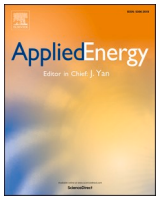
Applied Energy 400 (2025) 126590

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



An agent-based approach to modeling power firms' emission reduction strategies and market dynamics

Songyuan Liu [a], Peng Zhou [a,*], Mei Wang [b,*], Aobo Xu [a]

[a] School of Economics and Management, China University of Petroleum, No. 66 Changjiang West Road, Qingdao 266580, China
[b] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, No. 29 Jiangjun Avenue, Nanjing 211106, China

HIGHLIGHTS

- A comprehensive ABM for China's power-ETS system, capturing multi-market interactions and firm heterogeneity, is proposed.
- Advanced simulation using multi-agent deep reinforcement learning for optimized decision-making is employed.
- The model utilizes realistic data from over 3000 Chinese power firms, ensuring high evaluation accuracy.

ARTICLE INFO

Keywords:
Agent-based model
Multi-agent reinforcement learning
Heterogeneity
Power market
Emissions trading

ABSTRACT

This paper develops an agent-based model to study the impact of China's national carbon market on the decision-making behaviors of power generation firms by accounting for their multidimensional heterogeneity. The multi-agent deep deterministic policy gradient algorithm is employed to optimize firm strategies in production, low-carbon technology adoption, and allowance trading. An application study is conducted by using the real data from over 3000 Chinese power firms and the real price data from the China national carbon market. The modeling results show that tightening emission reduction targets leads to higher carbon and power prices, greater renewable energy generation, and increased adoption of low-carbon technologies. Additionally, our study highlights the critical role of allowance allocation methods, with auction-based rule providing more stable and higher carbon price signals that incentivize earlier emission reductions. The research also identifies a disparity between large-scale and small-medium-scale firms in terms of participation in allowance trading and low-carbon technology adoption, with larger firms leading in both areas. The findings offer valuable insights for enhancing the cost-effectiveness and incentive mechanisms of carbon market.

1. Introduction

The emission trading system (ETS) has emerged as a crucial policy tool for effectively managing carbon emissions and facilitating the widespread adoption of low-carbon technologies [1–3]. Under the ETS, the government first sets the carbon emission cap for a determined compliance period and allocate the emission allowance to firms according to certain rules [4]. Meanwhile, firms can flexibly adopt three types of emission reduction measures, including allowance trading strategies, production decision or technological innovation [5]. At the end of the compliance period, the government penalizes firms that exceed the emission by reducing their emission allowances for the following year and requiring them to pay fines.

Currently, 36 emissions trading systems are in force worldwide [6]. China has initiated regional ETS pilots in eight provinces since 2013 and officially launched the national carbon market in July 2021. Numerous studies have demonstrated that carbon markets play a crucial role in helping China achieve its emission reduction targets and promoting low-carbon technology innovation [7–9]. For instance, [10,11] have shown that the implementation of ETS significantly reduces carbon intensity per unit of GDP and fosters the development and application of low-carbon technologies. Nevertheless, the national carbon market in China has exhibited phenomena such as loose allowance allocation, sluggish trading, and price volatility during its actual operation [12]. These issues may undermine the cost-effectiveness of the carbon market. Specifically, the loose allocation of allowances leads to market supply-demand imbalances, insufficient trading activity, and frequent price fluctuations, thereby weakening the incentive effect of the carbon market on corporate emission reduction behaviors [13–15].

* Corresponding authors.
E-mail addresses: pzhou@upc.edu.cn (P. Zhou), wangmei1989@nuaa.edu.cn (M. Wang).

https://doi.org/10.1016/j.apenergy.2025.126590
Received 10 October 2024; Received in revised form 18 April 2025; Accepted 4 August 2025
Available online 9 August 2025

Nomenclature

Abbreviations

ABM	Agent-based model
ETS	Emission trading system
MARL	Multi-Agent Reinforcement Learning
MADDPG	Multi-agent deep deterministic policy gradient
CTDE	Centralized Training and Decentralized Execution
A3C	Advantage Actor-Critic
TD	Temporal-Difference

Indices

i	Index of the firm agent
T	Index of the compliance period
t	Index of the time scale of day
m	Index of the time scale of month
y	Index of the time scale of year
k	Index of the time scale of second
j	Index of the low-carbon technology
g	Index of the generation type

Variables

$q_{i,t}^{DA}$	Trading volume in the day-ahead market
$p_{i,t}^{DA}$	Trading prices in the day-ahead market
$q_{i,t}^{ID}$	Trading volume in the day-ahead market
$p_{i,t}^{ID}$	Trading prices in the intraday market
$q_{i,m}^{M}$	The monthly contract volume in power market
$q_{i,y}^{Y}$	The annual contract volume in power market
$x_{i,t,k}$	The trading volume of agent i in the carbon market
$p_{t,k}^{A}$,$p_{i,t,k}^{A.bid/ask}$	The trading price of agent i in the carbon market
$z_{i,j,t}$	The investment decisions of agent i on low-carbon technologies

The effectiveness of emission reductions in carbon markets is heavily dependent upon the design of carbon trading mechanisms and the expectations and behaviors of firms. Previous studies on the evaluation of ETS mechanism usually employ top-down models, such as Computable General Equilibrium models (CGE) and System Dynamics models (SD) [16,17]. However, these top-down models have certain limitations. First, they overlook the heterogeneity characteristics of firms, and are confined to the constraints of strong assumptions [18,19]. Second, these models are based on the assumption of complete information, which deviates from the behavioral dynamics in realistic market environment. In addition, these models also neglect the dynamic interaction process among firms when simulating the equilibrium solution [20].

Agent-based model (ABM) offers a powerful alternative by explicitly representing the micro-level behavior of heterogeneous firms, including their market trading, auction bidding, and low-carbon technology adoption decisions [21,22]. Besides, ABM has advantages in dealing with issues such as interactive effects, heterogeneity, bounded rationality and learning process [23,24]. Therefore, some studies have used the ABM models to investigate the impact of ETS on the behavior of firms within different industries, including steel, petrochemical, power and agriculture [25,26].

In the ABM model, agents are the individual entities involved in bottom-level and the core components of the complex adaptive system. Each agent can communicate and interact with other agents, and make decisions or actions according to their goals, rules and environment [11]. As to power firms covered in carbon market, we consider agents to possess the following features [27]:

(1) Agent Interaction: Agents can interact with other agents or with the environment through information exchange, resource sharing, cooperation, and competition [24]. From a micro perspective, interaction can be regarded as a game process, where each agent makes the optimal choice according to their own utility function and behavior strategy [10]; from a macro perspective, interaction can be regarded as a collective behavior, where multiple agents achieve a certain global goal through coordination or competition [28]. For example, [29] designed the production decision rules of thermal power firms from the perspective of uncertainty of production decision, and defined the interaction patten and feedback relationship in the ABM model.

(2) Heterogeneity: In the modeling and analysis of the carbon market utilizing ABM, the heterogeneity of agents embodies the diversity, complexity, and dynamism inherent in the market [30]. The heterogeneity of agents includes differences in firms' type [31], technical parameters [32], information acquisition [33], firms' goals [34], risk preference [35], bidding strategy [10], and behavior norms [36], which affect the decision and interaction of agents. For example, [37] explained the decision behavior of firms under three non-equilibrium frameworks by simulating the regulated firms as heterogeneous agents with different allowance requirements, emission reduction costs and technology preferences.

(3) Bounded Rationality: Many studies have considered the bounded rationality of agents in applying ABM models [18]. Theoretically, bounded rationality reflects the limitations and uncertainties of agents when facing complex problems [38], which could affect the carbon price and trading volume of the carbon market [35]. Depending on the research problem, assumption, and parameter setting, scholars have characterized the bounded rationality behavior of agents by designing the learning rate and inertia coefficient of multi-agents by using gradient algorithm or genetic algorithm [24,39]. For example, [24] analyzed how bounded rational firms coordinate three emission reduction decisions (output adjustment, low-carbon technology adoption and allowance trading) by following the set of "fast and frugal heuristics".

(4) Learning and Adaptation: The learning and adaptation mechanism reflects the dynamic characteristics and evolution process of agents [40]. In the power market, each firm agent can adaptively adjust their own production strategy according to various environmental factors, such as market demand, long-term and short-term price trends, and the behavior of competitors [41]. In the learning process, agents acquire knowledge and adapt to market changes by observing the market information such as price, demand, supply, etc. [24]; they can also learn and adapt to the influence of other agents by observing the behavior or strategy of other agents [36]. Moreover, agents can update learning-based strategies by using different methods, such as Q-Learning algorithm, which was employed by [41] to update the interaction and adaptation process of agents with real market data.

Recent research has also explored the synergistic potential of combining ABM with Multi-Agent Reinforcement Learning (MARL) in energy markets. Studies have employed improved MARL algorithms, such as MADDPG, to address multi-agent energy management optimization in complex market environments [39] and to investigate coordinated bidding strategies in multi-agent power systems [42]. These studies demonstrate the growing interest in combining ABM and MARL to capture the complex interactions and learning behaviors of agents in dynamic energy environments.

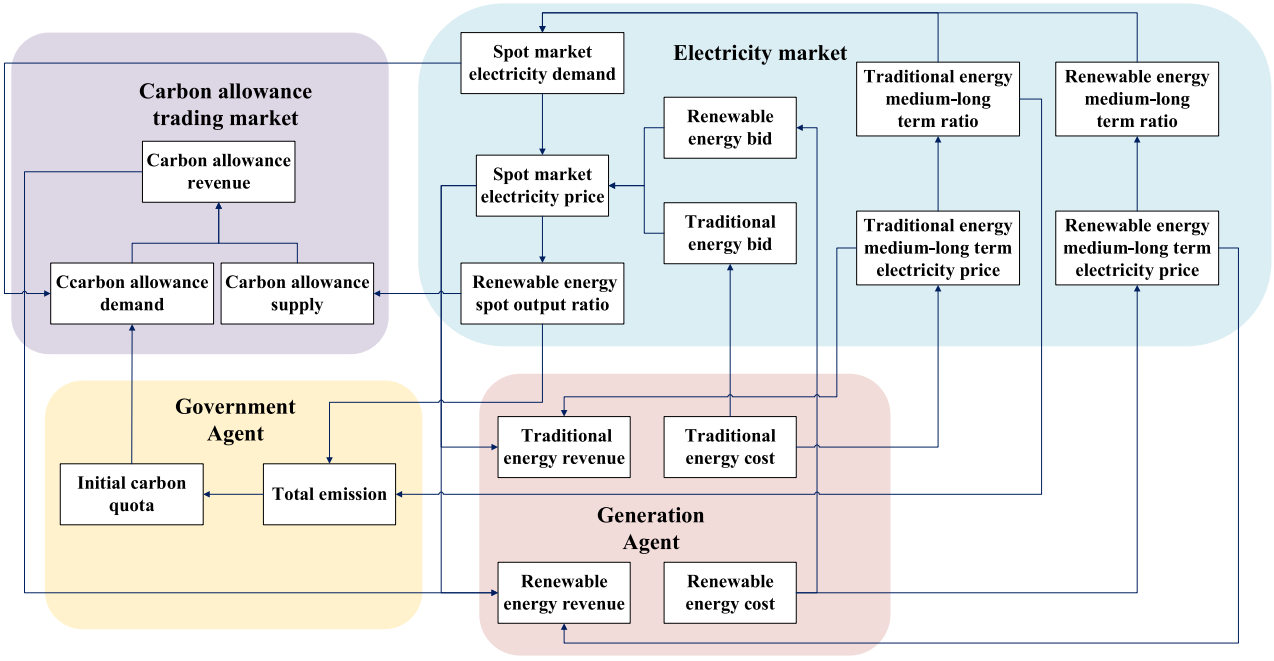


Fig. 1. The framework of the ABM. Note. Generation agents are heterogeneous power firms participating in electricity and carbon markets, differentiated by scale, cost structures, and technology attributes.

However, current models often fall short in realistically capturing these dynamics [43]. Much existing research on carbon markets focuses on the impact of policy changes on power firms within simplified, single, or static market environments, frequently oversimplifying firm behaviors and market dynamics. This simplification, coupled with reliance on small datasets and less sophisticated modeling techniques, limits the accuracy of policy evaluations. Critically, previous work often simplifies the complex and diverse decision-making processes of firms, neglects the heterogeneity of firm types, and overlooks the dynamics of real market price fluctuations. Furthermore, many studies also utilize a single temporal dimension, failing to adequately represent the interactive relationships between the allowance market and the power market over longer time horizons. This can significantly affect the accuracy of policy impact assessments.

This paper addresses these limitations by developing a novel ABM for the Power-ETS System that leverages real-world data from over 4000 power firms with multi-dimensional heterogeneity. The model explicitly accounts for the interactions among multiple markets and simulates the decision-making behaviors of power firms within this integrated framework. Furthermore, the model employs an improved MADDPG algorithm to enhance accuracy and predictive capability, enabling a more comprehensive understanding of system uncertainties and emergent properties. By simulating the complex interactions of a large and diverse population of agents, our model offers a more realistic representation of firm behaviors and market outcomes.

This paper is structured as follows. In Section 2, we provide a detailed introduction to the specification model, encompassing both the model framework and algorithm settings. Then, the results are provided and discussed in Section 3. Finally, we conclude in Section 4.

2. Model and algorithm design

2.1. Model framework

In this paper, we design an ABM that integrates the power and carbon markets to simulate the decision-making behavior of power firms under new market designs, including reducing emission budgets, changing allocation rule and covering more firms. The interaction between participating agents and multiple markets as previous studies [24,29,44], as shown in Fig. 1. The model aims to analyze the operational dynamics of power generation firms, encompassing power production, power pricing, allowance trading, carbon pricing, and low-carbon technology adoption.

In the presence of complete information, power firms develop carbon price expectations by analyzing historical data, environmental signals, and other relevant factors to inform their decisions on emission reduction [11]. Simultaneously, these agents adjust their behavioral strategies at each stage based on their own historical experience and market information using reinforcement learning algorithms in order to effectively respond to market changes and policy impacts, which are detailed in Section 3.4.

Similar to [24], we also incorporate a multi-level time framework into the model to refine different types of decisions made by firms, including yearly (T), Daily (t), and Secondly (k). Different time scales represent the types of differentiated decisions made by firms.

- The model takes the year as the longest time scale. For understanding and calibration simplicity, we assume that the compliance period is as long as one “year” (T). And each compliance period is further divided into several days (t):

$$T = \{1, 2, \ldots\} \tag{1}$$

- At each day (t), power firms conduct bidding transactions in the spot electricity market to determine the trading prices and volumes of the day-ahead market and intraday market ($q_{i,t}^{DA}, p_{i,t}^{DA}, q_{i,t}^{ID}, p_{i,t}^{ID}$), and achieve short-term emission reduction through production adjustments:

$$q_{i,t}^{DA}, p_{i,t}^{DA}, q_{i,t}^{ID}, p_{i,t}^{ID} \quad \forall i \in I, \forall t \in T \tag{2}$$

- In the power market, medium and long-term contracts can be established on a monthly (m) and annual (y) basis to determine the contract capacity and contract price:

$$q_{i,m}^{M}, p_{m}^{M}, q_{i,y}^{Y}, p_{y}^{Y} \quad \forall i \in I, \forall m \in M, \forall y \in Y \tag{3}$$

- The carbon market is subdivided into several seconds (k) at each day, reflecting the continuous double-sided auction mechanism for transactions:

$$x_{i,t,k}, p_{t,k}^{A} \quad \forall i \in I, \forall t \in T, \forall k \in K \quad K = \{1, 2, \ldots, 3600\} \tag{4}$$

- Firms can make investment decisions ($z_{i,j,t}$) on low-carbon technologies, reflecting their long-term expectations for future emissions reductions:

$$z_{i,j,t} \in \{0, 1\} \quad \forall i \in I, \forall j \in J, \forall t \in T \tag{5}$$

On this basis, firm decisions are made on a daily basis (such as production plans and technological investments), while carbon market transactions simulate a continuous bidding mechanism at a second-level frequency (reflecting the market's instantaneous liquidity). The two are coordinated through an asynchronous clock: daily decisions generate trading strategies, and the carbon market executes transactions at a second-level frequency within the day. And our ABM can be segmented into four distinct modules for individual modeling: power generation module, government module, electricity market module, and carbon market module, which is shown in Supplementary material.

2.2. Multi-agent reinforcement learning algorithm

The interaction and dynamic evolution process among diverse intelligent agents often occur within the state space and action space. Therefore, it is imperative to implement an adaptive strategy, such as fairness, cooperation, or competition strategy, in order to enhance collaborative learning among intelligent agents [45,46]. In order to comprehensively characterize the dynamic evolutionary traits of the coupling mechanism between electricity and carbon markets, this paper has made targeted enhancements and innovative applications to the MADDPG algorithm.

The decision-making processes of power generation firms, including production, technological adoption, and carbon trading decisions ($q_{i,t}^{DA}$, $q_{i,t}^{ID}$, $q_{i,t}^{M}$, $q_{i,t}^{Y}$, $z_{i,j}$, $x_{i,t,k}$), are informed by historical data. These decisions exhibit the Markov property, indicating that they are influenced by observations from previous time slots. From a mathematical perspective, the use of Markov games allows for the re-expression of optimal control as a MARL problem, encompassing sets of states, actions, rewards, and state transition probability distributions for agents [46].

Simultaneously, a Centralized Training and Decentralized Execution (CTDE) framework is adopted to adapt to the multi-agent environment [51]. Each agent has an independent actor network for decision-making, while they collectively share a global critic network to assess the effectiveness of their decision-making (as depicted in Fig. 2).

We define the state space and the local observation $s_{i,t} = \{o_{1,t}, o_{2,t}, \ldots, o_{n,t}\}$, as well as the action space ($act_{i,t}$) of the agent within the model. In the learning process based on the Actor-Critic framework, the policy network refines its output through learning and adjust parameters of the actor network (θ_i^{μ}), in order to select optimal actions within specific states for the purpose of maximizing long-term rewards.

$$a_{i,t} \sim \mathcal{N}_i\left(\mu_i(o_{i,t}|\theta_i^{\mu}), \sigma_i^2\right) \tag{6}$$

Eq. (6) indicates that, in state $o_{i,t}$, the actor network μ_i produces a deterministic action $\mu_i(o_{i,t}|\theta_i^{\mu})$. Then, it is perturbed by Gaussian noise $\mathcal{N}i$ with variance σ_i^2 to obtain the final stochastic action $act_{i,t}$. Besides, we also introduce Asynchronous Advantage Actor-Critic (A3C) to utilize multiple parallel actors for exploration and sampling. This approach accelerates the sample generation and policy learning processes, thereby expanding the DDPG algorithm framework.

π is the agent policy with parameter θ, $\pi_{\theta'}$ is the current policy, π_{θ} is the old policy, $\rho^{\pi_{\theta'}}$ is the status distribution under the current policy, $A^{\pi_{\theta}}(s, a)$ is the advantage function, and $J(\theta')$ is the cumulative expected reward of the i-th agent, so the gradient of the i-th agent can be expressed as[1]:

$$\nabla_{\theta'} J(\theta') = E_{s\sim\rho^{\pi_{\theta'}}, a\sim\pi_{\theta'}}[\nabla_{\theta'} \log \pi_{\theta'}(a|s) A^{\pi_{\theta}}(s, a)] \tag{7}$$

The attention mechanism can improve the information extraction and value function fitting ability of critic network. By combining the attention mechanism with MADDPG, critic network can more efficiently extract the interaction characteristics and importance information of multiple agents, thereby expediting the process of value function approximation and strategy evaluation. And the update of the value function (θ_i^{Q}) includes the following processes.

First, we define an attention weight, where f_s and g_i represent the coding function of state features and action features respectively, and the function score$(\cdot)$ represents the similarity scoring function of the two feature vectors, as Eq. (8) shows.

$$\alpha_{i,t} = \frac{\exp\left(\text{score}\left(f_s(s_t), g_i(a_{i,t})\right)\right)}{\sum_{j=1}^{N} \exp\left(\text{score}\left(f_s(s_t), g_j(a_{j,t})\right)\right)} \tag{8}$$

Moreover, the critic network receives all actions and observations of agents, and computes a joint value function ($\widetilde{Q}$) on the basis of the attention mechanism:

$$\widetilde{Q}(s_t, a_{1,t}, \ldots, a_{N,t}) = \sum_{i=1}^{N} \alpha_{i,t} Q_i(s_t, a_{1,t}, \ldots, a_{N,t}) \tag{9}$$

During the training process, the loss function ($\mathcal{L}(\theta_i^{Q})$) is calculated based on TD, where $r + \kappa\widetilde{Q}(s', a_{1,t+1}, \ldots, a_{N,t+1}; \theta_i^{Q-})$ denotes the TD target and $\widetilde{Q}(s, a_{1,t}, \ldots, a_{N,t}; \theta_i^{Q})$ represents the current value, which can be calculated as Eq. (10).

$$\mathcal{L}(\theta_i^{Q}) = \mathbb{E}_{s,a,r,s'}\left[\left(r + \kappa\widetilde{Q}(s', a_{1,t+1}, \ldots, a_{N,t+1}; \theta_i^{Q-}) - \widetilde{Q}(s, a_{1,t}, \ldots, a_{N,t}; \theta_i^{Q})\right)^2\right] \tag{10}$$

The parameters of the value function can be updated by gradient descent of the loss function, as Eq. (11) shows.

$$\theta_i^{Q} \leftarrow \theta_i^{Q} - \beta \nabla_{\theta_i^{Q}} \mathcal{L}(\theta_i^{Q}) \tag{11}$$

The workflow of MADDPG algorithm is summarized in Algorithm 1. We have developed an intelligent game optimization algorithm that considers heterogeneity, synergy, and generalization in order to address the equilibrium evolution path and key influencing factors of the electric-carbon system. Following multiple iterations and training

[1] Based on Temporal-Difference (TD), the single-step residuals are calculated, and the multi-step residuals are weighted and aggregated to obtain the generalized dominance function ($\widehat{A}_t$) as the dominance estimation of the strategy gradient.

sessions, the strategy of agents progressively converges towards a stable state, facilitating enhanced collaborative learning and interaction among agents.

Initialization: random weights for critic network, actor network, target critic network, target actor network, experience replay buffer, hyper-parameters.

For $episode = 1, 2, 3, \ldots, M$ **do**

1) Receive the initial state space and local observation: $s_{i,t} = \{o_{1,t}, o_{2,t}, ..., o_{n,t}\}$

$o_{i,t} = (p_{i,t}, q_{i,t}, a_{i,t}, e_{i,t}, A_t, D_t, GDP_t, \ldots)$

2) Initialize a random process for multi-agent energy management action exploration.

3) **For** $t = 1, 2, 3, \ldots, T$ **do**

a. Agent i selects actions based on the current state ($o_{i,t}$) through the actor network:

$a_{i,t} \sim \mathcal{N}_i(\mu_i(o_{i,t} \mid \theta_i^{\mu}),\ \sigma_i^2)$

b. Perform action $a_{i,t} = (q_{i,t}^{DA}, p_{i,t}^{DA}, q_{i,t}^{M}, p_{i,t}^{M}, x_{i,t}, z_{i,j,t})$, get a reward $r_{i,t}$ and the next observe $o_{i,t+1}$

c. Store the transition ($o_{i,t}, a_{i,t}, r_{i,t}, o_{i,t+1}$) in the experience replay buffer

d. Update the observation: $o_{i,t+1} \leftarrow o_{i,t}$

e. For $Agent = 1, 2, 3, \ldots, i$ do

i. Sample a random minibatch of transitions from the experience replay buffer: $(o_{i,t}^{s}, a_{i,t}^{s}, r_{i,t}^{s}, o_{i,t}^{s})$

ii. Calculate the target Q value y_i^s : $y_i^s = r_i^s + \gamma Q_{i,t\,\mathrm{arg}et}^{\mu'}\left(o_{i,t}^{s'}, a_1^{'}, a_2^{'}, \ldots, \boldsymbol{a}_{\boldsymbol{N}}^{'}\right)\Big|_{\boldsymbol{a}_{\boldsymbol{i}}^{'}=\mu_i^{'}(o_i^s)}$

iii. Update Critic network:

$\mathcal{L}(\theta_i^{Q}) = \mathbb{E}_{s,a,r,s'}[(r + \kappa\tilde{Q}(s', a1, t+1, \ldots, a_{N,t+1}; \theta_i^{Q-}) - \tilde{Q}(s, a_{1,t}, \ldots, a_{N,t}; \theta_i^{Q}))^2]$

iv. Update Actor network:

$$\nabla_{\theta} J_i^{\pi}(\theta) \approx \frac{1}{T}\sum_{t=1}^{T} \nabla_{\theta} \log \pi_{\theta}(a_{i,t} \mid s_t) \min(r_t(\theta)\hat{A}_t, \mathrm{clip}(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t)$$

End for

f. Update the weights for the target networks for each agent i: $\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}$, $\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}$

End for

End for

2.3. Data source

This paper applies ABM to the Chinese power industry covered in ETS. To provide a more realistic simulation, this paper selects over 3000 Million-KW power firms listed in the "*Global Power Database*" in China. And the distribution of installed capacity and generation of power firms are shown in Appendix A. Each firm possesses distinct technical levels characterized by different technical parameters while their resource endowment varies according to regional factors. This paper also offers comprehensive supplementary data on key technical parameters of various power firms in China, including coal consumption for power generation, auxiliary power ratio, and carbon emission intensity.

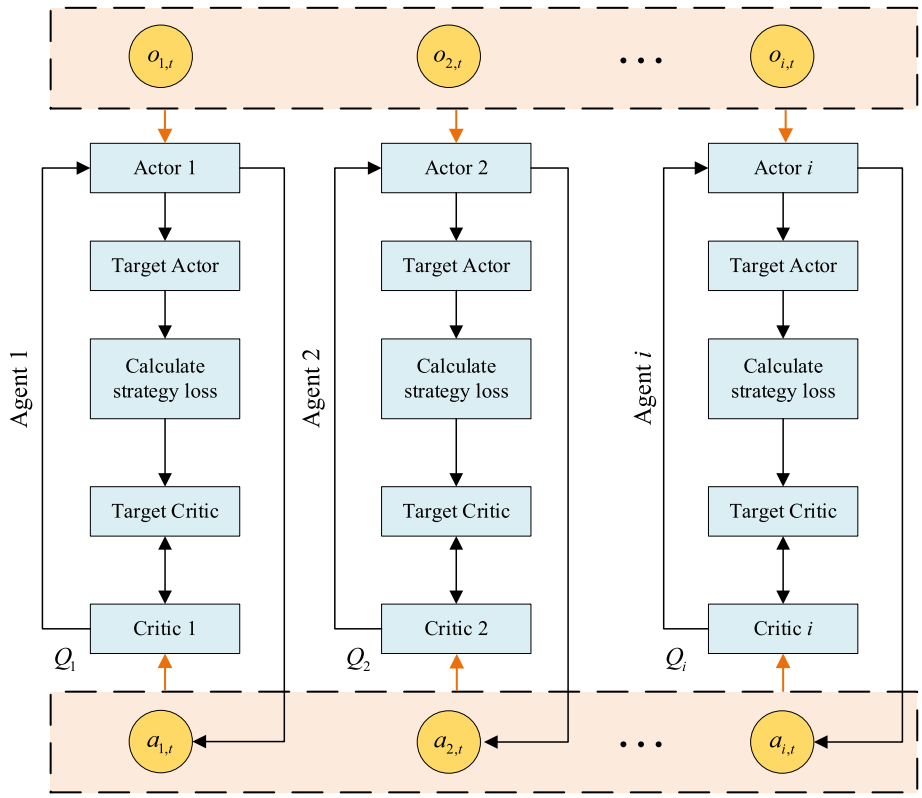


Fig. 2. The centralized training and decentralized execution framework of MADDPG.

2.4. Centralized training

The simulation program is executed on the NVIDIA GeForce RTX 4090 and implemented within the Python 3.12 framework. The training process and specific settings are provided in Appendix B. As introduced in Section 2.2, agents make decisions within a centralized training framework based on a common set of policy functions, yet exhibit heterogeneous behavioral parameters. The MADDPG algorithm orchestrates the learning process of multiple agents to optimize overall compliance.

As depicted in Fig. 3, the episodic total rewards presented demonstrate the progressive refinement of agents' strategies during the learning process. Actually, agents have acquired a stable energy management strategy through continuous interactive learning, leading to convergence of the reward value [47]. While the independent learning algorithms may struggle to adapt well to dynamic changes in multi-agent learning environments, resulting in unstable environmental conditions as strategies change.

Four aggregate results are selected to show the converging process of the training and convergence stage, including average power price, total power generation volume, average carbon price, and total allowance trading volume. With the learning stage proceeding, the results depicted in Fig. 4 demonstrate a more pronounced stable trend. The average power price and total generation volume exhibit reduced fluctuations, potentially attributed to the agents gradually discovering more effective energy management methods as they learn and adjust their strategies; while the results of the carbon market indicate that the system can progressively attain efficient and stable states. From these aggregated findings, it is evident that the MADDPG algorithm optimizes overall compliance by orchestrating the learning processes of multiple intelligent agents, thereby achieving improved price control and production adjustment. And the integration of MARL into ABM partially ensures the "reasonability (not optimality)" of simulation outcomes [24]. And the results indicate that employing MADDPG within a CTDE framework can enhance training stability. Furthermore, attention mechanisms can facilitate faster convergence and improve training quality by selectively integrating all relevant information from intelligent agents [48].

3. Results and discussions

3.1. Impact of the carbon market

While traditional LP/CGE models predict monotonic relationships between emission targets and carbon prices, our ABM reveals nonlinear interactions driven by firm heterogeneity.[2] Fig. 5 shows the price and trading volume in the carbon market in the short-term of simulation. The allowance price exhibits a significant fluctuation trend within the year. In the initial month of trading, the price increased from a low level to a peak (approximately 100 CNY/ton), potentially driven by heightened demand for carbon emissions rights from large-scale firms based on production forecasts and allowance holdings. Subsequently, the allowance price experienced a rapid decline and stayed around 40 CNY/ton, attributed to diminished market activity and trading frequency. By the end of the year, however, the allowance price swiftly escalated and ultimately reached 175 CNY/ton. Short-term price spikes (175 CNY/ton) are primarily driven by large scale firms at the end of the performance period (with 72 % of large firms trading within the last 30 days, compared to just 28 % of small and medium-scale firms), reflecting their limited risk hedging capacity.

Similarly, the findings of allowance trading volume in Fig. 5 also demonstrate "compliance transactions" in the current market, with firms showing a tendency to engage in trade towards year-end leading to dramatic fluctuations in carbon prices. The simulation results also show 23 % higher price volatility due to behavioral factors (with 62 % of small and medium-scale firms employing a "wait-and-see" heuristic strategy,

[2] More heterogeneity analysis results are shown in Appendix C.

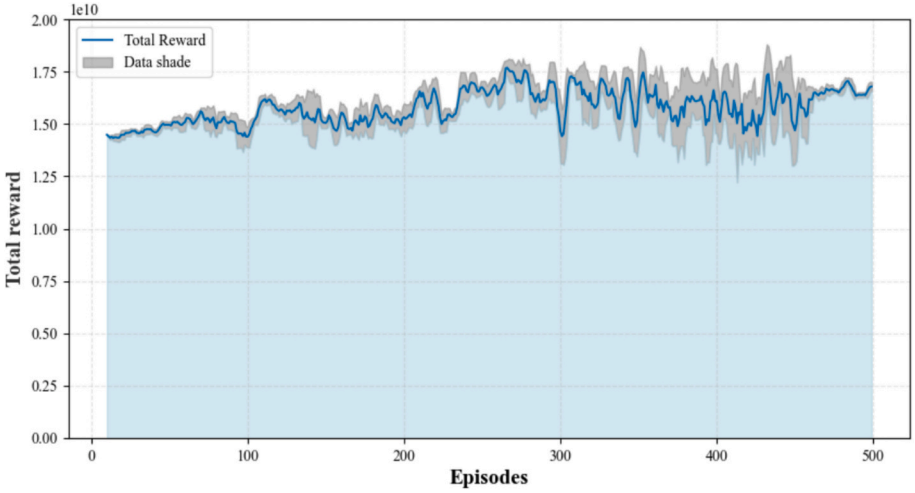


Fig. 3. The convergence curve of reward function based on the MADDPG algorithm.

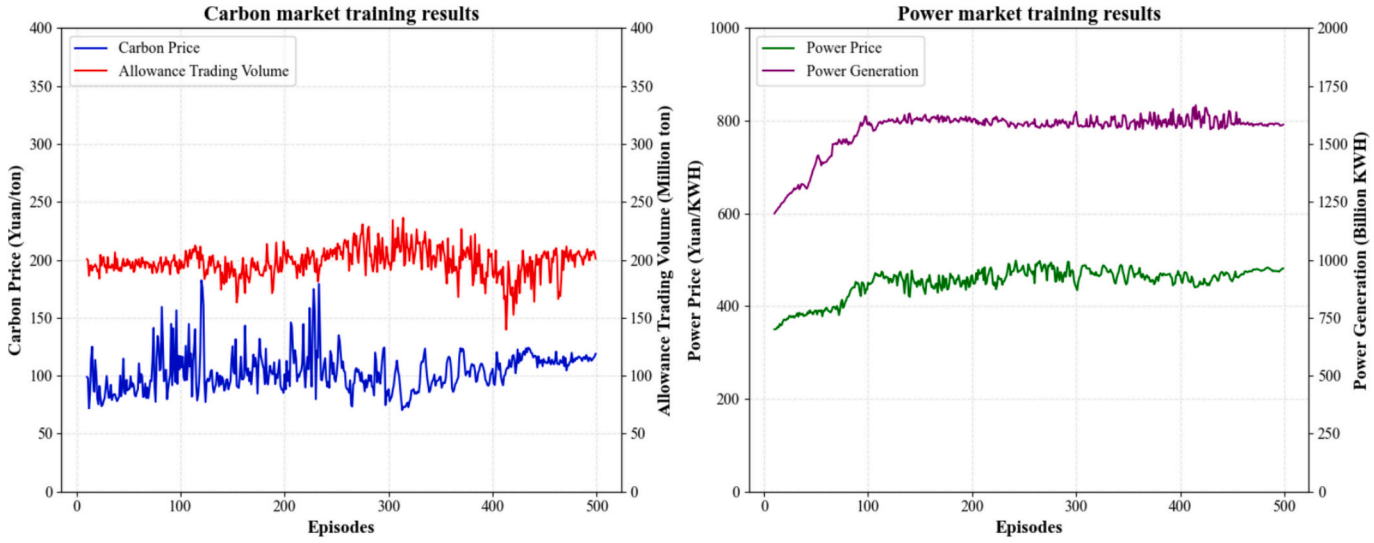


Fig. 4. Converging process of aggregate results.

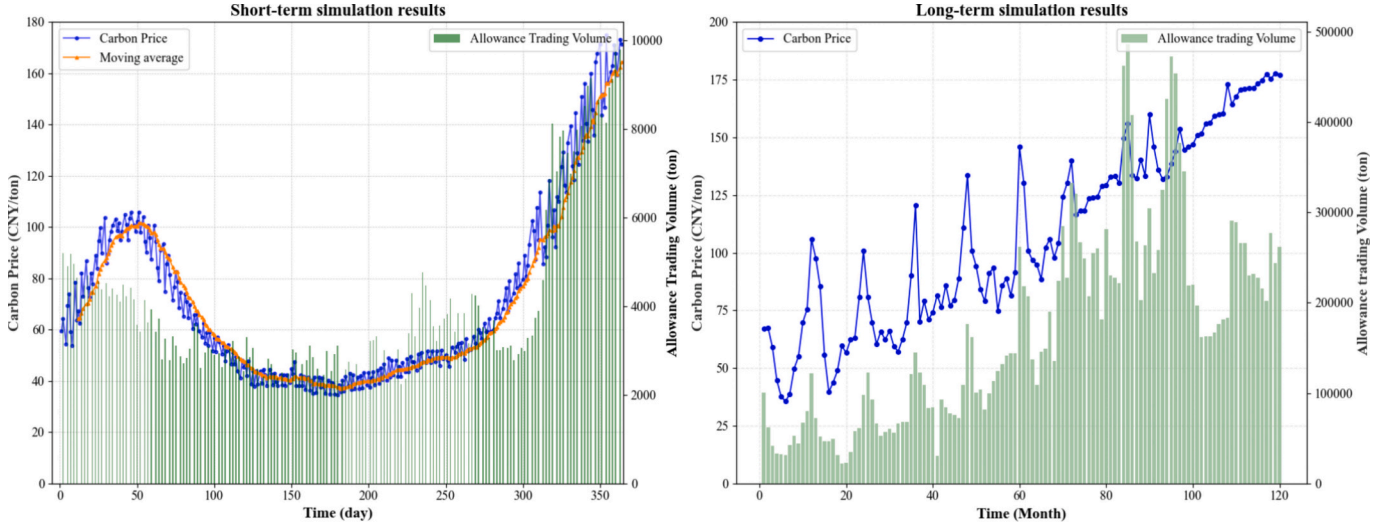


Fig. 5. Dynamics of carbon price and allowance trading volume in short-term scenario and long-term scenario.

compared to just 18 % of large firms).

Additionally, we conducted simulations on the long-term trading process of the carbon market. From the perspective of time, the carbon price demonstrates a fluctuating upward trajectory, with significant spikes observed at the end of each year as shown in Fig. 5. However, the price fluctuations gradually diminish as the simulation progresses, and the market price stabilizes at approximately 170 CNY/ton in the 10th year. Furthermore, Fig. 5 also illustrates variations in carbon trading volume. Transaction volume exhibits a steady increase during the initial eight years and displays a discernible temporal effect. Subsequently, there is a noticeable decline in trading volume which may be attributed to reduced demand resulting from technological upgrades implemented by firms.

3.2. Analysis of emission reduction targets

The increase in emission reduction targets can influence the product market, leading firms to reduce emissions by decreasing production and raising product prices. However, variations in emission reduction targets have a relatively minimal effect on the power market, with elasticities of 0.005 and − 0.007 for average product prices and average power generation respectively. Additionally, changes in emission reduction targets exhibit a clear and stable monotonic influence on product prices and production levels. And the increase in emission reduction targets contributes to the advancement of renewable energy production and constrain the expansion of thermal power generation, as shown in Fig. 6.

Fig. 7 presents the carbon market's trading price and volume under different emission reduction targets. The emission reduction target has monotonically increased impact on the average allowance price. This positive correlation arises because more stringent emission reduction targets decrease the overall supply of allowances within the carbon market. This scarcity intensifies competition among firms for the available allowances, consequently driving prices upwards. Meanwhile, the trading volume exhibits a U-shaped trend—initially decreasing and then increasing—as the emission reduction target rises. When the emission reduction target is below 5 %, the trading volume declines with stricter targets. At this stage, firms face relatively low decarbonization pressure, and most can meet compliance obligations by purchasing only a modest number of allowances. Meanwhile, the rising carbon price incentivizes some firms to adopt voluntary emission reductions. However, once the emission reduction target exceeds 5 %, the trading volume begins to increase alongside more stringent targets. Under these conditions, firms encounter significantly greater compliance pressure, and voluntary abatement alone becomes insufficient to meet regulatory requirements. Consequently, they must purchase substantial additional allowances to fulfill their obligations.

3.3. Analysis of allocation methods

Fig. 8 shows the dynamic trend and progression of carbon prices based on different allocation methods. Carbon prices derived from the auction method demonstrate a consistent upward trajectory throughout the compliance period. It is noteworthy that both the Benchmarking method and Historical intensity method exert similar influences on the carbon market price. The carbon price simulated by the two methods is lower than that of the auction method, and presents a typical L-shaped price trend. While, the auction method provides more clearly and higher price signal, which is more conducive to guiding firms in reducing emissions.

As shown in Fig. 8, we also compare trading volatility and market activity under different allowance allocation methods. The Benchmarking allocation method resulted in relatively low trading volume in the carbon market, with most of it maintained at 2000–4000 tons per day during the compliance period. Notably, the trading volume exhibited a sharp increase near the compliance period, which is similar to the reality of China's national carbon market trading. The results of the Historical intensity method exhibit certain similarities with Benchmarking, however, the historical intensity method leads to higher volatility. The auction mechanism facilitates transparent price discovery, thereby incentivizing increased firm trading at the beginning of the compliance period. As [8] posit, integrating auctions can enhance market prices to better reflect actual supply and demand relationships, consequently mitigating trading volume volatility.

Besides, the auction mechanism reduces compliance-driven volatility by 29 % (compared to benchmarking) but disproportionately burden small and medium-scale firms. Under the auction mechanism, small and medium-scale firms face 52 % higher compliance costs (due to limited liquidity) but accelerate the average adoption of low-carbon technologies by about 55 days. While large firms exploit auctions for arbitrage, 35 % of their trade occurs in the first three months, capturing 48 % of total annual trading profits. The full list of diverse policies for firms of varying scales is provided in Appendix C.

3.4. Analysis of firm decision

Fig. 9 shows the trading decisions in the carbon market among firms with different scales. Throughout the simulation period, the trading volume of large-scale power firms significantly surpasses that of other types of power firms and demonstrates a consistent fluctuation trend with the carbon market, as depicted in Fig. 5. This proves that current large-scale power firms serve as core participants in ETS, and their decision-making behavior can impact the operation and efficiency of the carbon market. Besides, these large-scale firms possess stronger financial capabilities and account for 75 % of speculative trading within the ETS. In contrast, capital constraints and production capacity force small and medium-scale firms to prioritize short-term compliance. More details show that large firms leverage financial reserves for strategic shifts. And 23 % invest in renewables preemptively, reducing their allowance dependence by 11 % at the end of the compliance period.

Fig. 9 also depicts the adoption of low-carbon technologies by firms, the total count of adoption is 420 over the compliance period. Most of them are adopted in the earlier periods, due to the diminishing abatement potential and increasing marginal cost associated with continued adoption [24]. Additionally, the adoption of low-carbon technologies is predominantly led by large-scale firms (approximate 59 %), while small and medium-scale firms contribute 41 % of technology adoptions and tend to begin embracing such technologies mid-year. This may be due to the limited funding and unfamiliarity with the technology faced by these firms in the early stages, causing them to be more cautious in their decision-making. Overall, large-scale firms possess a significant first-mover advantage in adopting low-carbon technologies, enabling them to respond more swiftly to market and policy changes in the early stages. Nonetheless, as the carbon market develops, small and medium-sized enterprises are gradually catching up in adopting low-carbon technologies.

4. Conclusions

In this paper, the ABM are adopted to explore the decision-making behavior of power firms under new ETS designs, including adjusting emission reduction targets, changing allocation rules and considering the heterogeneity of firm scale. We incorporate real-world data from over 4000 Chinese power firms and utilize actual carbon price and trading volumes data from China's national carbon market for parameter training. By applying the MADDPG algorithm, we provide an optimized decision-making framework for heterogeneous agents. Based on the ABM-MADDPG method developed in this paper, the key novel contributions of our work are as follows:

(1) Unique firm-level parameterization and heterogeneity. Our model captures the heterogeneity of power firms by

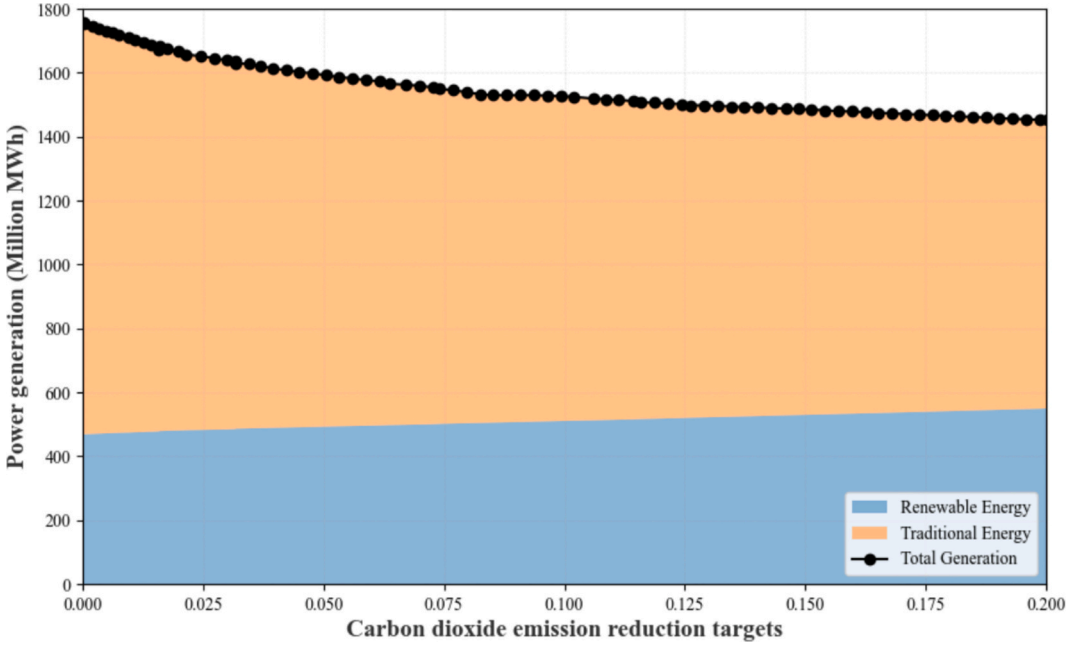


Fig. 6. The impact of emission reduction targets on power markets.

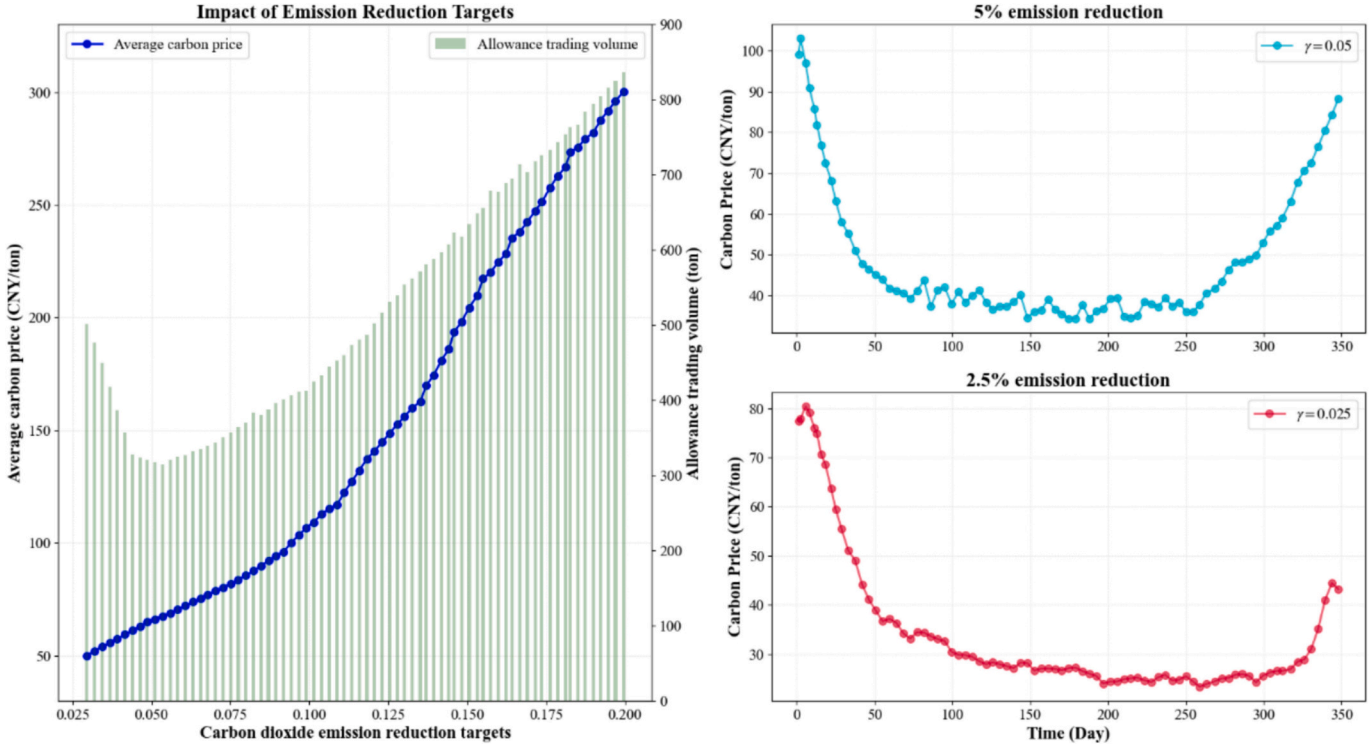


Fig. 7. The impact of emission reduction targets on carbon market.

incorporating real-world data on firm attributes, including scale, cost structures, and technological parameters. This allows us to analyze the distributional impacts of ETS policies across different types of firms, which is a significant advancement over traditional top-down models that often assume homogeneity among firms. For example, we find that large-scale firms dominate allowance trading and are more likely to adopt low-carbon technologies early, while small and medium-scale firms exhibit more cautious behavior due to financial and technological constraints.

(2) Dynamic interactions and emergent system behavior. The integration of ABM with MARL represents a methodological innovation in the study of carbon markets. It enables us to model the dynamic interactions among firms and markets, capturing emergent system behaviors such as price volatility, trading patterns, and technology adoption dynamics. For instance, our model effectively captures the emergent characteristics and

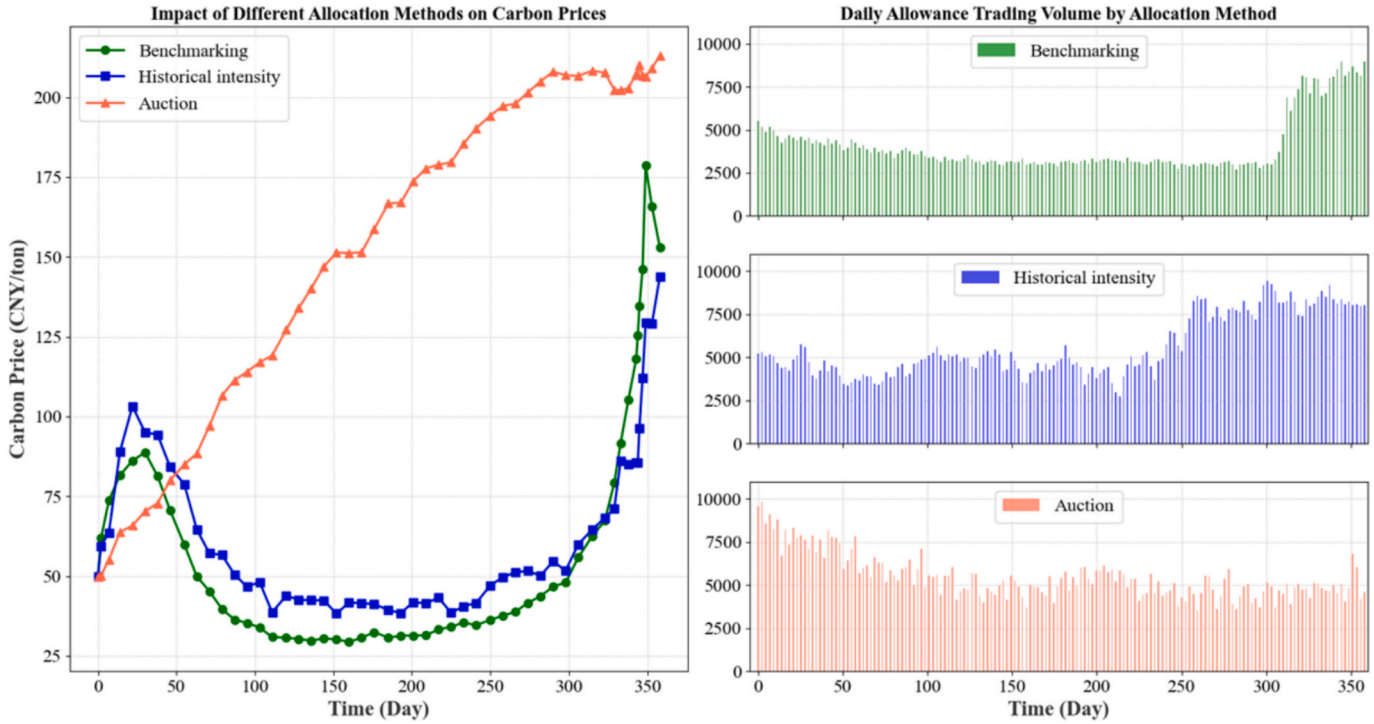


Fig. 8. The impact of different allocation methods on carbon market.

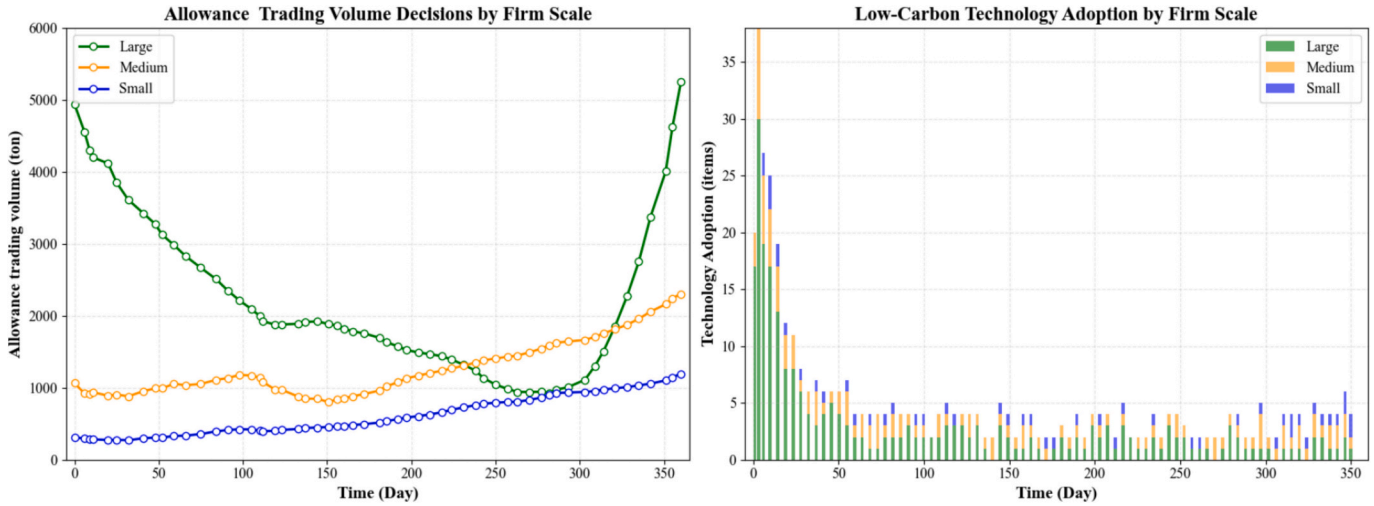


Fig. 9. The trading and low-carbon technology adoption decision of different scales of firms.

uncertainties of the system, as well as the nonlinear impact of emission reduction targets. Carbon prices exhibit significant fluctuations during the compliance period, driven by the trading behaviors of large-scale firms. This level of granularity and realism is not achievable with traditional modeling approaches.

Our findings highlight the necessity of differentiated and synergistic policy designs to analyze the heterogeneous behaviors of power firms and enhance the effectiveness of China's carbon market. For large-scale firms, policymakers should prioritize mechanisms that leverage their market dominance and technological leadership, such as mandating minimum auction-based allowance procurement, offering tax incentives for early adoption of advanced low-carbon technologies, and enhancing trading transparency to stabilize price signals. For small and medium-scale firms, targeted support is critical to overcoming financial and technical barriers; this includes establishing low-interest green loans, simplifying compliance through regional allowance-pooling platforms, and providing capacity-building programs to improve carbon market participation.

At the systemic level, a phased transition from free allocation to auction-based mechanisms—initially targeting large firms and gradually expanding to smaller entities—can balance equity and efficiency. Emission reduction targets should be dynamically adjusted to reflect regional disparities and firm-specific abatement potentials. Additionally, integrating carbon pricing with electricity market reforms can amplify policy synergies, such as introducing carbon-adjusted electricity tariffs. These recommendations aim to align firm-level strategies with national climate goals, fostering an inclusive and efficient transition towards China's dual carbon targets.

While our model provides valuable insights, it has some inevitable

limitations. Firstly, the model does not fully account for all technical parameters in the power generation process, and the construction time for technology adoption is simplified [24,49]. Secondly, the training process assumes a single global solution, but multiple equilibria could exist under different market conditions. Thirdly, the model's results are sensitive to initial carbon prices and firm risk preferences, highlighting the need for further validation. Due to computational constraints, our study only covers the power industry, which may be extended by considering other industries.

CRediT authorship contribution statement

Songyuan Liu: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Peng Zhou:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Mei Wang:** Writing – review & editing, Writing – original draft, Funding acquisition, Formal analysis, Conceptualization. **Aobo Xu:** Software, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to the financial support provided by the National Natural Science Foundation of China (nos. 72243012 & 72373065), and the UPC innovation fund project for graduate student of China University of Petroleum (no. 24CX04036A).

Appendix A. Data sources and parameter setting

This paper selects over 4000 Million-KW power firms listed in the "*Global Power Database*" in China. To enhance the realism of agent heterogeneity, we integrate field survey data from 120 power plants across China and employ Monte Carlo simulation to assign critical technical and operational parameters. Surveyed parameters include coal consumption, auxiliary power ratio, carbon intensity, O&M costs, annual output, and initial allowance allocation. Data spans five provinces (Shandong, Liaoning, Hebei, Tianjin and Inner Mongolia) and covers four plant types (coal-fired, hydro power, wind, and solar). For each parameter, we fit probability distributions based on survey data and industry benchmarks (e.g., *China Energy Statistical Yearbook*), as shown in Table A1.

Table A1
Parameter Distributions of the model.

Parameter	Distribution Type	Mean	Std.Dev	Source
Coal Consumption (g/kWh)	Lognormal	256	43.2	Plant operational reports
Auxiliary Power Ratio (%)	Beta ($\alpha = 2$, $\beta = 5$)	6.5	1.2	Survey data
Carbon Intensity (g/kWh)	Truncated Normal	810	50	CEIC Database
O&M Cost (10 k CNY)	Lognormal	4610	300	Survey data
Initial Allowance (Million ton)	Uniform	485.33	–	Provincial ETS guidelines
Annual value of production (10 k CNY)	Lognormal	109,843	1200	Survey data

Notes: The parameters of coal consumption are sampled from plant-specific distributions, with adjustments for technology vintage (e.g., ±5 % for ultra-supercritical units). The parameters of auxiliary power ratio are related to plant type and age, with older coal plants (up to 8 %) higher than new plants (5–6 %) and renewables (<2 %). The parameters of carbon intensity are directly linked to fuel type and efficiency, calibrated using IPCC emission factors. O&M Costs include labor, maintenance, and grid fees. Wind/solar costs decrease annually by 3 % to reflect learning curves.

Besides, this paper trains the agents using the real price data from the China national carbon market. The actual trading situation of the national carbon market is shown in Fig. A1.

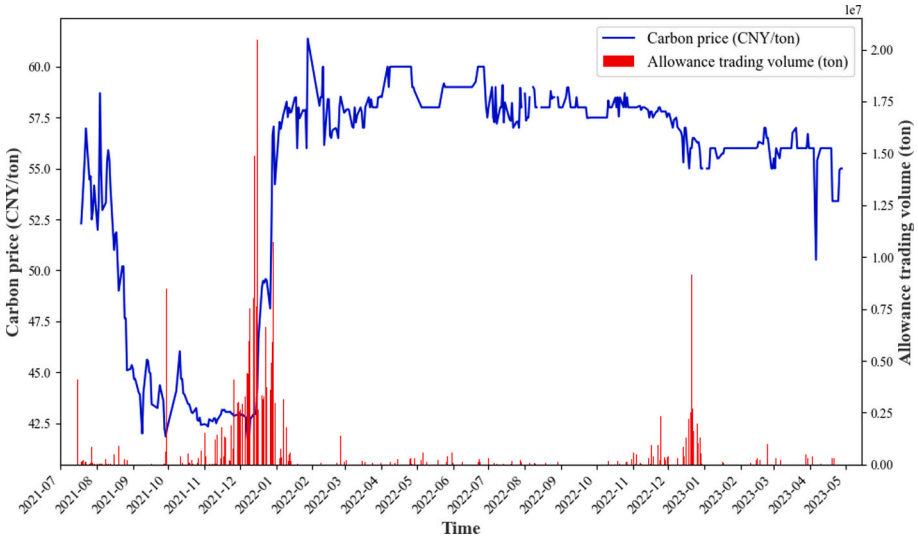


Fig. A1. The carbon price and allowance trading volume.

The initial system parameters of the model are detailed in Table A2. The parameters related to carbon trading in the model mainly refer to the relevant policy documents of China's carbon market. In addition, the total initial carbon quota is set to the total carbon emissions of the base year to ensure that the initial state of the model is consistent with the actual situation.

Table A2
System parameters of the model.

Parameter	Setting	Declare
N	4235	Number of firms
T	365	Number of days in the performance period
K	3600	Number of seconds in each day
γ	2 %	Emission reduction targets
p^f	610	Average fuel price (CNY)
ξ	1.5	Penalty factor for excess emissions
δ_t	5 %	The proportion of the auction quota
φ	100%	The reduction ratio of excess emissions

Appendix B. Algorithm Training

To ensure the effectiveness and stability of the algorithm, this paper has adopted a series of optimization measures. Firstly, through grid search and hyperparameter optimization, the parameter combination that can achieve the highest average reward value was determined. Secondly, the running average of the scenario rewards was used as the convergence evaluation criterion. This convergence evaluation method can effectively determine whether the training is stable and avoid overfitting. In addition, the model also adopted standard techniques such as experience replay buffer and target network to stabilize the training process, and selected the appropriate replay buffer size and target network update frequency based on the performance of the validation set. These measures jointly guarantee the stability and effectiveness of the algorithm and lay the foundation for obtaining reliable simulation results.

For both the actor and critic networks, a five-layer network structure was adopted, including an input layer, three hidden layers, and an output layer. To strike a balance between model performance and computational efficiency, the number of neurons in the three hidden layers was set to 128, 256, and 128, respectively. The Adam algorithm was chosen as the network optimizer, with a learning rate of 0.01 and a batch size of 1024, to enhance the stability and convergence speed of the training. To break the correlation between data and improve the efficiency of sample utilization, the model introduced an experience replay mechanism and set the replay capacity to 1 million to store the interaction experience data of the agent in the environment. To alleviate the bias in value estimation, a soft update strategy was adopted to update the target network, allowing its parameters to approach those of the online network at a small rate, thereby maintaining the stability of the training. The discount factor κ was set to 0.95 to fully consider the long-term returns of the agent and avoid divergence during training. To find the optimal hyperparameter combination, a grid search method was employed, and the highest average reward value was used as the basis for parameter setting. Additionally, to determine the convergence of the model, the average reward of each episode was calculated, and the moving average of 1000 episodes was further computed. When the increase in the moving average reward of 500 consecutive episodes was less than 0.1 %, it was considered that the model had reached a converged state and training could be stopped.

To ensure robustness, we conducted a grid search over key parameters (as shown in Table B1). During the training, we tested the effect of different initial strategies (random vs. history-driven) on the results. The results show that the initial strategy only affects the convergence rate (random strategy takes 2000 rounds vs. Historical strategy 1200 rounds), but there is no significant difference in the steady-state carbon price (170 ± 5 CNY/ton) and the dominance of large enterprises (transaction proportion > 60 %).[3] Such a design aims to ensure the effectiveness and stability of the algorithm and obtain reliable simulation results.

Table B1
Hyperparameter optimization.

Parameter	Tested range	Optimal value
Learning Rate	[0.001, 0.1]	0.01
Batch Size	[512, 2048]	1024
Discount Factor (κ)	[0.9, 0.99]	0.95

Additionally, Fig. B1 illustrates the variation in loss within the MADDPG algorithm under different network layer architectures. While the five-layer network architecture demonstrates a more rapid initial reduction in the critic's loss function compared to the two-layer network, visually assessing definitive convergence solely based on the loss curve for the five-layer network is challenging within the presented training duration. However, it is important to consider the convergence behavior across other key metrics. As depicted in Fig. 3, the episodic total rewards for the five-layer architecture show a clear and stable convergence trend, indicating that the agents are learning effective and consistent strategies. Furthermore, the aggregate market outcomes presented in Fig. 4, including average power price, total power generation volume, average carbon price, and total allowance trading volume, also exhibit stable convergence patterns under the five-layer network. These converging trends in rewards and aggregate

[3] It should be noted that the historical-driven strategy relies on historical data from China's carbon market, which may implicitly contain the characteristics of policy phases (such as the relatively high proportion of free quotas from 2021 to 2022). To eliminate deviations, we conduct 300 rounds of online fine-tuning for all strategies after pre-training, ensuring that the strategies adapt to the latest market rules.

market behavior suggest that despite the less visually conclusive loss curve, the five-layer architecture facilitates the learning of stable and effective policies that lead to convergence at the system level. This indicates that while the loss function might not exhibit a monotonic decrease to a near-zero value within the observed training period, the agents and the overall market dynamics reach a stable equilibrium. Further empirical validation with longer training horizons could provide additional insights into the long-term loss convergence of the five-layer network.

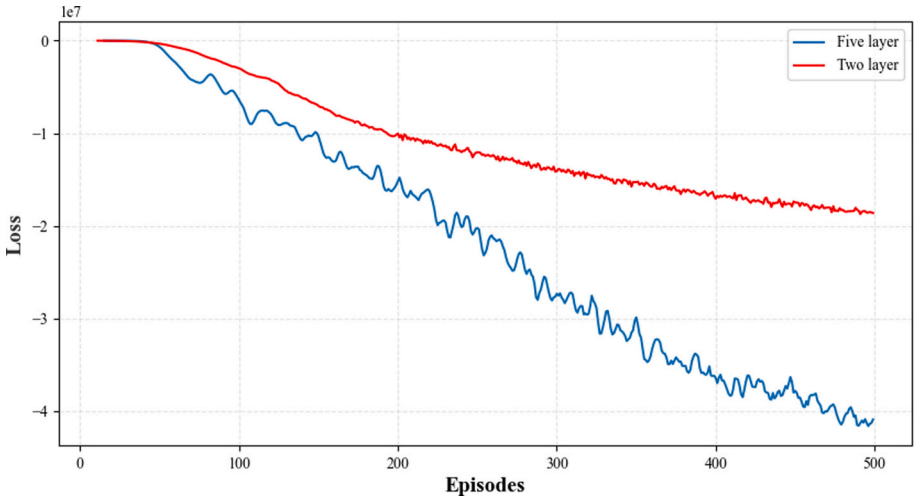


Fig. B1. The loss function curves of different layers based on the MADDPG algorithm.

Appendix C. Supplementary results

Table C1
Behavioral heterogeneity in ABM – Firm-scale dynamics in carbon markets.

	Large-scale firms	Small and medium-scale firms	Difference (Large − small and medium-scale firms)
Trading strategy (% of firms)	Speculative: 75 %	Speculative: 12 %	+63 %
Compliance result (% of firms)	95 %	88 %	+7 %
"Wait-and-see" strategy (% of firms)	18 %	62 %	**−44 %**
Market dominance (% of total trading volume)	77 %	23 %	+54 %
Carbon price sensitivity (% change in price per 1 % demand shock)	+0.8 %	+2.2 %	−1.4 %

Table C2
The impact of different policies on small, medium and large-scale firms.

Policy intervention	Impact on large-scale firms	Impact on small and medium-scale firms	System-level impact
5 % stricter emission reduction target	• Profits: +12 % • Speculative trading: +17 % • Low-carbon adoption: −10 days	• Profits: −9 % • Speculative trading: +10 % • Low-carbon adoption: −25 days	• Carbon price: +2.15 % • Renewable share: +2.78 %
Auction allocation (vs. Benchmarking)	• Compliance costs: +35 % • Trading profits: −12 % • Market dominance: 71 % • Time for low-carbon technologies adoption: 3 days	• Compliance costs: +52 % • Trading profits: −37 % • Market dominance: 29 % • Time for low-carbon technologies adoption: 55 days	• Carbon price: +135 % • Price volatility: −29 %
Historical intensity allocation (vs. Benchmarking)	• Compliance costs: +5 % • Trading profits: +12 % • Market dominance: 71 % • Time for low-carbon technologies adoption: −2 days	• Compliance costs: +11 % • Trading profits: −3 % • Market dominance: 29 % • Time for low-carbon technologies adoption: 5 days	• Carbon price: +3.25 % • Price volatility: +7.65 %

Data availability

The data utilized in this study are partially sourced from open databases (Global Power Database) and purchasable databases (Wind Database). Some supplementary data involve commercial and non-public data, which are available from the corresponding author upon reasonable request.

References

[1] Newell RG, Pizer WA, Raimi D. Carbon markets: past, present, and future. Ann Rev Resour Econ 2014;6(1):191–215. https://doi.org/10.1146/annurev-resource-100913-012655.

[2] Flora M, Variole T. Price dynamics in the European Union emissions trading system and evaluation of its ability to boost emission-related investment decisions. Eur J Oper Res 2020;280:383–94. https://doi.org/10.1016/j.ejor.2019.07.026.

[3] Zhou P, Wei Y, Loeschel A, Su B, Wang H. Special issue: transition management of energy systems towards carbon neutrality. Front Eng Manag 2022;9(3):355–7. https://doi.org/10.1007/s42524-022-0219-z.

[4] Wang M, Zhou P. Impact of permit allocation on cap-and-trade system performance under market power. Energy J 2020;41(6):215–31. https://doi.org/10.5547/01956574.41.6.mwan.

[5] Zhou P, Wen W. Carbon-constrained firm decisions: from business strategies to operations modeling. Eur J Oper Res 2020;281:1–15. https://doi.org/10.1016/j.ejor.2019.02.050.

[6] International Carbon Action Partnership. Emissions Trading Worldwide: 2024 ICAP Status Report. https://icapcarbonaction.com/en/publications/emissions-trading-worldwide-2024-icap-status-report;; 2024.

[7] Zhu J, Fan Y, Deng X, Xue L. Low-carbon innovation induced by emissions trading in China. Nat Commun 2019;10(1):1–8. https://doi.org/10.1038/s41467-019-12213-6.

[8] Chen X, Liu Y, Mcelroy M. Transition towards carbon-neutral electrical systems for China: challenges and perspectives. Front Eng Manag 2022;9(3):504–8. https://doi.org/10.1007/s42524-022-0220-6.

[9] Gallagher KS, Zhang F, Orvis R, Rissman J, Liu Q. Assessing the policy gaps for achieving China's climate targets in the Paris agreement. Nat Commun 2019;10(1):1256. https://doi.org/10.1038/s41467-019-09159-0.

[10] Tang L, Wu J, Yu L, Bao Q. Carbon allowance auction design of China's emissions trading scheme: a multi-agent-based approach. Energy Policy 2017;102:30–40. https://doi.org/10.1016/j.enpol.2016.11.041.

[11] Zhu R, Wei Y, Tan L. Low-carbon technology adoption and diffusion with heterogeneity in the emissions trading scheme. Appl Energy 2024;369:123537. https://doi.org/10.1016/j.apenergy.2024.123537.

[12] Wang B, Duan M. Consignment auctions of emissions trading systems: an agent-based approach based on China's practice. Energy Econ 2022;112:106187. https://doi.org/10.1016/j.eneco.2022.106187.

[13] Branger F, Quirion P. Reaping the carbon rent: abatement and overallocation profits in the European cement industry, insights from an LMDI decomposition analysis. Energy Econ 2015;47:189–205. https://doi.org/10.1016/j.eneco.2014.11.008.

[14] Naegele H, Zaklan A. Does the EU ETS cause carbon leakage in European manufacturing? J Environ Econ Manag 2019;93:125–47. https://doi.org/10.1016/j.jeem.2018.11.004.

[15] Dissanayake S, Mahadevan R, Asafu-Adjaye J. Evaluating the efficiency of carbon emissions policies in a large emitting developing country. Energy Policy 2020;136:111080. https://doi.org/10.1016/j.enpol.2019.111080.

[16] Romagnoli F, Barisa A, Dzene I, Blumberga A, Blumberga D. Implementation of different policy strategies promoting the use of wood fuel in the Latvian district heating system: impact evaluation through a system dynamic model. Energy 2014;76:210–22. https://doi.org/10.1016/j.energy.2014.06.046.

[17] Franco CJ, Castaneda M, Dyner I. Simulating the new British electricity-market reform. Eur J Oper Res 2015;245:273–85. https://doi.org/10.1016/j.ejor.2015.02.040.

[18] Nguyen LKN, Howick S, Megiddo I. A framework for conceptualising hybrid system dynamics and agent-based simulation models. Eur J Oper Res 2024;315:1153–66. https://doi.org/10.1016/j.ejor.2024.01.027.

[19] Li L, Wang J, Zhong X, Lin J, Wu N, Zhang Z, et al. Combined multi-objective optimization and agent-based modeling for a 100% renewable island energy system considering power-to-gas technology and extreme weather conditions. Appl Energy 2022;308:118376. https://doi.org/10.1016/j.apenergy.2021.118376.

[20] Richstein JC, Chappin EJ, de Vries LJ. Cross-border electricity market effects due to price caps in an emission trading system: an agent-based approach. Energy Policy 2014;71:139–58. https://doi.org/10.1016/j.enpol.2014.03.037.

[21] Hoekstra A, Steinbuch M, Verbong G. Creating agent-based energy transition management models that can uncover profitable pathways to climate change mitigation. J Complex 2017:1967645. https://doi.org/10.1155/2017/1967645.

[22] Araavind S, Samuli H, Fredy R, Jan S, Salla A, Annika W. Residential consumer enrollment in demand response: an agent based approach. Appl Energy 2024;374:123988. https://doi.org/10.1016/j.apenergy.2024.123988.

[23] Tesfatsion L. Agent-based computational economics: a constructive approach to economic theory. Handb Comput Econ 2006;2:831–80. https://doi.org/10.1162/106454602753694765.

[24] Yu S, Fan Y, Zhu L, Eichhammer W. Modelling the emission trading scheme from an agent-based perspective: system dynamics emerging from firms' coordination among abatement options. Eur J Oper Res 2020;286:1113–28. https://doi.org/10.1016/j.ejor.2020.03.080.

[25] Bakam I, Matthews RB. Emission trading in agriculture: a study of design options using an agent-based approach. Mitig Adapt Strateg Glob Chang 2009;14(8):755–76. https://doi.org/10.1007/s11027-009-9197-2.

[26] Ringler P, Keles D, Fichtner W. Agent-based modelling and simulation of smart electricity grids and markets–a literature review. Renew Sust Energ Rev 2016;57:205–15. https://doi.org/10.1016/j.rser.2015.12.169.

[27] Bunn DW, Oliveira FS. Agent-based analysis of technological diversification and specialization in electricity markets. Eur J Oper Res 2007;181(3):1265–78. https://doi.org/10.1016/j.ejor.2005.11.056.

[28] Jo H, Lee H, Suh Y, Kim J, Park Y. A dynamic feasibility analysis of public investment projects: an integrated approach using system dynamics and agent-based modeling. Int J Proj Manag 2015;33(8):1863–76. https://doi.org/10.1016/j.ijproman.2015.07.002.

[29] Wei Y, Liang X, Xu L, Kou G, Chevallier J. Trading, storage, or penalty? Uncovering firms' decision-making behavior in the Shanghai emissions trading scheme: insights from agent-based modeling. Energy Econ 2023;117:106463. https://doi.org/10.1016/j.eneco.2022.106463.

[30] Swinerd C, McNaught KR. Design classes for hybrid simulations involving agent-based and system dynamics models. Simul Model Pract Theory 2012;25:118–33. https://doi.org/10.1016/j.simpat.2011.09.002.

[31] Petrović M, Ozel B, Teglio A, Raberto M, Cincotti S. Should I stay or should I go? An agent-based setup for a trading and monetary union. J Econ Dyn Control 2020;113:103866. https://doi.org/10.1016/j.jedc.2020.103866.

[32] Chen H, Ma T. Technology adoption and carbon emissions with dynamic trading among heterogeneous agents. Energy Econ 2021;99:105263. https://doi.org/10.1016/j.eneco.2021.105263.

[33] Loomans N, Alkemade F. Exploring trade-offs: a decision-support tool for local energy system planning. Appl Energy 2024;369:123527. https://doi.org/10.1016/j.apenergy.2024.123527.

[34] Guo X, Zhang X, Zhang X. Incentive-oriented power-carbon emissions trading-tradable green certificate integrated market mechanisms using multi-agent deep reinforcement learning. Appl Energy 2024;357:122458. https://doi.org/10.1016/j.apenergy.2023.122458.

[35] Fang C, Ma T. Technology adoption with carbon emission trading mechanism: modeling with heterogeneous agents and uncertain carbon price. Ann Oper Res 2021;300:577–600. https://doi.org/10.1007/s10479-019-03297-w.

[36] Zuo Y, Zhao X. Effects of herding behavior of tradable green certificate market players on market efficiency: insights from heterogeneous agent model. Front Energy 2023;17:266–85. https://doi.org/10.1007/s11708-021-0752-1.

[37] Zhu L, Chen L, Yu X, Fan Y. Buying green or producing green? Heterogeneous emitters' strategic choices under a phased emission-trading scheme. Resour Conserv Recycl 2018;136:223–37. https://doi.org/10.1016/j.resconrec.2018.04.017.

[38] Manson SM. Bounded rationality in agent-based models: experiments with evolutionary programs. Int J Geogr Inf Sci 2006;20:991–1012. https://doi.org/10.1080/13658810600830566.

[39] Sun Q, Wang X, Liu Z, Mirsaeidi S, He J, Pei W. Multi-agent energy management optimization for integrated energy systems under the energy and carbon co-trading market. Appl Energy 2022;324:119646. https://doi.org/10.1016/j.apenergy.2022.119646.

[40] Zgonnikov A, Lubashevsky I. Unstable dynamics of adaptation in unknown environment due to novelty seeking. Adv Complex Syst 2014;17:1450013. https://doi.org/10.48550/arXiv.1305.3657.

[41] Rahimiyan M, Mashhadi HR. An adaptive Q-learning algorithm developed for agent-based computational modeling of electricity market. IEEE Trans Syst Man Cybern Syst 2010;40:547–56. https://doi.org/10.1109/TSMCC.2010.2044174.

[42] Qiu D, Chen T, Strbac G, Bu S. Coordination for multienergy microgrids using multiagent reinforcement learning. IEEE Trans Industr Inform 2023;19:5689–700. https://doi.org/10.1109/TII.2022.3168319.

[43] Wu J, Zheng X, Yu S, Yu L. Modeling multi-market coupling effects considering the consumption above quota trading market in renewable portfolio standards: an agent-based perspective. Energy Econ 2024;138:107826. https://doi.org/10.1016/j.eneco.2024.107826.

[44] Rego EE, Costa OLV, Ribeiro CDO, Lima F, Takada H, Stern J. The trade-off between demand growth and renewables: a multiperiod electricity planning model

under CO_2 emission constraints. Energy 2020;213:118832. https://doi.org/10.1109/ACCESS.2021.3103100.
[45] Liu D, Wang W, Li H, Shi M, Chen G, Xie Z, et al. Joint optimization of quota policy design and electric market behavior based on renewable portfolio standard in China. IEEE Access 2021;9:113347–61. https://doi.org/10.1109/ACCESS.2021.3103100.
[46] Harrold DB, Cao J, Fan Z. Renewable energy integration and microgrid energy trading using multi-agent deep reinforcement learning. Appl Energy 2022;318:119151. https://doi.org/10.1016/j.apenergy.2022.119151.
[47] Zhou Y, Ma Z, Shi X, Zou S. Multi-agent optimal scheduling for integrated energy system considering the global carbon emission constraint. Energy 2024;288:129732. https://doi.org/10.1016/j.energy.2023.129732.
[48] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. Neurocomputing 2021;452:48–62. https://doi.org/10.1016/j.neucom.2021.03.091.
[49] Fleiter T, Rehfeldt M, Herbst A, Elsland R, Klingler AL, Manz P, et al. A methodology for bottom-up modelling of energy transitions in the industry sector: the FORECAST model. Energ Strat Rev 2018;22:237–54. https://doi.org/10.1016/j.esr.2018.09.005.