Original Article | Open Access

AI Agent

Zhang *et al. AI Agent* **2026**, *2*, 21 **DOI:**10.20517/aiagent.2026.01

An energy-efficient scheduling approach for wind-solar-hydrogen systems based on distributed reinforcement learning

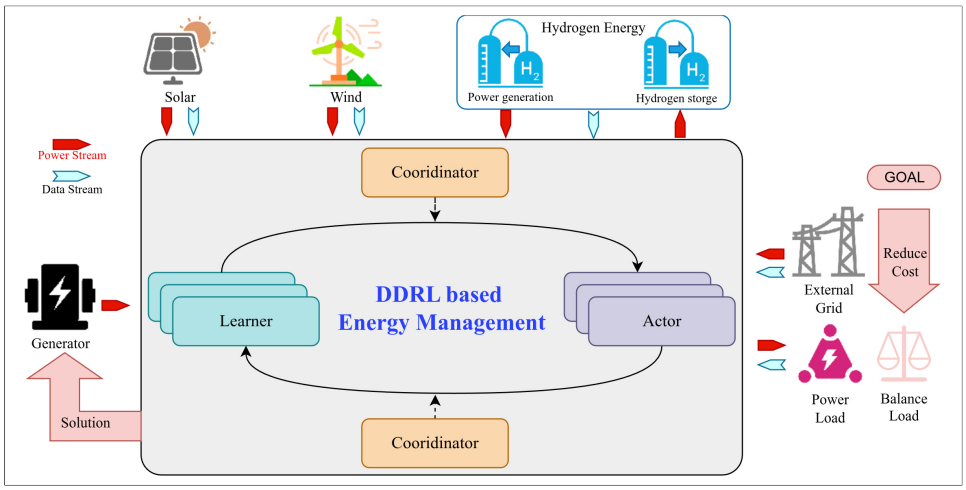
Bo Zhang[1], Conghao Wang[1], Yan Ma[1], Jingjing Xie[1], Liang He[1]

Keywords:
Carbon trading, multi-objective optimization, distributed reinforcement learning, wind-solar-hydrogen systems

Citation: Zhang, B.; Wang, C.; Ma, Y.; Xie, J.; He, L. An energy-efficient scheduling approach for wind-solar-hydrogen systems based on distributed reinforcement learning. *AI Agent* **2026**, *2*, 21. https://dx.doi.org/10.20517/aiagent.2026.01

Received: 22 Jan 2026
First Decision: 20 Mar 2026
Revised: 3 Apr 2026
Accepted: 12 May 2026
Published: 29 May 2026

Academic Editor:
Hao Li
Copy Editor:
Xing-Yue Zhang
Production Editor:
Xing-Yue Zhang



Abstract

This paper presents a comprehensive energy dispatch strategy based on distributed reinforcement learning to optimize the operation of integrated wind-solar-hydrogen systems. The proposed approach effectively reduces coal fuel costs and carbon emissions while ensuring precise load demand tracking. By implementing a distributed computing framework, the computational challenges associated with training the Deep Deterministic Policy Gradient algorithm on large-scale datasets are effectively addressed. This parallel architecture significantly enhances training efficiency and improves scalability for complex energy management tasks. Additionally, an efficient load pattern identification method, enhanced by Principal Component Analysis and K-means clustering, is developed to capture the salient characteristics of electricity load data. Furthermore, a high-fidelity representative scenario extraction approach, utilizing Dynamic Time Warping and Density-Based Spatial Clustering of Applications with Noise, is proposed to characterize the inherent uncertainties in wind and solar power generation. The integration of hydrogen-based energy storage is proposed as a flexible and sustainable solution to enhance system reliability and mitigate carbon emissions. Empirical simulation results demonstrate that the proposed methodology significantly reduces fuel costs and minimizes carbon emissions while exhibiting improved robustness and computational efficiency. By

School of Software, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China.

Correspondence to: Assoc. Prof. Yan Ma, School of Software, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China. E-mail: yma@nwpu.edu.cn

incorporating hydrogen storage systems and carbon trading mechanisms, the proposed approach optimally facilitates the integration of wind and solar power, thereby providing a comprehensive framework for the efficient operation of hybrid energy systems.

INTRODUCTION

The Paris Agreement establishes a critical global objective to significantly mitigate greenhouse gas emissions, targeting a temperature increase limit of well below 2 °C relative to pre-industrial levels. In alignment with this international framework, numerous nations have formulated Nationally Determined Contributions (NDCs), committing to achieve net-zero emissions by 2050[1]. As a pivotal participant in this global effort, China has pledged a comprehensive transition from fossil fuel-based power generation to renewable energy technologies, aiming for carbon neutrality by 2060[2]. This commitment was formally institutionalized during the United Nations General Assembly, where China unveiled its "3060 Dual Carbon" goal, outlining a strategic 40-year roadmap toward net-zero emissions[3]. Nevertheless, the nation encounters a formidable challenge due to its profound reliance on coal for thermal power generation, which accounts for approximately 40% of its total carbon footprint[4-6]. Consequently, the modernization of energy infrastructure is imperative for China to fulfill its rigorous emission reduction targets[7,8]. Driven by the accelerated deployment of renewable resources and advancements in energy efficiency, contemporary research has increasingly focused on the synergistic integration of heterogeneous renewable sources. Extensive studies highlight promising configurations, such as wind-solar hybrid systems[9-12] and multi-source architectures incorporating wind, solar, thermal, and energy storage units[13-16]. Among these architectures, wind-solar-hydrogen systems are recognized as essential for low-carbon power generation, offering enhanced reliability and economic efficiency compared to standalone power plants[17,18]. However, the large-scale integration of wind and solar energy - projected to be China's dominant electricity sources - requires addressing critical challenges related to inherent intermittency[19], load fluctuations, and seasonal supply-demand mismatches. To navigate these complexities, the hydrogen energy storage system (HESS) has emerged as a viable solution, facilitating long-duration and seasonal energy balancing within high-renewable grids. Hydrogen storage enhances the temporal flexibility of clean energy utilization and supports decarbonized electricity production while enabling strategic participation in carbon markets. Consequently, an assessment of the spatiotemporal complementarity between wind-solar resources and their alignment with electricity demand is vital for developing high-performance, renewable-dominated power systems. Existing research suggests that the decarbonization of power generation can be effectively pursued through hybrid wind-solar-hydrogen systems, although carbon footprints vary significantly depending on the specific generation strategies employed[20-23]. This variation underscores the urgent necessity of optimizing CO_2 reduction strategies within integrated wind-solar-hydrogen frameworks. Beyond traditional optimization, autonomous AI agents are emerging as powerful tools for accelerating scientific discovery and conducting end-to-end research tasks in the hydrogen energy sector[24].

The intermittent nature of renewable energy sources, specifically wind and solar power, presents significant operational challenges due to their inherent dependence on environmental and climatic conditions. This inherent variability in power generation not only exacerbates the risk of system failures but also introduces substantial uncertainty into energy dispatch modeling[25-27]. Contemporary energy dispatch approaches are broadly categorized into uncertainty-based and deterministic dispatch models. The former category addresses renewable variability by simulating probabilistic power output scenarios to achieve environmentally differentiated generation through scenario feature identification[28]. Conversely, deterministic models rely on forecasted generation curves for long-term planning while maintaining sufficient reserve capacity to accommodate fluctuations, making them suitable for conventional units, demand response systems, electric vehicles, and energy storage applications. To better characterize wind

power uncertainty, advanced statistical methods have been developed to enhance modeling accuracy. For instance, the Wasserstein distance metric enables the derivation of optimal discrete distributions for wind power generation[29], while the Frank-Copula function facilitates the construction of joint probability distributions. These techniques, integrated with roulette wheel selection methods, generate comprehensive base scenario sets that can be further refined through improved spectral clustering algorithms incorporating noise filtering and distance correlation strategies. This systematic refinement process yields representative scenario sets that support two-level collaborative optimization models for integrated energy systems[30]. Recent methodological advancements include the quantile regression technique combined with dimensionality reduction clustering techniques, which leverage historical statistical data to transform deterministic prediction sequences into probabilistic output scenario sets[31]. Alternative approaches employ Latin hypercube sampling for scenario generation followed by fast antecedent elimination techniques to enhance scenario reduction efficiency[32]. In the present study, we implement a Principal Component Analysis (PCA)-enhanced K-means clustering approach for efficient load pattern identification[33]. Furthermore, we develop a high-fidelity representative day extraction method based on Dynamic Time Warping (DTW) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for annual renewable generation data. These advanced techniques achieve significant data dimensionality reduction while preserving critical temporal characteristics, thereby providing a robust foundation for subsequent energy dispatch strategy implementation. The methodological framework maintains a rigorous representation of seasonal variations in wind and solar power generation patterns, ensuring the practical applicability of the resulting dispatch solutions.

Advanced energy dispatch methodologies are instrumental in enhancing the operational performance and reliability of multi-source power supply architectures. Current energy dispatch approaches for such systems can be categorized into three main types: rule-based, optimization-based, and learning-based methods[34-36]. Rule-based strategies, while straightforward to implement, face significant limitations due to their dependence on predefined human knowledge. These approaches struggle with diverse input scenarios and require substantial time and resources to maintain as operational conditions evolve[34,37]. Optimization-based methods have demonstrated significant efficacy in resolving problems defined by complex sub-problem interdependencies and nested optimal substructures. Robust optimization algorithms have been successfully implemented across both wind and solar energy frameworks[38,39], whereas particle swarm optimization has yielded promising results for hybrid renewable configurations, facilitating the development of coordinated control strategies for energy storage[40]. The Sequential Least Squares Programming (SLSQP) algorithm has been utilized to develop integrated dispatch frameworks for multi-source energy management, markedly enhancing resource utilization within wind-solar-hydrogen architectures. Although Convex Optimization based SLSQP implementations have been validated for augmenting dispatch efficacy and satisfying load requirements, they are often characterized by significant computational overhead, necessitating extensive iterations to achieve convergence[35]. Several challenges persist in current dispatch approaches. The inherent uncertainty of wind and solar generation remains inadequately addressed, and power adjustment mechanisms across supply-side components require further refinement. Additionally, while multi-energy, multi-device architectures require hourly resolution for modeling accuracy, the resulting multi-scale dispatch problems often exhibit prohibitive computational complexity, complicating the trade-off between real-time efficiency and long-term system stability. These computational demands stem from the need to simultaneously optimize multiple energy sources while accounting for their temporal variability and operational constraints. A comprehensive assessment of environmental costs in wind-solar-hydrogen systems requires multi-dimensional analysis, particularly examining the interplay between integrated energy-carbon pricing mechanisms and the system's operational dynamics in relation to carbon emissions. Although current research underscores the pivotal role of energy storage systems in facilitating low-carbon transitions[41], contemporary economic evaluations suggest that carbon trading frameworks primarily

incentivize emission reductions through operational enhancements in energy efficiency, rather than catalyzing foundational industrial transformations. This insight carries important implications for the economic viability of large-scale power generation facilities pursuing de-carbonization strategies[42]. Furthermore, the dual approach of enhancing energy efficiency while optimizing industrial structures has been shown to augment the complementary emission mitigation capacity within emissions trading schemes, suggesting promising pathways for more effective climate mitigation in the power sector.

Existing deep reinforcement learning (DRL) applications in energy dispatch frequently rely on centralized or single-agent architectures, which are susceptible to the curse of dimensionality and exhibit limited scalability when managing large-scale, multi-source systems[35,36]. Furthermore, prior economic formulations typically overlook complex market interactions such as carbon emission trading mechanisms. To address these critical gaps, this work proposes a comprehensive, system-level energy dispatch framework for a grid-connected wind-solar-hydrogen integrated energy system that incorporates thermal backup. As shown in Figure 1, the framework includes four components. First, a PCA-enhanced K-means clustering model identifies typical load patterns from annual data, reducing computation burdens while maintaining accuracy. Second, a DTW-DBSCAN method extracts representative days from renewable generation data, capturing seasonal and uncertain variations. Third, a DDPG algorithm optimizes dispatch strategies through continuous agent-environment interaction, allowing real-time adjustments and improving long-term performance. Fourth, HESS is integrated to provide long-duration and seasonal energy balancing. HESS enables bidirectional conversion between electricity and hydrogen. Although its efficiency is nonlinear and power-dependent, a constant round-trip efficiency is assumed for system-level modeling. Distributed learning improves training speed and supports the inclusion of carbon trading signals in dispatch decisions. The main contributions of this paper are summarized as follows: (1) Develops a joint data-driven pipeline integrating PCA-enhanced K-means clustering and DTW-DBSCAN to mitigate computational complexity while preserving high-fidelity representation of renewable generation uncertainties; (2) Proposes a highly scalable distributed training architecture that overcomes the dimensionality and computational limitations of conventional centralized DRL methods, significantly accelerating training speed and enabling efficient, stable scheduling in complex wind-solar-hydrogen systems; (3) Formulates a comprehensive economic model that systematically internalizes coal fuel consumption, carbon emission costs under trading mechanisms, and electricity procurement expenses into the multi-objective optimization process.

The remainder of this paper is organized as follows. Section EXPERIMENTAL presents the experimental methodology, including model development for the wind-solar-hydrogen systems and the design of the distributed reinforcement learning methods. Section RESULTS AND DISCUSSION provides the results and discussion across various scenarios. Finally, Section CONCLUSIONS concludes the paper.

EXPERIMENTAL

Efficient load pattern identification: A PCA-enhanced and K-means clustering model

Clustering power load data is essential for identifying representative load patterns, which are fundamental to load forecasting, demand response programs, and efficient grid management. Traditional clustering techniques, such as Gaussian Mixture Models (GMM), have been extensively utilized due to their capacity to characterize complex data distributions via probabilistic assignments. However, GMM exhibits significant limitations: it is computationally burdensome for high-dimensional datasets and relies heavily on the assumption of underlying Gaussian distributions, which frequently fails to align with empirical load profiles. Furthermore, GMM requires meticulous initialization and shows high sensitivity to hyper-parameter configurations, including the number of components and covariance structures. To circumvent these challenges, we propose a framework integrating PCA with K-means clustering. PCA effectively reduces dimensionality by extracting dominant features, thereby enhancing computational throughput and

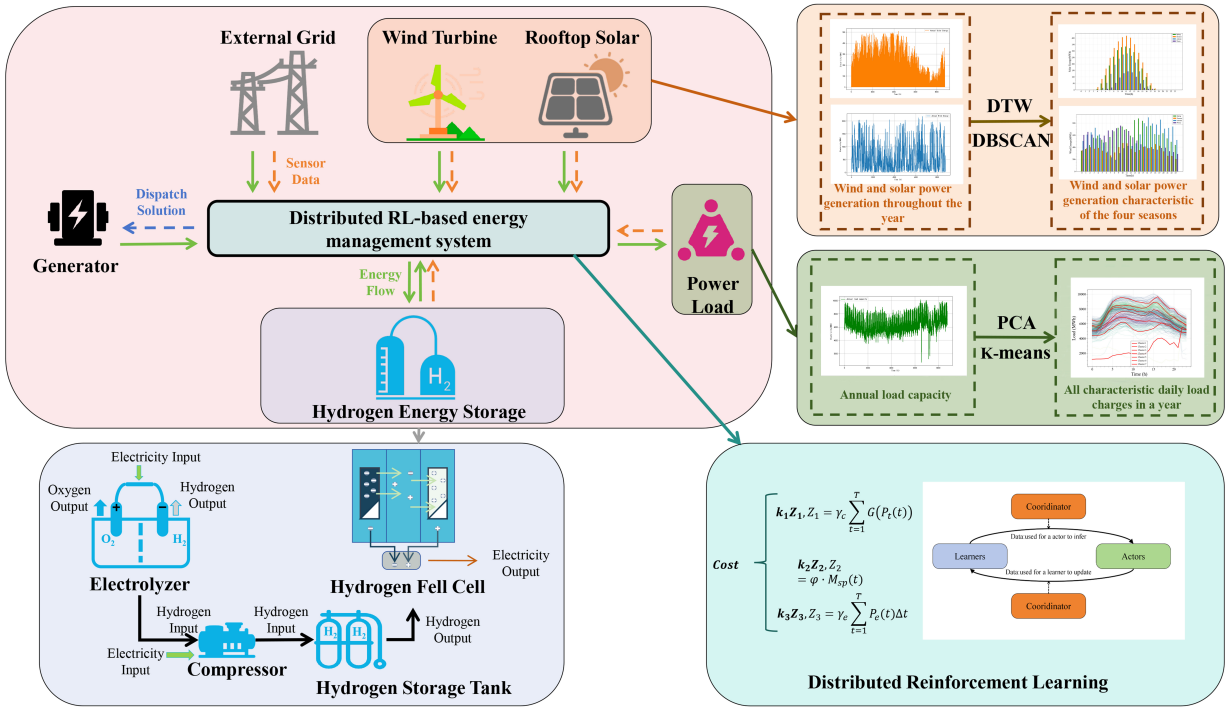


Figure 1. The framework includes four components: (1) a PCA-enhanced K-means clustering model for identifying load patterns; (2) a DTW-DBSCAN method for extracting representative renewable generation days; (3) a DDPG algorithm for optimizing dispatch strategies; and (4) HESS integration for long-duration energy balancing. PCA: Principal Component Analysis; DTW: Dynamic Time Warping; DBSCAN: Density-Based Spatial Clustering of Applications with Noise; DDPG: Deep Deterministic Policy Gradient; HESS: hydrogen energy storage system.

mitigating the impact of noise. When applied to the reduced feature space, K-means clustering offers a rapid and interpretable methodology for partitioning load profiles into distinct, meaningful cohorts. Compared to GMM, the proposed PCA-K-means approach demonstrates enhanced scalability, reduced computational overhead, and greater ease of implementation, making it highly suitable for large-scale utility data analytics. This synergistic combination not only preserves the intrinsic structural properties of the data but also facilitates the robust identification of representative power consumption patterns.

Let $X \in \mathrm{R}^{n \times 24}$ represent the matrix of normalized daily load profiles, where each row denotes a specific day and each column corresponds to an hourly interval. PCA is subsequently employed to project this data into a lower-dimensional feature space while retaining the maximum variance. The optimal number of principal components, m, is determined such that the cumulative explained variance exceeds a predefined threshold, τ. In this study, the threshold τ is established at 0.95 to ensure high information retention. Following dimensionality reduction, K-means clustering is applied to the transformed dataset $Z \in \mathrm{R}^{n \times m}$. The objective is to partition the dataset into k distinct clusters, C_i, by minimizing the total intra-cluster variance.

$$\underset{C_1,\ldots,C_k}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{z_j \in C_i} \| z_j - \mu_i \|^2$$

where z_j denotes the j-th load profile in the reduced feature space, and μ_i represents the centroid of cluster C_i. For each cluster, a representative day is identified as the original profile whose lower-dimensional representation is closest to the respective cluster centroid.

$$x_i^* = \underset{x_j \in C_i}{\operatorname{argmin}} \| z_j - \mu_i \|$$

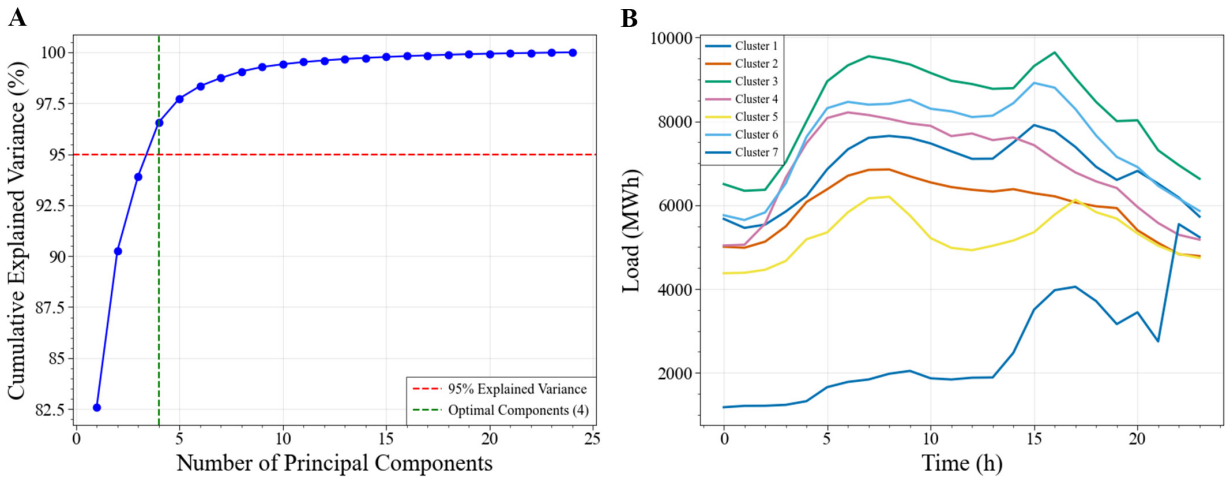


Figure 2. Cluster results. (A) Cumulative explained variance; (B) Representative days load.

where x_j signifies the original 24-hour load curve corresponding to the reduced data point z_j. The resulting set $\{x_i^*,\ldots,x_k^*\}$ effectively captures the characteristic variability of annual demand with reduced computational complexity. This extracted set provides a robust foundation for subsequent multi-objective dispatch optimization.

Figure 2A illustrates the cumulative explained variance as a function of the dimensionality of the retained principal subspace. The analysis reveals that a specific quantity of principal components is essential to sustain a cumulative variance threshold of at least 95%. This provides a rigorous theoretical justification for dimensionality reduction, ensuring that the intrinsic characteristics of the original data are preserved while simplifying the data structure. As demonstrated by the results, the cumulative explained variance exceeds 95% when the principal component count is established at four. This configuration ensures the retention of the majority of informational content without incurring significant computational overhead from excessive dimensionality. Figure 2B presents the load profiles of seven representative days extracted via the K-means clustering algorithm. Each individual curve characterizes the typical load pattern of a specific cluster, reflecting temporal variations over a 24-hour horizon. A comparative analysis across these clusters reveals distinct differences, particularly in the timing of peak loads and the magnitude of fluctuations. These representative profiles offer crucial insights for load forecasting and power system scheduling, facilitating the identification of diverse consumption patterns and their operational implications for the grid.

Figure 3A illustrates the distribution of the dataset within the reduced PCA space, projected onto the first two principal components. In this visualization, distinct colors signify the clustering results obtained via the K-means algorithm. The clear segregation of color-coded points demonstrates the effective identification of a robust clustering structure within the reduced-dimensional feature space. Furthermore, Figure 3B exemplifies the percentage of explained variance for the first four principal components in a bar chart format. Quantitative values for explained variance are explicitly annotated above each bar, facilitating a direct comparison of the relative contribution of each individual component. By evaluating the explained variance, the significance of each principal component in preserving the original data variability can be rigorously assessed. This analysis consequently determines the optimal dimensionality to retain during the feature extraction process.

Figure 3C depicts the load profiles of all days within each cluster, with the representative day's load profile highlighted in red. Transparent curves in different colors represent the load variations of individual days

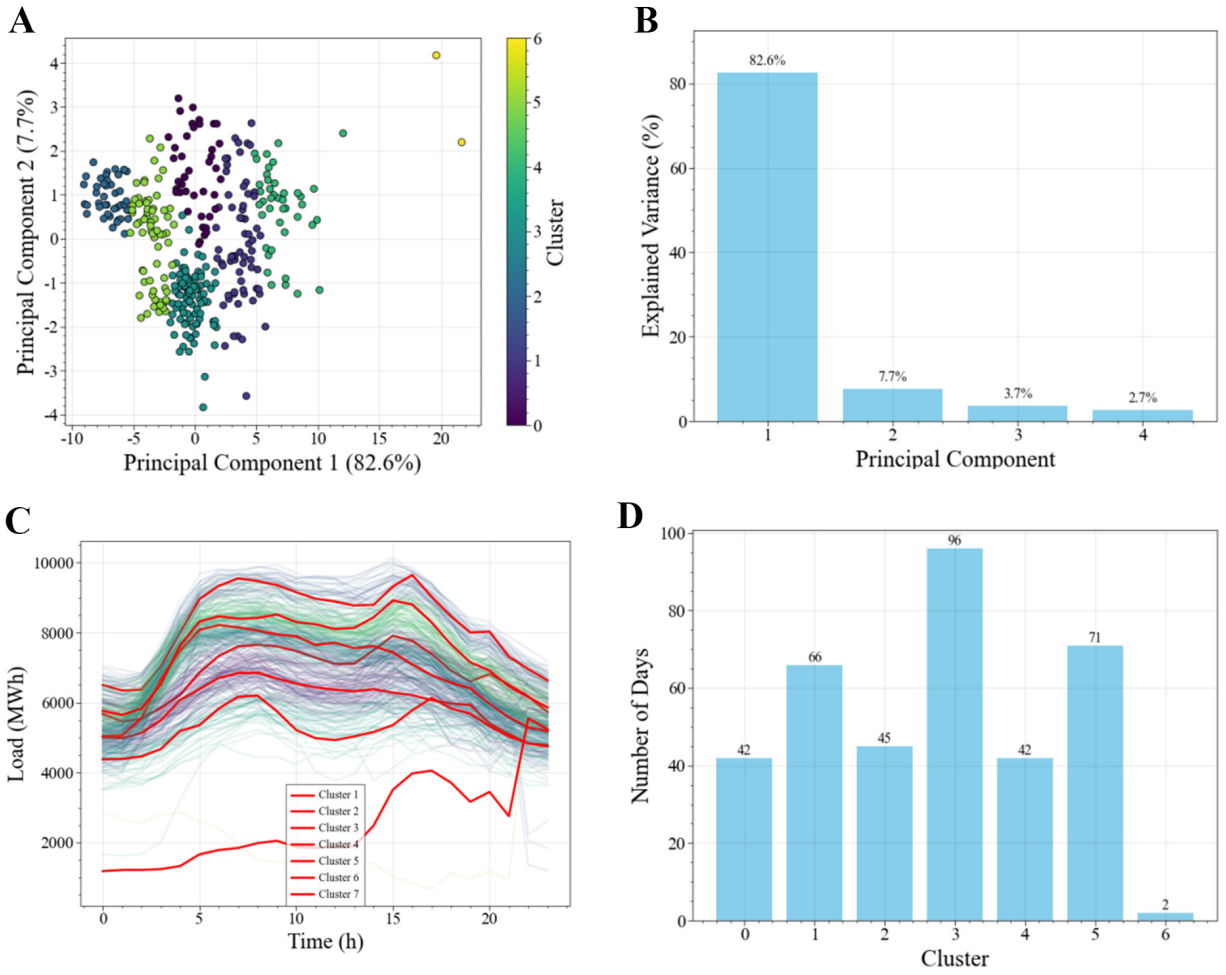


Figure 3. Cluster analysis. (A) PCA + K-means clustering (first two principal components); (B) Explanatory variance for the first four dimensions; (C) Load profiles for all days in each cluster; (D) Number of days in each cluster. PCA: Principal Component Analysis.

within the same cluster, while the red curve signifies the typical load pattern of the cluster. By comparing the load profiles within a cluster, the consistency of load patterns can be observed, and the red curve emphasizes the core characteristics of the cluster, and Figure 3D shows the size of the data in each cluster, and makes a special cluster for the anomalous data.

Renewable generation pattern identification: a high-fidelity DTW-DBSCAN model

Traditional methods for extracting representative days from wind and solar power generation time series often utilize Euclidean distance-based clustering techniques, such as K-means or arithmetic averaging. Although computationally efficient, these methods possess notable limitations, specifically their reliance on linear temporal alignment and their inability to capture phase shifts, peak duration variations, or other nonlinear temporal distortions inherent in renewable energy profiles. In contrast, the DTW-DBSCAN framework excels at capturing local temporal distortions and can autonomously filter noise through density thresholds, thereby eliminating the need for subjective pre-specification of cluster counts. This study introduces a hybrid methodology that integrates DTW and DBSCAN. The proposed approach segments annual generation data into 24-hour daily subsequences on a seasonal basis, constructs a DTW-based shape similarity distance matrix to quantify morphological alignment, and performs unsupervised clustering via DBSCAN to categorize days with similar fluctuation patterns. Unlike conventional centroid-based methods that rely on Euclidean assumptions, representative days in this study are selected as real-world samples

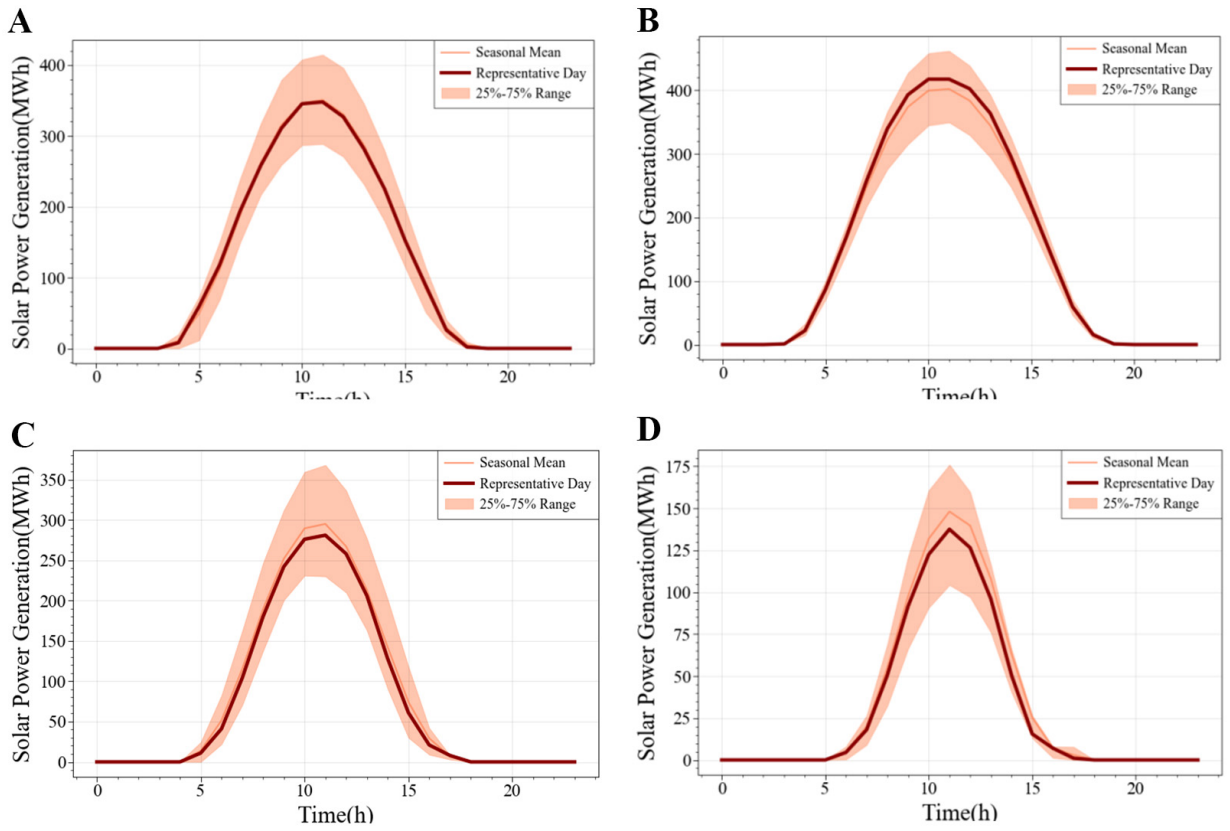


Figure 4. Solar power generation profiles during the year. (A) Spring; (B) Summer; (C) Autumn; (D) Winter.

located closest to the DTW-geometric centroids within each cluster. Given a set of daily generation curves $\{x_1,x_2,\ldots,x_n\}$, where each $x_i \in \mathrm{R}^T$ denotes the generation series for a day with T time intervals, the DTW distance between two sequences x_i and x_j is defined as follows:

$$DTW(x_i,x_j) = \min_{w \in W} \sum_{(p,q) \in w} \| x_i^p - x_j^q \|$$

In this context, W denotes the set of all valid warping paths, while (p, q) represents the aligned index pair between sequences. A pairwise DTW distance matrix $D \in \mathrm{R}^{n \times n}$ is then constructed and utilized as input for the DBSCAN clustering algorithm. DBSCAN effectively categorizes days with similar temporal patterns without the necessity of specifying a predefined number of clusters. The algorithm identifies clusters based on a neighborhood radius ε and a minimum number of points minPts required to establish a dense region. For each resulting cluster C_k, the most representative day is identified as

$$x_k^* = \arg\min_{x_i \in C_k} \sum_{x_j \in C_k} DTW(x_i,x_j).$$

The optimal selection process yields representative curves $\{x_1^*,\ldots,x_k^*\}$ by minimizing intra-cluster DTW distances, thereby effectively preserving the characteristic temporal patterns and inherent variability of wind and solar generation profiles. This methodology provides high-fidelity inputs for the planning and operational simulation of renewable-dominated energy systems, ensuring robustness against the temporal uncertainty and nonlinear dynamics intrinsic to wind and solar resources. The seasonal profiles for solar and wind power, along with their corresponding representative days, are illustrated in Figures 4 and 5, respectively. These profiles accurately capture the characteristic seasonal energy levels.

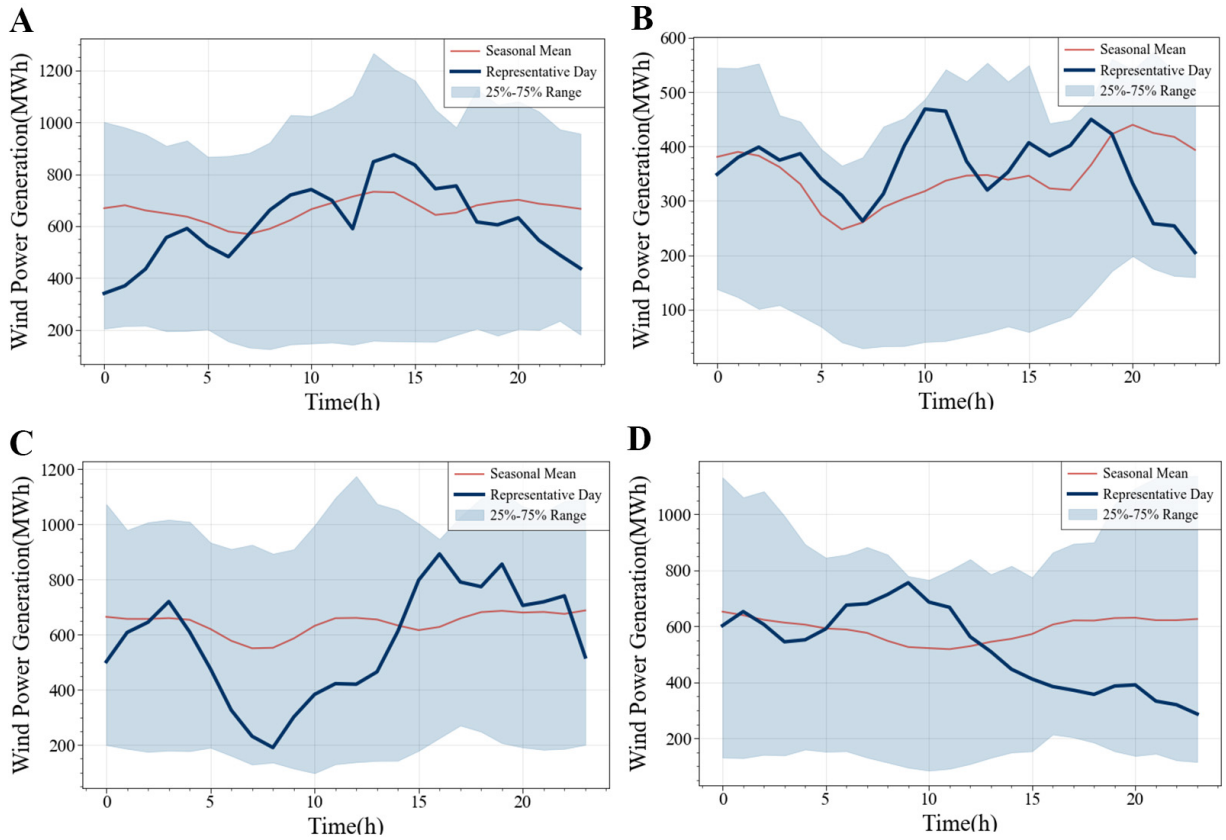


Figure 5. Wind power generation profiles during the year. (A) Spring; (B) Summer; (C) Autumn; (D) Winter.

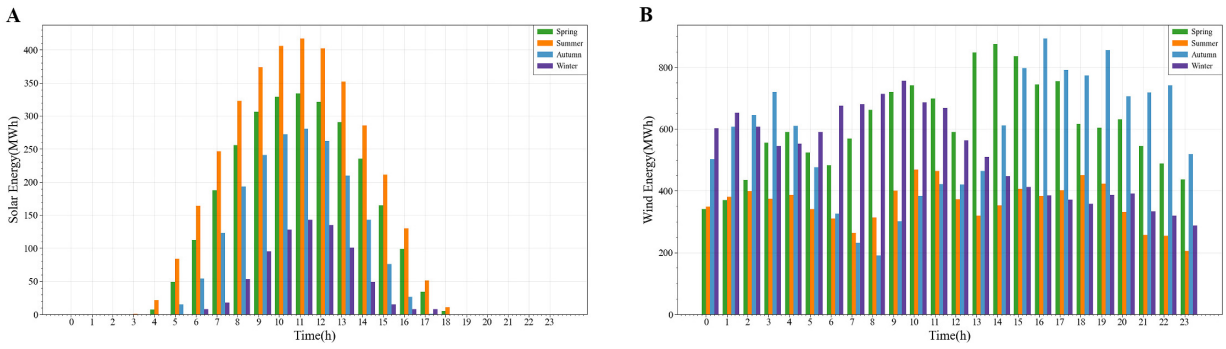


Figure 6. Diurnal power generation profiles across all seasons. (A) Solar power generation profiles; (B) Wind power generation profiles.

The representative power generation profiles for solar and wind resources under seasonal variations are depicted in Figure 6A and B, respectively. Observations indicate that solar generation typically peaks during daylight hours in spring and summer, whereas wind power yields higher output during the day in spring and autumn and transitions to nighttime peaks during winter. These results demonstrate that renewable energy generation is not stochastic but adheres to predictable diurnal and seasonal cycles. The presence of these inherent temporal patterns provides a robust foundation for employing clustering methods to effectively identify and characterize typical operational scenarios.

Hydrogen energy storage system model

The HESS configuration investigated in this research comprises an electrolyzer, a hydrogen storage vessel, and a proton exchange membrane fuel cell (PEMFC). This system serves exclusively as an energy conversion and storage medium, where the electrolyzer converts surplus renewable generation into stored hydrogen.

Conversely, the fuel cell regenerates electrical power by consuming the stored hydrogen during periods of supply deficit. In industrial practice, the energy conversion efficiency of hydrogen storage is influenced by various nonlinear parameters, including operating power, thermal conditions, and current density. While high-fidelity empirical models can precisely characterize these behaviors, they introduce substantial computational complexity and non-convexity to the optimization problem. To maintain mathematical tractability and computational efficiency, this study utilizes a linearized constant efficiency approximation. The operational efficiencies for the charging and discharging cycles are assumed to be fixed coefficients, represented by $\eta_{H,c}$ and $\eta_{H,d}$, respectively. For illustrative purposes, both efficiencies are assigned a value of 0.5[43], resulting in an aggregate round-trip efficiency of 25%. To model the hydrogen storage model to fit the system, we specify the evolution of the State of Charge (SoC) model of the HESS as

$$E_{H,t+1} = E_{H,t} + \Delta t \cdot \left(\eta_{H,c} \cdot P_{H,c,t} - \frac{1}{\eta_{H,d}} \cdot P_{H,d,t} \right)$$

where $E_{H,t}$ is the hydrogen storage state of charge at time t, $P_{H,c,t}$ and $P_{H,d,t}$ are the charging and discharging power, representing the power input to the electrolyzer and the power output from the fuel cell, respectively, $\eta_{H,c}$ and $\eta_{H,d}$ are the constant charging and discharging efficiencies, Δt is the duration of the time interval. The discharging term is expressed as $P_{H,d,t}/\eta_{H,d}$ because the electrical energy output from the fuel cell is only a fraction of the actual energy extracted from the hydrogen storage. Specifically, to deliver $P_{H,d,t}$ units of electricity to the system, the fuel cell must consume $P_{H,d,t}/\eta_{H,d}$ units of hydrogen energy.

This simplified model provides a practical balance between modeling fidelity and computational efficiency, and is particularly suitable for long-term scheduling, where large-scale dispatch problems must be solved repeatedly with limited computational resources.

Efficient optimal scheduling method for wind-solar-hydrogen systems

Optimizing the dispatch strategy for hybrid systems to achieve socio-economic benefits calls for a comprehensive consideration of grid requirements, maximization of renewable energy penetration, and minimization of carbon emission costs. With the advancement of distributed reinforcement learning (DRL), distributed frameworks have demonstrated significant efficacy in large-scale optimization tasks. As illustrated in Figure 7, the fundamental architecture of these algorithms typically comprises three core components: the Actor, the Learner, and the Coordinator. The Actor interacts with the environment to generate experience data, the Learner processes this data to update policy network parameters, and the Coordinator facilitates seamless communication between these components to ensure efficient data transfer. Parallel execution across multiple Actors significantly enhances data throughput, thereby providing a robust data foundation for training complex decision-making policies. In light of the temporal uncertainties inherent in wind and solar generation, this study proposes a multi-faceted power management policy tailored for wind-solar-hydrogen systems. The approach leverages deep reinforcement learning to govern an autonomous agent that maximizes cumulative rewards through continuous interaction with its environment. This decision-making process is guided by optimized policies that determine the most effective actions based on the current system states.

In DRL, two common paradigms exist: synchronous and asynchronous training. In synchronous algorithms, all actor processes must wait for global policy updates before continuing their interactions with the environment. This strict synchronization often leads to idle time, as actors or learners must pause until others complete their tasks, resulting in reduced system efficiency. In contrast, asynchronous algorithms allow actors to interact with their environments and update local models independently, without waiting for global synchronization. This approach significantly improves data throughput and computational efficiency, making it especially suitable for large-scale and complex tasks such as energy dispatch optimization in

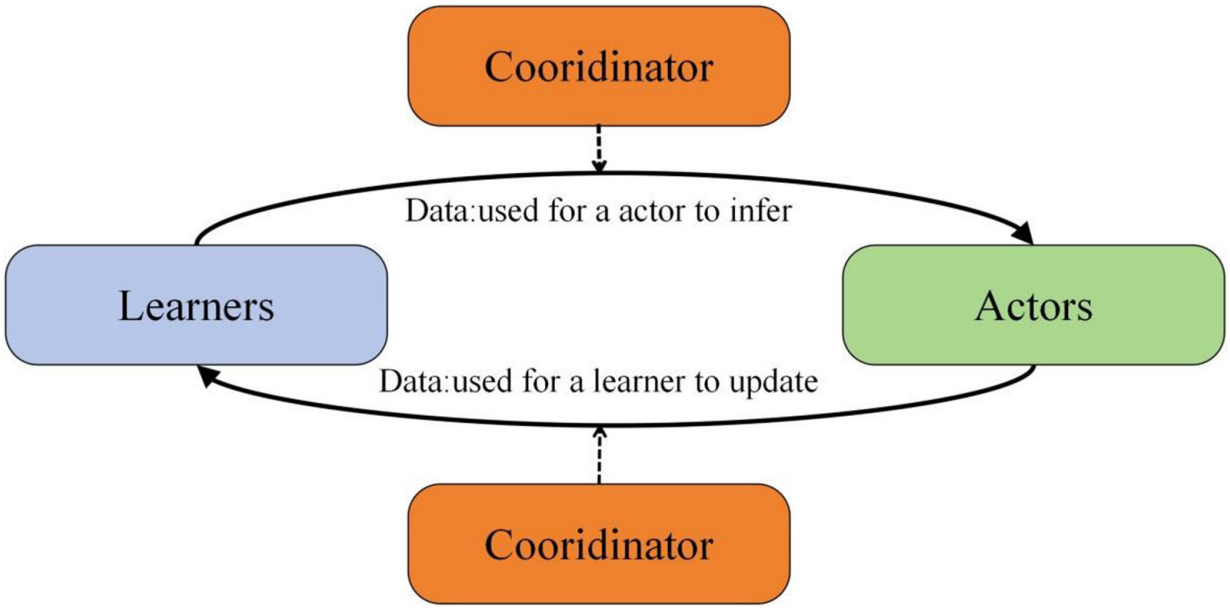


Figure 7. Basic framework of distributed reinforcement learning.

wind-solar-hydrogen systems. The distributed extension of the DDPG algorithm adopts this parallelism by enabling multiple actor processes to interact with independent environments simultaneously, each maintaining a local policy $\mu_i(s|\theta_i^\mu)$ and value function $Q_i(s,a|\theta_i^Q)$. To ensure consistency, these local parameters are periodically aggregated to update global networks. The aggregation rule averages the parameters across all actors:

$$\theta^\mu \leftarrow \frac{1}{N}\sum_{i=1}^{N}\theta_i^\mu, \quad \theta^Q \leftarrow \frac{1}{N}\sum_{i=1}^{N}\theta_i^Q$$

where N denotes the number of distributed actors. Each actor samples experiences (s_i,a_i,r_i,s'_i) and contributes to a shared replay buffer $D=U_{i=1}^{N}D_i$, which the learner uses to train global networks. The distributed framework significantly improves data throughput and training stability. In asynchronous implementations, actors continue to collect data without waiting for global parameter updates, further enhancing scalability for large-scale energy dispatch optimization. In the distributed implementation, a centralized-learner and multi-actor architecture is facilitated by the MATLAB Parallel Server. Specifically, three parallel actor nodes are deployed, each assigned to a distinct segment of the annual load-pattern dataset and maintaining a local copy of the target policy network to interact with its respective environmental instance. To balance computational efficiency with training stability, the synchronization and aggregation interval is established at $N_{sync}=10$. Consequently, the system enforces a global synchronization every 10 parameter iterations by broadcasting the latest Actor network weights to all parallel nodes. This mechanism ensures that each actor receives timely feedback from the global policy, effectively mitigating the gradient lag phenomenon inherent in DRL. Distributed storage and asynchronous update mechanisms further optimize training efficiency. The update rule for the online Q-network is formulated as

$$\theta^Q \leftarrow \theta^Q + \alpha_Q \delta \cdot \nabla_{\theta^Q} Q(s,a|\theta^Q)$$

Here, θ^Q denotes the parameters of the online Q-network, while α_Q represents the learning rate. The temporal difference error is symbolized by δ. The term $\nabla_{\theta^Q} Q(s,a|\theta^Q)$ refers to the gradient of the Q-value function with respect to the network parameters, characterizing how variations in θ^Q impact the Q-value. This gradient directs the parameter updates to minimize the temporal difference error δ. The online policy network is updated as

$$\theta^{\mu} \leftarrow \theta^{\mu} + \alpha_{\mu} \nabla_{\theta^{\mu}} \mu(s|\theta^{\mu}) \nabla_{a} Q(s,a|\theta^{Q})|_{a=\mu(s|\theta^{\mu})}$$

In this context, θ^{μ} represents the parameters of the online policy network, whereas α_{μ} denotes the associated learning rate. The expression $\nabla_{\theta^{\mu}} \mu(s|\theta^{\mu}) \nabla_{a} Q(s,a|\theta^{Q})|_{a=\mu(s|\theta^{\mu})}$ signifies the gradient of the expected Q-value with respect to the policy network parameters, computed via the deterministic policy gradient theorem. This gradient provides the direction for updating θ^{μ} to maximize the expected return. The target networks are updated as

$$\begin{cases} \theta^{Q'} \leftarrow \tau\theta^{Q} + (1\text{-}\tau)\theta^{Q'} \\ \theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1\text{-}\tau)\theta^{\mu'} \end{cases}$$

where τ denotes the soft update rate, typically a small constant significantly less than 1, which governs the update frequency of the target networks. The symbols $\theta^{Q'}$ and $\theta^{\mu'}$ designate the parameters of the target Q-network and the target policy network, respectively. By adjusting the target networks incrementally, this soft update mechanism enhances training stability and facilitates convergence.

The agent-environment interaction in the DDPG algorithm's reinforcement learning framework is formally modeled as a Markov Decision Process (MDP), represented by the five-tuple (s, a, r, p, γ). The state s encapsulates comprehensive environmental information that serves as input to the neural network, while the action a constitutes the network's output that directly influences and modifies the environment. The reward signal r provides the optimization direction for the neural network, driving parameter updates toward higher expected returns. The state transition probability p, representing the environment's intrinsic physical dynamics, determines the likelihood of subsequent state transitions. The discount factor γ governs the temporal value perspective, ensuring that the agent balances immediate operational costs with the long-term flexibility of energy storage over the scheduling cycle. Each of these fundamental MDP components will be rigorously defined and mathematically characterized in the subsequent subsections.

State and action space definitions

The DDPG algorithm is implemented to derive an optimal control policy within real-valued state and action domains. The reinforcement learning architecture primarily comprises two interacting entities: the environment and the agent. At each discrete time step t, the agent perceives the current system state $s(t)$ and subsequently determines an action $a(t)$=[$P_t(t)$] within the bounded action space [$-\Delta P_t$,ΔP_t] (MWh). To fulfill the prescribed control objectives, the state vector s(t) must encapsulate comprehensive environmental information necessary for informing effective decision-making. The constituent variables of the state space are defined as follows:

$$s(t) = [P_t(t), P_w(t), P_s(t), P_e(t), P_{bsc}(t), P_{dem}(t)]$$

Within this framework, $P_t(t)$ signifies the active power output of the thermal unit at time t, which fluctuates in response to operational requirements and demand variations. The variables $P_w(t)$ and $P_s(t)$ characterize the power generation from wind and solar energy sources at time t, respectively. Furthermore, $P_e(t)$ denotes the quantity of external power acquired at time t to compensate for the gaps in the thermal unit's generation. The term $P_{bsc}(t)$ identifies the power contribution from the HESS, while $P_{dem}(t)$ represents the instantaneous load demand.

To explicitly clarify the system dynamics, the variables within this framework are categorized into exogenous and controllable variables. The exogenous (uncontrollable) variables include the wind power output $P_w(t)$, solar power output $P_s(t)$, and regional electrical load demand $P_{dem}(t)$, which are fundamentally driven by

meteorological and stochastic operational processes. Conversely, the controllable variables encompass the active power output of the thermal unit $P_t(t)$, the electrical power exchanged with the external grid $P_e(t)$, and the power contribution from the hydrogen energy storage system $P_{bsc}(t)$. Specifically, the agent directly dictates the dispatch of the thermal unit, while the grid exchange and HESS operations dynamically ensure system-wide load balance. To ensure a high-fidelity simulation of thermal power dynamics, the rate of change in $P_t(t)$ is strictly governed by the thermal unit's maximum allowable ramp rate. In instances where the thermal power output exceeds these operational boundaries, the current training episode is terminated.

Reward function design

The optimization of the energy dispatch strategy for thermal generation is primarily aimed at securing economic advantages for both societal and industrial stakeholders. Consequently, it is imperative to satisfy system load requirements, maximize the penetration of renewable energy sources, and minimize total operational expenditures. This research utilizes thermal power units and energy storage systems as synergistic components to facilitate the seamless integration of wind and solar power. Furthermore, a carbon trading mechanism is integrated into the dispatch framework, simultaneously addressing coal consumption costs and CO_2 emission levels. Such an integrative approach seeks to establish a harmonized equilibrium between economic performance and environmental sustainability through carbon reduction.

(1) **Coal fuel cost:** To minimize the expenditures associated with coal fuel consumption, the cost function r_1 is formulated as follows:

$$r_1(t) = \mathrm{Y}_c G(P_t(t))$$

$$G(P_t(t)) = a_1(P_t(t))^2 + b_1(P_t(t)) + c_1$$

where each time step t corresponds to a specific interval for decision-making. Y_c is the total coal combustion cost, set at 115 \$/MWh, and serves as the basis for evaluating the economic impact of fuel consumption. The coal consumption function $G(P_t(t))$ estimates the coal usage of the thermal unit based on its power output $P_t(t)$, the parameters a_1, b_1, and c_1 are used in computing the thermal unit's output and corresponding coal consumption. Since non-fossil energy sources do not directly emit CO_2, the model includes carbon costs only for thermal power generation.

(2) **Carbon emission expenditure:** Since non-fossil energy sources produce zero direct emissions, the model allocates carbon-related expenditures solely to the thermal power component. The total volume of CO_2 emissions is quantified as follows:

$$M(P_t(t)) = a_2(P_t(t))^2 + b_2(P_t(t)) + c_2$$

Here, $M(P_t(t))$ signifies the CO_2 emissions associated with the thermal unit output $P_t(t)$ at time t. The emission volume is determined via parameters a_2, b_2, and c_2, which mathematically characterize the emission characteristics of the specific unit. Accounting for carbon market mechanisms, this study assumes that a predefined quota of CO_2 emission allowances is allocated gratis. These free-of-charge allowances are determined by the generation capacity and the specific category of the thermal unit, formulated as follows:

$$M_{fr}(t) = \beta P_t(t)\Delta t$$

where $M_{fr}(t)$ represents the free CO_2 emission allowance allocated to the thermal unit at time t. The coefficient β defines the amount of allowance granted per unit of electricity generated. Under this allocation scheme, any emissions exceeding the free allowance must be offset through the purchase of additional allowances, resulting in a CO_2 emission cost. To penalize excessive emissions and encourage low-carbon generation, the CO_2 emission cost function $r_2(t)$ is defined as

$$M_{sp}(t) = M(P_t(t)) - M_{fr}(t)$$

$$r_2(t) = \begin{cases} \varphi \cdot M_{sp}(t), & \text{if } M_{sp}(t) > 0 \\ 0, & \text{otherwise} \end{cases}$$

where $M_{sp}(t)$ denotes the payable CO_2 emissions at time t, representing the amount by which actual emissions exceed the allocated free allowance. The carbon price is denoted by φ, which is set to 11.5 \$/ton[44]. This price is used to calculate the monetary cost of excess emissions within the carbon trading framework.

(3) **Electricity purchase cost:** To ensure that the power supply meets the load demand, an electricity purchase cost is introduced based on the power deficit. The total power is given by

$$P(t) = P_t(t) + P_w(t) + P_s(t).$$

If an energy storage system is considered, $P(t)$ also includes the power supplied by storage, $P_{bsc}(t)$. The electricity purchase cost $r_3(t)$ is defined as

$$r_3(t) = \begin{cases} Y_e \cdot (P_{\text{dem}}(t) - P(t)), & \text{if } P_{\text{dem}}(t) > P(t) \\ 0, & \text{otherwise} \end{cases}$$

Here, Y_e denotes the electricity purchase price, established at 327 \$/MWh, which serves as a benchmark for comparing external grid procurement costs with internal generation expenses. The term Δt represents the dispatch interval, fixed at 1 h, which determines the temporal resolution of the simulation model. Throughout the interaction process, the agent acquires rewards by progressively modifying the environmental state. The instantaneous reward $r(t)$ is formulated as follows:

$$r(t) = -(r_1(t) + r_2(t) + r_3(t)).$$

The control policy is continuously updated by the agent based on the value of the reward function. Given that the energy management of the wind-solar-hydrogen system involves time-coupled variables, such as the SoC of the HESS, the scheduling task is formulated as a finite-horizon MDP. Accordingly, the discount factor γ is set to 1. This configuration ensures that the agent yields an unbiased cumulative reward over the entire daily scheduling cycle, effectively balancing immediate operational costs with the long-term flexibility of energy storage.

Binding conditions

To guarantee the stable and reliable operation of the integrated wind-solar-hydrogen system, several physical hard constraints must be strictly adhered to. The mathematical formulations of these constraints are presented as follows:

(1) **Load balance constraint:** To ensure load balance at each time t, the total power supply must be no less than the load demand. The supply includes wind power $P_w(t)$, solar power $P_s(t)$, purchased electricity $P_e(t)$, storage discharge power $P_{bdch}(t)$, and subtracts storage charging power $P_{bch}(t)$. The load balance constraint is given by

$$P_t(t) + P_w(t) + P_s(t) + P_e(t) + P_{bdch}(t) - P_{bch}(t) \geq P_{dem}(t).$$

(2) **Climbing capacity constraint:** To accurately replicate the operational behavior of an actual thermal power unit, a ramp rate limitation is imposed on the power output fluctuations. The governing climbing capacity constraint is formulated as follows:

$$-\Delta P_t \leq (P_t(t + \Delta t) - P_t(t)) / \Delta t \leq \Delta P_t$$

Here, ΔP_t signifies the ramp rate limitation of the thermal power generating unit. This parameter establishes the maximum permissible increment or decrement in power output within a discrete time interval.

(3) **Generation output constraints:** The active power outputs produced by the thermal, solar, and wind units are strictly governed by their individual rated capacities. These operational boundaries are mathematically defined as follows:

$$\begin{cases} 0 \leq P_t(t) \leq P_{t,max} \\ 0 \leq P_w(t) \leq P_{w,max} \\ 0 \leq P_s(t) \leq P_{s,max} \end{cases}$$

In this formulation, $P_{t,\max}$ signifies the maximum generation capacity assigned to the thermal power unit. Analogously, $P_{w,\max}$ and $P_{s,\max}$ designate the peak output limitations for the wind and solar generating units, respectively.

(4) **Storage capacity constraints:** The operational capacity of the energy storage system is governed by specific physical thresholds to ensure safety and longevity. These capacity limitations are mathematically formulated as follows:

$$\begin{cases} P_{bsc}(t\text{-}1) - P_{bdch}(t) + P_{bch}(t) = P_{bsc}(t) \\ 0 \leq P_{bsc}(t) \leq P_{bsc,max} \end{cases}$$

In this formulation, $P_{bsc}(t)$ denotes the power supplied by the HESS at time t. As defined above, $P_{bch}(t)$ and $P_{bdch}(t)$ represent the charging and discharging power of the HESS at time t, respectively.

(5) **Charging and discharging constraints:** The operational logic presiding over the charging and discharging cycles of the energy storage system is articulated as follows:

$$\begin{cases} P_{bch}(t)\Delta t + P_{bsc}(t\text{-}1) \leq P_{bsc,max} \\ P_{bdch}(t)\Delta t\text{-}P_{bsc}(t\text{-}1) \leq 0 \\ 0 \leq P_{bch}(t) \leq P_{bch,max} \\ 0 \leq P_{bdch}(t) \leq P_{bdch,max} \end{cases}$$

where $P_{bdch,\max}$ and $P_{bch,\max}$ denote the maximum discharging and charging power of the energy storage system, respectively, $P_{bsc,\max}$ represents the maximum energy storage capacity of the system.

The neural network takes power system data as input and outputs adjustment values for thermal power generation. Through this continuous input-output interaction, the operation of the thermal unit is progressively optimized, improving overall generation efficiency within each sampling interval.

RESULTS AND DISCUSSION

This study utilizes annual time-series data for electricity load, wind power, and solar generation, sourced from the Open Power Systems Data platform[45]. The dataset provides hourly-resolution information essential for power system modeling, encompassing electricity prices, load profiles, and renewable generation

capacities aggregated across various European bidding zones and neighboring regions. For the simulation, seven representative days are extracted from the annual records to capture seasonal variations: spring (Days 2 and 3), summer (Days 4 and 5), autumn (Day 6), and winter (Days 1 and 7).

To facilitate a comprehensive comparative analysis, ten distinct experimental scenarios are defined within this research, with their respective scheduling profiles illustrated in Figure 8. Scenario 1 utilizes the Sequential Least Squares Programming (SLSQP) algorithm in the absence of an energy storage system. Scenario 2 implements the SLSQP algorithm integrated with a hydrogen-based energy storage system. Scenario 3 adopts the Deep Deterministic Policy Gradient (DDPG) optimization framework without storage capabilities. Scenario 4 applies the DDPG algorithm coupled with a hydrogen-based storage system. Scenario 5 employs a DRL approach without energy storage. Scenario 6 integrates the DRL algorithm with a hydrogen-based storage system. Scenario 7 extends the configuration of Scenario 6 by incorporating wind and solar generation data processed through an First-order Autoregressive [AR(1)] noise model. Scenarios 8, 9, and 10 are specifically designed to conduct sensitivity analyses based on the architectural setup of Scenario 7. In the baseline configuration, uniform weights of unity are assigned to the fuel consumption cost, carbon emission penalty, and electricity procurement expenditure. Scenario 8 assesses the impact of fuel price volatility by doubling the weight attributed to coal fuel consumption. Scenario 9 simulates more stringent environmental regulations by increasing the carbon emission weight to 2. Finally, Scenario 10 evaluates the sensitivity to market price fluctuations by adjusting the electricity purchase weight to 2. To address multi-objective energy dispatch within continuous action domains, the DDPG algorithm leverages an Actor-Critic architectural framework. Within this structure, the Actor network $\mu(s|\theta^{\mu})$ proposes control actions, while the Critic network $Q(s,a|\theta^{Q})$ evaluates the effectiveness of these decisions. Detailed system parameters utilized in this study are summarized in Table 1.

Performance validation

A detailed analysis of the economic performance across the ten scenarios is presented in Table 2. The results systematically assess the impact of different optimization algorithms, the inclusion of a HESS, and varying cost sensitivities on the system's operational strategy and total cost.

Firstly, a comparison between scenarios with and without the HESS reveals its significant economic benefits. For instance, comparing Scenario 1 (SLSQP without HESS) to Scenario 2 (SLSQP with HESS), the total cost is marginally reduced. Similarly, the introduction of the HESS in the DDPG algorithm (Scenario 4 *vs.* Scenario 3) and the DRL algorithm (Scenario 6 *vs.* Scenario 5) leads to total cost reductions of 0.023×10^7 and 0.107×10^7 \$ respectively. This consistently demonstrates that integrating the HESS improves energy utilization and reduces overall operational costs, regardless of the optimization algorithm used.

Secondly, among the different optimization strategies, the DRL algorithm combined with HESS (Scenario 6) achieves the lowest total cost of 5.360×10^7 \$. Comparing Scenario 6 with Scenario 5 explicitly demonstrates the mechanism behind this improvement: although the integration of HESS leads to a slightly higher coal fuel cost (2.823×10^7 *vs.* 2.629×10^7 \$), this controlled increase in steady thermal generation enables the system to drastically scale back on costly external power purchases (2.910×10^6 *vs.* 5.334×10^6 \$). Furthermore, the energy time-shifting capabilities reduce reliance on high-emission peaking periods, lowering overall carbon emission costs (2.246×10^7 *vs.* 2.305×10^7 \$). This favorable trade-off confirms it is the most effective strategy for balancing coal fuel consumption, CO_2 emissions, and electricity purchasing to achieve economic optimality. Scenario 7, which introduces noise to the renewable generation data, results in a slightly higher total cost (5.379×10^7 \$) than Scenario 6, reflecting the economic trade-offs required to manage system operations under uncertainty.

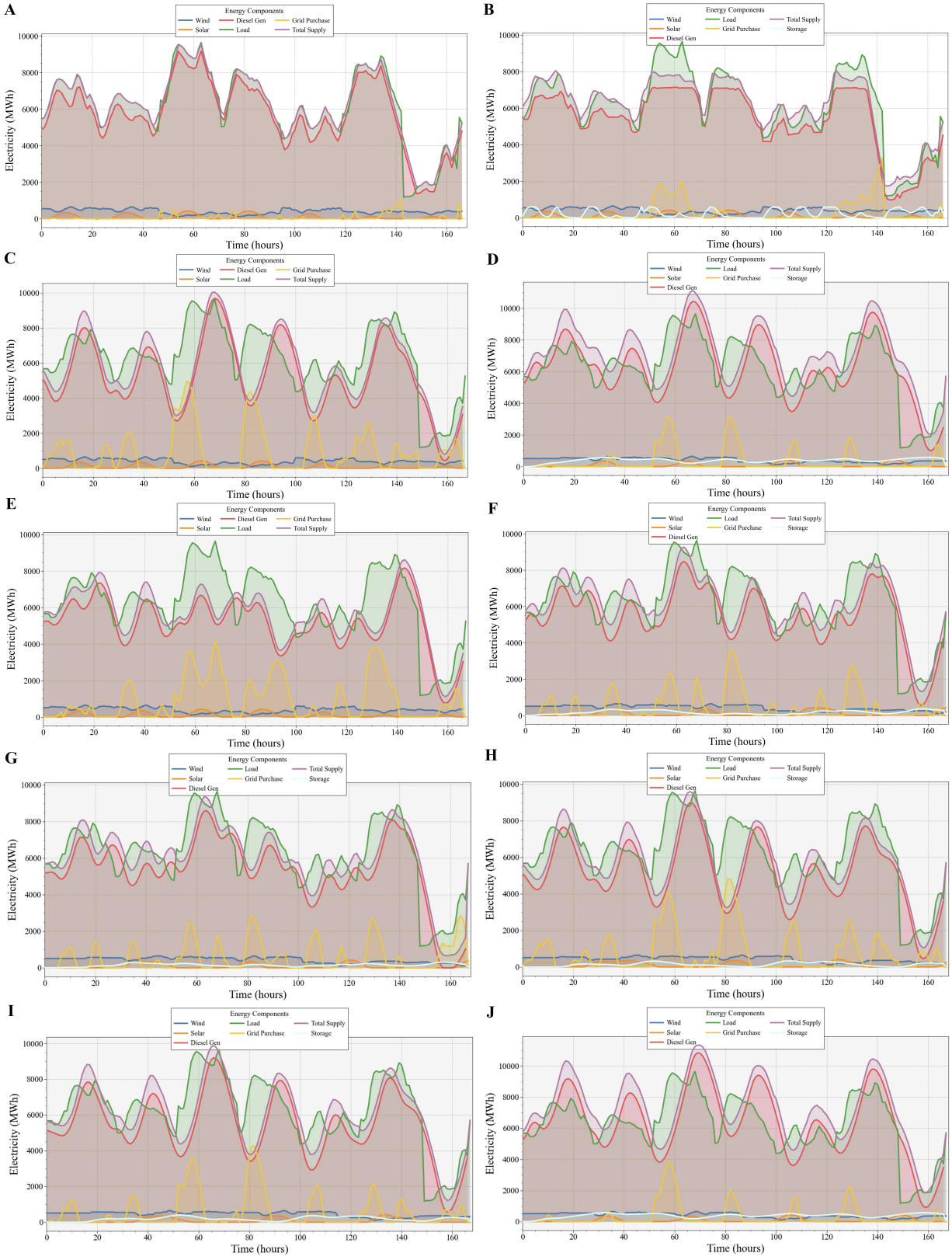


Figure 8. Experiment result. (A) System power scheduling profile for Scenario 1; (B) System power scheduling profile for Scenario 2; (C) System power scheduling profile for Scenario 3; (D) System power scheduling profile for Scenario 4; (E) System power scheduling profile for Scenario 5; (F) System power scheduling profile for Scenario 6; (G) System power scheduling profile for Scenario 7; (H) System power scheduling profile for Scenario 8; (I) System power scheduling profile for Scenario 9; (J) System power scheduling profile for Scenario 10.

Table 1. Systems parameters

Parameter	Value	Parameter	Value	Parameter	Value
Y_c ($/MWh)	115	Y_e ($/MWh)	327	$P_{t,max}$ (MWh)	1×10^4
a_1	1×10^{-5}	a_2	2×10^{-5}	$P_{w,max}$ (MWh)	2,000
b_1	1×10^{-4}	b_2	2×10^{-4}	$P_{s,max}$ (MWh)	1,500
c_1	0	c_2	600	$P_{bsc,max}$ (MWh)	3,000
I	2×10^6	β	0.1	$P_{bch,max}$ (MWh)	300
φ ($/ton)	11.5	k	24	$P_{bdch,max}$ (MWh)	300
Actor learning rate	1×10^{-4}	Critic learning rate	0.001	Gradient threshold	1
L2 Regularization	1×10^{-4}	Noise variance	0.6	Decay rate	1×10^{-5}
Replay buffer size	1×10^6	Max training episodes	1×10^4		

Table 2. Comparative evaluation of operational costs across different scenarios

Scenario	Coal fuel cost ($) ($r_1$)	Carbon emission cost ($) ($r_2$)	External power purchase cost ($) ($r_3$)	Aggregate operational cost ($)
1	2.986×10^7	2.520×10^7	1.925×10^6	5.696×10^7
2	2.789×10^7	2.316×10^7	5.627×10^6	5.668×10^7
3	2.702×10^7	2.335×10^7	5.317×10^6	5.569×10^7
4	2.932×10^7	2.398×10^7	2.159×10^6	5.546×10^7
5	2.629×10^7	2.305×10^7	5.334×10^6	5.467×10^7
6	2.823×10^7	2.246×10^7	2.910×10^6	5.360×10^7
7	2.626×10^7	2.287×10^7	4.654×10^6	5.379×10^7
8	2.607×10^7	2.224×10^7	6.376×10^6	5.469×10^7
9	2.418×10^7	2.151×10^7	8.909×10^6	5.460×10^7
10	2.965×10^7	2.490×10^7	4.231×10^5	5.497×10^7

Finally, the sensitivity analysis (Scenarios 8-10), based on the optimal framework of Scenario 7, provides critical insights into the system's response to different cost priorities. In Scenario 8, doubling the weight of the coal fuel cost forces the system to reduce fuel consumption (from 2.626×10^7 to 2.607×10^7 $) and instead rely more on purchasing electricity (cost increased from 4.654×10^6 to 6.376×10^6 $). In Scenario 9, a higher penalty on CO_2 emissions prompts an even more significant shift away from thermal generation, achieving the lowest fuel (2.418×10^7 $) and emission ($2.151 \times 10^7$ $) costs at the expense of having the highest electricity purchase cost (8.909×10^6 $). Conversely, in Scenario 10, when the electricity purchase cost is penalized, the system minimizes its reliance on the grid, leading to the lowest purchase cost (4.231×10^5 $) but increasing its own thermal generation, which in turn raises both fuel and emission costs. This demonstrates the model's ability to intelligently adapt its energy dispatch strategy in response to varying economic and environmental signals.

Robustness verification

Scenario 7 was specifically designed to validate the robustness of the optimal dispatch strategy (developed in Scenario 6) against the inherent uncertainties of renewable energy generation. To achieve this, we introduced stochastic fluctuations, modeled via an AR(1) process, to the wind and solar power data to simulate real-world forecasting errors.

The results demonstrate strong performance stability. The total operational cost for Scenario 7 was 5.379×10^7 $ [Table 2], representing a marginal increase of only 0.019×10^7 $ - or approximately 0.35% - compared

Table 3. Parallel training performance evaluation

Threads number	Total wall clock training time (h)	Speedup ratio
1 (Standard)	72.5	1.00x
2	37.2	1.95x
4	19.1	3.80x
6	11.5	5.50x

to the 5.360×10^7 $ cost of its deterministic counterpart, Scenario 6. This minimal cost deviation, despite the introduction of significant input noise, signifies that the DRL agent has acquired a resilient and adaptive dispatch policy. This resilience is critical for practical implementation, as it confirms the model's capability to maintain near-optimal economic performance even when faced with imperfect environmental forecasts, thereby validating its robustness for real-world deployment.

Comparison of training speeds

To assess the computational efficiency of the proposed algorithm, we evaluate the performance of multi-threaded parallel training. All experiments are conducted on a single workstation equipped with an 11th Gen Intel Core i7-11800H CPU (8 Cores, 16 Threads) and 32 GB of RAM. The performance metrics, summarized in Table 3, demonstrate a significant reduction in the total wall-clock training time as the number of parallel worker threads increases.

The baseline single-threaded training required 72.5 h to complete. By scaling the number of workers to 2, 4, and 6, the training time was dramatically reduced to 37.2, 19.1, and 11.5 h, respectively. This corresponds to empirical speedup ratios of 1.95x, 3.80x, and 5.50x. The speedup is observed to be nearly linear up to 4 threads, indicating efficient parallelization with minimal communication overhead. Beyond this point, a slight drop-off in linear scaling efficiency is noted. The results validate that the multi-threaded training architecture effectively leverages multi-core CPUs to accelerate the learning process. The observed scalability confirms that employing a moderate number of parallel workers presents an optimal trade-off between computational resource allocation and training time reduction for this task.

CONCLUSIONS

This paper develops a DRL-based scheduling framework for integrated wind-solar-hydrogen systems, incorporating advanced data-driven techniques for load and scenario characterization. The integration of DTW-DBSCAN and PCA-enhanced K-means clustering allows the framework to effectively navigate the inherent stochasticity of renewable generation and load demand. Experimental evaluations indicate that the proposed methodology yields an approximately 6% reduction in total annual operational expenditures compared to the baseline Scenario 1, down from 5.696×10^7 to 5.360×10^7 $. This economic performance is primarily driven by the agent's ability to execute a sophisticated energy time-shift arbitrage strategy, effectively converting surplus renewable power into hydrogen storage. The proposed control architecture successfully transforms hydrogen-based energy storage from a passive backup component into a proactive, economically versatile asset for high-renewable grids. Furthermore, the proposed scalable distributed training architecture significantly accelerates the training speed and substantially improves data throughput, highlighting a robust computational capacity for large-scale energy dispatch optimization. Future work will focus on extending this framework to multi-agent collaborative scheduling and incorporating more complex dynamics from evolving carbon-linked energy markets. Additionally, we plan to investigate the optimization effects of employing a non-linear semi-empirical behavior model for hydrogen storage, comparing its efficacy against the current constant efficiency model in dynamic microgrid scheduling.

DECLARATIONS

Authors' contributions

Conceptualization, methodology, software, writing - original draft: Zhang, B.
Data curation, investigation, formal analysis: Wang, C.
Supervision, writing - review and editing, project administration, funding acquisition: Ma, Y.
Validation, visualization: Xie, J.
Supervision, resources, investigation: He, L.

Availability of data and materials

The data generated during the current study are available from the corresponding author on reasonable request. The publicly available input time-series data used in this study were obtained from Open Power System Data (OPSD)[45], Time Series, available at URL (https://doi.org/10.25832/time_series/2020-10-06).

Financial support and sponsorship

The work has been supported in part by the Natural Science Foundation of Shaanxi Province, China (No. 2023-JC-QN-0687), in part by the Central University Basic Research Fund of China.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool Gemini (version 3.1, released 2026-02-19) was used solely for language editing. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Wang, C.; Ma, Y.; Xie, J.; Ouyang, Q. Multi-objective energy dispatch with deep reinforcement learning for wind-solar-thermal-storage hybrid systems. *J. Energy. Storage.* **2025**, *105*, 114635. DOI
2. Hui, Y.; Wang, M.; Guo, S.; et al. Comprehensive review of development and applications of hydrogen energy technologies in China for carbon neutrality: technology advances and challenges. *Energy. Convers. Manag.* **2024**, *315*, 118776. DOI
3. Liu, Y.; Xie, X.; Wang, M. Energy structure and carbon emission: analysis against the background of the current energy crisis in the EU. *Energy* **2023**, *280*, 128129. DOI
4. Mallapaty, S. How China could be carbon neutral by mid-century. *Nature* **2020**, *586*, 482-3. DOI PubMed
5. Cheng, Y.; Zhang, N.; Kirschen, D. S.; Huang, W.; Kang, C. Planning multiple energy systems for low-carbon districts with high penetration of renewable energy: an empirical study in China. *Appl. Energy.* **2020**, *261*, 114390. DOI
6. Liu, H.; Zhu, H.; Li, F.; et al. Economic operation strategy of electricity-gas-thermal-hydrogen integrated energy system taking into account the cost of carbon. *Electr. Power. Constr.* **2021**, *42*, 21-29. (in Chinese). DOI
7. Zhuo, Z.; Du, E.; Zhang, N.; et al. Cost increments of electricity supply to achieve carbon neutrality in China. *Nat. Commun.* **2022,**. *13*, 3172. DOI PubMed PMC
8. Jin, J.; Zhou, P.; Li, C.; Guo, X.; Zhang, M. Low-carbon power dispatch with wind power based on carbon trading mechanism. *Energy* **2019**, *170*, 250-60. DOI
9. Zhang, Y.; Li, J.; Ji, X.; Yang, M.; Ye, P. Optimal scheduling of electricity-gas-heat integrated energy system with coordination of flexibility and reliability. *Sustain. Energy. Technol. Assess.* **2024**, *71*, 103968. DOI

10. Kumar, C. M. S.; Singh, S.; Gupta, M. K.; et al. Solar energy: a promising renewable source for meeting energy demand in Indian agriculture applications. *Sustain. Energy. Technol. Assess.* **2023**, *55*, 102905. DOI
11. Gangopadhyay, A.; Seshadri, A.; Sparks, N.; Toumi, R. The role of wind-solar hybrid plants in mitigating renewable energy-droughts. *Renew. Energy.* **2022**, *194*, 926-37. DOI
12. Huang, K.; Liu, P.; Ming, B.; Kim, J.; Gong, Y. Economic operation of a wind-solar-hydro complementary system considering risks of output shortage, power curtailment and spilled water. *Appl. Energy.* **2021**, *290*, 116805. DOI
13. Liu, P.; Quan, F.; Gao, Y.; Alotaibi, B.; Alsenani, T. R.; Abuhussain, M. Green energy forecasting using multiheaded convolutional LSTM model for sustainable life. *Sustain. Energy. Technol. Assess.* **2024**, *63*, 103609. DOI
14. Shoaei, M.; Hajinezhad, A.; Moosavian, S. F. Design, energy, exergy, economy, and environment (4E) analysis, and multi-objective optimization of a novel integrated energy system based on solar and geothermal resources. *Energy* **2023**, *280*, 128162. DOI
15. Reddy, S. S. Optimal scheduling of thermal-wind-solar power system with storage. *Renew. Energy.* **2017**, *101*, 1357-68. DOI
16. Li, Y.; Choi, S. S.; Vilathgamuwa, D. M.; Yao, D. L. An improved dispatchable wind turbine generator and dual-battery energy storage system to reduce battery capacity requirement. In *2016 IEEE 2nd Annual Southern Power Electronics Conference (SPEC)*, Auckland, New Zealand, December 5-8, 2016; IEEE: New York, NY, USA, 2016; pp 1-6. DOI
17. Sgouridis, S.; Carbajales-dale, M.; Csala, D.; Chiesa, M.; Bardi, U. Comparative net energy analysis of renewable electricity and carbon capture and storage. *Nat. Energy.* **2019**, *4*, 456-65. DOI
18. Van Leeuwen, R.; De Wit, J.; Smit, G. Review of urban energy transition in the Netherlands and the role of smart energy management. *Energy. Convers. Manag.* **2017**, *150*, 941-8. DOI
19. Li, J.; Lu, T.; Yi, X.; An, M.; Hao, R. Energy systems capacity planning under high renewable penetration considering concentrating solar power. *Sustain. Energy. Technol. Assess.* **2024**, *64*, 103671. DOI
20. Özkan, O.; Alola, A. A.; Adebayo, T. S. Environmental benefits of nonrenewable energy efficiency and renewable energy intensity in the USA and EU: examining the role of clean technologies. *Sustain. Energy. Technol. Assess.* **2023**, *58*, 103315. DOI
21. Haddad, A.; Ramadan, M.; Khaled, M.; Ramadan, H. S.; Becherif, M. Triple hybrid system coupling fuel cell with wind turbine and thermal solar system. *Int. J. Hydrogen. Energy.* **2020**, *45*, 11484-91. DOI
22. Zrelli, M. H. Renewable energy, non-renewable energy, carbon dioxide emissions and economic growth in selected Mediterranean countries. *Environ. Econ. Policy. Stud.* **2016**, *19*, 691-709. DOI
23. Asumadu-sarkodie, S.; Owusu, P. A. The causal effect of carbon dioxide emissions, electricity consumption, economic growth, and industrialization in Sierra Leone. *Energ. Source. Part. B.* **2016**, *12*, 32-9. DOI
24. Yao, T.; Yang, Y.; Yan, Y.; et al. Knowledge-extractor: a self-evolving scientific framework for hydrogen energy research driven by AI agents. *AI. Agent.* **2025**, *1*, 7. DOI
25. Adetokun, B. B.; Oghorada, O.; Abubakar, S. J. Superconducting magnetic energy storage systems: prospects and challenges for renewable energy applications. *J. Energy. Storage.* **2022**, *55*, 105663. DOI
26. Jahanger, A.; Ozturk, I.; Chukwuma Onwe, J.; Joseph, T. E.; Razib Hossain, M. Do technology and renewable energy contribute to energy efficiency and carbon neutrality? Evidence from top ten manufacturing countries. *Sustain. Energy. Technol. Assess.* **2023**, *56*, 103084. DOI
27. Xie, J.; Ma, Y.; Wang, C.; Wang, Y.; Yang, S.; Ouyang, Q. Spatio-temporal multi-head graph attention network for power forecasting of regional photovoltaic plants. *Solar. Energy.* **2026**, *304*, 114202. DOI
28. Fu, Y.; Liu, M. Scenario decomposition method for multi-objective stochastic dynamic economical dispatch problem. *Autom. Electr. Power. Syst.* **2014**, *38*, 34-40. (in Chinese). DOI
29. Wang, S.; Luo, F.; Dong, Z. Y.; Ranzi, G. Joint planning of active distribution networks considering renewable power uncertainty. *Int. J. Electr. Power. Energy. Syst.* **2019**, *110*, 696-704. DOI
30. Zhang, X.; Wang, S.; Qian, Y.; et al. Two-level optimal allocation method of integrated energy system considering source and load uncertainty and carbon trading. In *2023 IEEE 6th International Electrical and Energy Conference (CIEEC)*, Hefei, China, May 12-14, 2023; IEEE: New York, NY, USA, 2023; pp 2351-6. DOI
31. Zhang, Juntao, Cheng, Chuntian, Shen, Jianjian, Li, G, Li, X, Zhao, Z. Short-term joint optimal operation method for high proportion renewable energy grid considering wind-solar uncertainty. *Proc. CSEE.* **2020**, *40*, 5921-32. (in Chinese). DOI
32. Wu, M.; Xu, J.; Shi, Z. Low carbon economic dispatch of integrated energy system considering extended electric heating demand response. *Energy* **2023**, *278*, 127902. DOI
33. Zheng, L.; Li, Y.; Wei, C.; Bai, X. A data-driven method for operation pattern analysis of the integrated energy microgrid. *Energy. Convers. Manag. X.* **2021**, *11*, 100092. DOI
34. Salpakari, J.; Lund, P. Optimal and rule-based control strategies for energy flexibility in buildings with PV. *Appl. Energy.* **2016**, *161*, 425-36. DOI

35. He, Z.; Liu, C.; Wang, Y.; Wang, X.; Man, Y. Optimal operation of wind-solar-thermal collaborative power system considering carbon trading and energy storage. *Appl. Energy.* **2023**, *352*, 121993. DOI
36. Yang, T.; Zhao, L.; Li, W.; Zomaya, A. Y. Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning. *Energy* **2021**, *235*, 121377. DOI
37. Sanjari, M. J.; Karami, H.; Gooi, H. B. Analytical rule-based approach to online optimal control of smart residential energy system. *IEEE. Trans. Ind. Inf.* **2017**, *13*, 1586-97. DOI
38. Cai, T.; Zhao, C.; Xu, Q. Energy network dispatch optimization under emergency of local energy shortage. *Energy* **2012**, *42*, 132-45. DOI
39. Liao, G. A novel evolutionary algorithm for dynamic economic dispatch with energy saving and emission reduction in power system integrated wind power. *Energy* **2011**, *36*, 1018-29. DOI
40. Tian, X. Energy storage complementary control method for wind-solar storage combined power generation system under opportunity constraint. *The. Journal. of. Engineering.* **2023**, *2023*, e12256. DOI
41. Lei, D.; Zhang, Z.; Wang, Z.; Zhang, L.; Liao, W. Long-term, multi-stage low-carbon planning model of electricity-gas-heat integrated energy system considering ladder-type carbon trading mechanism and CCS. *Energy* **2023**, *280*, 128113. DOI
42. Lin, B.; Huang, C. Analysis of emission reduction effects of carbon trading: Market mechanism or government intervention? *Sustain. Prod. Consum.* **2022**, *33*, 28-37. DOI
43. Qi, N.; Huang, K.; Fan, Z.; Xu, B. Long-term energy management for microgrid with hybrid hydrogen-battery energy storage: a prediction-free coordinated optimization framework. *Appl. Energy.* **2025**, *377*, 124485. DOI
44. Xiao, J.; Li, G.; Xie, L.; Wang, S.; Yu, L. Decarbonizing China's power sector by 2030 with consideration of technological progress and cross-regional power transmission. *Energy. Policy.* **2021**, *150*, 112150. DOI
45. Wiese, F.; Schlecht, I.; Bunke, W.; et al. Open Power System Data - frictionless data for electricity system modelling. *Appl. Energy.* **2019**, *236*, 401-9. DOI