International Journal of Robotics and Automation,, Vol. 45, No. 10, 2025
An Open Access Paper

MULTI-REGIONAL ECONOMIC DISPATCHING AND DEMAND RESPONSE STRATEGIES OF SMART GRID BASED ON REINFORCEMENT LEARNING PERSPECTIVE

Zhiqing Zhou*

Abstract

The study addresses smart grid challenges (diversity, randomness, flexibility from electric vehicles, and smart buildings) by constructing a two-layer optimization model for dynamic retail pricing and load-unit demand response. The modified model uses reinforcement learning to learn the optimal electricity price, thereby balancing grid stability and user energy consumption. Through numerical simulation verification experiments, it was found that the power company's total revenue value and the load's total cost value in the research model were 7424.6 and 4152.8, respectively. Compared to the random parameter method, the difference was the smallest and better than the other two algorithms. Experiments have shown that price-based demand response models based on reinforcement learning can effectively solve related problems in unknown electricity market environments, and have important application value in maximizing social welfare in unknown market environments.

Key Words

Reinforcement learning; Smart grid; Economic dispatch; Demand response; Load

1. Introduction

In recent years, computer, network communication, and control technologies have driven significant changes in human society, leading to the information age marked by digitization, networking, and intelligence. The Power Grid (PG) is evolving into a Smart Grid (SG) due to this shift [1]. With increasing energy demand and depleting fossil fuels, traditional PGs face inefficiencies and maintenance challenges, failing to meet growing energy needs and environmental goals. SG is an emerging power technology that is replacing traditional PG to achieve sustainable development [2]. Operating SG requires flexible integration of distributed and renewable energy sources to ensure high-quality power services and grid security. The integration of information in SG, especially electric vehicles, has made Demand Response (DR) a key research area [3, 4]. DR management adjusts electricity consumption patterns through pricing or incentives to manage PG operations effectively. This study employs non-model-based Reinforcement Learning (RL) algorithms to solve price-based DR problems in SG.

* School of Humanities and Management, Xi'an Traffic Engineering Institute, Xi'an, 710300, China; e-mail: ZhiqingZh@outlook.com

(DOI: 10.2316/J.2025.203-0618)

2. Related Works

The inclusion criteria for this literature search focus on SG's multi-regional economic dispatch and DR strategy, as well as research based on the RL perspective. Exclusion criteria include unrelated topics or research methods that do not meet the criteria. The search databases cover authoritative platforms such as IEEE Xplore and ACM Digital Library. The search terms include "smart grid", "multi-regional economic dispatch", "demand response strategy", "reinforcement learning", etc., to ensure the comprehensiveness and accuracy of the literature review and the transparency and repeatability of the research.

Currently, SG technology is maturely applied in power distribution, generation, and consumption. Ullah et al. designed an energy optimization strategy using a multi-objective genetic algorithm, achieving 24% and 28% reductions in operating costs and carbon emissions with/without DR plans [5]. Kumari et al. proposed a secure DR model using RL and Ethereum blockchain to reduce energy consumption and costs [6]. Apostolopoulo et al. introduced a 2-stage algorithm, formulating the DR problem into a game theory framework to determine optimal electricity consumption and pricing [7]. Aladdin et al. proposed a multi-agent RL approach for efficient DR in SGs, reduc-

ing costs and peak-to-average ratio while maintaining user satisfaction [8]. Hafeez et al. proposed a wind-driven bacterial foraging algorithm to schedule Internet of Things-supported residential device electricity consumption, lowering peak power costs and enhancing user comfort [9]. Salazar et al. proposed an RL-based DR pricing and strategy, combining price and incentive-based DR management, effectively managing consumer demand [10]. Gharbi et al. found Dutch-style auctions superior in a supply-driven electricity market [11]. Reka et al. established a privacy-based DR model using machine learning for residential consumers in a cloud fog-based SG environment [12]. Alharbi proposed an SG DR framework based on Internet of Things house integration, optimizing quantitative DR supply [13]. Zarei et al. proposed a multi-objective optimization for optimal power flow in disaster recovery, achieving peak shaving/valley filling with 20% DR [14]. Yu et al. proposed a lightweight authentication protocol for DR management in resource-limited environments, ensuring secure mutual authentication and anonymity [15]. Bagherpur et al. utilized RL to improve DR plans in SG, enhancing system reliability and adjusting pricing models based on market strategies [16].

Jalali Khalil Abadi Z et al. developed a fuzzy logic-based scheduling algorithm for PG task scheduling and resource management, offering low computational complexity. However, due to the dependence of fuzzy logic rules on expert experience, this algorithm was subjective and might reduce the accuracy of scheduling [17]. Navarro González F J et al. proposed an irrigation scheduling algorithm for photovoltaic irrigation networks to save energy. However, its performance might be affected by the instability of photovoltaic power generation, and the stability of irrigation scheduling under different lighting conditions needs further research [18]. Zhang S et al. proposed an urban electric vehicle parking system Internet for vehicle-to-grid dispatching to improve public services in smart cities, but it might face data security and privacy issues, and large-scale deployment costs were high [19]. Das P et al. proposed a two-stage electric vehicle charging and discharging scheduling model to reduce transmission network pressure. However, this model might not fully consider regional PG differences and the diversity of user charging habits, which poses practical challenges [20]. Hajari S et al. studied the impact of photovoltaics, wind energy, gas turbines, and energy storage systems on distribution network reliability, finding that distributed generation integration enhances reliability [21]. Yadav AK et al. researched the contact line model in interconnected synchronous phasor networks for grid observability and reliability, identifying the most reliable deployment locations for phasor measurement units [22]. Mahajan V et al. proposed using discrete Markov chains for multi-state modeling of renewable energy and energy storage devices, overcoming the shortcomings of traditional methods, and found that the application of this method reduced active power losses and improved power flow [23]. Mahajan A K Y V et al. discussed optimizing the number of deployed phasor measurement units and handling interruptions during anomalies, considering zero injection busbars and network observability/reliability [24].

In summary, various algorithm strategies proposed in this field, such as the multi-objective genetic algorithm and RL algorithm, have demonstrated their effectiveness. However, there has been a lack of in-depth exploration into the differences in applicable scenarios, performance comparisons, and potential conflicts among these algorithms. The effectiveness of different algorithms under different grid structures and operating conditions has not been thoroughly analyzed, nor has the possibility of collaboration or substitution between algorithms been considered. Furthermore, the gap between experimental environments and actual SG applications has also not been addressed. In addition, current research faces critical gaps (lack of integrated multi-regional economic scheduling-DR strategies), limitations (poor user privacy protection, dynamic adaptability, and real-grid generalizability), and contradictions (inconsistent peak load reduction outcomes across studies). Future directions should focus on developing unified multi-region strategies, enhancing algorithm synergy in complex grids, and validating through large-scale field experiments. Based on RL, this study designs a dual-layer optimized dynamic pricing model consisting of dynamic retail pricing of power companies and optimal DR of load units, enabling it to adapt to the dynamic electricity market environments.

This Q-learning-based dual-layer dynamic pricing model outperforms traditional static/game theory methods. Key innovations:

(1) Real-time adaptive pricing captures nonlinear electricity price demand relationships through Q-learning, replacing predefined rules;
(2) Unlike single-stage game equilibrium, the dual-layer framework achieves bidirectional optimization between utility pricing and user load adjustment through RL;
(3) Data-driven approach eliminates reliance on perfect information assumptions in game theory. Compared with two-stage games, its advantages include cross-cycle optimization through discount factors, dynamic strategy adaptation through exploration mechanisms, and parallel computing relative to Nash equilibrium iterations. Therefore, this model not only improves the flexibility and robustness of the power system but also provides an effective demand management tool for power companies, which helps promote the utilization of renewable energy and the sustainable development of the power system.

3. Construction of SGPDR Model Based on RL

3.1 SGPDR Model Design

The price-based DR objective is to coordinate the energy consumption of a limited number of load units within a certain period in response to dynamic retail unit prices, thereby achieving the weighted sum goal of profit and load comprehensive cost for the power company. Based on this, this study establishes a Price-based Demand Re-

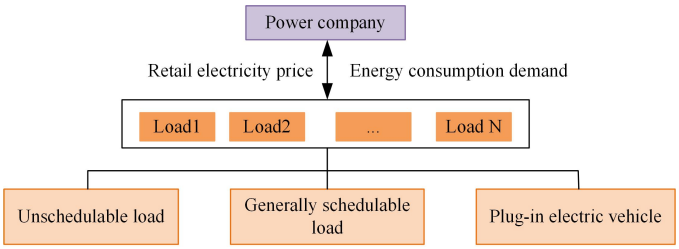

Figure 1. Price-based DR model for SG.

sponse model for the Smart Grid (SGPDR) model (Figure 1). The model has two levels: the upper target is to maximize the power firm's profit, and the lower is to minimize the load comprehensive cost.

The retail electricity market in Figure 1 includes two levels: power companies (upper level) and loads (lower level), with information flowing in both directions. Among them, the load (lower layer) transmits energy demand information to the power company, and the power company (upper layer) releases dynamic retail price information to the load. This study analyzes the operational data and load unit characteristics of power companies, combined with expert discussions, technical documents, and historical data, to collect and verify customer needs and technical requirements, avoiding traditional surveys. Based on the impact of demand on user satisfaction, the Kano model (basic, expected, and exciting attributes) is used to classify the demand, ensuring coverage in both business and technical domains. Normally, loads are divided into two categories: Non-schedulable Load (NSL) N_N and schedulable load N_d This study proposes a General Schedulable Load (GSL) G_g and innovatively considers more flexible ones like Plug-in Electric Vehicle (PIEV) v, as shown in (1).

$$N_d = G_g \cup v \tag{1}$$

At this point, the energy consumption of the GSL is shown in (2).

$$p_{n,t}^{G} = e_{n,t}^{G}\left(1+\varepsilon_t\frac{\eta_{n,t}^{G}-\lambda_t}{\lambda_t}\right), \forall t \in T \tag{2}$$

In (2), T represents a certain period of time. ε_t represents the price elasticity coefficient of the schedulable load at time t which is usually negative. $p_{n,t}^{G}$ and $e_{n,t}^{G}$ respectively represent the actual energy consumption and Energy Consumption Demand (ECD) of the GSL n at time t . $\eta_{n,t}^{G}$ and λ_t represent the retail and wholesale electricity prices at time t. When the energy consumption of the GSL n is $p_{n,t}^{G}$, the remaining required energy cannot be satisfied, leading to dissatisfaction of load unit n. The dissatisfaction function for this level of dissatisfaction is shown in (3).

$$\varphi_{n,t}^{G} = \frac{1}{2}\alpha_n^{G}\left(e_{n,t}^{G}-p_{n,t}^{G}\right)^2+\beta_n^{G}\left(e_{n,t}^{G}-p_{n,t}^{G}\right) \tag{3}$$

In (3), α_n^{G} and β_n^{G} represent the satisfaction coefficients related to the load unit. This function indicates that a significant decrease in demand will lead to higher levels of dissatisfaction. In practical situations, the reduction in demand for GSL n is limited, as shown in (4).

$$DR_n^{\min} \le \boldsymbol{e}_{n,t}^{G} - p_{n,t}^{G} \le DR_n^{\max} \tag{4}$$

In (4), $DR_n^{\min}$ and $DR_n^{\max}$ respectively represent the Upper and Lower (U-L) limits of the demand reduction for GSL n. If both are constant, the energy consumption range of $p_{n,t}^{G}$ can be determined. PIEVs are flexible distributed units due to the characteristics of their onboard batteries, and their energy consumption is shown in (5).

$$\begin{cases} p_{n,t}^{v} = e_{n,t}^{v}\left(1+\varepsilon_t\dfrac{\eta_{n,t}^{v}-\lambda_t}{\lambda_t}\right), n \in v, \forall t \in T \\ -p_n^{-v} \le p_{n,t}^{v} \le p_n^{-v} \end{cases} \tag{5}$$

In (5), $p_{n,t}^{v}, e_{n,t}^{v}$, and $\eta_{n,t}^{v}$ represent the actual energy consumption, ECD, and received Retail Electricity Price (REP) of electric vehicle n at time t.$p_n^{-v\succ 0}$ represents the rated power of PIEV n. The Charging and Discharging (C/D) power of electric vehicles cannot exceed their rated power at any time. The calculation of PIEVs considering battery self-consumption and C/D point efficiency is shown in (6).

$$\begin{cases} \sum\limits_{i=1}^{T}\mu_{n,t}p_{n,t}^{V} = E_n^{T} - E_n^{0} \\ C_n^{m} \le E_n^{0} + \sum\limits_{i=1}^{t}\mu_{n,t}p_{n,t}^{V} \le C_n^{M}, \forall t \in T \end{cases} \tag{6}$$

In (6), E_n^{T} and E_n^{0} represent the battery level at the end and beginning of PIEV charging, respectively. $\mu_{n,t}$ represents the C/D efficiency of PIEV n at t.C^m and C_n^{M} are the U-L limits of PIEV battery capacity. The battery capacity of an electric vehicle must not exceed its maximum capacity at any time. The dissatisfaction function of electric vehicle owners is shown in (7).

$$\varphi_{n,t}^{v} = \frac{1}{2}\alpha_n^{v}\left(e_{n,t}^{v}-p_{n,t}^{v}\right)^2+\beta_n^{v}\left(e_{n,t}^{v}-p_{n,t}^{v}\right), \forall t \in T \tag{7}$$

In (7), α_n^{v} and β_n^{v} represent the satisfaction coefficients related to electric vehicle n. However, frequent C/D of electric vehicles can shorten their lifespan, so the definition for quantifying this impact is shown in (8).

$$c_{n,t}^{v,\deg} = k\left|p_{n,t}^{v}\right|, \forall t \in T \tag{8}$$

In (8), k represents the degradation coefficient.

The energy need for NSL should be met at all times, and their energy consumption is shown in (9).

$$p_{n,t}^{non} = e_{n,t}^{non} \forall \forall t \in T \tag{9}$$

In (9), $p_{n,t}^{non}$ and $e_{n,t}^{non}$ respectively represent the actual energy consumption and ECD of non-schedulable n at time t. From the perspective of load, it is expected to determine its optimal energy consumption to minimize the overall cost, as shown in (10).

$$\min_{p} \sum_{t=1}^{T} \left[\sum_{n \in N_n} \eta_{n,t}^{non} p_{n,t}^{non} + \sum_{n \in N_s} \left(\eta_{n,t}^{G} p_{n,t}^{G} + \varphi_{n,t}^{G} \right) + \sum_{n \in V} \left(C_{n,t}^{V,\,\mathrm{deg}} + C_{n,t}^{V,\,\mathrm{net}} + \varphi_{n,t}^{V} \right) \right] \tag{10}$$

In (10), p is the energy consumption vector of all joining loads all in the period. The three parts in the formula correspond to the three kinds of loads mentioned above.

The goal of the power company is to establish the optimal REP to achieve maximum profit, and its solving model is shown in (11).

$$\max_{\eta} \sum_{t=1}^{T} \left(\sum_{n \in N \text{ in}} \eta_{n,t}^{non} p_{n,t}^{non} + \sum_{n \in N_g} \eta_{n,t}^{G} p_{n,t}^{G} + \sum_{n \in V} \eta_{n,t}^{V} p_{n,t}^{V} - \lambda_t c_t \right) \tag{11}$$

In (11), η represents the REP gained by overall loads. c_t represents the total electricity purchased at time t. The first three parts of the formula correspond to the revenue generated by the power company from selling electricity to NSL, GSL, and PIEV loads, respectively. The last part represents the cost of electricity bought by power companies from higher-level grid operators. Among them, $c_t = \sum_{n \in N_n} p_{n,t}^{non} + \sum_{n \in N_s} p_{n,t}^{G} + \sum_{n \in V} p_{n,t}^{V}$, and the profit of the power company at t is recorded as $U_t = \sum_{n \in NN_n} \eta_{n,t}^{non} p_{n,t}^{non} + \sum_{n \in N_g} \eta_{n,t}^{G} p_{n,t}^{G} + \sum_{n \in V} \eta_{n,t}^{V} p_{n,t}^{V} - \lambda_t c_t$. Therefore, (11) can be simplified into (12).

$$\begin{cases} \max_n \sum_{t=1}^{T} U_t \\ \eta_{-n} \le \eta_{n,t} \le \bar{\eta}_n, \forall n \in N, \forall t \in T \end{cases} \tag{12}$$

In (12), η_{-n} and $\bar{\eta}_n$ are the U-L limits of REPs. N represents the set of all loads. The study assumes a linear relationship between electricity price changes and demand reduction, supported by price elasticity theory showing short-term near-linear user responses to small price fluctuations (±20% benchmark). Real-grid data confirm a significant linear correlation ($R^2 > 0.7$) between load changes and price adjustments within this range. The mathematical expression of the SG price DR model is shown in (13).

$$\max_{n,p} \sum_{t=1}^{T} [\sigma U_t - (1-\sigma) C_t] \tag{13}$$

In (13), σ represents the weight coefficient, $\sigma \in [0, 1]$.

3.2 Solution of SGPDR Model Based on RL

At present, there are some mature methods for solving the PG-DR model, such as model predictive control, two-stage stochastic programming, and robust optimization, but all of them have limitations. Therefore, this study chooses the RL method to solve the SGPDR model. The RL method can adaptively determine the optimal strategy without relying on accurate prediction of uncertain parameters and accurate load models. Figure 2 shows the basic principle of RL.

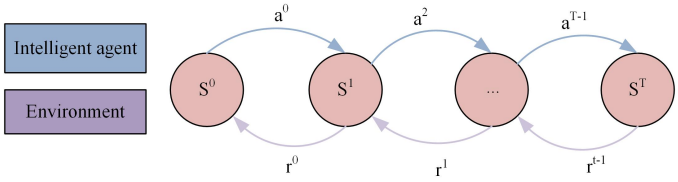

Figure 2. Basic principle of RL.

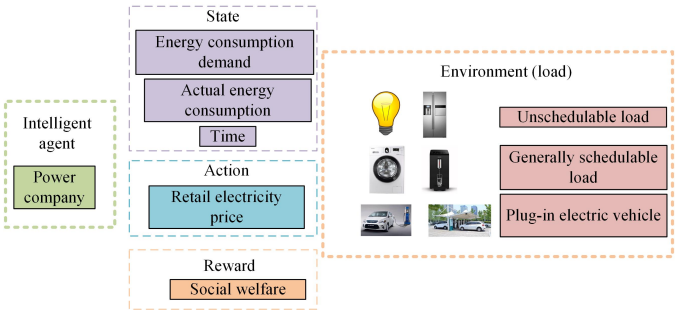

Figure 3. SGPDR based on RL framework.

In Figure 2, RL explores unfamiliar environments by taking continuous actions and optimizing behavior plans based on the rewards given by the environment, ultimately finding the optimal strategy with the highest cumulative reward. This study establishes an RL-based SGPDR model to determine the optimal retail pricing, as shown in Figure 3.

In Figure 3, all loads serve as the environment, and the power company considers them as intelligent agents. The REP represents the actions taken by the intelligent agents towards the environment. The actual energy consumption, energy demand, and time indicators denote the state, and the combined revenue and load cost of the power company are considered rewards. Then, this study further adopts a Markov decision process to model the dynamic retail price. Among them, this study defines the mapping from state to action as strategy ∂, and the goal of the dynamic retail price is to search the best ∂^* which is able to maximize cumulative returns, as shown in (14).

$$\begin{cases} \partial : S \to A \\ \partial^* = \arg g \max_{\partial} \sum_{t=1}^{T} r^t \left(s^t, \partial \left(s^t \right) \right) \end{cases} \tag{14}$$

In (14), S and A are the state and action sets in Markov decision process. r^t and s^t represent the reward and environmental status at time t.

Finally, this study takes the Q-learning to analyze the selection of REP for power companies. According to the basic rule of Q-learning, the optimal action value function represents beginning with state s, taking a, and then using the maximum cumulative discount of ∂^* for return. It follows the Bellman optimal equation, as shown in (15).

$$Q^*(s,a) = \varsigma \left[r(s,a) + \gamma \max Q_{a'}^{*} \left(s', a' \right) \right] \tag{15}$$

```
def generate_actions(current_price, demand_elasticity, time_period):
    # Define baseline price and thresholds
    baseline_price = 0.8  # USD/kWh
    core_lower = baseline_price * 0.8
    core_upper = baseline_price * 1.2

    # Dynamic action space generation
    if core_lower <= current_price <= core_upper:
        if demand_elasticity > 0.3:  # High elasticity
            actions = [round(-0.01 * i, 2) for i in range(1, 3)] + + \
                  [round(0.01 * i, 2) for i in range(1, 3)]
        else:  # Normal
            actions = [-0.02, -0.01, 0, +0.01, +0.02]
    else:  # Extreme price zone
        if time_period == 'peak':
            actions = [-0.05, -0.02, 0, +0.02]  # Cap +5% during peak
        else:
            actions = [-0.1, -0.05, 0, +0.05, +0.1]

    return actions

# Q-learning update with dynamic actions
state = observe_grid_state()
actions = generate_actions(state['price'], state['elasticity'], state['period'])
action = epsilon_greedy_select(Q_table, state, actions)
next_state, reward = execute_action(action)
Q_update(Q_table, state, action, reward, next_state)
```

Figure 4. Flow chart of the pseudo-code for the multi-granularity adaptive discretization algorithm.

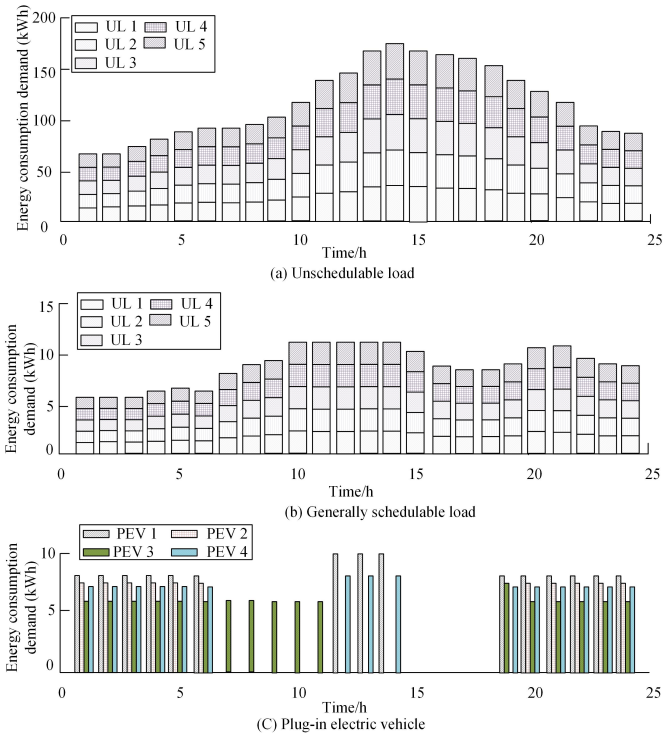

Figure 5. ECD of NSLs and schedulable loads.

In (15), s' represents the state at the next moment. a' represents the action taken in state s'. ς represents the expected operator for the next state and reward randomness. Therefore, as long as $Q^*(s,a)$ is obtained, the optimal strategy can be determined, as shown in (16).

$$\partial^*(S) = \arg g \max_{a \in A} Q^*(s,a) \tag{16}$$

To better reflect the dynamic power market, a multi-granular adaptive discretization algorithm is then proposed. Figure 4 shows the algorithm pseudo-code.

The study proposes a dynamic electricity pricing strategy: prices are divided into a core range (benchmark $\pm 20\%$) with fine-grained $1-2\%$ adjustments and an extreme range using $5-10\%$ coarse adjustments, based on historical data showing 90% fluctuations within $\pm 15\%$. Pricing actions adapt dynamically: core zones use $\pm 2\%$ steps, while near limits shift to $\pm 5\%$. During peak demand, price hikes are capped ($+5\%$ max) to reduce user resistance, while renewable surplus periods employ finer price cuts (0.5% steps) to boost consumption.

4. Performance Testing of SGPDR Model Based on RL

In the numerical simulation verification experiment of the SGPDR model based on RL, nine schedulable loads (including 5 GSLs and 4 PIEVs) and five NSLs are considered as DR problems, as show in Table 1 [25].

The study uses an IEEE 14-node power market simulator with PJM market data for parameter tuning. Grid search and cross-validation optimize learning rate, discount factor (for convergence), and action step size (to boost revenue). Key metrics include power company revenue (¥/day), user satisfaction (load demand fulfillment rate, %), and algorithm convergence speed (iterations). The comparison results before and after adjustment are shown in Table 2 [26].

In Table 2, all evaluation indicators have been effectively improved in the optimized parameter combination. The daily average income, load satisfaction rate, convergence iteration times, and standard deviation of electricity price fluctuations have increased by $+17.1\%, +7.7\%, -37.5\%$, and 33.3%, respectively.

The NSL and schedulable energy consumption data are sourced from open-source data of a certain gas and power company, as shown in Figure 5.

In Figure 5, the ECD peak of NSL occurs from 13:00 to 17:00. The ECD trend of GSL is basically consistent, with two peak demands occurring between 10:00-15:00 and 19:00-22:00, respectively. The ECD of PIEV is concentrated in two time periods: 1:00-7:00 and 19:00-24:00. If the schedulable loads' actual energy consumption coordinates wrong, the electricity burden on the PG will greatly increase, thereby affecting the normal operation of the electricity market economy. In the simulation experiment, wholesale prices are decided by the PG operator.

In the simulation, the REP reduces or grows by 0.1 times in each iteration, indicating that the action space is discrete. The results of three optimal retail electricity pricing strategies for load days are shown in Figure 6.

In Figure 6, the trend of wholesale and REPs is similar, which is in line with maximizing social welfare. Among them, there are two sharp price drops at 12:00 and 16:00.

Table 1
Setting of the Experimental Parameters.

Parameter category	The parameter name	The parameter name
The Q-learning parameters	Learning rate	0.2
	Discount factor	0.9
	Probe rate	0.3
Parameters of electricity price model	Basic electricity price (yuan / kWh)	0.7
	Price elasticity coefficient	-0.5
Demand response parameters	Load adjustment upper limit of (%)	Core: ±2% Edge: ±5%
	Response Delay Time (hours)	1
Monte Carlo simulations	Sample size	1000
	Random perturbation of the standard deviation	5% Benchmark load

Table 2
Test Results Before and After Parameter Tuning.

Index	Before tuning	After tuning	Improve the range
Daily average income (yuan)	24,500	28,700	+17.1%
Load satisfaction rate (%)	82.3	88.6	+7.7%
Convergence iteration times	1,200	750	-37.5%
Electricity price fluctuation standard deviation	0.18	0.12	-33.3%

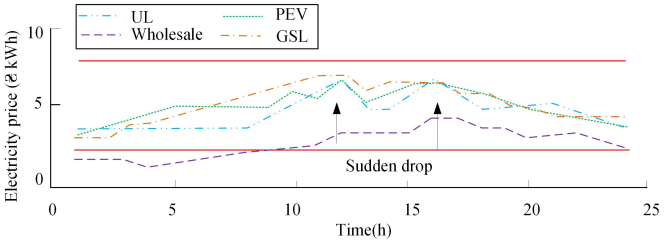

Figure 6. Optimal REP strategy for day 5.

The p-value of the price reduction in both periods is less than 0.05 , which is statistically significant. The price difference of each load unit during non-peak, mid-peak, and peak periods is as follows: non-peak>mid-peak>peak period. This is related to the fact that the REP coefficient during peak hours is smaller than mid-peak and off-peak. After gaining the optimal REPs for all loads, the optimal energy consumption for each load unit can be directly calculated, as shown in Figure 7.

In Figure 7 (a), the optimal ECD peak of GSL occurs in two time periods: 10:00-15:00 and 19:00-22:00. The GSL's actual energy consumption peak also occurs during these two time periods, and the overall trend of the two is consistent. Figure 7 (b) shows that the optimal ECD peak of PIEV occurs in two time periods: 11:00-14:00 and 18:00-24:00. To alleviate power pressure and increase its own profits, PIEV chooses to discharge during this peak period.

Figure 8 shows the decrease in total ECD for each schedulable load. Compared to other schedulable loads, GSL6 has a smaller reduction in ECD volume. This is because load units with a larger satisfaction coefficient tend to have smaller demand reductions, otherwise, it will increase the cost of dissatisfaction. To determine whether the Q-value of the model converges to the max-value, this study selects the $\mathrm{Q}\left(\mathrm{s}^{8}, \mathrm{a}^{7}\right)$ of each load, as exhibited in Figure 9.

In Figure 9, at the beginning of the iteration, the power company does not know how to establish a REP that can bring a greater reward. As the iteration progresses, the Q of the three loads grows orderly and converges to the max at last, because the power company acknowledges the dynamic response of the loads by trials and errors. To discuss the influence of the profit weight coefficient, the study uses the Monte Carlo method to capture the trend of the average REP, the power company's total income, and the total load cost with the change of the profit weight coefficient, as shown in Figure 10.

Figure 10 shows that as the profit weight coefficient rises, the average REP increases from $\sim$ 4¥/kWh to 5.1¥/kWh, power company revenue grows, and total load cost surges from 0¥ to 9,000¥. Higher weight prioritizes maximizing utility profits over minimizing user costs, driving price hikes. The price increase has reduced energy consumption and slowed down the overall growth rate.

The study tests the model under three scenarios: Scenario 1 (low: 100 load curves, ±5% price fluctuation for residential areas), Scenario 2 (moderate: 500 curves, ±10% for commercial zones with solar), and Scenario 3 (high: 1000 curves, ±20% for industrial areas with wind volatility). Evaluation metrics include grid-side peak-valley difference and revenue volatility, user-side cost changes and response rates, and algorithm stability/convergence speed. The test results are shown in Table 3.

Table 3 reveals a 38% reduction in the peak-valley differential rate but a 130% increase in yield volatility due to aggressive price adjustments. User response rates decline with higher fluctuations, suggesting extreme pricing discourages participation. This algorithm maintains stability in continuous action space, but requires 2.5 times the training time, making it suitable for residential applications (with a cost advantage of -3.2%). Based on the non-price-based DR scenario, the daily energy consump-

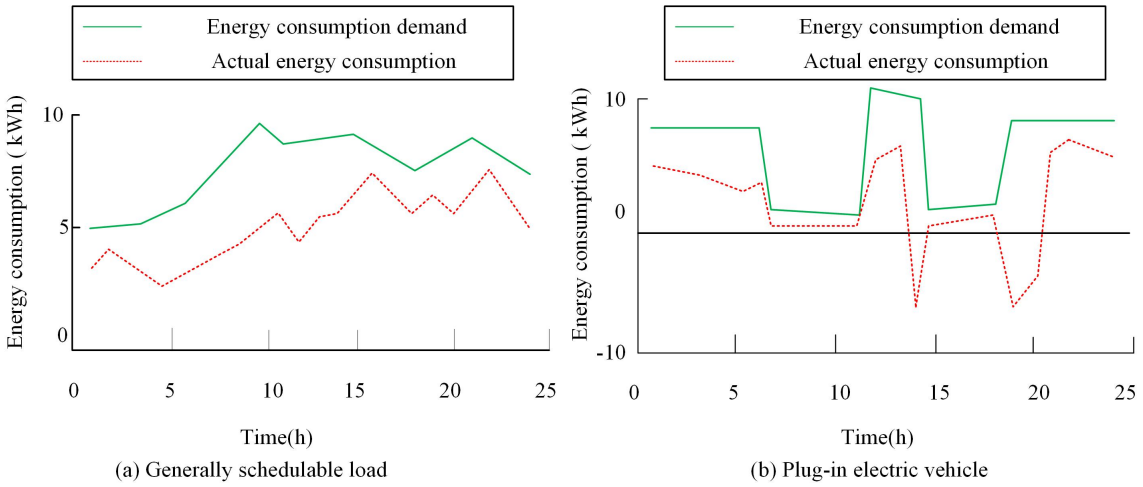

Figure 7. Optimal daily energy consumption.

Table 3
Model Performance Results in Different Price Fluctuation Scenarios.

Scene	Low volatility	In the fluctuation	High volatility
Load/price fluctuations	100, ±5%	500, ±10%	1000, ±20%
Bee valley difference rate	15.20%	12.70%	9.40%
Return fluctuation rate	8.10%	12.30%	18.60%
Changes in electricity bills	-3.20%	-5.80%	-8.10%
Response ratio	68%	54%	41%
Convergence rate	1,500	2,200	3,800

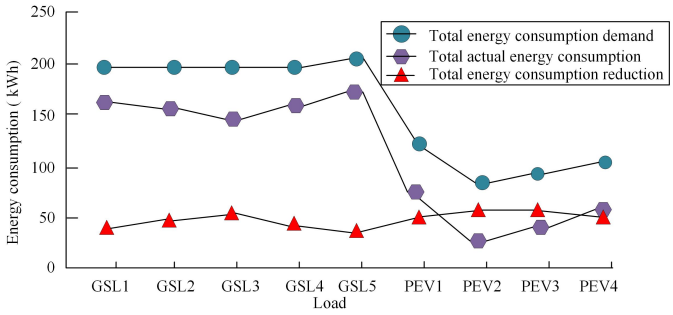

Figure 8. Reduction in total ECD for schedulable loads.

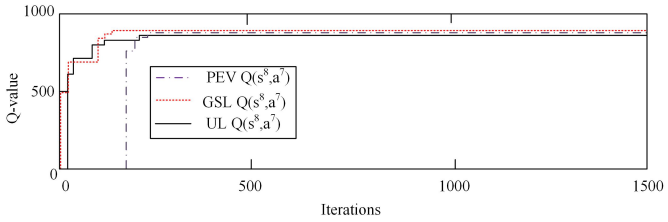

Figure 9. Q-value convergence situation.

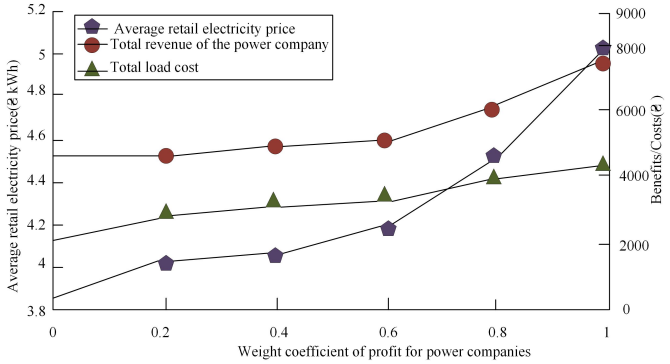

Figure 10. The impact of parameters on average REP, total revenue of power companies, and total load cost.

tion comparison of all loads in the two scenarios is shown in Figure 11.

Figure 11 shows that price-based DR adjusts energy consumption during high electricity price periods by coordinating load units, smoothing the energy consumption curve, and improving social welfare, thereby reducing energy consumption fluctuations. Compared to DDD, PDP, and RPM methods in Figure 12, the proposed model demonstrates higher efficiency and stability in balancing grid demands and economic benefits.

In the simulation, this study sets the initial Lagrange multiplier to 0 . In PDP, $\alpha = 0.05, \sigma = 2.5$, $\varepsilon_{feas} = \varepsilon_{\nu} = 10^{-2}$, and $\beta = 0.2$. The original random dual-vector is 0.5. The REP trends obtained by the four methods are the

Table 4
Test Results of the Model in Realistic Scenarios.

Index	Traditional method	Research model	Improve the range
Mean daily peak-valley difference rate	28.5%	19.2%	-32.6%
Wind power utilization rate	68%	82%	±20.6%
Residents participation rate	45%	73%	±62.2%
Business user savings	E120 per month	E185 / month	±54.2%
Decisions delayed	200ms	90ms	-55%
Abnormal situation treatment success rate	72%	89%	+23.6%

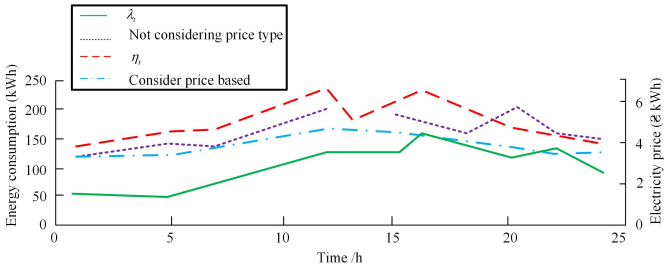

Figure 11. Comparison results of daily energy consumption for all loads.

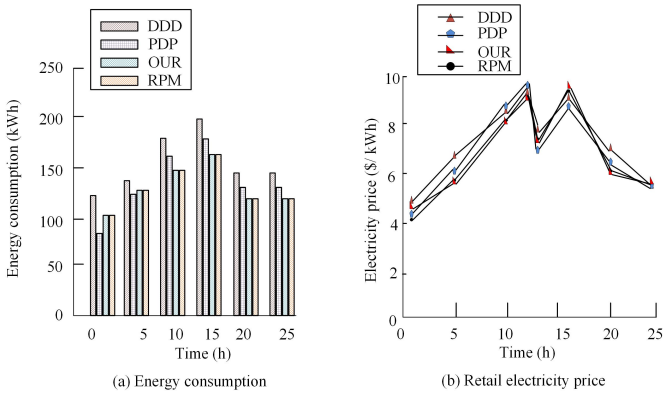

Figure 12. Comparison results of four algorithms.

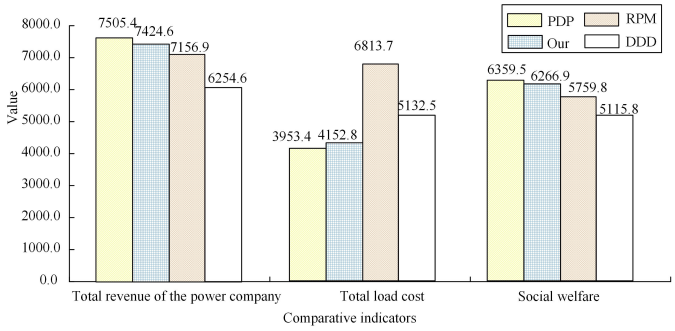

Figure 13. Numerical comparison results of algorithms.

same. Compared to the energy consumption curves of the DDD algorithm and PDP algorithm, the energy consumption curve obtained by the research model is smoother. Meanwhile, the REP and energy consumption curve obtained from the research model are closest to those obtained from the random parameter method. The numerical comparison results of total revenue, social welfare, and total load cost for power companies under four algorithms are shown in Figure 13.

Figure 13 compares four algorithms: the RPM model achieves the highest social welfare (6359.5), while the proposed RL-based price-based DR model ranks second (6266.9), outperforming DDD and PDP. The model's power company revenue (7424.6) and load cost (4152.8) are closest to RPM's results. Practical validation using Aachen University's 2023-2024 SG project (2,800 households, 150 commercial users across three communities) confirms its effectiveness in real-world scenarios. The research model is applied to the grid and the results with the traditional methods of the grid are shown in Table 4.

Table 4 demonstrates the model's advantages: the average daily peak-valley difference rate decreases by 32.6%(28.5% → 19.2%), the utilization rate of wind power generation increases by 20.6% to 82%, the resident participation rate increases by 62.2% to 73%, commercial electricity costs are saved by 54.2%, and decision-making delays are reduced by 55%. The success rate of handling abnormal situations has increased by 23.6%, reflecting the optimization of resource allocation, faster response speed, and stable PG.

5. Conclusion

A dynamic pricing model based on RL was designed to address the uncertainty of load unit electricity prices and energy consumption behavior in the electricity market, which can adapt to flexible changes in load and dynamic electricity market conditions. The research results indicated that price-based electricity price compensation effectively coordinated the energy consumption of load units and improved the welfare of the retail electricity market. Compared with DDD, PDP, and RPM, the REP/energy consumption curve of this model was closest to RPM, with social welfare of 6266.9 (the smallest difference from RPM), power company revenue of 7424.6, and load cost of 4152.8, which were better than DDD and PDP. This study shows that the RL-based price-driven DR model effectively addresses the DR issue in unknown electricity markets. The innovation of this study lies in three aspects. First, the RL-based dynamic pricing model provides a novel way to manage load and market uncertainties. Second, the price-based electricity price compensation mechanism optimizes energy consumption and boosts market welfare. Third, the successful application of the RL-based DR model in unknown markets paves the way for future research and practical use. However, this study also has limitations, such as model validation relying solely on numerical simulations rather than real-world prototype testing. Future work should focus on large-scale prototype design and integration of advanced artificial intelligence and data analysis technologies. Real-time data can be used to optimize parameters and improve responsiveness to market and load changes, operational efficiency, and social welfare.

Conflicts of Interest

The author reports there are no competing interests to declare.

Funding

The research is supported by Key Scientific Research 2024 Study on strategy of multi-regional economic dispatching and demand response of smart grid based on reinforcement learning, Project No.: 2024KY-43.

References

[1] R. Cai, B. Xia, X. Zhu, L. Wang, J. Gu, and J. Tang, "Design of a risk model and analytical decision information system for power operation in the context of smart grid," *International Journal of Power and Energy Systems*, vol. 44, no. 10, 2024.

[2] W. Mao, S. Yang, and X. Gao, "Modelling and optimisation of green energy systems based on complementary integration of renewable energy," *International Journal of Power and Energy Systems*, vol. 44, no. 10, 2024.

[3] W. Lei, H. Chongqi, and Y. Bin, "Optimal operation analysis of integrated community energy system considering the uncertainty of demand response," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3681–3691, 2021.

[4] S. Zhao and L. Zhao, "Forecasting long-term electric power demand by linear semiparametric regression," *Advances In Industrial Engineering And Management*, vol. 11, no. 1, 2022.

[5] U. Kalim, A. Sajjad, K. T. Ahmad, K. Imran, J. Sadaqat, S. A. Ibrar, and H. Ghulam, "An optimal energy optimization strategy for smart grid integrated with renewable energy sources and demand response programs," *Energies*, vol. 13, no. 21, pp. 5718–5722, 2020.

[6] K. Aparna and T. Sudeep, "A reinforcement-learning-based secure demand response scheme for smart grid system," *IEEE Internet of Things Journal*, vol. 9, no. 3, pp. 2180–2191, 2021.

[7] A. P. Athanasios, T. E. Eleni, and P. Symeon, "Demand response management in smart grid networks: A two-stage game-theoretic learning-based approach," *Mobile Networks and Applications*, vol. 26, no. 2, pp. 548–561, 2021.

[8] S. Aladdin, E.-T. Samah, F. M. M, and E. S. Adly, "Marlasg: Multi-agent reinforcement learning algorithm for efficient demand response in smart grid," *IEEE access*, vol. 1, no. 8, pp. 210626–210639, 2020.

[9] H. Ghulam, W. Zahid, K. I. Ullah, K. Imran, and S. Zeeshan, "Efficient energy management of iot-enabled smart homes under price-based demand response program in smart grid," *Sensors*, vol. 20, no. 11, pp. 3155–3169, 2020.

[10] J. Salazar, Eduardo, M. Jurado, and E. Samper, Mauricio, "Reinforcement learning-based pricing and incentive strategy for demand response in smart grids," *Energies*, vol. 16, no. 3, pp. 1466–1482, 2023.

[11] A. Gharbi, M. Ayari, and E. Yahya, Abdulsamad, "Demand-response control in smart grids," *Applied Sciences*, vol. 13, no. 4, pp. 2355–2369, 2023.

[12] S. Reka, Sofana, P. Venugopal, V. Ravi, and T. Dragicevic, "Privacy-based demand response modeling for residential consumers using machine learning with a cloud–fog-based smart grid environment," *Energies*, vol. 16, no. 4, pp. 1655–1669, 2023.

[13] W. Alharbi, "Integrating internet-of-things-based houses into demand response programs in smart grid," *Energies*, vol. 16, no. 9, pp. 3699–3715, 2023.

[14] A. Zarei and N. Ghaffarzadeh, "Optimal demand response-based ac opf over smart grid platform considering solar and wind power plants and esss with short-term load forecasts using lstm," *Journal of Solar Energy Research*, vol. 8, no. 2, pp. 1367–1379, 2023.

[15] Y. S. Jin, P. K. Sung, and L. J. Young, "Privacy-preserving lightweight authentication protocol for demand response management in smart grid environment," *Applied Sciences*, vol. 10, no. 5, pp. 1758–1774, 2020.

[16] B. Reza, M. Nasser, and B. Babak, "Improving demand-response scheme in smart grids using reinforcement learning," *International Journal of Energy Research*, vol. 45, no. 15, pp. 21082–21095, 2021.

[17] K. Jalali, Z. Abadi, and N. Mansouri, "A comprehensive survey on scheduling algorithms using fuzzy systems in distributed environments," *Artificial Intelligence Review*, vol. 57, no. 1, pp. 4–18, 2024.

[18] J. Navarro-González, Francisco, A. Pardo, M, and E. Chabour, Houssem, "An irrigation scheduling algorithm for sustainable energy consumption in pressurised irrigation networks supplied by photovoltaic modules," *Clean Technologies and Environmental Policy*, vol. 25, no. 6, pp. 2009–2024, 2023.

[19] S. Zhang, S. Zhang, and K. Yeung, Lawrence, "Urban internet of electric vehicle parking system for vehicle-to-grid scheduling: Formulation and distributed algorithm," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 1, pp. 67–79, 2023.

[20] P. Das and P. Kayal, "An advantageous charging/discharging scheduling of electric vehicles in a pv energy enhanced power distribution grid," *Green Energy and Intelligent Transportation*, vol. 3, no. 2, pp. 100170–100184, 2024.

[21] S. Hajari, A. Yadav, and N. Singh, "Impact of PV, WT, GTG, and ESS on the Reliability of Distribution System," *Majlesi Journal of Electrical Engineering*, vol. 17, no. 2, 2023.

[22] K. Yadav, Atul and V. Mahajan, "Tie-line modelling in interconnected synchrophasor network for monitoring grid observability, cyber intrusion and reliability," *Engineering Review*, vol. 42, no. 2, pp. 114–132, 2022.

[23] V. Mahajan, S. Mudgal, and K. Yadav, Atul, "Reliability modeling of renewable energy sources with energy storage devices," in *Energy Storage in Energy Markets*, pp. 317–368, Academic Press, 2021.

[24] V. Mahajan and K. Yadav, Atul, "Cyber-attack and reliability monitoring of the synchrophasor smart grid network," *Jurnal Kejuruteraan*, vol. 34, no. 6, pp. 1149–1168, 2022.

[25] L. Tightiz, M. Dang, L, and J. Yoo, "Novel deep deterministic policy gradient technique for automated micro-grid energy management in rural and islanded areas," *Alexandria Engineering Journal*, vol. 82, pp. 145–153, 2023.

[26] L. Tightiz and J. Yoo, "A robust energy management system for korean green islands project," *Scientific Reports*, vol. 12, no. 1, p. 22005, 2022.

Biographies

Zhiqing Zhou, received her graduate degree in Accounting from Xi'an Jiaotong University. She is currently a full-time instructor in the Financial Management Department of the School of Humanities and Management, Xi'an Traffic Engineering Institute. Her research area is the application of financial management in the financial industry.