,a_t)\right.\\ \left.clip\left(\dfrac{\pi'_\theta(a_t|s_t)}{\pi_\theta(a_t|s_t)},1-\varepsilon,1+\varepsilon\right)A_{\pi_\theta}(s_t,a_t)\right)\end{matrix} \tag{56}$$

where D_k denotes the set of preserved trajectories, defined as $D_k=\{\tau_i\}$.

(2) Critic network construction

The aim of the Critic network is to precisely assess the state - action value function. To learn this value function, the gradient descent algorithm can be employed to minimize the mean squared error, and its neural network parameter update process is as follows:

$$\phi_{k+1}=\arg\min_{\phi}\frac{1}{|D_k|T}\sum_{\tau\in D_k}\sum_{t=0}^{T}\left(V_\phi(s_t)-\hat{R}_t\right)^2 \tag{57}$$

(3)Interaction Process between the Actor Network and the Critic Network

Figure 1 depicts the interaction procedure between the Actor neural network and the Critic neural network, the inputs of both networks are the system state values, i.e., $\{P_{load,t},h_{load,t},P_{PV,t},c_{soc,t-1}\}$.Actor network The state information is extracted from the features, and the mean value μ_θ and log standard deviation $\log(\sigma_\theta)$ are output from the neural network to form a normal distribution, which is sampled to obtain the policy π_θ, which generates the action a_t for the dynamic optimal scheduling of the IES. Critic network: the output state value $V_\phi(s)$ is used to estimate the dominance function A_π to evaluate and improve the current Actor network strategy π_θ.

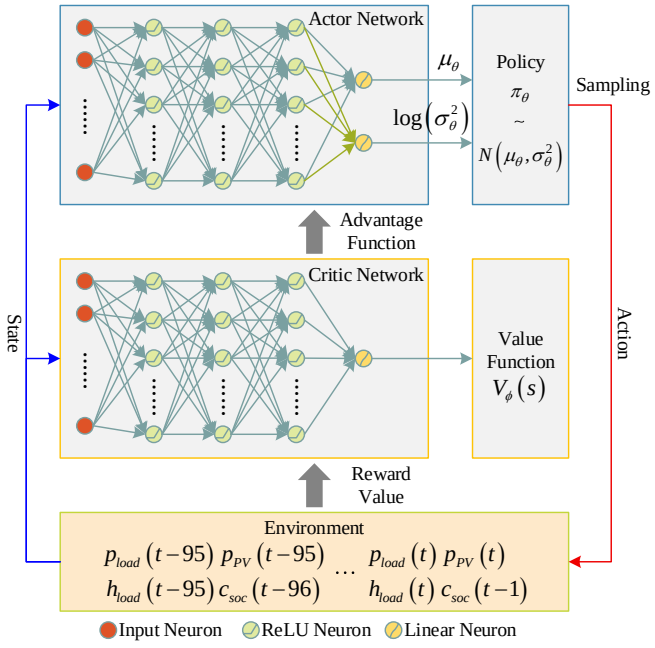


Figure 1: Interaction process between Actor network and Critic network

3.2.2 Model Parameter Training Tuning

The Deep Q - network (DQN) algorithm is a Q - learning algorithm founded on neural networks within the realm of deep reinforcement learning. Its purpose is to address reinforcement learning challenges that involve high - dimensional state and action spaces. The training procedure of the DQN algorithm can primarily be broken down into the subsequent two steps:

(1) Policy evaluation. A neural network is used to estimate the maximum cumulative payoff (i.e., Q-value) that can be obtained by taking each action in each state. Specifically, for each state s and action a, the neural network outputs the corresponding Q value $Q(s,a)$, i.e.:

$$Q(s,a) = f\_\theta(s,a) \tag{58}$$

where $f\_\theta(s,a)$ is the output of the neural network and θ is the parameter of the neural network.

(2) Strategy improvement. The ε-greedy $(\varepsilon - greedy)$ strategy is used to select actions. That is, at each selection of an action, an action is randomly selected with a probability of ε and the action with the largest current Q value is selected with a probability of $1-\varepsilon$. Where ε is a positive number less than 1, it is employed to regulate the equilibrium between exploration and exploitation.

In the process of strategy improvement,The parameters of the neural network are adjusted by minimizing the mean squared error of the -value function. In particular, a stochastic gradient descent (SGD) algorithm is employed to reduce the mean squared error between the target value (that is, the TD objective) and the output of the neural network, namely:

$$L(\theta) = E\left[\left(r + \gamma \max Q'(s',a') - Q(s,a)\right)^2\right] \tag{59}$$

where θ is the parameter of the neural network, r is the reward in the current state, γ is the discount factor, s' is the next state, a' is the optimal action in the next state, and Q' is the output of the target network, i.e., a fixed-parameter neural network is used to compute the Q value.

4 Numerical simulations

Against the backdrop of the dual-carbon goal, establishing a novel power system centered around new energy is not just a significant path for the transformation and improvement of the power system; it is also a crucial means to achieve the dual-carbon objective. For the electric power grand model, relying on different types of dynamic scheduling strategies to give the key industries more diversified power resources, so as to enhance the effect of the key industries to achieve the dual-carbon goals.

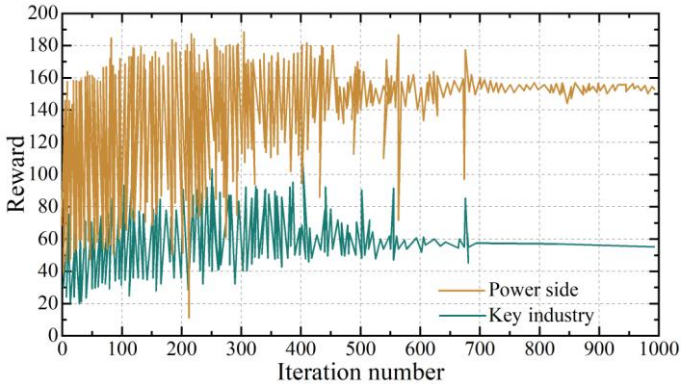
4.1 Algorithm performance and optimization comparison results

4.1.1 Dynamic optimization algorithm performance

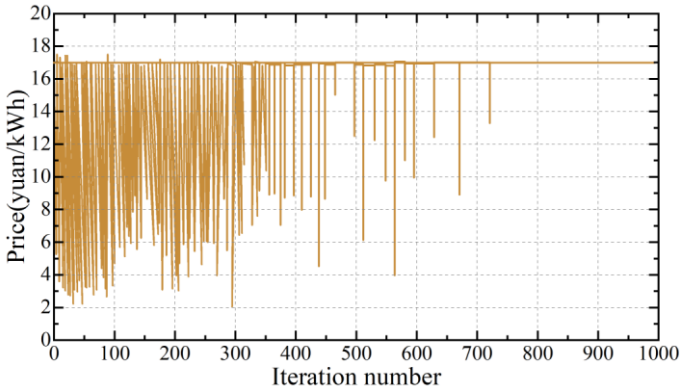
To confirm the convergence of the dynamic optimal scheduling algorithm, we assess whether it is continuously stable within the preset thresholds by observing the reward changes in the power supply side and the key enterprises during the optimization of the objective function.

Specifically, Experiments on reward iteration and price iteration were conducted in the power supply sector and major enterprises. The simulation outcomes are presented in Figure 2. Specifically, Figures 2(a) and 2(b) respectively illustrate the convergence of rewards and the convergence of electricity prices.

In the integrated grid dynamic dispatch strategy of key enterprises under the power grand model, the dual carbon and tariff decisions of the power supply side and key industries are not independent, but interact with each other. Specifically, the demand behavior of the focus influences the decision of the supply side, and the decision of the supply side (pricing, generation strategy) influences the behavior of the focus industry. The whole interaction process is then analyzed in conjunction with Fig. 2(a) and (b). In the initial stage, both intelligences are in the exploratory phase, and the lack of sufficient information to determine which tariff will bring greater rewards leads to large fluctuations in rewards and tariffs. However, with about 600+ iterations of exploratory learning and adapting under changing load conditions, the intelligences gradually stabilized their rewards and tariffs. The iterative curves of the electricity price are largely consistent with the fluctuations of the rewards of the focus industry and the supply side, which further reflects the real-time interaction interactions between the focus industry and the supply side. This result not only validates the convergence of the algorithm, but also reveals the importance of synergy between intelligences in a dynamic environment. Through continuous learning and adaptation, the intelligences are able to find the optimal scheduling strategy in uncertain environments, thus improving the overall performance and stability of the system.



(a) Reward convergence



(b) Electricity price convergence

Figure 2: Dynamic optimization of algorithm performance

4.1.2 Optimization results of different algorithms

To confirm the efficacy of the dynamic scheduling optimization algorithm presented in this paper, which is designed for the dynamic scheduling optimization of the dual - carbon target in key industries, DQN algorithm, Counterfactual Multi-Intelligent Agents (COMA) Deep Reinforcement Learning Algorithm, DDPG scheduling method, and Model Predictive Control (MPC)-based scheduling method are selected for comparison. Among them, the Actor of each intelligent body of COMA is a recurrent neural network featuring 2 hidden layers, each housing 64 neurons, governs the system. The connections between these layers are established via a fully - connected neural network. The centralized Critic network in COMA, the neural network of DQN, and the neural network of the agents in DDPG share an identical configuration. They consist of 3 hidden layers, with each layer having 180 neurons. The activation function for these hidden layers is the Rectified Linear Unit (ReLU). As for the model predictive control, a fully - connected neural network having 2 hidden layers serves as the predictive model component.

To prevent the imprecise optimization outcomes that stem from the unpredictability of the training results, this paper randomly selects the data of three key enterprises in an industrial park in July 2025 as inputs, and Table 1 gives the statistical data of the average daily running costs after optimization by different methods, which is intended to show that the difference in optimization results the root cause lies in the techniques employed rather than chance occurrences.

Specifically, the optimization outcomes derived from the COMA algorithm and the DDPG - based algorithm are comparable to the results of the methods put forward in this paper. When compared to the methods presented in this paper, the average daily operating cost shows an increase of 3.12% and 3.65% respectively for these two algorithms. In contrast, the average daily running cost of the DQN - based algorithm and the MPC - based algorithm rises by 6.33% and 8.55% respectively relative to the method proposed in this paper.By analyzing the principles of various algorithms, it becomes evident that traditional optimal scheduling methods are more significantly influenced by the forecasting accuracy of renewable energy unit output and load. In the DQN approach, the demand response volume and the output of energy storage are restricted to predefined discrete values. This limitation results in the selection of actions that cannot encompass the entire action space, and the actions that are selected are highly likely to be sub - optimal.The DDPG algorithm utilizes a single intelligent agent to schedule all the key enterprises simultaneously. As a result, the sets of states and actions are quite large, which leads to the selection of sub-optimal actions by the intelligent body, and it requires a large amount of communication.The COMA algorithm assumes that the strategies of other intelligences are unchanged by a certain intelligence during counterfactual estimation in the policy update process, thus leading to non-optimal actions being selected. Evidently, the scheduling approach presented in this paper is capable of more readily identifying the optimal action within the action space compared to the other two algorithms. Moreover, it is better suited for addressing the issue of integrated optimization and coordinated scheduling of power supply under the dual - carbon goal in key industries.

Table 1: Comparison of optimization results with different methods

Method	Average daily operating cost/yuan			
	Enterprise A	Enterprise B	Enterprise C	Total
Ours	538.41	732.85	1021.76	2293.02
COMA	565.38	747.59	1051.48	2364.45
DDPG	563.96	756.17	1056.69	2376.82
DQN	574.72	784.98	1078.52	2438.22
MPC	598.65	795.46	1094.87	2488.98

4.2 Emission Reduction Effect of Electricity Dispatch in Key Industries

4.2.1 Expenses related to carbon trading across various scenarios

To conduct a more in - depth comparison of the rationality and efficacy of the proposed model, the dynamic optimal scheduling scenario (Scenario S1) put forward in this paper is juxtaposed with the subsequent three scenarios. (1) Considering the dynamic optimal scheduling strategy under load determination (Scenario S2). (2) Short-sighted optimization scenario (Scenario S3) considering algorithms based on Scenario S2. (3) Social welfare under time-sharing pricing mechanism (Scenario S4) based on four scenarios. To better demonstrate the reasonableness of the proposed dynamic optimal scheduling model when taking load uncertainty into account, it is hypothesized that the four scenarios share the same basic parameters. In a typical day, Table 2 presents the values of the model metrics across four distinct scenarios.

The simulation results presented in the table indicate that the social welfare values of the dynamic optimization scenarios proposed in this paper are similar to those of the other three scenarios, and at the same time, the social welfare values under dynamic optimization are always better than those under time-of-day pricing. Although the real-time pricing under uncertainty reduces the welfare value compared with the deterministic case, the real-time pricing under uncertainty can more closely match the actual electricity consumption of the key industries, i.e., the proposed dynamic optimization strategy achieves a better social welfare value under the guarantee of the model robustness, and the comparison of different scenarios serves to confirm the effectiveness and plausibility of the suggested model.

Table 2: Model index values in four scenarios

Scene	Social welfare value	Priority industry benefits	Power supply benefit value	Carbon transaction cost/yuan
S1	2306.72	5036.72	1237.45	-268.25
S2	2348.59	5114.58	1368.71	-273.57
S3	2023.68	4683.61	1118.86	-151.38
S4	2159.47	4827.85	1205.64	-115.86

4.2.2 The role of carbon mitigation mechanisms

The carbon trading mechanism is set to exert a more substantial influence on the power supply structure of the crucial industries, that is to say, the direct effect is on thermal power units with higher carbon emissions, in addition to the green certificate trading mechanism, which may also be on the new energy units in the key industries. On this basis, this research paper computes the emission - reduction advantages of major industries by integrating the carbon trading mechanism and the green certificate trading mechanism. In the comprehensive scenario, the trading of China Certified Emission Reductions (CCER) is not taken into account. Instead, only the carbon quota trading cost and the green certificate trading cost are considered in the dynamic optimization model of major industries.An industrial park is selected as the research subject. The trend of the annual total power generation of each type of units in the area under the combined effect of the carbon trading mechanism and the green certificate trading mechanism is analyzed, as presented in Figure 3. Additionally, the carbon emissions in the comprehensive scenario are shown in Figure 4.

As can be seen from the figure, with the strengthening of the policy of carbon trading mechanism and green certificate trading mechanism, the proportion of power generation from coal-fired units gradually decreases, during the final years of the planning phase, there has been nearly no fresh power generation from coal - fired power plants. As the installed capacity of

renewable energy facilities grows, the electricity output from wind and solar photovoltaic power units has increased. Simultaneously, under the combined influence of two mechanisms, the carbon emission reduction benefits in the area have become more pronounced. In the latter part of the planning period, because of a substantial decline in the power - generating capacity of coal - fired units, the change in carbon emissions has been relatively stable,or even a decline in the phenomenon. The total carbon emissions in the planning period decreased by about 10.27% or so compared to the time when there was no carbon emission reduction policy.

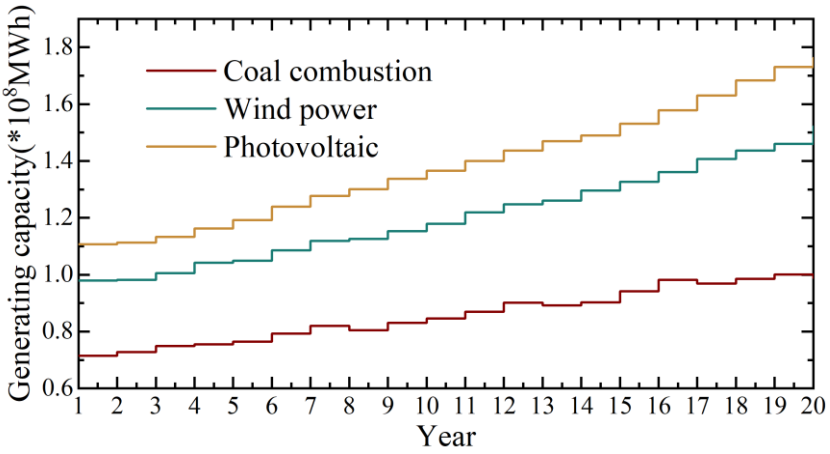


Figure 3: The power generation capacity of the omitted type of units

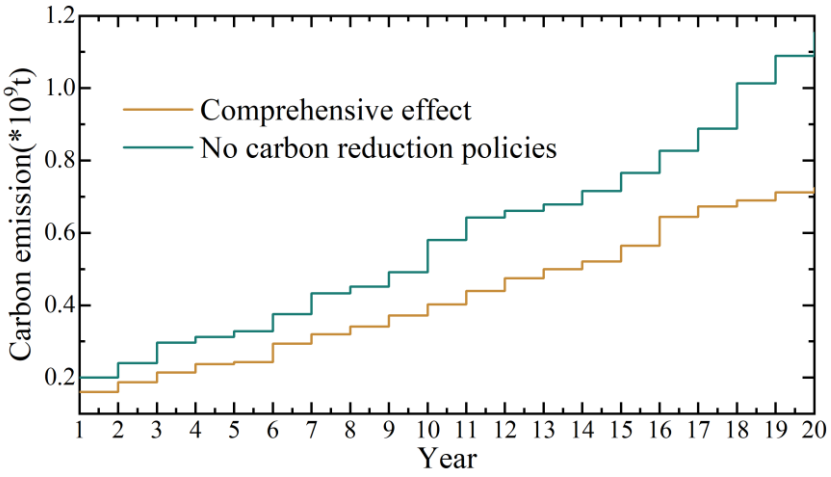


Figure 4: Carbon emissions under comprehensive circumstances

4.2.3 Impact of carbon trading prices on power planning

The carbon trading price is set into three levels, high, normal and low, to conduct an analysis of how sensitive the expansion of key enterprise units is to the influence of varying carbon trading prices. There is also a big difference in the carbon cost of key enterprises at different carbon trading prices. Taking an industrial park as an example, the carbon trading benefits at different carbon trading prices are shown in Figure 5.

In the initial planning period, the key enterprises have higher carbon benefits when carbon trading prices are at a high level. With the passage of time, the market demand for electricity gradually increased, the carbon emissions of key enterprises also increased, more than the market allocation of carbon emissions, the need to purchase from the market beyond the part of the carbon emission rights, the same key enterprises at the elevated price of carbon trading, the

carbon gain earlier than the other two cases to become a negative gain. Currently, when carbon trading prices are high, efforts are being made to lower carbon - related expenses, key enterprises began to invest in new energy units, with the hydropower, wind power and other green energy units put into use, the carbon costs of key enterprises from the twelfth year to gradually shrink. In the other two carbon trading price cases, due to the smaller loss of carbon cost, the key enterprises start to realize the importance of investing in new energy in the 12th year, which can be seen that the new energy can be invested and operated earlier when the price of carbon trading rises, key enterprises are more inclined to make early investments in new - energy units. This is because a higher carbon trading price prompts them to take such actions. By investing in new - energy units, these enterprises can effectively cut down on carbon emissions. As a result, they can further lower their development costs. This situation is highly beneficial for key enterprises in their pursuit of achieving the dual - carbon goal.

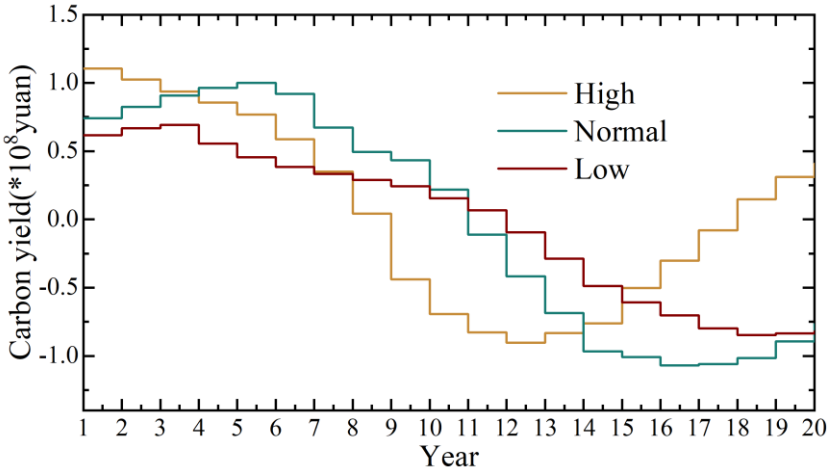


Figure 5: Carbon trading gains at different carbon trading prices

5 Conclusion

According to the power provider model, this paper considers a dynamic optimal dispatch model with dual-carbon objectives for key industries under grid output uncertainty, which takes grid losses and environmental costs as objective functions and sets multiple types of constraints.To address the model, this study presents a dynamic optimal scheduling algorithm founded on the Actor - Critic framework. This algorithm is integrated with the DQN algorithm for the purpose of tuning and training model parameters. An industrial park is selected as the subject of research, and numerical simulation analysis is conducted.The findings indicate that the algorithm put forward in this paper can achieve the optimal solution for the average daily operating cost of major enterprises. Moreover, in various scenarios, the cost - benefit of carbon trading is more favorable. Under the influence of multiple carbon emission reduction mechanisms, the emission reduction effect of major industries can be notably improved. Additionally, the carbon trading price will also exert a certain influence on the carbon emission strategy of major industries. Therefore, fully exploring the electricity consumption patterns and trends of key industries in the power model can help key enterprises better implement emission reduction strategies and promote key enterprises to realize the dual-carbon goal.

About the Author

Zeqi Zhang, male, 25, is a Junior Engineer at Shenzhen Power Supply Bureau Co., Ltd. He is currently a postgraduate student at Sun Yat-sen University, and his primary research focuses on electricity-carbon coupling integrated with artificial intelligence.

Qingge Ji, born in 1966, PhD, associate professor, is a senior member of China Computer Federation. His main research interests include computer vision, computer graphics and virtual reality.

Enhong Wu, graduated with a bachelor's degree from Wuhan University of Technology in 2020 and obtained a master's degree in Pedestrian Trajectory Prediction from Sun Yat-sen University in 2025. He currently serves as an R&D Engineer at Pony.ai, where he is responsible for the deployment and optimization of computing systems. He possesses independent research and technology implementation capabilities in the field of artificial intelligence.

References

[1] Lin, B., & Li, Z. (2022). Towards world's low carbon development: The role of clean energy. Applied Energy, 307, 118160.

[2] Tian, J., Yu, L., Xue, R., Zhuang, S., & Shan, Y. (2022). Global low-carbon energy transition in the post-COVID-19 era. Applied energy, 307, 118205.

[3] Sengupta, P., Choudhury, B. K., Mitra, S., & Agrawal, K. M. (2020). Low carbon economy for sustainable development. Encyclopedia of renewable and sustainable materials, 3, 551-560.

[4] Zhou, S., Tong, Q., Pan, X., Cao, M., Wang, H., Gao, J., & Ou, X. (2021). Research on low-carbon energy transformation of China necessary to achieve the Paris agreement goals: A global perspective. Energy Economics, 95, 105137.

[5] Yang, W., Zhao, R., Chuai, X., Xiao, L., Cao, L., Zhang, Z., ... & Yao, L. (2019). China's pathway to a low carbon economy. Carbon balance and management, 14(1), 14.

[6] Ahmed, F., Al Kez, D., McLoone, S., Best, R. J., Cameron, C., & Foley, A. (2023). Dynamic grid stability in low carbon power systems with minimum inertia. Renewable Energy, 210, 486-506.

[7] Abdilahi, A. M., Mustafa, M. W., Abujarad, S. Y., & Mustapha, M. (2018). Harnessing flexibility potential of flexible carbon capture power plants for future low carbon power systems. Renewable and Sustainable Energy Reviews, 81, 3101-3110.

[8] Rajaperumal, T. A., & Columbus, C. C. (2025). Transforming the electrical grid: the role of AI in advancing smart, sustainable, and secure energy systems. Energy Informatics, 8(1), 51.

[9] Padmanaban, S., Nasab, M. A., Samavat, T., Zand, M., & Nasab, M. A. (2023). Artificial intelligence techniques for smart power systems. In IoT and Analytics in Renewable Energy Systems (Volume 1) (pp. 107-123). CRC Press.

[10] Lin, H. Y. (2022). Large-scale artificial intelligence models. Computer, 55(05), 76-80.

[11] Edwards, R. E., New, J., Parker, L. E., Cui, B., & Dong, J. (2017). Constructing large scale surrogate models from big data and artificial intelligence. Applied energy, 202, 685-699.

[12] Zhang, J., Hua, X. S., Huang, J., Shen, X., Chen, J., Zhou, Q., ... & Zhao, Y. (2019). City brain: practice of large-scale artificial intelligence in the real world. IET Smart Cities, 1(1), 28-37.

[13] Dorri, A., Kanhere, S. S., & Jurdak, R. (2018). Multi-agent systems: A survey. Ieee Access, 6, 28573-28593.

[14] Asaad, R. R., Saeed, V. A., & Abdulhakim, R. M. (2021). Smart Agent and it's effect on Artificial Intelligence: A Review Study. Icontech International Journal, 5(4), 1-9.

[15] Bai, X., Huang, S., Wei, C., & Wang, R. (2025). Collaboration between intelligent agents and large language models: A novel approach for enhancing code generation capability. Expert Systems with Applications, 269, 126357.