Electronics and Electrical Engineering

AIMS Electronics and Electrical Engineering, 9(4): 448–475.
DOI: 10.3934/electreng.2025021
Received: 09 April 2025
Revised: 03 August 2025
Accepted: 15 August 2025
Published: 01 September 2025

https://www.aimspress.com/journal/ElectrEng

Research article

Multi-objective optimization method for charging and discharging of electric vehicles via Q-learning-based grey wolf algorithm

Zhi Zhang[1], Taijun Guo[1], Yefeng Liu[2], Xinfu Pang[3,*] and Yi Zhang[4]

[1] School of Mechanical Engineering and Automation, Shenyang Institute of Technology, Fushun 113122, China
[2] School of Automation and Electrical Engineering, Linyi University, Linyi 276005, China
[3] Key Laboratory of Energy Saving and Controlling in Power System of Liaoning Province, Shenyang Institute of Engineering, Shenyang 110136, China
[4] Liaoning Urban Construction Technical College, Shenyang 110122, China

* **Correspondence:** Email: pangxf@sie.edu.cn.

Abstract: With the widespread adoption of electric vehicles (EVs), their large-scale integration into the power grid may lead to uncoordinated charging issues, resulting in a significant increase in peak power loads. This may adversely affect the safe, stable operation and economic efficiency of the power grid. To address these challenges, this paper proposed a multi-objective optimization method for EV charging and discharging via Q-learning-based grey wolf optimizer (GWO). The proposed method follows a systematic approach: First, the system architecture was established, defining the roles of EVs and flexible devices in the charging and discharging scheduling process. Second, a multi-objective optimization dispatch model was formulated, with the objectives of minimizing system operation cost, load mean square error, and load peak-to-valley difference. The model was subject to constraints related to power balance and the operational limits of various devices. Third, an advanced multi-objective grey wolf optimizer with integrated Q-learning was developed to address the dispatch model. This optimizer aimed to achieve the Pareto front. Parameters were optimized via orthogonal experiments. The optimizer's effectiveness was measured using metrics assessing distribution and convergence. To address the problem, a case study was used to assess the performance of the proposed Q-GWO algorithm compared to benchmark algorithms, analyzing optimization outcomes and algorithm strengths. Case studies demonstrated 17.3% cost reduction and 31.5% lower peak-valley difference versus benchmark algorithms. Q-GWO's convergence speed (validated via IGD/HV metrics) outperformed GWO by 15.2%, proving efficacy for real-world V2G

dispatch. Additionally, the TOPSIS multi-criteria decision analysis approach was employed to identify an optimal solution.

Keywords: grey wolf algorithm; multi-objective optimization; Q-learning; time-of-use electricity price; vehicle-to-grid (V2G)

1. Introduction

1.1. Literature review

Traditional energy sources have been the cornerstone of human societal and industrial activities, dominating the global energy structure [1]. However, as environmental pollution intensifies and energy shortages worsen, the traditional fossil fuel–based energy system is increasingly unable to meet the demands of sustainable development [2]. In this context, distributed renewable energy sources, characterized by their cleanliness, renewability, and wide availability, have emerged as a critical solution to address the dual challenges of energy and environmental sustainability [3]. As the proportion of renewable energy generation in power systems continues to rise, the inherent randomness and intermittency of renewable sources significantly increase the complexity of ensuring safe and stable power system operations. This issue also negatively impacts the economic efficiency and stability of power systems [4]. To address these challenges, vehicle-to-grid (V2G) technology has emerged as a promising solution. By enabling bidirectional energy exchange between electric vehicles and the grid [5], V2G technology can effectively alleviate power system fluctuations, reduce the grid's impact from renewable energy integration, and fully leverage the advantages of clean energy, thereby promoting green transformation and efficient operation of power systems.

Global scholars have extensively explored the coordinated optimization dispatch of electric vehicles and microgrids. To address electric vehicle optimization dispatch challenges, researchers primarily employ intelligent algorithms and mathematical solvers. To optimize EV charging and discharging, a mathematical model must first be established, followed by the selection of an appropriate optimization decision-making method. Optimization methods can be broadly categorized into mathematical methods and intelligent algorithms. Mathematical methods include linear programming [6], integer programming [7], and dynamic programming [8], which typically rely on the convexity, differentiability, and linearity of the objective function. Existing studies have proposed various mathematical approaches for EV dispatch, such as collaborative multi-level frameworks [9], hierarchical operation strategies [10], day-ahead dispatch using multi-objective ranking [11], and techniques emphasizing user convenience with V2G options [12]. Furthermore, methods addressing uncertainties like wind power through specific objective functions [13] and frameworks based on mixed-integer linear programming [14] have been developed. However, the high dimensionality, complexity, and constraints inherent in EV charging and discharging optimization often render mathematical methods unable to find precise solutions efficiently or handle non-ideal problem properties.

Given these limitations, intelligent algorithms are increasingly adopted. These algorithms utilize iterative search and optimization strategies to identify near-optimal dispatch solutions within large solution spaces. Their key advantage lies in not imposing strict requirements on the properties of the

objective function or constraints, making them highly adaptable to system variations and uncertainties. Existing research has explored various intelligent algorithms for EV dispatch, including two-phase approaches with advanced genetic algorithms [3], bi-level frameworks using generalized Nash equilibrium models [15], and long-term strategies incorporating user range anxiety and rolling optimization via non-dominated sorting genetic algorithms [16].

To further enhance the efficiency of intelligent algorithms, researchers have explored the integration of reinforcement learning techniques, particularly Q-learning, to improve algorithm performance [17]. Q-learning excels by learning optimal actions through reward-guided trial-and-error. Integrating Q-learning with meta-heuristics like the grey wolf optimizer (GWO) offers significant potential to refine its search mechanism. Key anticipated improvements include adaptively balancing exploration and exploitation during iterations based on feedback, dynamically adjusting critical algorithm parameters like movement coefficients for efficient convergence, intelligently selecting search operators based on performance history, and strategically guiding exploration to escape local optima. Successful Q-learning integrations across diverse algorithms, including optimizing GA parameters [18], selecting operators in swarm intelligence [19], enhancing ABC efficiency [20], configuring GA frameworks [21], and combinatorial optimization [22], demonstrate this potential. However, applying Q-learning to enhance specific aspects of GWO's search mechanism, such as guiding its social hierarchy, hunting behavior, or parameter adaptation, for solving EV charging/discharging optimization remains unexplored, constituting a critical research gap.

In conclusion, EV charging and discharging optimization decision-making is a complex dispatch problem involving multi-objective trade-offs and the handling of various uncertainties. Based on the analysis of existing research on EV charging and discharging optimization methods, future research should focus on two main directions: system modeling and optimization decision-making. In system modeling, flexibility resources such as energy regulation devices, energy storage systems, and demand response should be fully utilized to enhance system adaptability [23]. In optimization decision-making, grey wolf algorithms and Q-learning algorithms demonstrate significant potential, as they can efficiently search and optimize in large solution spaces, thereby improving system dispatch efficiency and performance.

1.2. Motivation and contributions

Based on a thorough review of existing literature on electric vehicle optimization dispatch, we contend that the influence of flexible resource allocation on the performance of EV optimization systems must be considered. Furthermore, most studies on multi-objective optimization dispatch of EVs tend to use weighted summation or hierarchical methods for multiple objective functions, while directly solving the dispatch problem to obtain the Pareto front using intelligent algorithms still requires further investigation. Moreover, the parameters of intelligent algorithms are often arbitrarily set without a scientific basis, necessitating explanations and justifications for their configuration. Additionally, the existing literature on multi-objective EV dispatch optimization rarely assesses the convergence and distribution performance of multi-objective algorithms using metrics such as inverse generational distance (IGD), hypervolume (HV), and maximum spread (MS). Similarly, there is limited exploration of employing multi-objective decision-making methods to identify the most suitable compromise solution from the Pareto-optimal set for EV dispatch applications. The main contributions of this study are reflected in the following aspects:

(1) This study presents a comprehensive multi-objective optimization dispatch model designed to tackle the intricate challenges of electric vehicle dispatch. Central to this model are three pivotal considerations: load mean square deviation, load peak-to-valley difference, and system operation cost. To optimize energy management efficiency, the model implements dynamic pricing strategies on the demand side, offering a framework to enhance the effectiveness of energy consumption patterns. Additionally, it explores the potential of V2G technology, emphasizing its crucial role in facilitating the integration of renewable energy sources into the grid. By leveraging V2G technology, the model promotes efficient energy utilization and storage, thereby supporting grid stability and sustainable energy practices.

(2) To optimize electric vehicle dispatch, a method combines a multi-objective grey wolf optimizer with Q-learning to enhance algorithm performance and identify the Pareto front, balancing objectives effectively. The integration of Q-learning allows the optimizer to adapt dynamically, enhancing its ability to navigate the search space effectively. To refine the algorithm's performance, orthogonal experiments are utilized for parameter calibration, ensuring robustness and efficiency. The algorithm's effectiveness is assessed comprehensively using metrics such as IGD, HV, and MS, providing a thorough evaluation from multiple perspectives. To determine the most suitable solution among multiple criteria, the TOPSIS method is utilized to identify the optimal compromise solution by comparing each alternative's proximity to the ideal solution, thereby guiding informed dispatch decisions. This approach enhances the decision-making process by ensuring that the chosen solution is well-informed and effective. This approach not only advances the optimization process but also ensures practical applicability in real-world scenarios.

(3) The proposed optimization dispatch method is validated through specific case studies, and the experimental results highlight its effectiveness. The dispatch strategy successfully reduces EV owners' charging costs while minimizing grid load fluctuations, ensuring both economic and stable operation of microgrids. Additionally, the Q-GWO algorithm demonstrates superior solution performance compared to benchmark algorithms, further underscoring its advantages in addressing the optimization challenges of EV dispatch systems.

The rest of this paper is structured as follows: Section 2 presents the system architecture and the corresponding mathematical formulation; Section 3 elaborates on the development of a solution algorithm tailored for the multi-objective optimization dispatch methodology applied to electric vehicles; Section 4 demonstrates the effectiveness of the proposed method through simulation results; and Section 5 highlights the key research contributions of the study. Additionally, important research trends are summarized in Table 1.

Table 1. Literature comparison table.

Reference	Direction of power flow		Objective function			Solution method	Parameter optimization	Pareto	Evaluating indicator	Decision-making method
	One-way	Both-ways	Economic goal	Load goal	Others					
Shaheen et al. (2024)		√	√	√		MOWOA		√		
Yin et al. (2024)		√	√		√	CPLEX				

Zhao et al. (2022)		√	√		√	MOGA		√		
Jia et al .(2024)	√		√	√		Lagrangian-BCQ				
Zohre et al. (2024)		√	√			NSGA-II				
Zhao et al. (2022)		√	√		√	GA				
LV et al. (2022)		√	√	√		IGWO				
Hou et al. (2022)		√	√		√	GA				
Yin et al. (2023)		√	√			CPLEX				
This paper		√	√	√		Q-GWO	Orthogonal test	√	IGD HV MS	TOPSIS

2. Problem description

2.1. System structure

The structure of the system enabling electric vehicles to participate in charging and discharging optimization is depicted in Figure 1. On the power supply side, the system primarily consists of photovoltaic power generation systems, wind turbine generators, micro gas turbines, and energy storage devices, which collectively provide stable electrical energy to the system. On the load side, the system covers both electric vehicle loads and conventional loads.

Electric vehicles are grouped into three categories according to dispatch needs: (1) EVs participating in discharging dispatch, which are capable of supplying power back to the grid during discharging operations; (2) EVs participating in charging dispatch, which are scheduled for charging based on grid conditions and demand; and (3) independently operated EVs, which do not participate in charging or discharging dispatch management due to the specific travel demands of their users.

2.2. Objective function

(1) Objective 1: Minimize the system operation cost

Taking into full consideration the incentive mechanisms for electric vehicle load demand-side management, it is evident that the deeper the response of electric vehicle users to V2G technology, the greater the benefits they derive. The objective function is to minimize the dispatch cost of the microgrid-integrated V2G system. By defining the reference quantities based on the positive and negative values of the dispatch center's revenue and expenditure, the objective function for minimizing the operational cost can be formulated as follows:

$$F_1 = C_{\mathrm{MT}} + C_{\mathrm{BA}} + C_{\mathrm{EV}} + C_{\mathrm{Grid}} + C_{\mathrm{pv}} + C_{\mathrm{wind}} - G_{\mathrm{load}} - G_{\mathrm{EV}} - G_{pv} - G_{\mathrm{MT}} - G_{\mathrm{BA}} \tag{1}$$

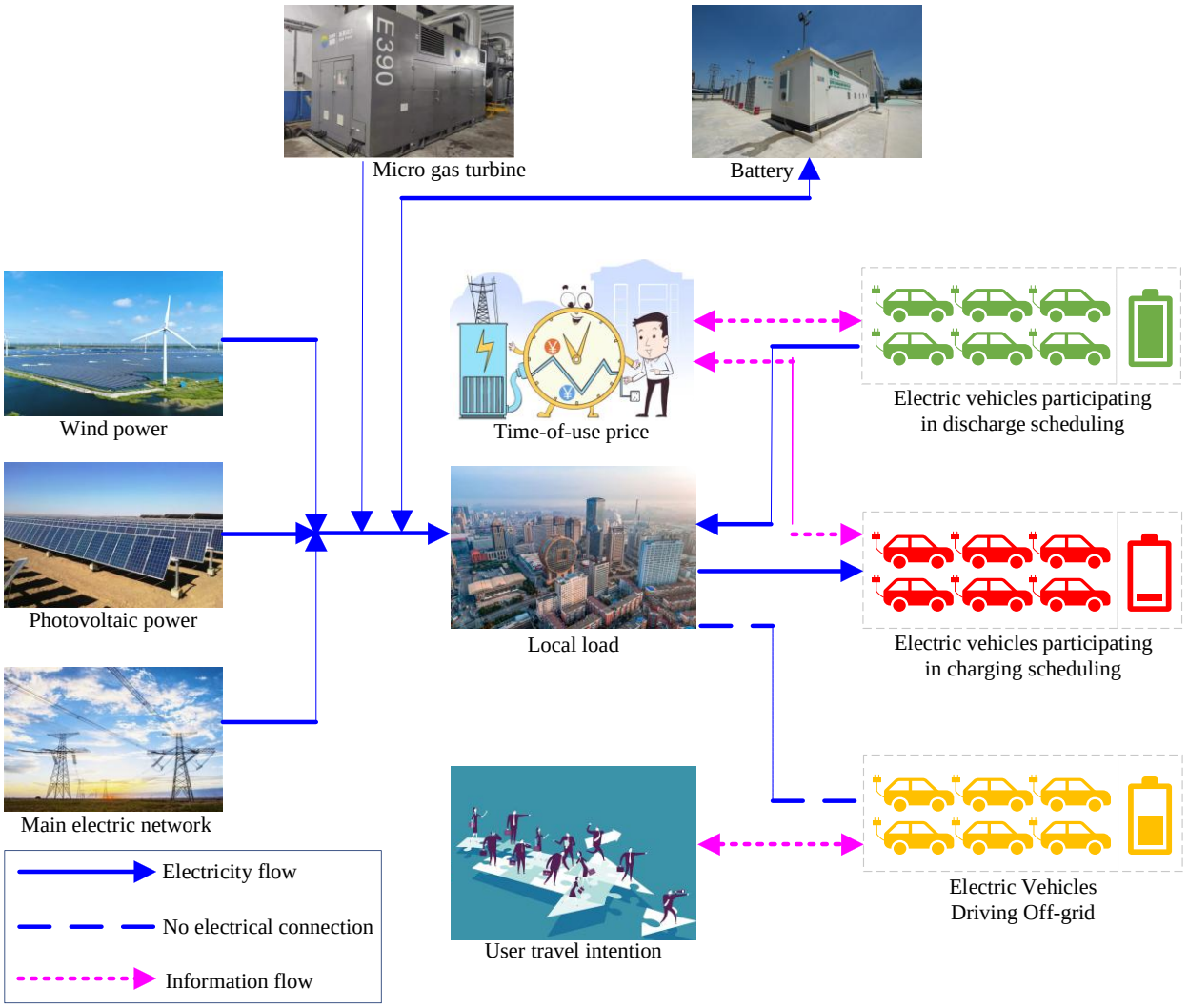

Figure 1. Charging and discharging system structure diagram of electric vehicles.

Where F_1 represents the total operation cost of the microgrid system integrated with the V2G system. A positive value indicates a loss, while a negative value represents a gain. C_{Grid} denotes the transaction cost between the dispatch center and the grid. C_{MT} represents the operation cost of the micro-gas turbine. C_{EV} represents the compensation cost for charging/discharging electric vehicle batteries. C_{BA} represents the depreciation cost of the battery storage. C_{pv} represents the operation cost of photovoltaic generation. G_{load} represents the operation cost of wind turbine generation. G_{EV} represents the revenue generated by the dispatch center from selling electricity to the load. G_{EV} represents the revenue from electric vehicle users during charging. G_{pv} represents the revenue from photovoltaic generation. G_{MT} represents the revenue from micro-gas turbine generation. G_{BA} represents the revenue from the discharge battery storage. Since the revenues G_{pv}, G_{MT}, and G_{BA} overlap with C_{Grid}, G_{EV}, and G_{load}, the values of G_{pv}, G_{MT}, and G_{BA} are set to zero.

$$C_{\text{Grid}} = \sum_{i=1}^{24} \left(P_{\text{Grid},t} \times Pr_{\text{Grid},t} \right) \tag{2}$$

$$P_{\text{Grid},t} = P_{\text{pv},t} + P_{\text{wind},t} + P_{\text{MT},t} - P_{\text{load},t} - P_{\text{BA},t} - \sum_{n=1}^{N} P_{\text{EV},n,t} \tag{3}$$

$$n_{\text{BA}} = \begin{cases} n_{\text{BA\_charge}}, P_{\text{BA}} > 0 \\ 1/\, n_{\text{BA\_discharge}}, P_{\text{BA}} \leqslant 0 \end{cases} \tag{4}$$

$$n_{\text{EV}} = \begin{cases} n_{\text{EV\_charge}}, P_{\text{EV}} > 0 \\ 1/\, n_{\text{EV\_discharge}}, P_{\text{EV}} \leqslant 0 \end{cases} \tag{5}$$

$$Pr_{\text{Grid},t} = \begin{cases} P_{\text{buy},t}, P_{\text{pv,t}} + P_{\text{BA},t} + P_{\text{MT},t} - P_{\text{load},t} - \sum_{n=1}^{N_t} P_{\text{EV},n,t} < 0 \\ P_{\text{sell},t}, \text{else} \end{cases} \tag{6}$$

Where t represents the time point, n represents the nth vehicle, and C_{Grid} represents the sum of income and expenditure from transactions between the microgrid and the grid, considered as cost. $P_{\text{pv},t}$ represents the photovoltaic output, $P_{\text{BA},t}$ represents the battery output, $P_{\text{MT},t}$ represents the micro-turbine output, $P_{\text{load},t}$ represents the electric load, $P_{\text{EV},n,t}$ represents the EV charging/discharging power, $Pr_{\text{Grid},t}$ represents the buy-sell electricity price, and N_t represents the number of dispatchable EVs at time t.

$$C_{\text{MT}} = \sum_{i=1}^{24} P_{\text{MT},t} \times Pr_{\text{MT}} \tag{7}$$

$$C_{\text{BA}} = \sum_{i=1}^{24} P_{\text{BA},t} \times Pr_{\text{BA}} \tag{8}$$

$$C_{\text{load}} = \sum_{i=1}^{24} P_{\text{load},t} \times Pr_{\text{Grid},t} \tag{9}$$

Where Pr_{MT} represents the cost per unit of power generation for the micro-turbine, Pr_{BA} represents the depreciation cost per unit for the battery, and $Pr_{\text{Grid},t}$ represents the electricity price at hour t when purchasing from the grid.

(2) Objective 2: Minimize the mean squared error of grid-connected power

It is observed that the peak, off-peak, and flat rate periods are time intervals, each consisting of multiple hours, during which the grid electricity price remains constant. If the optimization objective is solely based on economic considerations, the charging and discharging powers of electric vehicles and batteries may exhibit random fluctuations within these intervals of constant electricity prices. Consequently, the resulting values of the economic objective function may remain identical across these intervals. This indicates that a single economic objective may not impose sufficient constraints on load variations. To achieve smoother load variations and mitigate fluctuations, an additional objective function is introduced, focusing on minimizing the mean squared error of grid-connected power.

$$F_2 = \sum_{i=1}^{24} \left(P_{\text{Grid},t} - \sum_{i=1}^{24} P_{\text{Grid},t} / 24 \right)^2 \tag{10}$$

Where $P_{\text{Grid},i}$ represents the microgrid-grid power exchange.

(3) Objective 3: Minimize the peak-to-valley difference of grid-connected power

The objective function for the peak-to-valley difference of grid-connected power is as follows:

$$F_3 = \max\left(P_{\text{Grid},t}\right) - \min\left(P_{\text{Grid},t}\right) \tag{11}$$

2.3. Constraints

(1) Electric vehicle charging and discharging power constraints

$$P_{\text{EVmin}} \leqslant P_{\text{EV},n,t} \leqslant P_{\text{EVmax}}\ ,\quad 0 \leqslant n \leqslant N\ ,\quad P_{\text{EV},n,t} = Z \times P_{\text{EVgap}} \tag{12}$$

(2) Micro-turbine power generation constraints

$$P_{\text{MTmin}} \leqslant P_{\text{MT},t} \leqslant P_{\text{MTmax}}\ ,\quad P_{\text{MT},t} = Z \times P_{\text{MTgap}} \tag{13}$$

(3) Battery charging/discharging power constraints

$$P_{\text{BAmin}} \leqslant P_{\text{BA},t} \leqslant P_{\text{BAmax}}\ ,\quad P_{\text{BA},t} = Z \times P_{\text{BAgap}} \tag{14}$$

Where N represents the total number of electric vehicles; P_{EVmin} represents the discharge power upper limit of the ith EV; P_{EVmax} represents the charging power upper limit; $P_{\text{EV},n,t}$ represents the load power of the EV; $P_{\text{MT},t}$ represents the power generation capacity of the micro-turbine; P_{MTmin} represents the minimum output of the micro-turbine; P_{MTmax} represents the maximum output of the micro-turbine; $P_{\text{BA},t}$ represents the power generation capacity of the battery; P_{BAmax} represents the maximum charging power of the battery; P_{BAmin} represents the maximum discharge power of the battery; and P_{EVgap}, P_{MTgap}, and P_{BAgap} represent the power levels of the EV battery, micro-turbine, and battery, respectively. Z is an integer.

(4) Electric vehicle battery capacity constraints

$$S_{\text{EV,min}} \leqslant S_{\text{EV},n,t} \leqslant S_{\text{EV,max}},\ S_{\text{EV},n,t} = S^{\text{ini}}_{\text{EV},n,t} + P_{\text{EV},n,t} \times t \tag{15}$$

(5) Electric vehicle state of charge constraints at the end of charging

$$S_{\text{EV},n\_\text{end}} = S_{\text{EV,max}},\ S_{\text{EV},n\_\text{end}} = S_{\text{EV},n,t} + P_{\text{EV},n,t} \times t \tag{16}$$

(6) Battery capacity constraints

$$S_{\text{BA,min}} \leqslant S_{\text{BA},t} \leqslant S_{\text{BA,max}},\ S_{\text{BA},t} = P_{\text{BA},t} \times t \tag{17}$$

Where n represents the index of the nth schedulable electric vehicle; $S_{\text{EV},n,t}$ represents the stored energy of the EV; $S_{\text{EV,min}}$ represents the minimum energy level of the EV; $S_{\text{EV,max}}$ represents the maximum energy level of the EV; $S_{\text{EV},n\_\text{end}}$ represents the energy level at the time of retrieval; $S_{\text{BA},t}$ represents the stored energy of the battery; $S_{\text{BA,min}}$ represents the minimum energy level of the battery; and $S_{\text{BA,max}}$ represents the maximum energy level of the battery.

(7) Equality constraints ensuring the constancy of the charging load of electric vehicle aggregators pre- and post-dispatch

$$\sum_{t=1}^{T}\sum_{n=1}^{N} P^{\text{ini}}_{\text{EV},n,t} = \sum_{t=1}^{T}\sum_{n=1}^{N} P_{\text{EV},n,t} \tag{18}$$

(8) Inequality constraints ensuring system load stays within maximum allowable limits

$$P_{\text{load},t} = \sum_{t=1}^{T} P_{dt,t} + \sum_{t=1}^{T} P_{\text{pv,t}} + \sum_{n=1}^{N}\sum_{t=1}^{T} P_{\text{EV},n,t} + \sum_{t=1}^{T} P_{\text{MT},t} + \sum_{t=1}^{T} P_{\text{BA},t} \leq P_{\text{load max}} \tag{19}$$

(9) Inequality constraints on the bounds of state of charge for electric vehicles

$$SOC_{\min} \le SOC_{\mathrm{EV},n,t} \le SOC_{\max}, SOC_{\mathrm{EV},n,t} = \frac{S_{\mathrm{EV},n,t} + P_{\mathrm{EV},n,t} \times t}{E_{\mathrm{EV}}} \tag{20}$$

(10) Inequality constraints on the bounds of depth of discharge for electric vehicle batteries

$$0 \le D_{\mathrm{DOD}} \le D_{\mathrm{DOD\_max}}, D_{DOD} = \frac{E_{EV} \cdot \left(SOC_{\mathrm{init},n} - \min_t SOC_{EV,nt}\right)}{E_{EV}} \times 100\% \tag{21}$$

Where $P_{\mathrm{EV},n,t}^{\mathrm{ini}}$ represents the charging load of electric vehicles before dispatch; $P_{\mathrm{EV},n,t}$ denotes the charging load after dispatch; $P_{\mathrm{load},t}$ refers to the total system load; $P_{\mathrm{load\,max}}$ represents the maximum allowable system load; $P_{dt,t}$ and $P_{\mathrm{pv,t}}$ denote wind power and photovoltaic power, respectively; $SOC_{\mathrm{EV},n,t}$ is the state of charge of the *n*th electric vehicle at time *t*; $SOC_{\max}$ and $SOC_{\min}$ represent the maximum and minimum allowable SOC, respectively; $E_{\mathrm{EV},n}$ indicates the electric quantity of the *n*th electric vehicle; E_{EV} is the rated electric quantity of electric vehicles; D_{DOD} represents the depth of discharge of electric vehicles; and $D_{\mathrm{DOD\_max}}$ denotes the maximum allowable depth of discharge. $SOC_{\mathrm{init},n}$ is the initial SOC of the nth EV when it reaches the charging point (grid connection). $\min_t SOC_{EV,nt}$ is the minimum SOC reached by the nth EV during its current grid connection session.

2.4. Decision variable

The decision variables of the proposed model are as follows: charging/discharging power of electric vehicles ($P_{\mathrm{EV},n,t}$), grid-connected power of the system ($P_{\mathrm{Grid},t}$), power generation of the micro gas turbine ($P_{\mathrm{MT},t}$), and charging/discharging power of the battery storage ($P_{\mathrm{BA},t}$).

3. Algorithm design

3.1. Optimize strategy structure

An optimization strategy for electric vehicle charging and discharging has been developed, which integrates flexibility devices and is depicted in Figure 2.

The system dispatch is configured with three optimization objectives: load mean square deviation, load peak-to-valley difference, and system operation cost. Heuristic methods are employed to handle system constraints, including energy balance, power generation unit output, and device operation limits. The optimization process for the system employs a multi-objective direct solution methodology, which aims to identify the Pareto optimal front. The Q-GWO algorithm is used as the solution algorithm, and its performance is compared with commonly used optimization algorithms in the dispatch field, including GWO, MOPSO, and MODE. In the study, orthogonal experiments are utilized to optimize algorithm parameters, and the performance of the algorithm is evaluated using metrics such as IGD, HV, and MS to validate its effectiveness. Finally, the TOPSIS method, based on ideal points, is applied to derive the best dispatch solution.

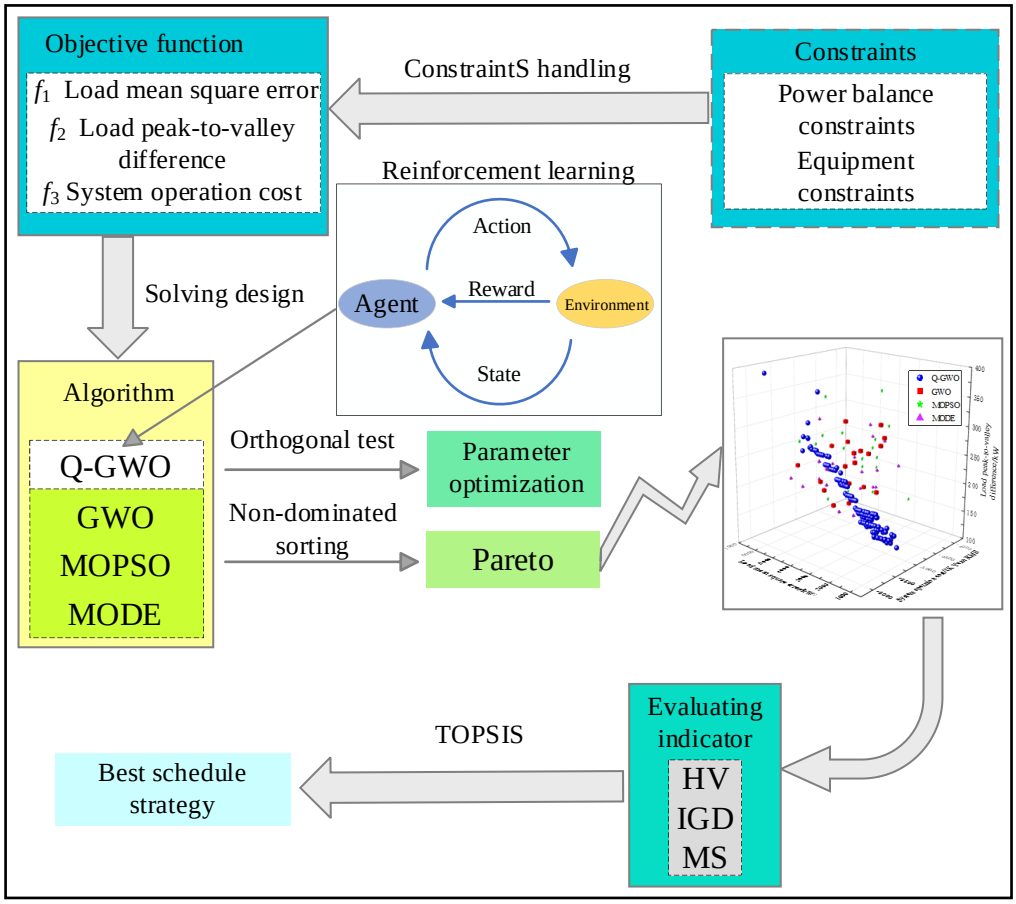

Figure 2. V2G optimal scheduling strategy diagram.

3.2. Solution algorithm design

In designing algorithms to solve V2G dispatch systems, the encoding issue is indeed crucial for constructing effective solutions [24]. Each individual in the Q-GWO algorithm is encoded as a T-dimensional vector, where T represents the number of time intervals within the dispatch period. The dimension of this vector is $2m + 2$, where m denotes the number of electric vehicles participating in the dispatch. The vector contains output plan information for different time intervals within the dispatch period, specifically electric vehicle charging power, electric vehicle discharging power, micro-turbine operating power, and battery storage operating power. This structure allows the algorithm to precisely control and adjust power output at each time point, ensuring dynamic balance between the supply side and load side. The dispatch of the V2G system aims to regulate the output of each component on the supply side, ensuring it remains balanced with the load at all times. Therefore, in designing the Q-GWO algorithm to solve the dispatch model, each individual is represented as a T-dimensional vector to store the output plan within a dispatch period. Each vector is composed in the order of electric vehicle charging power, electric vehicle discharging power, micro-turbine operating power, and battery storage operating power, with a vector dimension of $2m + 2$. The specific representation is shown in Table 2.

Table 2. Representation of the solution.

EV charging power			EV discharging power			Micro gas turbine output	Battery output
$P_{e,11}^{c}$	...	$P_{e,1m}^{c}$	$P_{e,1(m+1)}^{d}$	⋯	$P_{e,1(2m)}^{d}$	$P_{e,1}^{m}$	$P_{e,1}^{b}$
$P_{e,21}^{c}$	...	$P_{e,2m}^{c}$	$P_{e,2(m+1)}^{d}$	⋯	$P_{e,2(2m)}^{d}$	$P_{e,2}^{m}$	$P_{e,2}^{b}$
...	...	⋯	⋯	⋯	⋯	⋯	⋯
$P_{e,T1}^{c}$	...	$P_{e,Tm}^{c}$	$P_{e,T(m+1)}^{d}$	⋯	$P_{e,T(2m)}^{d}$	$P_{e,T}^{m}$	$P_{e,T}^{b}$

The traditional multi-objective grey wolf algorithm exhibits a fixed implicit linear relationship between the convergence factor '*a*' and the iteration process, which imposes certain limitations in practical applications. While the algorithm is theoretically designed to set an initial value close to 2 to enhance global exploration capability and gradually reduce it to near 0 for faster convergence, actual operational results demonstrate relatively weak global optimization performance and slow convergence speed. This leads to suboptimal boundary precision and distribution sparsity of the Pareto front.

Additionally, the search factor '*w*' in traditional grey wolf algorithms employs a linear decrement strategy, where *w* decreases linearly from 2 to 0 as the optimization process progresses. This single adjustment method may fail to adapt to complex and dynamic optimization requirements in practical applications. Furthermore, the complexity and variability of problems may render the simple linear decrement adjustment insufficiently flexible.

Through Q-learning, the search factor '*w*' typically adopts a linear decrement strategy, whereas the improved algorithm performs adaptive adjustments based on environmental feedback to more precisely adapt to the current state of the search space, thereby enhancing flexibility at different stages. In the Q-learning algorithm, the Q-table is updated through state space, action space, selection mechanisms, and reward mechanisms. Therefore, Figure 3 showcases the flowchart of a Q-learning algorithm specifically designed to optimize electric vehicle charging and discharging dispatch, illustrating an innovative solution for dispatch optimization.

(1) State space

Based on the relationship between the diversity coefficient and convergence coefficient of the population, it can be divided into four states: evolution, decline, deepening, and exploration, as shown in Table 3. By dynamically adjusting the different states within the state space, the algorithm showcases a strong ability to effectively balance exploring new solution spaces and utilizing existing high-quality solutions, thereby enhancing the overall performance of the algorithm. The diversity coefficient of the population is defined as follows: a larger diversity coefficient indicates better population diversity, providing a broader search space and facilitating the exploration of more potential solutions. The convergence coefficient of the population is defined as follows: a larger convergence coefficient indicates higher similarity among individuals, which aids the population in quickly converging to the optimal solution.

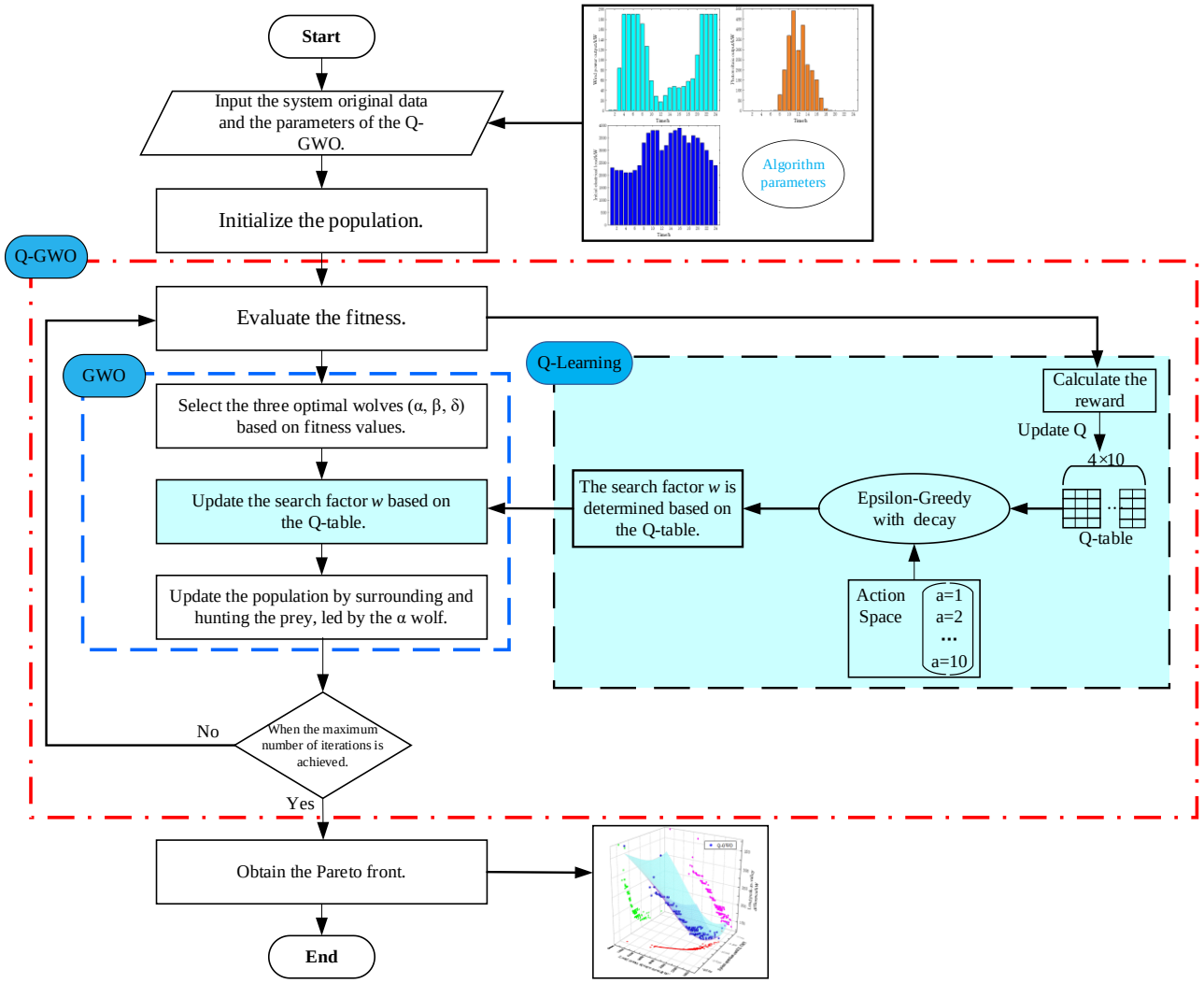

Figure 3. V2G optimal scheduling solution process.

$$DC = \frac{f_{avg} - f_{\min}}{f_{\max} - f_{\min}} \tag{22}$$

$$CC = \frac{1}{\sqrt{\frac{\sum_{i=1}^{M}(f_i - f_{avg})^2}{M-1}}} \tag{23}$$

Where $f_{avg}, f_{\max}, f_{\min}, f_i$ represent the average population fitness, the maximum fitness value, the minimum fitness value, and the fitness value of the *i*th particle, respectively.

Table 3. State space.

State	Diversity coefficient	Convergence coefficient
1 Evolutionary	$DC(gen+1) \geq DC(gen)$	$CC(gen+1) \geq CC(gen)$
2 Exploration	$DC(gen+1) \geq DC(gen)$	$CC(gen+1) < CC(gen)$
3 Intensification	$DC(gen+1) < DC(gen)$	$CC(gen+1) \geq CC(gen)$
4 Decay	$DC(gen+1) < DC(gen)$	$CC(gen+1) < CC(gen)$

(2) Action space

In the iterative process of the algorithm, agents dynamically adjust the search factor 'w' by selecting different actions, thereby enabling adaptive balancing between global exploration and local exploitation. This mechanism not only helps overcome issues related to local optima but also significantly enhances the algorithm's optimization efficiency. The search factor 'w' is defined within the range of 0–2, divided into 10 subintervals with a step size of 0.2. In practical applications, based on a predefined action selection mechanism, specific values are randomly sampled within each subinterval, as illustrated in Figure 4.

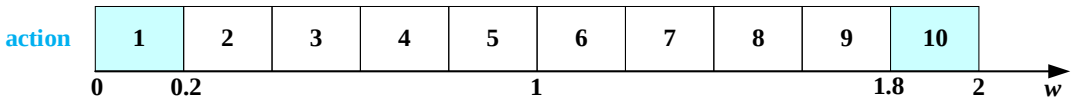

Figure 4. Action space.

(3) Action selection strategy

By dynamically adjusting the exploration rate, the agent can determine the action selection strategy based on the epsilon-greedy selection mechanism, thus achieving a dynamic balance between exploration and exploitation in the learning process. This enhances the algorithm's learning efficiency and global optimization capability. During the initial phase, the agent focuses more on exploring the environment, while in the later phase, it prioritizes utilizing existing knowledge. The dynamic adjustment of the exploration rate follows the rules outlined below.

$$if\ \varepsilon > \varepsilon_{\min}, then\ \varepsilon = \varepsilon \times \varepsilon_d \tag{24}$$

Where $\varepsilon, \varepsilon_{\min}, \varepsilon_d$ represent the exploration rate, minimum exploration rate, and exploration decay factor, respectively.

(4) Reward mechanism

In the Q-GWO algorithm, a reward mechanism is implemented to provide immediate feedback to agents, guiding them to select appropriate search factors '*w*' during each generation's evolutionary process, thereby enhancing the population's fitness performance. The reward mechanism is categorized into optimal fitness rewards and average fitness rewards. If the fitness of the new generation improves, agents are rewarded, encouraging them to continue selecting similar operations. Conversely, if fitness does not improve, corresponding operations are punished, reducing the likelihood of agents choosing such operations in subsequent iterations.

$$R_{gen}^1 = \begin{cases} +1.5, & f_{\min}(gen+1) < f_{\min}(gen) \\ -0.5, & f_{\min}(gen+1) = f_{\min}(gen) \\ -1.0, & f_{\min}(gen+1) > f_{\min}(gen) \end{cases} \tag{25}$$

$$R_{gen}^2 = \begin{cases} +1.5, & f_{avg}(gen+1) < f_{avg}(gen) \\ -0.5, & f_{avg}(gen+1) = f_{avg}(gen) \\ -1.0, & f_{avg}(gen+1) > f_{avg}(gen) \end{cases} \tag{26}$$

$$R_{gen} = R_{gen}^1 + R_{gen}^2 \tag{27}$$

(5) Q-Table update mechanism

The Q-table is updated using the following equation:

$$Q_{gen+1}\left(s_{gen},a_{gen}\right)\leftarrow Q_{gen}\left(s_{gen},a_{gen}\right)+LR\cdot\left[R_{gen}+DF\cdot\max_{a}Q_{gen}\left(s_{gen+1},a\right)-Q_{gen}\left(s_{gen},a_{gen}\right)\right] \quad (28)$$

Where LR represents the learning rate, DF represents the discount factor, s represents the state space, a represents the action space, and $Q(s, a)$ represents the Q-table.

3.3. Comprehensive decision-making

The Pareto frontier obtained through multi-objective algorithms is a set of solutions. To make the results practically valuable, comprehensive decision-making is required to select the optimal solution from the solution set [25]. The multi-objective optimization problem of V2G is also a multi-attribute comprehensive decision-making problem. To address the complexities of multi-objective decision-making, the TOPSIS method offers a systematic approach to finding the best compromise in scheduling. This method involves the decision maker setting an ideal solution, which may not always be achievable. By employing TOPSIS, decision-makers can effectively navigate the intricacies of multi-criteria optimization, ensuring a balanced and practical solution. It involves establishing a distance measure to quantify how closely each solution in the solution set aligns with this ideal solution. The decision-making process involves identifying a set of viable options that most closely align with the ideal solution or, conversely, deviate the least from the undesired outcome, ultimately serving as the final course of action.

4. Simulation analysis

4.1. System initial data

The EVs are modeled with a capacity range of 18–60 kWh, a maximum charging/discharging power of 10 kW, and a charging/discharging efficiency of 0.9. The operational cost per unit of charging/discharging is set at 0.3 CNY/kWh. For natural gas, the heat value is 10.1 kWh/m^3, with a unit price of 2.9 CNY/m^3. The natural gas unit is characterized by a start-up power of 100 kW, a maximum output power of 200 kW, and a power adjustment step of 20 kW. The generation efficiency of the natural gas unit is 0.274, with a unit generation cost of 0.79 CNY/kWh. Similarly, the unit generation costs for wind turbines and photovoltaic systems are set at 0.1 CNY/kWh and 0.15 CNY/kWh, respectively. The detailed parameters of the electric vehicles participating in the dispatch are listed in Table 4. The wind power output, photovoltaic output, and initial load data are provided in Figures 5, 6, and 7, respectively. Additionally, the electric vehicle travel data and time-of-use electricity prices are illustrated in Figures 8 and 9.

Table 4. Conception and scope of logistics.

Parameter	Numerical value
Quantity of EV	500
Rated charging power	12 kW
Rated discharge power	12 kW

EV battery capacity	60 kWh
Expected SOC of the user	0.8
Maximum discharge depth	0.5
Power consumption per kilometer	0.15
Battery charge efficiency	90%
Battery discharge efficiency	90%

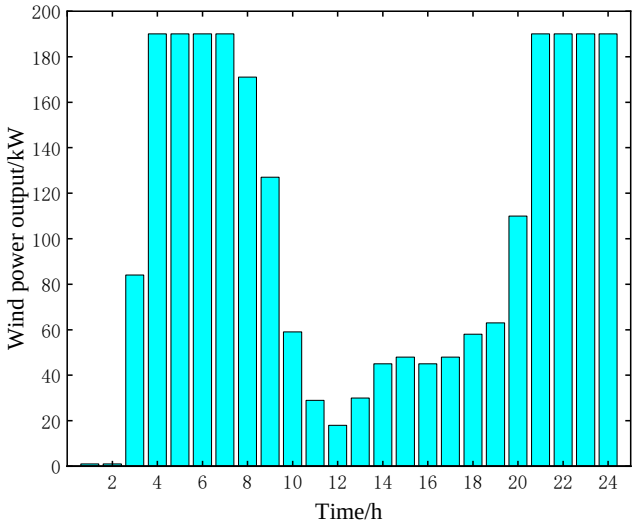

Figure 5. Wind power output.

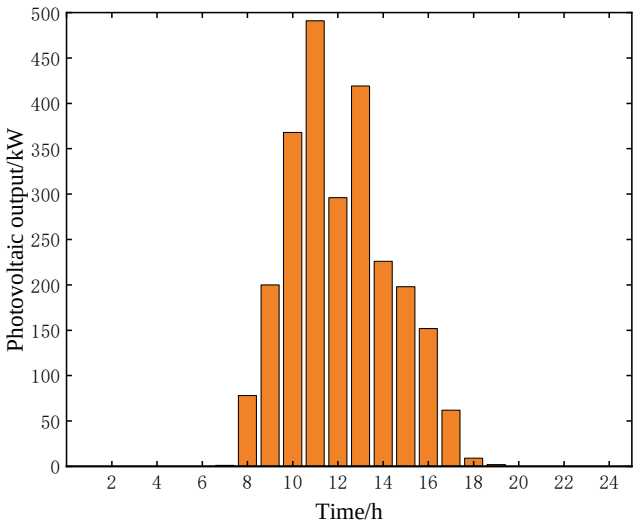

Figure 6. Photovoltaic output.

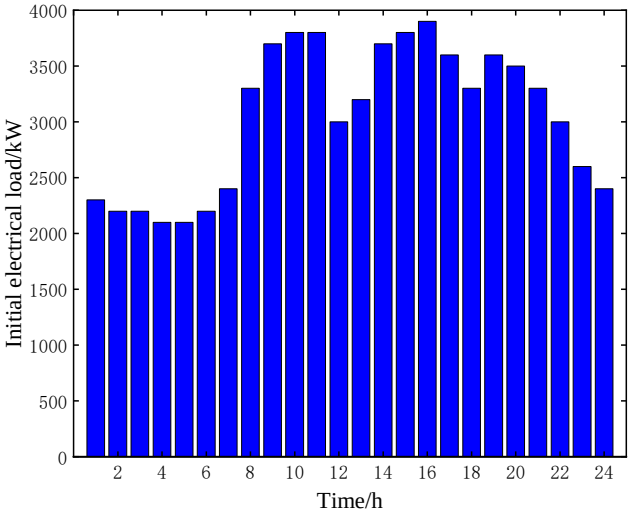

Figure 7. Initial electrical load.

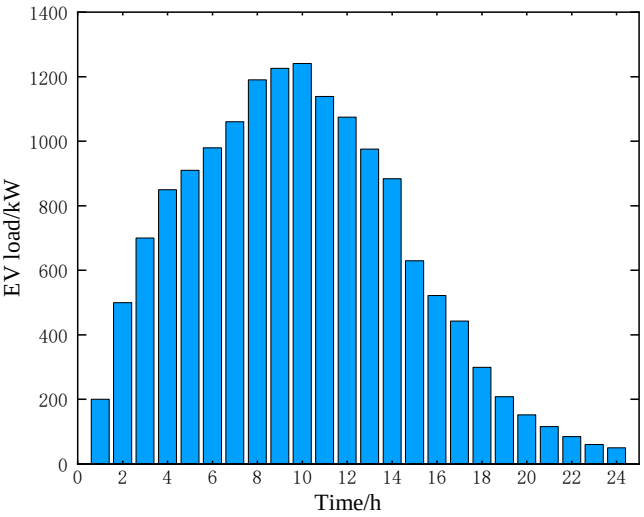

Figure 8. Initial load of electric vehicle.

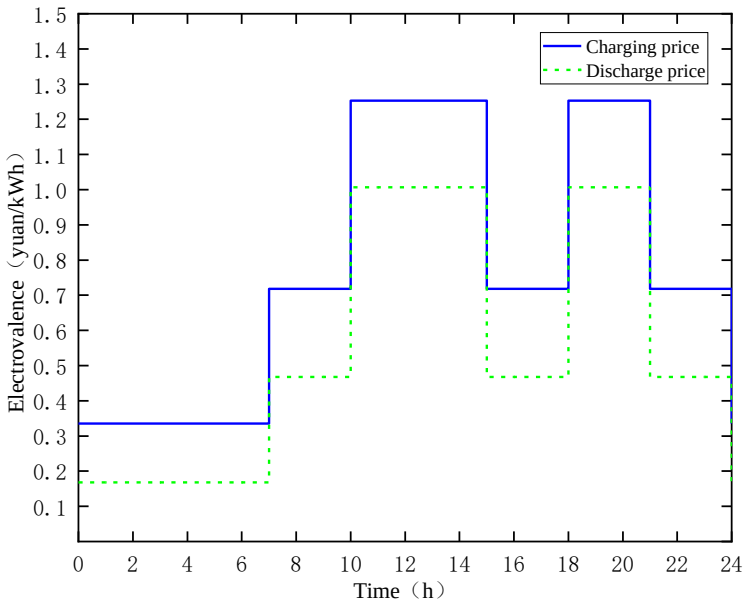

Figure 9. Time-of-use price.

4.2. Simulation environment

All experiments were conducted on a 64-bit Windows 10 system equipped with a 5th-generation Intel(R) Core (TM) i5-8265U processor (1.60 GHz) and 8 GB of memory. Simulations were performed using MATLAB R2019b.

4.3. Algorithm parameter setting and optimization

The Q-GWO algorithm is characterized by several key parameters: population size (N), external archive capacity (P), number of iterations (T), and learning rate (LR). Each parameter is assigned to four distinct levels, and the orthogonal experimental method is employed to assess their impact across different levels. Using the L9(34) orthogonal design framework, the experimental setup is detailed in Table 5. Three response variables are considered: system operation cost, load peak-to-valley difference, and load mean square difference. The total normalization value method is used to compute the total normalized OD value for the parameters. All RV values are normalized within the interval [0,1] to ensure consistency. For parameter optimization, the geometric mean of the OD values serves as the evaluation metric. It is important to note that a higher OD value reflects enhanced algorithmic performance. The corresponding calculation formulas are provided below:

$$d_i = \frac{F_i^{\max} - F_i}{F_i^{\max} - F_i^{\min}} \tag{29}$$

$$OD = \sqrt[l]{(d_1 d_2 \cdots d_l)} \tag{30}$$

Table 5. Q-GWO algorithm parameter level table.

Parameter	Level		
	1	**2**	**3**
N	50	100	150
P	30	40	50
T	300	400	500
LR	0.25	0.5	0.75

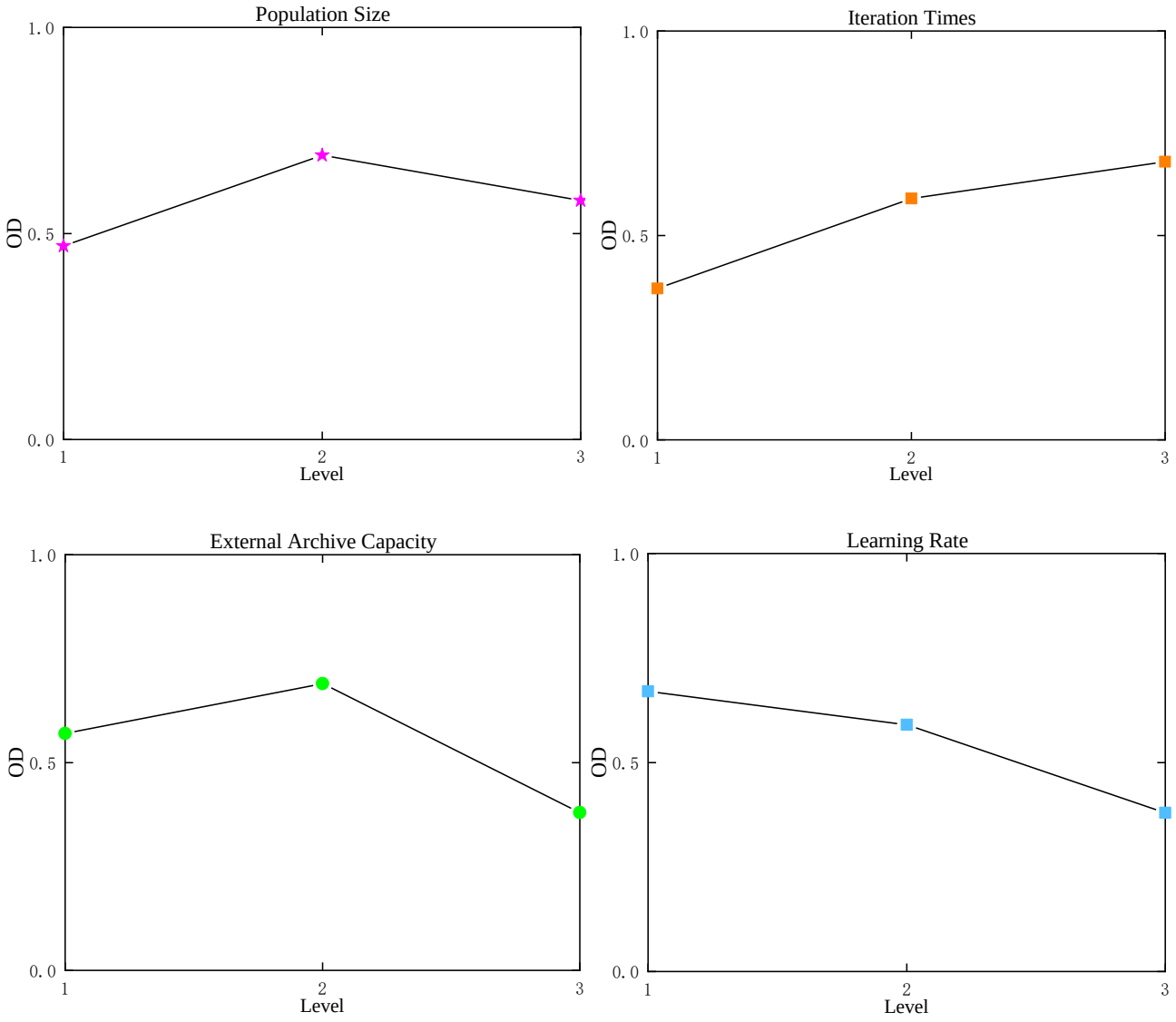

Figure 10. Q-GWO factor level trend chart.

The trend of parameter level changes for the Q-GWO algorithm is shown in Figure 10. It can be observed that the Q-GWO algorithm achieves the best performance when the parameters are set as follows: population size (N = 100), external archive capacity (P = 40), iteration times (T = 500), and learning rate (LR = 0.25). Additionally, the parameters of the GWO, MOPSO, and MODE algorithms will be set to the same values as those used in the Q-GWO algorithm for a comparative analysis of their performance and solution quality.

The parameter level table and the results of the orthogonal test are presented in Tables 5 and 6, respectively. A total of four parameters were set, each with three levels. Orthogonal tests were conducted to confirm which level of each parameter has a more significant impact on the simulation results. A total of nine experiments were performed, and the results of each experiment were recorded in Table 6.

Table 6. Q-GWO orthogonal test results table.

Experiment number	Factor				*OD*
	N	***P***	***T***	***LR***	
1	1	1	1	1	0.4671
2	1	2	2	2	0.7360
3	1	3	3	3	0.5610
4	2	1	2	3	0.4431
5	2	2	3	1	0.7980
6	2	3	1	2	0.6345
7	3	1	3	2	0.6569
8	3	2	1	3	0.6030
9	3	3	2	1	0.6772

4.4. Simulation results analysis

To evaluate the proposed optimization dispatch strategy, a standard daily load profile of a regional microgrid was selected. The strategy was implemented with a 24-h cycle and 1-h intervals to ensure thorough assessment. After confirming the initial data of the dispatch system and the parameter settings of the dispatch algorithm, four algorithms—Q-GWO, GWO, MOPSO, MODE, and NSGA-II—were employed to solve the model. The results of the four algorithms are presented in Figure 11, while the Pareto front obtained using the Q-GWO algorithm is shown in Figure 12.

It can be observed from the Pareto results that the three key indicators of operation cost, grid-connected power mean square difference, and peak-to-valley difference exhibit a belt-like distribution in the three-dimensional space. This distribution reflects the convergence characteristics of the model solutions in the multi-objective optimization space. During the iteration process, as the number of algorithm operations increases, the optimal solutions gradually approach a specific region and eventually stabilize. The dense distribution of boundary points indicates the algorithm's precise adjustment during exploration, ensuring the accurate depiction of the Pareto front. Furthermore, the uniformity of the overall point distribution further verifies the algorithm's effectiveness, demonstrating that the solutions fully cover the entire Pareto front interval without obvious gaps or clusters. This implies that the particles in the repository (i.e., potential solutions) are almost distributed across all parts of the final Pareto front, ensuring the diversity and representativeness of the solutions.

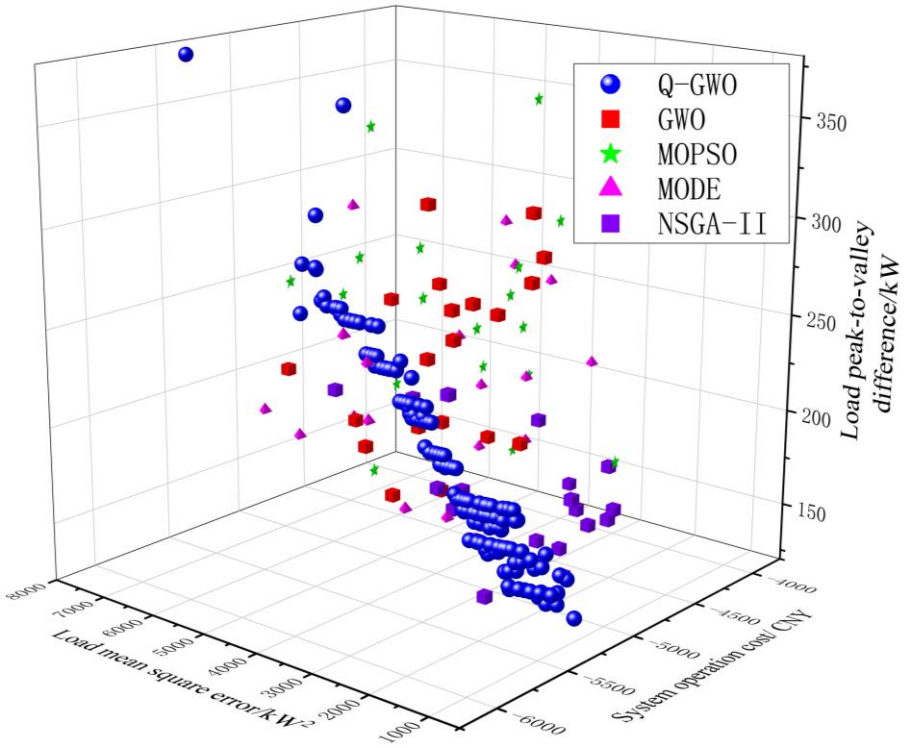

Figure 11. Pareto front of four algorithms.

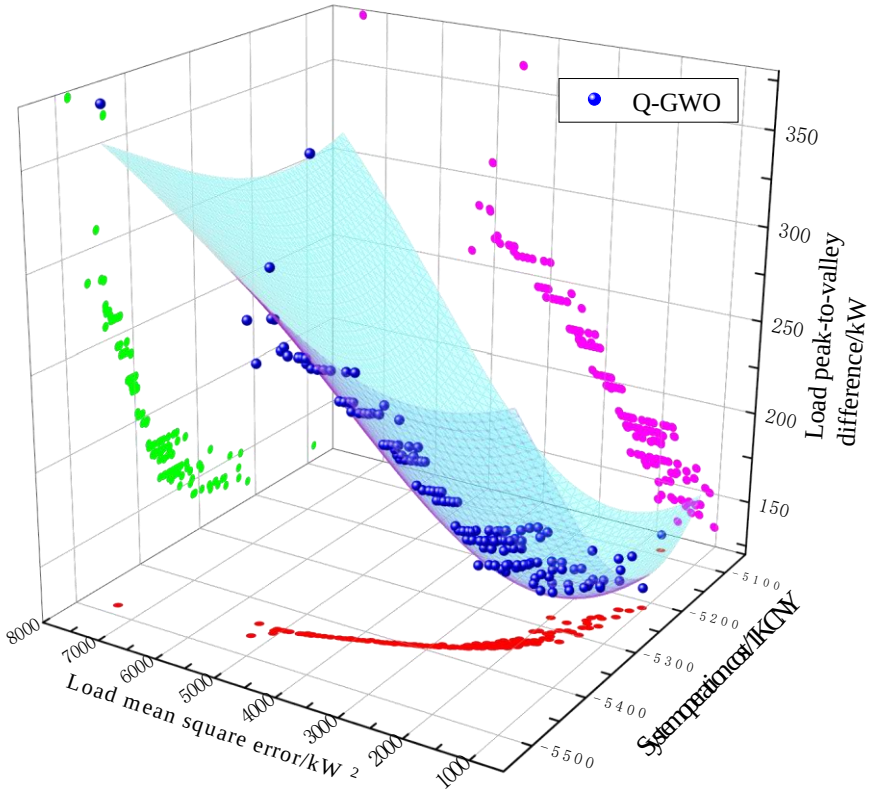

Figure 12. Pareto frontier surface and three-dimensional projection solved by Q-GWO.

Figures 13–15 illustrate the convergence curves of the objective functions in the system. Through comparative analysis, it is evident that the Q-GWO algorithm exhibits superior performance during the optimization process. Specifically, the Q-GWO algorithm demonstrates outstanding performance across three optimization objectives: operation cost, load mean square deviation, and load peak-to-valley difference. It not only achieves faster convergence but also attains higher convergence precision in the final results. Compared to other comparative algorithms, the optimization effects of the Q-GWO algorithm are significantly more pronounced, which fully demonstrates its notable superiority in handling the proposed model.

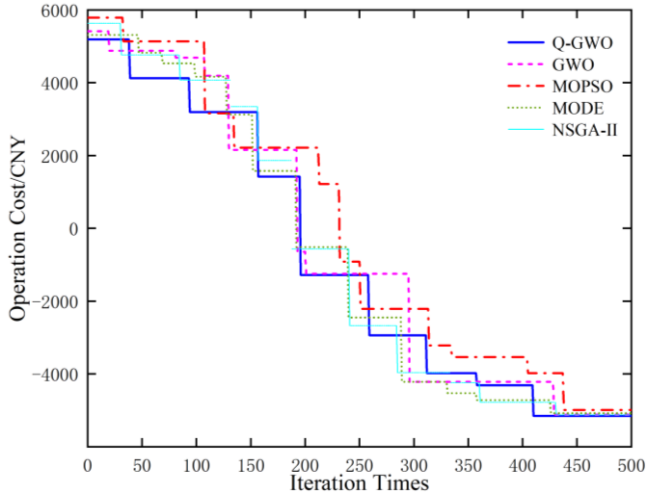

Figure 13. Operation cost convergence curves.

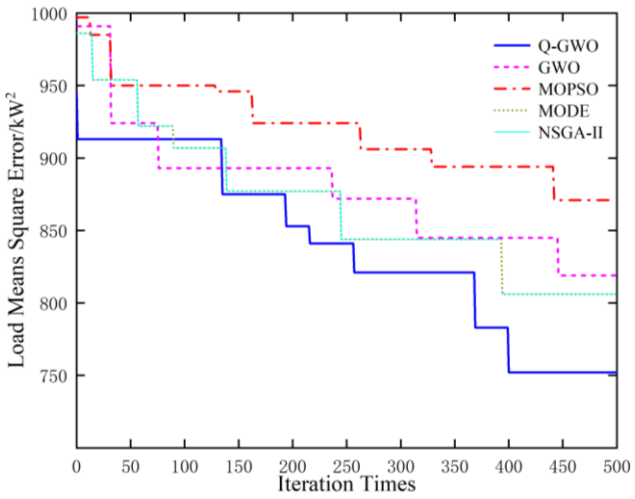

Figure 14. Load mean square error.

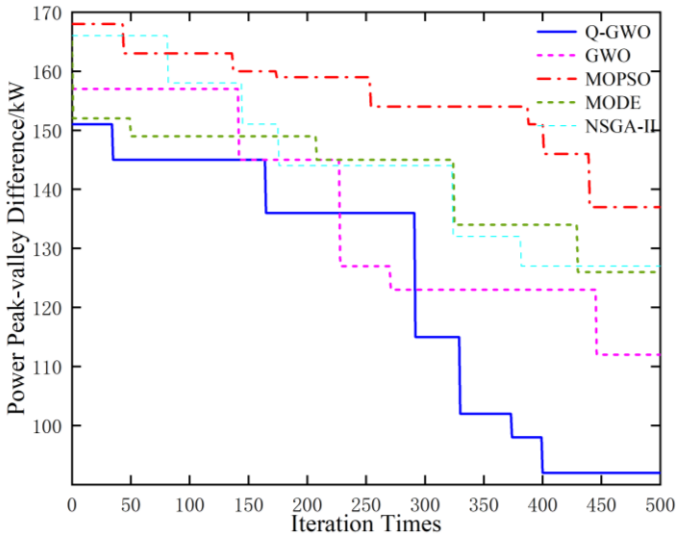

Figure 15. Power peak-valley difference convergence curve of the four algorithms.

To evaluate the GWO algorithm performance integrated with reinforcement learning for local optimization, five test functions with multiple local optima were selected for analysis. The details of these functions are listed in Table 7. It is evident that the QGWO algorithm performs exceptionally well for local optimization tasks.

Table 7. Benchmark functions with multiple local optimal solutions.

Function name	Function equation	Dimensions	Range	Theory minimum
Rastrigin	$f(x) = 20 + sum\left(x^2 - 10\cos(2\pi x)\right)$	2	[-5, 5]	0
Dropwave	$f(x) = -\frac{1+\cos\left(12\sqrt{x_1^2+x_2^2}\right)}{2\left(x_1^2+x_2^2\right)+1}$	2	[-5, 5]	-1
Griewank	$f(x) = \frac{1}{4000}\left(x_1^2+x_2^2\right) - \cos\left(\sqrt{x_1}\right)\cos\left(\frac{\sqrt{x_1}}{\sqrt{2}}\right)+1$	2	[-5, 5]	0
Labyrinth	$f(x) = \sin(x_1)\cos(x_2) + \sin(x_2)\cos(x_1) + \sin(x_1+x_2)\cos(x_1-x_2)+1$	2	[-5, 5]	-1
Rosenbrbrock	$f(x) = (1 - x_1)^2 + 100\left(x_2 - x_1^2\right)^2$	2	[-5, 5]	0

Training a Q-table is central to the QGWO algorithm, but its time complexity increases sharply with problem dimensionality, leading to significant computational overhead. This was evaluated using three benchmark functions from each category: single-mode, multi-modal, and composite-mode, as listed in Table 8. Calculational times for QGWO, GWO, MOPSO, and MODE were measured and compared and are shown in Table 9.

Table 8. Benchmark functions.

ID	Function equation	Dimensions	Range	Theory minimum
	Single-mode functions			
1	$f(x) = \sum_{i=1}^{N} x_i^2$	30	[-5, 5]	0
2	$f(x) = \sum_{i=1}^{N} \lvert x_i \rvert + \prod_{i=1}^{N} \lvert x_i \rvert$	30	[-5, 5]	0
3	$f(x) = \sum_{i=1}^{N} \left[\sum_{i=1}^{N} x_j \right]^2$	30	[-5, 5]	0
	Multi-modal functions			
4	$f(x) = \sum_{i=1}^{N} x_i \sin\left(\sqrt{\lvert x_i \rvert}\right)$	30	[-5, 5]	0
5	$f(x) = \sum_{i=1}^{N} \left[x_i^2 - 10\cos(2\pi x_i) + 10 \right]$	30	[-5, 5]	0
6	$f(x) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{N} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{N}\cos(2\pi x_i)\right) + 20$	30	[-5, 5]	0
	Composite modal functions			
7	$f(x) = 20 + sum\left(x^2 - 10\cos(2\pi x)\right)$	30	[-5, 5]	0

8	$f(x)=\frac{1}{4000}\left(x_1^2+x_2^2\right)-\cos\left(\sqrt{x_1}\right)\cos\left(\frac{\sqrt{x_1}}{\sqrt{2}}\right)+1$	30	[-5, 5]	0
9	$f(x)=(1-x_1)^2+100\left(x_2-x_1^2\right)^2$	30	[-5, 5]	0

Table 9. Benchmark function calculation time comparison.

	Algorithm	1	2	3	4	5	6	7	8	9
Run time /s	GWO	0.85521	0.84212	1.65244	0.95511	0.91221	0.96332	0.91664	0.84553	0.77123
	QGWO	1.01933	0.92665	1.72223	1.01887	0.95772	1.08423	1.03751	0.94882	0.82117
	MOPSO	0.97657	0.91327	1.95766	0.96321	0.96573	0.91231	0.96574	0.89456	0.84574
	MODE	0.95364	0.90345	1.83258	0.95634	0.92133	0.92345	0.93234	0.90356	0.81232

After obtaining the Pareto solution set, a decision-making process is required. Table 10 presents the decision results obtained by four algorithms using a multi-objective decision-making method based on the ideal solution. It can be observed that the Q-GWO algorithm demonstrates superior solution results compared with the other algorithms. Combined with the previously provided convergence curves, this sufficiently proves that the improved Q-GWO algorithm exhibits more outstanding optimization effects in the multi-objective optimization dispatch problem of V2G.

Table 10. Decision results of four algorithms.

Algorithms	**System operation cost/CNY**	**Load means square error /kW^2**	**Load peak-to-valley difference /kW**
Q-GWO	-5152	7.52×10^5	92
GWO	-5119	8.19×10^5	112
MOPSO	-4987	8.71×10^5	137
MODE	-5075	8.58×10^5	126

This paper also presents the dispatch results of the micro-turbine output, battery output, electric vehicle output, and grid-connected power obtained by the Q-GWO algorithm and three comparative algorithms for the multi-objective optimization dispatch problem of V2G. These results are illustrated in Figures 16–19. The results demonstrate that by adjusting the output, the grid-connected power of the system can be maintained within a certain range with slight fluctuations, thereby enhancing the system's power supply stability. In a disordered charging scenario, electric vehicles exhibit random charging and discharging patterns, causing a substantial rise in grid peak load. Conversely, implementing an ordered charging and discharging strategy enables electric vehicles to charge during low-demand periods and discharge during high-demand periods, thereby reducing grid load pressure. This approach not only meets the high-power needs of users during peak times but also considerably reduces load peaks, ensuring the stability of the power system.

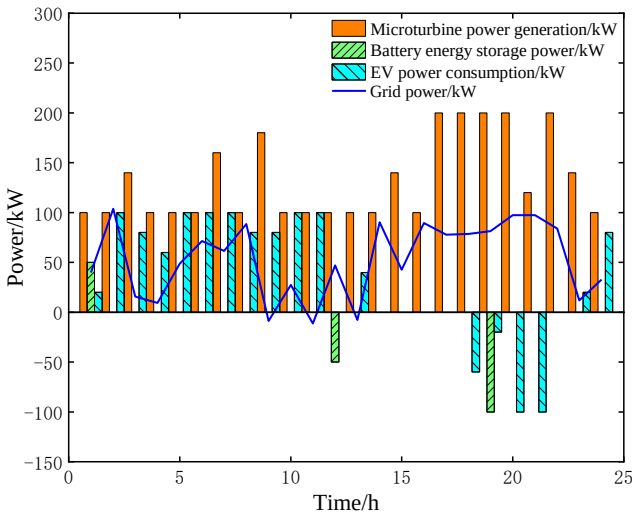

Figure 16. Results solved by Q-GWO.

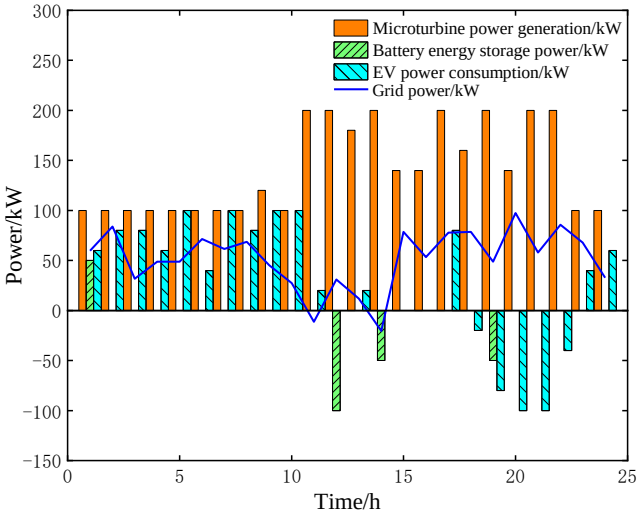

Figure 17. Results solved by GWO.

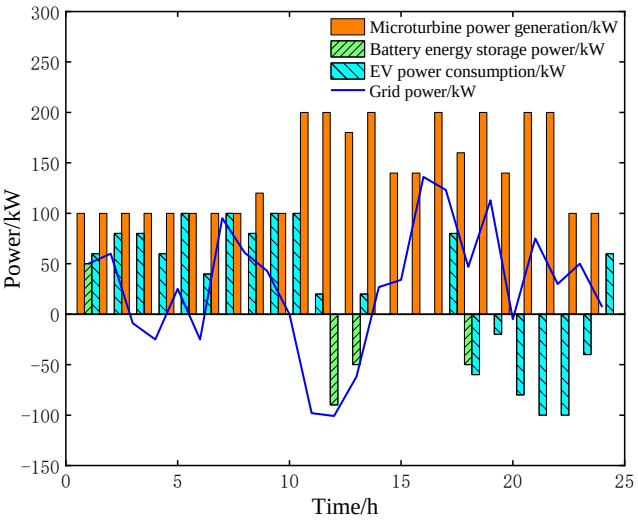

Figure 18. Results solved by MOPSO.

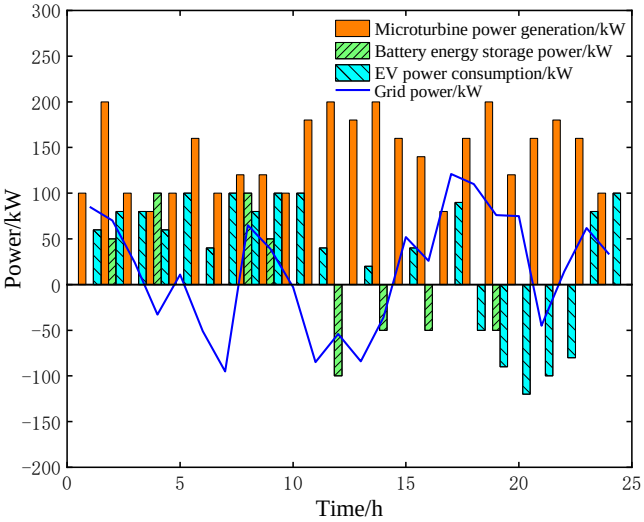

Figure 19. Results solved by MODE.

4.5. Algorithm analysis

Once intelligent algorithms are employed to solve the model, evaluating their performance becomes essential. In the realm of multi-objective optimization, solution quality is typically assessed through measures of convergence and diversity. Convergence indicates how closely the obtained solutions approach the Pareto front, while diversity reflects how thoroughly the solution space is explored and how densely the solutions are distributed along the Pareto front. Commonly used metrics for evaluation include the hypervolume metric, inverse generational distance, and maximum spread.

To compute the inverse generational distance and maximum spread, it is necessary to have prior knowledge of the true Pareto front for the optimization problem. However, in practical scenarios, the true Pareto front is often unknown. Consequently, an alternative approach can be employed to

estimate the true Pareto front. This involves running all algorithms involved in solving the dispatch problem independently 10 times and recording the Pareto-optimal solutions from each run. These solutions are then merged into a new solution set. Subsequently, non-dominated sorting based on crowding distance is performed on this new set. Finally, the non-dominated solutions are used to approximate the true Pareto front.

The assessment outcomes presented in Table 11 highlight the superior performance of the Q-GWO algorithm. In terms of the IGD metric, Q-GWO achieved a value of 2.692×10^{-3}, significantly outperforming the other three algorithms. Furthermore, the hypervolume analysis demonstrates that Q-GWO exhibited the most exceptional performance, with an average value surpassing those of the competing methods. Based on the metric of maximum spread, the Pareto front obtained through the Q-GWO algorithm demonstrates a more extensive distribution, which highlights a notable advantage of the Q-GWO algorithm in terms of both distribution and convergence of the solution set.

Table 11. Evaluation results.

Algorithm	IGD	HV	MS
Q-GWO	2.692×10^{-3}	0.409×10^{-2}	0.9473
GWO	2.943×10^{-3}	0.384×10^{-2}	0.8869
MOPSO	3.425×10^{-3}	0.368×10^{-2}	0.8017
MODE	3.347×10^{-3}	0.371×10^{-2}	0.8329

To further validate the performance of the Q-GWO algorithm in multi-objective optimization dispatch, this study employs a comparative experimental approach. Specifically, the four algorithms mentioned above are used to independently solve the multi-objective dispatch model for 10 trials. The performance of each algorithm is scientifically evaluated using the relative percentage increase (RPI) metric. Additionally, a multi-factor analysis of variance (ANOVA) is conducted on the optimization objectives with a 95% confidence level. The comparison results are shown in Figures 20–22. As can be observed from these figures, the Q-GWO algorithm demonstrates superior optimization performance compared to the other algorithms.

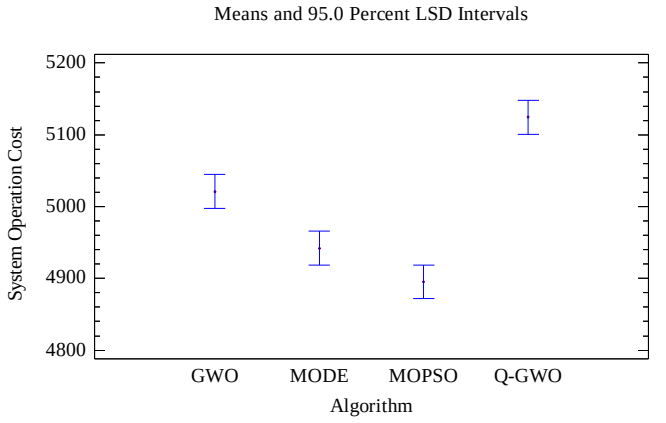

Figure 20. Comparison of operation cost.

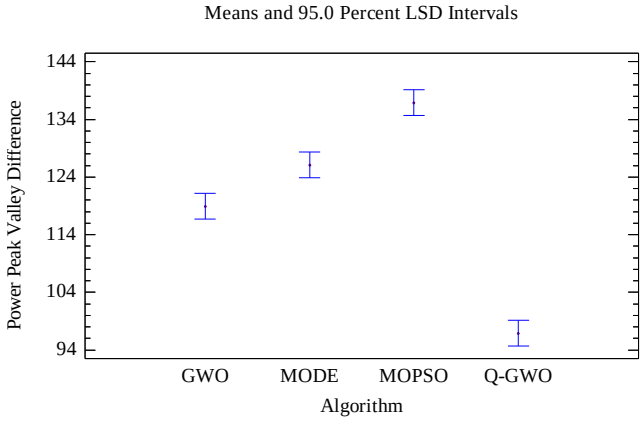

Figure 21. Comparison of power peak-valley difference.

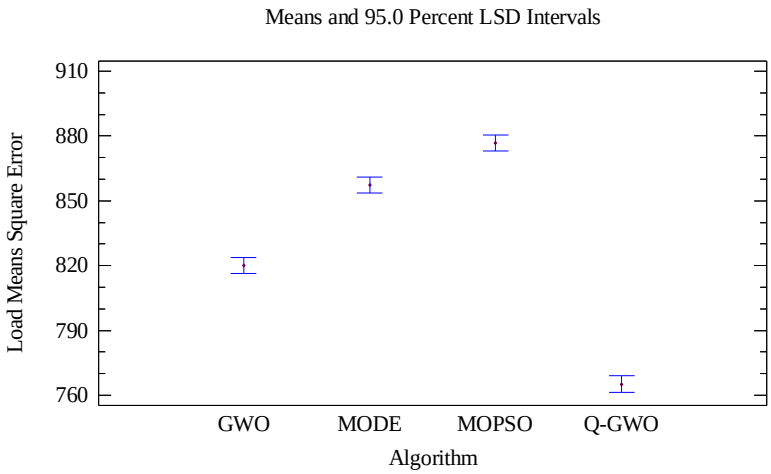

Figure 22. Comparison of different algorithms in load means square error.

5. Discussion

The primary advantage of the proposed Q-learning-enhanced GWO method lies in its adaptive search capability. Q-learning dynamically guides the GWO's exploration and exploitation phases, allowing it to effectively adjust to varying demands throughout the optimization process, leading to improved convergence and solution quality for this complex, constrained problem. Furthermore, the systematic parameter tuning via orthogonal experiments enhances robustness. However, a potential disadvantage is the increased computational complexity compared to the standard GWO or simpler heuristics, due to the overhead of the Q-learning mechanism. While effective, the parameter tuning process itself also requires additional computational efforts.

6. Conclusions

Through rigorous case studies, the proposed method demonstrated significant effectiveness. Key findings are as follows:

(1) A comprehensive multi-objective dispatch model is developed, with operational costs, load mean square deviation, and load peak-to-valley difference serving as primary evaluation criteria. The framework integrates the impact of time-of-use pricing from a demand-side perspective, thereby enhancing the role of V2G technology in advancing renewable energy integration.

(2) To address the optimization requirements of the EV dispatch system, an enhanced GWO algorithm integrated with Q-learning has been developed to achieve the Pareto front of the proposed dispatch model directly. This advanced algorithm leverages Q-learning to enhance its search capabilities and adapt to the varying demands of different optimization stages. To optimize the algorithm's performance, orthogonal experiments are utilized to determine the most suitable parameters. Following this, the TOPSIS method is employed to identify the optimal compromise solution, offering valuable insights for dispatch decision-making.

(3) The proposed multi-objective optimization dispatch method was thoroughly assessed via specific case studies, ensuring robust testing and validation. Experimental results demonstrate the effectiveness of the dispatch strategy, showing significant reductions in system operational costs and

effective mitigation of grid load fluctuations. These outcomes ensure both economic operation and enhanced stability of the power system.

Notably, this enhanced performance comes with marginally increased computational complexity, a trade-off for higher solution precision. Looking ahead, future work will explore scalability testing under higher EV penetration with stochastic renewable generation, integration of deep reinforcement learning to optimize computational efficiency, and real-world validation through hardware-in-the-loop simulations incorporating battery degradation models. These extensions will strengthen the strategy's practical applicability for next-generation smart grids. Collectively, this research provides a robust framework for EV-grid synergy, balancing economic and stability objectives through adaptive intelligence while charting actionable pathways for sustainable energy integration.

Author contributions

Conceptualization, X.P.; methodology, Z.Z. and X.P.; software, Z.Y.; validation, Z.Z.; formal analysis, Z.Z.; investigation, Z.Z. and X.P.; resources, X.P.; data curation, Z.Z.; writing—original draft, Z.Z. and X.P.; writing—review and editing, Y.L.; visualization, Z.Z. and T.G.; supervision, Y.L.; project administration, Y.L.; funding acquisition, Z.Z. and X.P. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

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

Acknowledgments

This work was supported by the Natural Science Foundation of Liaoning Province of China (2023JH2/101700068, 2024-MS-217), Department of Education of Liaoning Province of China (LJ222411632035, LJ212411632075), and the Shenyang Science and Technology Plan Project (24-213-3-29).

Conflicts of interest

The authors declare that they have no conflicts of interest in this paper.

References

1. Shaheen HI, Rashed GI, Yang B, Yang J (2024) Optimal electric vehicle charging and discharging scheduling using metaheuristic algorithms: V2G approach for cost reduction and grid support. *J Energy Storage* 90: 111816. https://doi.org/10.1016/j.est.2024.111816
2. Yin W, Jia L, Ji J (2024) Energy optimal scheduling strategy considering V2G characteristics of electric vehicle. *Energy* 294: 130967. https://doi.org/10.1016/j.energy.2024.130967

3. Huang Z, Guo Z, Ma P, Wang M, Long Y, Zhang M (2022) Economic-environmental scheduling of microgrid considering V2G-enabled electric vehicles integration. *Sustainable Energy, Grids and Networks* 32: 100872. https://doi.org/10.1016/j.segan.2022.100872
4. Jia R, Pan H, Zhang S, Hu Y (2024) Charging scheduling strategy for electric vehicles in residential areas based on offline reinforcement learning. *J Energy Storage* 103: 114319. https://doi.org/10.1016/j.est.2024.114319
5. Mosammam ZM, Ahmadi P, Houshfar E (2024) Multi-objective optimization-driven machine learning for charging and V2G pattern for plug-in hybrid vehicles: Balancing battery aging and power management. *J Power Sources* 608: 234639. https://doi.org/10.1016/j.jpowsour.2024.234639
6. Khezri R, Steen D, Wikner E, Tuan LA (2024) Optimal V2G scheduling of an EV with calendar and cycle aging of battery: An MILP approach. *IEEE T Transp Electr* 10(4): 10497–10507. https://doi.org/10.1109/TTE.2024.3384293
7. Wang Y, Shen X, Xu Y (2023) Joint planning of active distribution network and EV charging stations considering vehicle-to-grid functionality and reactive power support. *CSEE J Power Energy Syst* 10(5): 2100–2113. https://doi.org/10.17775/CSEEJPES.2023.03930
8. Zhang S, JAMES JQ (2021) Electric vehicle dynamic wireless charging system: Optimal placement and vehicle-to-grid scheduling. *IEEE Internet of Things Journal* 9(8): 6047–6057. https://doi.org/10.1109/JIOT.2021.3109956
9. Saner CB, Trivedi A, Srinivasan D (2022) A cooperative hierarchical multi-agent system for EV charging scheduling in presence of multiple charging stations. *IEEE T Smart Grid* 13(3): 2218–2233. https://doi.org/10.1109/TSG.2022.3140927
10. Yu H, Tu J, Lei X, Shao Z, Jian L (2024) A cost-effective and high-efficient EV shared fast charging scheme with hierarchical coordinated operation strategy for addressing difficult-to-charge issue in old residential communities. *Sustain Cities Soc* 101: 105090. https://doi.org/10.1016/j.scs.2023.105090
11. Kandpal B, Pareek P, Verma A (2022) A robust day-ahead scheduling strategy for EV charging stations in unbalanced distribution grid. *Energy* 249: 123737. https://doi.org/10.1016/j.energy.2022.123737
12. Chai YT, Che HS, Tan C, Tan WN, Yip SC, Gan MT (2023) A two-stage optimization method for Vehicle to Grid coordination considering building and Electric Vehicle user expectations. *Int J Electr Power Energy Syst* 148: 108984. https://doi.org/10.1016/j.ijepes.2023.108984
13. Yin W, Wen T, Zhang C (2023) Cooperative optimal scheduling strategy of electric vehicles based on dynamic electricity price mechanism. *Energy* 263: 125627. https://doi.org/10.1016/j.energy.2022.125627
14. Triviño-Cabrera A, Aguado JA, de la Torre S (2019) Joint routing and scheduling for electric vehicles in smart grids with V2G. *Energy* 175: 113–122. https://doi.org/10.1016/j.energy.2019.02.184
15. Wang Z, Zhang L, Tang W, Ma Z, Huang J (2024) Equilibrium configuration strategy of vehicle-to-grid-based electric vehicle charging stations in low-carbon resilient distribution networks. *Applied Energy* 361: 122931. https://doi.org/10.1016/j.apenergy.2024.122931
16. Hou H, Wang Y, Chen Y, Zhao B, Zhang L, Xie C (2022) Long-time scale vehicle-to-grid scheduling strategy considering psychological effect based on Weber-Fechner law. *Int J Electr Power Energy Syst* 136: 107709. https://doi.org/10.1016/j.ijepes.2021.107709

17. Yu Y, Zhong Q, Sun L, Han Y, Zhang Q, Jing X, et al. (2025) A Self-adaptive two stage iterative greedy algorithm based job scales for energy-efficient distributed permutation flowshop scheduling problem. *Swarm Evol Comput* 92: 101777. https://doi.org/10.1016/j.swevo.2024.101777
18. Cheng L, Tang Q, Zhang L, Yu C (2022) Scheduling flexible manufacturing cell with no-idle flow-lines and job-shop via Q-learning-based genetic algorithm. *Comput Ind Eng* 169: 108293. https://doi.org/10.1016/j.cie.2022.108293
19. Zhao Y, Ding Y, Pei Y (2024) Adaptive Optimization in Evolutionary Reinforcement Learning Using Evolutionary Mutation Rates. *IEEE Access* 12: 165384–165394. https://doi.org/10.1109/ACCESS.2024.3493198
20. Wang J, Lei D, Cai J (2022) An adaptive artificial bee colony with reinforcement learning for distributed three-stage assembly scheduling with maintenance. *Applied Soft Computing* 117: 108371. https://doi.org/10.1016/j.asoc.2021.108371
21. Huang JP, Gao L, Li XY (2024) A hierarchical multi-action deep reinforcement learning method for dynamic distributed job-shop scheduling problem with job arrivals. *IEEE T Autom Sci Eng* 22: 2501–2513. https://doi.org/10.1109/TASE.2024.3380644
22. Karimi-Mamaghan M, Mohammadi M, Pasdeloup B, Meyer P (2023) Learning to select operators in meta-heuristics: An integration of Q-learning into the iterated greedy algorithm for the permutation flowshop scheduling problem. *Eur J Oper Res* 304(3): 1296–1330. https://doi.org/10.1016/j.ejor.2022.03.054
23. Ma Y, Lu Y, Yin Y, Lei Y (2024) Pricing strategy of V2G demand response for industrial and commercial enterprises based on cooperative game. *Int J Electr Power Energy Syst* 160: 110051. https://doi.org/10.1016/j.ijepes.2024.110051
24. Shariatzadeh M, Antunes CH, Lopes MA (2024) Charging scheduling in a workplace parking lot: Bi-objective optimization approaches through predictive analytics of electric vehicle users' charging behavior. *Sustainable Energy, Grids and Networks* 39: 101463. https://doi.org/10.1016/j.segan.2024.101463
25. Triviño A, López A, Yuste AJ, Cuevas JC (2024) Decentralized EV charging and discharging scheduling algorithm based on Type-II fuzzy-logic controllers. *J Energy Storage* 93: 112054. https://doi.org/10.1016/j.est.2024.112054

AIMS Press