Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.36888

Real-Time Dynamic Dispatch Strategy for High-Penetration Wind-Solar Power Systems Based on Deep Reinforcement Learning

Yan Wang

North China Electric Power University, Beijing, China
wangyan_1019@outlook.com

Abstract. High-penetration wind and photovoltaic integration is driving smart grids from conventional deterministic dispatch toward real-time dynamic dispatch under random fluctuation and fast response requirements. Wind and solar outputs are strongly affected by weather conditions, and forecast errors may lead to rapid net-load ramping, insufficient reserve capacity, branch overload, and limited renewable accommodation. To address the operational requirements of high-penetration wind-solar power systems, a real-time dynamic dispatch strategy based on deep reinforcement learning is constructed. Nodal load, wind-solar forecast errors, conventional generator output, battery state of charge, branch loading ratio, and real-time electricity price are incorporated into the state space, while the SAC algorithm outputs continuous dispatch actions for coordinated thermal generation adjustment, battery charging and discharging, and interruptible load response. Experiments are conducted on a modified IEEE 118-bus system, where MATPOWER is used for AC power-flow verification and Python is used to build the reinforcement learning environment. PPO, DDPG, and rolling economic dispatch are selected as comparative methods. The results show that SAC reduces the daily operation cost to 45,820 USD, decreases the renewable curtailment rate to 4.82%, and lowers load shedding to 0.71 MWh under 70% wind-solar penetration. It also effectively controls branch-overload events in most medium forecast-error scenarios. These results indicate that deep reinforcement learning can form a stable dispatch balance among economic efficiency, renewable accommodation, and security constraints, providing a feasible computational solution for real-time control in high-renewable smart grids.

Keywords: smart grid, deep reinforcement learning, real-time dynamic dispatch, high renewable penetration, security-constrained optimization

1. Introduction

High-penetration wind and photovoltaic integration is reshaping the operational structure of smart grids, shifting power dispatch from controllable conventional generation toward real-time control under uncertainty, rapid fluctuation, and multi-resource coordination. Wind and solar outputs are highly affected by weather conditions, and short-term forecast errors may cause fast net-load

DOI: 10.54254/2755-2721/2026.36888

ramping, insufficient reserve capacity, and local branch overload. Very short-term renewable forecasting studies have shown that minute-ahead and hour-ahead prediction are increasingly important for renewable integration under fast operating conditions [1]. Existing research has introduced machine learning and reinforcement learning into renewable accommodation, microgrid energy management, and economic dispatch, while many applications still focus on local systems or limited objective settings. Dynamic energy dispatch in IoT-driven microgrids indicates that learning from historical operation data can support intelligent coordination among distributed generation, photovoltaic units, and batteries [2]. Around real-time dynamic dispatch in smart grids, a deep reinforcement learning strategy can incorporate nodal load, wind-solar forecast errors, generator output, battery SOC, and branch loading into the state space. A DRL-based active power dispatch framework has also demonstrated the feasibility of obtaining near-optimal dispatch actions under unseen load scenarios [3]. The research content therefore focuses on continuous control of thermal generation, storage charging and discharging, and flexible load response, aiming to reduce operating cost while improving renewable accommodation and operational security.

2. Literature review

2.1. Renewable power forecasting

Renewable power forecasting provides essential state information for real-time dispatch because it transforms meteorological changes into short-term power signals that can be interpreted by the dispatch system. Time-series models capture historical trends, machine learning methods identify nonlinear input-output relationships, and deep learning models strengthen the representation of multivariate features such as wind speed, solar irradiance, temperature, and cloud cover. A systematic review of deep learning for renewable energy forecasting highlights that data preprocessing, deterministic prediction, probabilistic prediction, and model comparison directly affect the reliability of forecasting outputs [4]. As the forecasting horizon moves from day-ahead planning to minute-level and hour-level control, prediction accuracy has a more immediate influence on dispatch security. Even small errors may become significant power deviations when wind and solar penetration is high. Recent work on multivariate wind and solar forecasting further shows that spatial correlation, temporal dependence, and forecast reconciliation should be considered when renewable sources are jointly integrated into energy systems [5]. These findings indicate that forecasting should provide predicted values together with uncertainty-related information, allowing the dispatch policy to arrange reserves, storage operation, and flexible loads according to the level of forecast deviation.

2.2. Reinforcement learning dispatch

The reinforcement learning dispatch considers the power system operation as a sequential decision-making problem in which the policy can be obtained by observing states, performing actions, and receiving rewards. As opposed to the optimization approaches based on mathematical models, deep reinforcement learning is better to be used in dynamic environments affected by the renewable energy production, load changes, and variable electricity prices. DDPG provides deterministic continuous actions, thus it can be used for the output control of generators and battery powers [6]. PPO solves the problem of training instability by limiting the change of policies, which is important for dispatch actions to stay in feasible zones. SAC adds entropy regularization into the policy update, and it becomes adaptable to high-dimensional state space and continuous action space [7].

DOI: 10.54254/2755-2721/2026.36888

The current works have already shown the possibility of using the DRL in microgrid energy management and active power dispatch. However, there are additional challenges in learning at the level of transmission system, including dealing with the limits of branch flows, security constraints, limitations on storage use, and multi-resource coordination. Thus, dispatch actions should be closely coupled with the power system constraints.

2.3. Security-constrained control

Security-constrained control plays an essential part in deciding whether the deep reinforcement learning-based dispatch policy can shift from simulation to engineering implementation. Power balance, branch flow constraints, bus voltage constraints, generator ramping constraints, and storage SOC constraints should be considered for real-time dispatch due to the fact that cost-driven actions might generate local overloading and device limit constraints. MATPOWER offers steady-state power flow and optimal power flow modules which are appropriate for embedding physical validation into simulation-based dispatch scenarios [8]. The significance of the application of security constraint in high wind-solar penetration cases would be larger since renewable power fluctuation could rapidly change power flow patterns. An approach that is suitable for high wind-solar penetration is the combination of power flow, action boundary correction, and penalty design in the reinforcement learning framework. In this way, the agent not only obtains economic rewards but also learns the consequence of illegal actions [9]. With such design of the connection between reward design and physical validation, the dispatch policy could achieve fast response time and low infeasibility decisions in the process of intelligent dispatch.

3. Experimental methods

3.1. Simulation system setting

The simulation system is built on a modified IEEE 118-bus system, preserving the original network topology, branch impedance, and load distribution. Ten generation buses are retained as thermal units, six generation buses are connected to wind farms, four generation buses are connected to photovoltaic stations, and three battery energy storage systems are placed in load-intensive areas. The system base capacity is set to 100 MVA. The dispatch horizon is 24 h with a 15 min interval, producing 96 real-time dispatch periods. Wind-solar installed penetration is set to 70%, and wind and photovoltaic forecast errors are controlled within 5%–25%. The storage capacity is 50 MWh, the maximum charging or discharging power is 10 MW, and the SOC range is limited to 0.2–0.9. MATPOWER conducts AC power-flow verification, while Python encapsulates the reinforcement learning environment, including state updating, action execution, and reward feedback.

3.2. State-action design

The state space consists of nodal load, wind-solar forecast output, forecast error, current thermal generation, battery SOC, branch loading ratio, and real-time electricity price. The action space includes thermal power adjustment, battery charging or discharging power, and interruptible load regulation [10]. Before each action is executed, the system power-balance constraint must be satisfied, as shown in Equation (1).

$$\sum_{i=1}^{N_g} P_{g,i}^{t}+P_{w}^{t}+P_{s}^{t}+P_{dis}^{t}-P_{ch}^{t}+P_{shed}^{t}=P_{L}^{t} \tag{1}$$

DOI: 10.54254/2755-2721/2026.36888

Here, $P_{g,i}^{t}$ denotes the output of thermal unit i, P_{w}^{t} and P_{s}^{t} denote wind and photovoltaic output, P_{dis}^{t} and P_{ch}^{t} denote battery discharging and charging power, P_{shed}^{t} denotes load shedding, and P_{L}^{t} denotes system load. The reward function integrates generation cost, storage degradation, renewable curtailment, load shedding, and security violations into one objective, as shown in Equation (2).

$$r_t = -\left(C_g^t + C_b^t + \lambda_1 P_{cur}^t + \lambda_2 P_{shed}^t + \lambda_3 V_{vio}^t + \lambda_4 L_{vio}^t\right) \tag{2}$$

Here, C_g^t is generation cost, C_b^t is battery degradation cost, P_{cur}^t is renewable curtailment, V_{vio}^t and L_{vio}^t are voltage-violation and branch-overload penalties, and λ_1 to λ_4 are weighting coefficients.

3.3. Algorithm training procedure

SAC is selected as the core algorithm, while PPO, DDPG, and rolling economic dispatch are used as comparative methods. Each training episode starts from a randomly selected operating day. The environment reads the load profile, wind-solar forecast values, initial generator outputs, and initial storage SOC. The agent then outputs continuous dispatch actions, and MATPOWER performs power-flow calculation to return the constraint status. If branch overload, voltage deviation, or SOC violation occurs, the environment updates the feedback through reward penalties and action-boundary correction. The transition sample (s_t, a_t, r_t, s_{t+1}) is stored in the replay buffer, and mini-batch samples are used to update the Q-network and policy network [11]. The SAC policy update objective is shown in Equation (3).

$$J_\pi(\theta) = E_{s_t \sim D, a_t \sim \pi_\theta}\left[\alpha \log \pi_\theta(a_t|s_t) - Q_\phi(s_t, a_t)\right] \tag{3}$$

Here, θ denotes policy-network parameters, ϕ denotes Q-network parameters, D denotes the replay buffer, and α denotes the entropy temperature coefficient. Training stops when the reward curve stabilizes or the maximum number of episodes is reached. Testing is conducted on unseen high-volatility wind-solar scenarios to evaluate generalization.

4. Results

4.1. Economic dispatch results

Economic results are calculated from 96 intra-day dispatch intervals in the modified IEEE 118-bus system under 70% wind-solar penetration. As shown in Table 1, the daily operation cost of SAC is 45,820 USD, which is approximately 6.34% lower than the 48,920 USD obtained by rolling economic dispatch. This result indicates that SAC can coordinate thermal generation and battery charging or discharging more efficiently. The renewable curtailment rate of SAC is 4.82%, lower than 5.41% for PPO and 6.08% for DDPG, indicating stronger renewable accommodation during high wind-solar output periods. The equivalent battery cycle number of SAC is 1.74, close to PPO and DDPG, showing that the cost reduction is not achieved through excessive battery cycling.

DOI: 10.54254/2755-2721/2026.36888

Table 1. Economic comparison of different dispatch methods

Method	Daily cost/USD	Fuel cost/USD	Curtailment rate/%	Load shedding/MWh	Battery equivalent cycles
Rolling ED	48,920	43,610	7.36	1.84	1.21
DDPG	47,860	42,780	6.08	1.25	1.69
PPO	46,970	41,940	5.41	0.96	1.63
SAC	45,820	40,880	4.82	0.71	1.74

4.2. Secure operation results

Security results further examine branch-overload risk under the joint variation of wind-solar penetration and forecast error. As shown in Figure 1, when wind-solar penetration increases from 40% to 80% and forecast error increases from 5% to 25%, the average daily branch-overload events increase from 0.4 to 10.8. This result indicates that high penetration and high forecast error jointly amplify power-flow security pressure. Under 70% wind-solar penetration, overload events increase from 2.9 to 8.5 when forecast error rises from 10% to 25%, showing that forecast deviation becomes a key risk source in real-time dynamic dispatch. With security penalties and action-boundary correction, SAC keeps overload events below 5 in most medium-error scenarios, indicating that it can identify stressed power-flow states and reduce local overload risk through coordinated generator and storage actions.

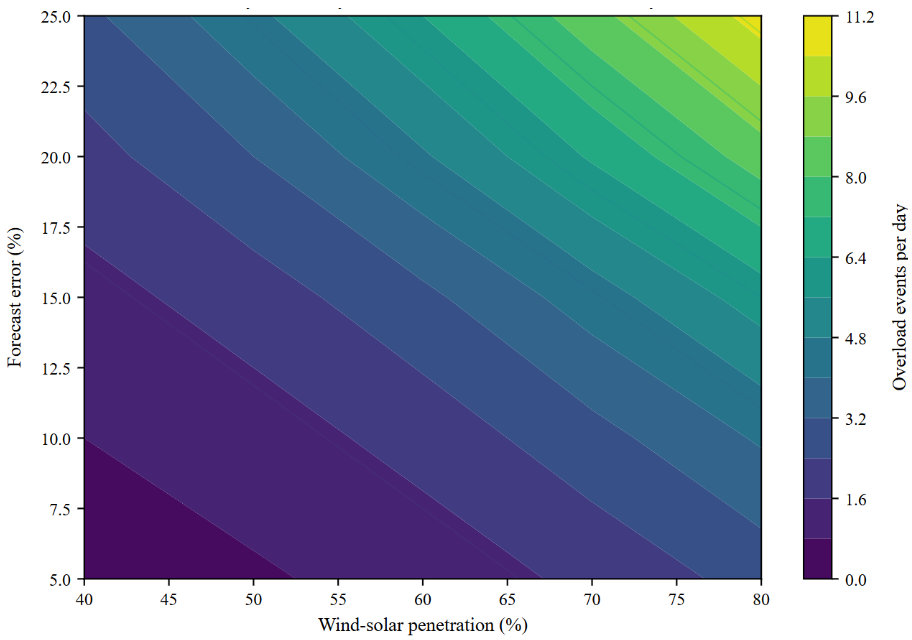

Figure 1. Security sensitivity surface under renewable uncertainty

5. Discussion

The economic results show that SAC achieves stronger continuous-action optimization than rolling economic dispatch, DDPG, and PPO. Its cost reduction mainly comes from coordinated thermal

DOI: 10.54254/2755-2721/2026.36888

generation regulation and battery charging or discharging schedules, rather than excessive storage cycling. The lower renewable curtailment rate indicates that the policy can absorb wind and solar power more effectively during high-output periods and release storage flexibility during load peaks. The security results further show that the simultaneous increase in wind-solar penetration and forecast error significantly raises branch-overload risk. What it implies is that in situ dispatch cannot depend solely on economic motivations; power flow security and action boundary have to be incorporated in action feedback. SAC produces less overload cases in most medium errors, which proves the validity of security cost and action-boundary adjustments. Nevertheless, in extreme situations like an 80% penetration level and 25% forecasting error, the risk of overload remains high, which implies that the influence of scenario representation, reserve setup, and security validation persists.

6. Conclusion

A deep reinforcement learning-based wind-solar-storage coordinated control framework is developed for real-time dynamic dispatch in high-penetration renewable power systems. The simulation system is built on a modified IEEE 118-bus network, combining MATPOWER power-flow verification with a Python reinforcement learning environment. Wind-solar forecast errors, thermal generator output, battery SOC, branch loading ratio, and real-time electricity price are used as core state variables, and SAC is applied to learn continuous dispatch actions. The experimental results show that the proposed strategy can reduce daily operation cost, renewable curtailment rate, and load shedding, while controlling branch-overload risk in most operating scenarios. SAC algorithm shows better performance than rolling economic dispatch and other reinforcement learning methods in terms of economic efficiency, stability, and accommodation of renewables. Further research may consider introducing carbon emission restrictions, demand-side response elements, extreme weather cases, and coordination between multiple regions for increasing flexibility and practical significance of the intelligent dispatch strategy.

References

[1] Pinciroli, L., et al. (2023). Optimal operation and maintenance of energy storage systems in grid-connected microgrids by deep reinforcement learning. *Applied Energy, 352*, 121947.

[2] Domínguez-Barbero, C., et al. (2024). Energy management of a microgrid considering nonlinear losses in batteries through deep reinforcement learning. *Applied Energy, 368*, 123435.

[3] Mansouri, M., et al. (2024). A cloud-fog computing framework for real-time energy management in multi-microgrid system utilizing deep reinforcement learning. *Journal of Energy Storage, 97*, 112912.

[4] Hedayatnia, A., et al. (2024). Two-stage data-driven optimal energy management and dynamic real-time operation in networked microgrid based on a deep reinforcement learning approach. *International Journal of Electrical Power & Energy Systems, 160*, 110142.

[5] Jayaraj, S., et al. (2024). Reinforcement learning for energy management of an islanded microgrid: Analysis of state, state space and reinforcement function. *Sustainable Energy, Grids and Networks, 38*, 101362.

[6] Jonban, M. S., et al. (2024). A reinforcement learning approach using Markov decision processes for battery energy storage control within a smart contract framework. *Journal of Energy Storage, 86*, 111342.

[7] Bui, V.-H., et al. (2025). A critical review of safe reinforcement learning strategies in power and energy systems. *Engineering Applications of Artificial Intelligence, 143*, 110091.

[8] Panda, S. P., et al. (2024). Methods for mitigating uncertainty in real-time operations of a connected microgrid. *Sustainable Energy, Grids and Networks, 38*, 101334.

[9] Santosuosso, L., et al. (2024). Stochastic economic model predictive control for renewable energy and ancillary services trading with storage. *Sustainable Energy, Grids and Networks, 38*, 101373.

DOI: 10.54254/2755-2721/2026.36888

[10] Sørensen, M. L., et al. (2023). Recent developments in multivariate wind and solar power forecasting. *Wiley Interdisciplinary Reviews: Energy and Environment, 12*(2), e465.

[11] Campbell, A. M., et al. (2024). A multi-decadal hourly coincident wind and solar power production dataset for the contiguous United States. *Scientific Data, 11*(1), 1121.