ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Spatiotemporal Feature Mining and Intelligent Decision-Making for Power System Dispatch Optimization under Uncertainty

Qiang Li[1], *Shuo Yu*[1,*], *Hongyu Tang*[2] *and Guoliang Zhang*[2]

[1] Inner Mongolia Power (Group) Co., Ltd., Saihan District, Hohhot 010020, Inner Mongolia, China

[2] Beijing Tsintergy Technology Co., Ltd., Haidian District, Beijing 100084, China

SUMMARY: *Adding more renewable energy to the power system, it brings many uncertainties in generation and load that makes traditional dispatch optimization hard. This paper puts forward a dispatching optimization method based on uncertainty of power system by integrating spatiotemporal features with deep reinforcement learning. First we establish multi-time-scale dispatch model to include wind, solar production and load demand uncertainties. 2nd GC-LSTM for obtaining spatiotemporal correlations from Renewable Energy Output; 3rd Design Attention Based MA- DDPG Algorithm for Smart Dispatch Decision Making; Finally simulation results indicate that the enhanced IEEE118-node system can confirm the efficacy of the suggested approach. The result indicates that this approach can precisely capture the spatio-temporal coupling of the output from renewable energy sources, resulting in a 12.3% reduction in operating costs as compared to conventional stochastic optimization methods, along with a decrease in wind and solar curtailment rates by 35.6% as opposed to traditional approaches, and it is faster by 42.8%, with strong technical support to ensure safe and economic operation at higher levels of renewable energy.*

KEYWORDS: *power system dispatching; Uncertainty; Spatiotemporal feature mining; Graph neural network; Deep reinforcement learning*

1 Introduction

The penetration rates of renewable energy sources such as wind and solar PV in the power system are always growing thanks to global energy transition, “Dual Carbon” goals. According to data from the National Energy Administration, China’s installed capacity for renewable energy reached more than 1.4 billion kilowatts at the end of 2025, accounting for over half of all installed capacities [1]. However since renewables can't output constantly and they go up and down a lot and have peak clipping effect it makes everything much harder to tell what we have available on supply side or how much demand we'll get on electrical grid-really hard times keeping things safe & steady while being able to actually afford running this whole thing properly.

Traditional power system dispatching optimization is mainly done through deterministic models or simplified uncertain methods like robust and stochastic optimization. However, there are also shortcomings: (1) It's unrealistic to assume a probability distribution for the uncertainty variables as they do not accurately reflect how renewable energy sources behave over time and space [2-4]. (2) Optimization Models have a lot of computational weight

*etxpublic02@163.com
https://doi.org/10.65102/is2026803
Published: 30 April, 2026

making it hard for timely decision-making in real-time dispatch. (3) They cannot explore deeply into the internal pattern of historical data, thus poor intelligent of decision-making[5].

In recent years, the development of Artificial Intelligence Technology has been extremely fast and many new ways to do power system dispatching optimization have been added. Deep learning is very good at predicting spatiotemporal sequences and extracting features; reinforcement learning is amazing at adapting to learn complicated decisions [6-8]. To combine AI technology with traditional optimization methods deeply, discover spatial-temporal evolution patterns of uncertain factors, and realize intelligent data-driven dispatch decision-making, it's an important research area in power system engineering.

The uncertainty-based scheduling optimization has 3 types of studies now: (1) Stochastic Optimization Methods, which use scenario generation and pruning techniques to represent uncertainties but must make a trade-off between the number of scenarios and computational accuracy. (2) Robustness optimize method forms the uncertainty set that makes sure system is safe under worst condition although it's very careful about its result [9].(3)Distributed robust optimization method combine all methods said above and they also need to build fuzzy sets based on experience, there are some problem while solving the solution in terms of computation.

Spatiotemporal feature extraction CNN and LSTM are applied in the renewable energy power forecasting. GNNs have some benefits on grid topological info extracting because they can deal with non-Euclidean space well. However, most of the existing studies only focus on extracting features from one aspect, temporal or spatial scale, but not both together. Intelligent decision-making part is mainly applied to deep reinforcement learning (DRL), especially DQN, DDPG, PPO algorithm applied in unit scheduling and economic dispatch task of power system. And currently, researches often consider the power system as a black box that does not take into account its physical limitations and operation; single-agent framework can't meet the needs of distributed decision making in large-scale systems. To solve these problems, this paper carries out the following research work: 1) We propose a graph Convolutional LSTM model, it is possible to extract spatiotemporal coupling features of renewable energy output through a combination of the spatial feature extraction ability of graph convolutional networks and the time series modeling advantage of LSTMs; 2Cybersecurity constrained MA-DDPG, a attention-based multi-agent deep deterministic policy gradient method for intelligent scheduling decision-making that has improved the algorithmic explainability and convergence performance of the original one.[10];3) A unified "Prediction-Optimization-Decision-Making" framework that integrates spatio-temporal feature mining and scheduling optimization. The method was verified with a modified IEEE 118 node system, and it worked well.

2 Methods

2.1 Problem Description and Mathematical Model

2.1.1 Uncertainty Scheduling Optimization Framework

Take a power system with wind, solar, thermal power units as an example to build a 2-stage dispatch optimization model of the day-ahead and intra-day periods. Day ahead: do your best to ensure that every unit has its on/off schedule according to what you expect about how much power is going to be generated from renewables based on the base output curve, Intra-day: there are some uncertain things when we do ultra-short-term forecasts in real life [11].

System is to reduce the overall cost as much as possible. The cost includes fuel costs for thermal power units, start-up and shutdown costs, penalties for curtailing renewable energy, and penalties for reducing loads according to equation (1).

$$min \sum_{t-1}^{T} \left[\sum_{i=1}^{N_G} (C_{i,t}^{gen} + C_{i,t}^{su} + C_{i,t}^{sd}) + \sum_{j=1}^{N_W} c_j^{curt} P_{j,t}^{w,curt} + \sum_{k=1}^{N_L} c_k^{shed} P_{k,t}^{shed} \right] \quad (1)$$

In the formula: T denotes N_G N_G N_W N_L the scheduling cycle; and Nl represent the number $C_{i,t}^{su}$ of thermal power units, wind farms, $P_{k,t}^{shed}$ and load nodes, respectively; Ci,t indicates the generation cost of unit i during period t; Cs and Cw represent start-up/shutdown costs; while Ww and Lw denote curtailed wind power generation and load shedding quantities, respectively.

2.1.2 Constraints

We take wind, solar, thermal power units as a kind of power system to build a 2-stage dispatch optimization model for the day ahead and intra-day periods. Day before hand: try to have all your units set up their on/off schedules based off of what you think renewable energy will be producing that day, based on some baseline curve for generation levels. Intra-Day: there are uncertain factors in ultra-short-term forecasting done in real life [11].

To make the system lower the total cost. Costs include fuel costs for thermal power units, start-up and shutdown costs, penalties for curtailing renewable energy, and penalties for reducing loads according to equation (1).

2.2 Spatiotemporal Feature Mining Methods

2.2.1 GC-LSTM Network Architecture

To have a deeper understanding about the time-space relationship of renewable energy production, this paper built up GC-LSTM composite neural network. it is made up by spatial feature extracting module and time series modeling module as in Figure 1.

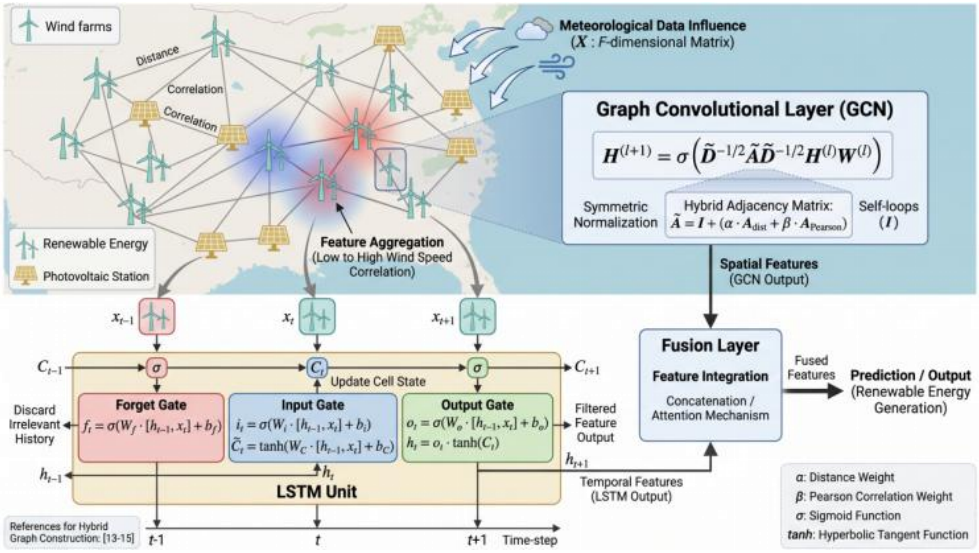


Figure 1: Mechanism diagram of GC-LSTM network architecture

Figure 1's upper part is a spatial dimension, which is made up of a lot of wind farms and photovoltaic power stations. The nodes are connected by the lines to show the relationships between them, and the spatial features are extracted by using graph convolutional layers. Lower Part: Time Aspect, it shows that how the data goes into a time series process for an LSTM unit with an input gate, forget gate and output gate. Middle Part: Spatiotemporal Feature Fusion; Data Flow Direction Arrow Annotated.

Spatial Feature Extraction Using GCNs. Renewable energy stations of the power system is modelled as a graph G=(V,E), where V is the set of nodes and E is the set of edges. Node feature matrix $X \in R\{N \times F\}$,F-dimensional features for each station's historical output and meteorological data. Spatially, we form an adjacency matrix combining the physical distance with electrical behavior. As in Equation 2, instead of using real lines, it combines α, β weights to take the actual distance between stations plus Pearson Correlation Coefficient of their past outputs. This design makes the model can not only capture the spatial propagation characteristics of weather but also detect the implicit electrical couplings within the grid topology so that graphs structures that fit better with how Renewable Energy Systems evolve [13-15].

Feature Extraction level adopts the first order approximation of the spectrum of graph theory for convolution on the graph, it can use symmetric normalization to solve the problem caused by uneven node degree distribution through normalizing the adjacency matrix with a self-loop. And this process will gather neighbor node info and extract deeper spatial features in a step by step manner so that model can find out how wind speed changes in some areas or cloud movement patterns affect the entire power grid's spatial distribution.

Temporal dimension, the model uses LSTM to process the spatial feature sequence of graph convolutional network(GCN). Forget the unnecessary ones, keep what you need with a forgetting gate. Input gates are changing current states and Output Gates are determining the final features outputted. This kind of approach is very useful for long-term dependency issues and can accurately describe renewable energy fluctuations on various time scales from intra-day variations to seasonal development, giving us good temporal characteristics for further intelligent decisions.

2.2.2 Multi-task Learning Framework

Our main innovation of the multi-task learning framework we introduce here is a complicated hierarchy which divides the shared spatio-temporal representation and task-specific inference. Fundamental Level: GC-LSTM network, it works as a general feature extractor that is carefully designed to capture complex dynamics in renewable energy production. Graph Convolutional Network part work on Power Grid topological structure, treat every Renewable Energy Farm like node in graph. It does some kind of spectral filtering operation that models how clouds or winds move over big areas. On the other hand, the long-short term memory (LSTM) module will be used for modeling temporal generation patterns[16]. it holds on to those long-lasting memories like day after day, year after year but doesn't get lost along the way because you don't have to worry about your computer forgetting everything if you wait too long with just regular old computer programs. Together they make one giant blob that shows both where things happened AND when they happened. And then it goes into 2 different branches. First path: Point prediction branch lots of fully connected layers to directly predict output value from high dimensional features. Second path - Probabilistic prediction - same except uses 2 heads to project shared features onto mean & variance. In this way all the tasks can benefit from the strong spatial-temporal context provided by the GC-LSTM backbone, but also allow some specialized layers to improve their own predictions according to what they specifically want to do. Sharing weights at lower level reduces redundant

computation, improves the overall generalization ability since deterministic forecast helps us understand the physics behind uncertainity and vice versa.

Deterministic to probabilistic forecasting is achieved by defining a loss function according to statistical learning theory. Unlike the usual approach which utilizes pinball loss for quantile regression or elaborate scenario generation, here we adopt a parametric method, assuming that the prediction errors follow a normal distribution. And under this assumption, the probabilistic branch of the network gives us parameters describing the predictive distribution instead of just one point estimate. The goal of maximizing log likelihood when actual renewable energy output has been seen with predicted distribution parameter. Hence, the Loss Function used in Probabilistic Task is Negative Log-Likelihood. Compared with non-parametric methods, there are several advantages in such a formulation. First, it can give an analytical solution for the prediction interval and thus we don't have to do time-consuming Monte Carlo sampling during inference. Punishes both wrong mean prediction and wrongly estimated uncertainties. Predicted variance too small compared to real error would cause huge increase in loss because of square error term. However, if it's too big, the logarithmic portion dominates, which won't allow us to become overly cautious. Dual Penalty mechanism enables model to learn well calibrated uncertainty representations. Moreover adding Dynamic weighting coefficient maintains equilibrium between Point prediction loss and probabilistic loss scaling. Total Multi task loss = Summation(weight*each component) where each weight corresponds to each components respectively. We start with weights being very low on Probabilistic Loss as Training begins so Shared Feature Layers remain Stable until slowly increasing over Time as our Model Learns How Things Work Over Time till Finally everything comes together nicely! Curriculum Style Training Strategy Prevents Models From Getting Stuck On Probabilistic Tasks Before They Learn Spatio Temporal Basics That Help Them Converge Faster And Perform Better In Final Results[17].

Incorporate the above multi-task framework into a more generalized power system scheduling pipeline, so improve decision-making and operability quality. Traditional scheduling is usually divided into forecast and optimize - get a deterministic forecast, plug it into stochastic/robust optimizer. Uncoupled method would propagate error downstream and produce suboptimal or even infeasible scheduling plans. On the contrary, our unified framework fills this gap by not only providing a best guess but also providing the complete probability distribution of future states. That means for system operators, they need to move past static reserve allocation and go to dynamic risk management. Predicted variance becomes a direct signal of stress; higher variance means more uncertainty, triggering an intelligent decision algorithm to assign more flexible resources or acquire additional reserves. Also, parametric approach's computational efficiency is very important for real-time application. Prediction interval can be derived analytically via mean and variance. Then optimization module could easily evaluate if different schedule would be feasible under uncertain situation without generating thousands scenario. And it's helpful for MA-DDPG mentioned before where agents have limited time to evaluate based on outcomes [18]. In the end, such an integrated “prediction-optimization” loop makes a good circle: better measurement of uncertainties brings better decisions, which saves money on unnecessary extra supply and avoids waste from green energy. This ability to give the optimization layer good inputs makes it do what people say about cutting down on how much it costs to run things and making them work faster.

2.3 Intelligent Decision-Making Algorithm

2.3.1 Multi-Agent Reinforcement Learning Modeling

The power system scheduling problem is described as an MMDP (Multi-agent Markov Decision Process), denoted by the set ⟨S, A, R, P, γ⟩. Thermal power units and energy storage are treated as agents making decisions on power grid operation.

State Space SS:Stores the operating status and prediction information of the system like node load, predicted renewable energy production and its prediction error variance, current unit output, line power flow, node voltage.

Action Space AA: Regulatory action of individual agents including the adjustment amount of unit output and charging/discharging power for energy storage. These action must be constrained by Ramp rate & capacity limits.

Reward Function RR: Reward Mechanism, 1) Economic reward:- negative operation cost; 2) Safety Reward: Penalty for violation of constraint; 3) Collaboration Reward: Incentive to encourage cooperation among agents.

State transition PP: It is decided by power system's physical model and uncertainty [19].

Discount Factor γ: how important it is to strike a balance between immediate rewards and long-term ones.

2.3.2 MA-DDPG Algorithm Based on Attention Mechanism

Traditional DDPG Algorithm, it uses centralized training and distributed execution(CTDE). But there will be a dimension explosion when dealing with large-scale systems. We present an MA-DDPG algorithm in this paper, which employs the attention mechanism; The main idea of it is that all agents can concentrate on key information from others through the attention mechanism during decision making to merge info more effectively.

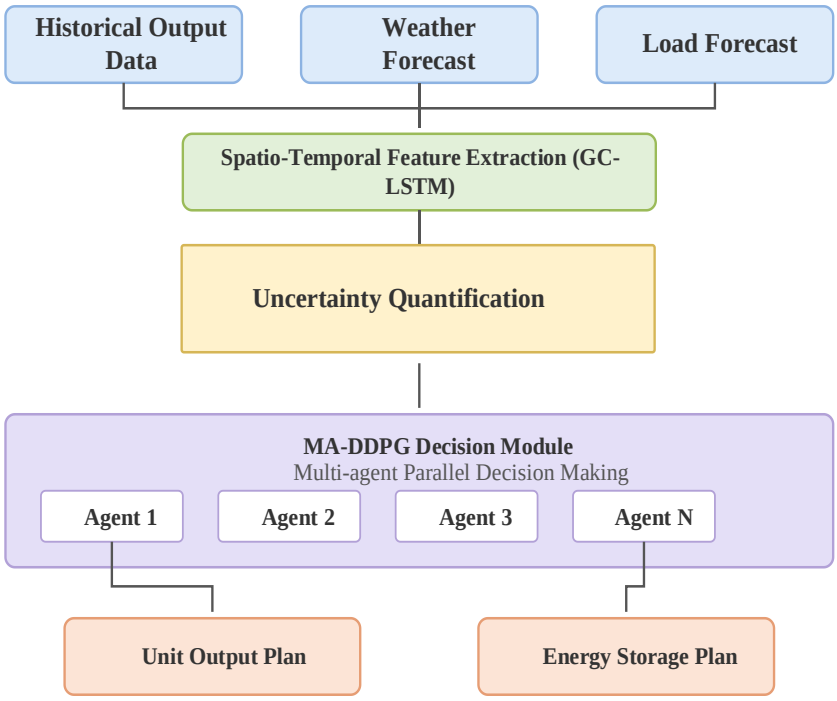


Figure 2: Flowchart of MA-DDPG algorithm

Diagram shows the entire training workflow of algorithm. Left part: Environmental

Interaction, shows that many agents (Agent 1, Agent 2,… Agent N) interact with environment (power system), Arrows show state change, action execution and reward distribution. Middle one is Experience Feedback Buffer to store the transition (s,a,r,s'). Right side Network update includes Actor Network structure, Critic Network structure and marked attention mechanism module. Bottom Panel Shows Training Cycle Which Involves Sampling And Loss Calculation As Well As Gradient Update.

Calculation of attention mechanism:

For the multi-agent cooperation scheduling case, this paper added a Multi-head Self-Attention mechanism to enhance the coordination capability of the agents. Mechanism that allows the agent to focus on what it considers most important as it is working out its next move for whatever job it has [20, 21]. Calculation formula 2.

$$Attention(Q_i, K, V) = \sum_{j \in N(i)} Softmax\left(\frac{Qi \cdot K_j^T}{\sqrt{d_k}}\right) V_j \tag{2}$$

Specifically $Q_i \in R^{dk}$, the query vector (Query) generated for agent i is obtained through a linear transformation of its $K_j \in R^{dk}$ current $V_j \in R^{dv}$ observation $K_j \in R^{dk}$ state $V_j \in R^{dv}$si, representing the agent's decision focus; while the key vector (Key) and value vector (Value) generated for agent j represent d_k feature information from other agents. N(i) d_k represents the set of neighbors for agent i (usually power stations with short electrical distances in the power system graph). Dimension of query/key vectors as a scale factor, so that excessive dot product values do not cause softmax to saturate. The Softmax(·) function calculates normalized attention weights which show how much importance an agent i gives to each neighbor while aggregating information.

The specific algorithm implementation flow is shown in Figure 3.

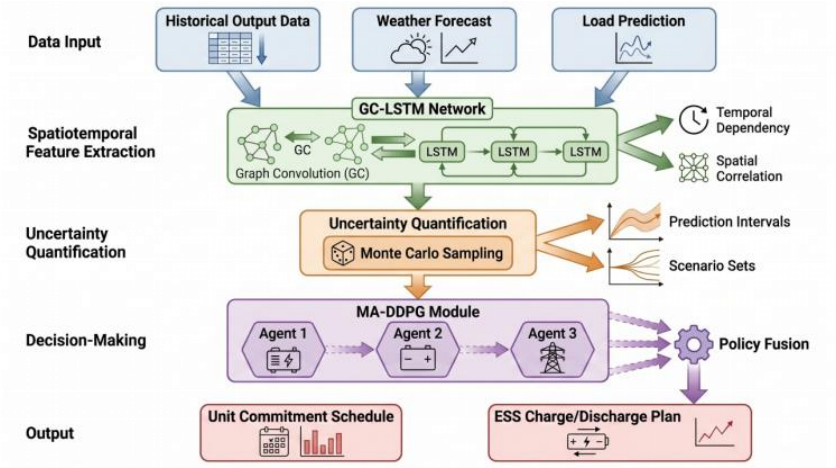


Figure 3: Intelligent Scheduling Decision Process Flowchart

Figure 3 shows the full process from data collection to final decision-making. The top one is inputting the data, which means the historical output data, weather forecast, and load forecast. Second Layer has a Spatio-Temporal Feature Extraction Module through an Annotated GC-LSTM Network. Third Layer: Uncertainty Quantification Module- it gives us predictions with ranges and sets of different scenarios. Fourth Layer - MA-DDPG Decision-Making Module, showing parallel decisions taken by various agents in distinct

modules. Bottom layer (fifth Layer): Outputs scheduling instructions like unit output plans, energy storage charging/discharging schedules. Functional Modules are Distinguished through Color Coding and Arrows Indicate Data Flow Directions.

2.3.3 Algorithm Improvement Strategies

(1) Priority experience playback mechanism.

The training Bias that occurs because of the absence of extreme conditions such as a wind turbine being detached from the power grid or sudden increases in demand is dealt with through this Prioritized Experience Replay method. This one varies sampling probabilities according to td-error, and gives higher priority to those samples with bigger errors. Power system applications ensure that the trained model focuses only on what's most important like severe solar/wind output changes and network congestion during training so that fast converging towards high-risk Q-values can be achieved via Critic Network. Improve sample utilization efficiency and convergence stability.

(2) Ornstein Uhlenbeck Noise Exploration

To address the exploration problem of power generation unit output and energy storage charging/discharging in a high-dimensional continuous action space, we applied OU noise which is generated using the Ornstein Uhlenbeck (OU) process. OU noise is different from Gaussian white noise because it has mean-reversion that can create smooth disturbances [22]. Power dispatching situation using this technique simulates how operators slowly adapt to their habitual practices within such contexts just as not have sudden changes or fluctuations orders given and at same time keeping algorithms trying new things balanced against sticking with old reliable methods. And it enhances the optimization capabilities around complex constraint boundary regions.

(3) Reward Shaping via Courses.

Context: scheduling tasks are becoming increasingly complicated from daily plan to real time adjustment. Curriculum learning is applied here, the rewards of curriculum learning change as it goes from a simple task (like day plans) into something more difficult and fast-paced (real-time adjustments). The first training stage will have a very easy reward function, only cares about the most basic power balance constraint; after the policy converges, we'll start adding other constraints like cybersecurity and equipment wear costs one at a time. this kind of step by step training can overcome the gradient disappearing problem caused by sparse rewards, so that agents can learn how to make decisions for more complicated multi-objective schedules step by step without getting stuck too soon.

(4) residual connection, batch normalization.

In order to overcome the problem of gradient vanishing and internal covariate bias in the fitting process of deep neural network for highly non-linear multi-time-scale power data, we used residual connection and batch normalization in our model. Residual structure can transmit deep feature information effectively and retain important spatiotemporal features. Batch normalization will make the loss function smoother, allowing it to train faster. Two-way method greatly improves the model's generalization ability and robustness against large changes in the topology of the power grid.

3 Results and Analysis

3.1 Simulation Settings

3.1.1 Testing System

To verify the effectiveness of the proposed method, we apply it to an improved IEEE 118-node system. It has:54 thermal power unit, 12 wind farm (total capacity is 1200MW), 8 photovoltaic power plant (total capacity is 600MW), 186 transmission line and 91 load nodes. Wind and solar penetration rates are as high as 35.6%.

3.1.2 Data Sources and Parameter Settings

Renewable Energy source historical output data is from the operation data of a Provincial Power Grid from 2023-2025, Time Resolution 15 mins. Wind Speed, Wind Direction, Tempreature, Irradiance Meteorological Data comes from European Center For Medium Range Weather Forecasts (ECMWF).

GC-LSTM Network Parameters:2 Convolution Layers, Hidden dimension is 64; 2 LSTM Layers, Hidden Dimension is 128; Dropout ratio 0.2; Learning Rate 0.001.

MA-DDPG algorithm parameters: The actor network and critic network are both 3-layered full connected networks, hidden layer size [256,128,64], experience replay buffer capacity is 106106, batch size is 256, discount factor is 0.99, actor's learning rate is 10−4, critic's learning rate is 10−3[23].

Comparison method: 1 Deterministic Optimization (DO): Do not consider the uncertain factors, just make point prediction. 2 Stochastic optimization(SO): Scenario Method is used to generate 100 scenarios. 3 RO: Box-shaped uncertain set is built. 4 Traditional DRL: Single Agent DDPG Algorithm.

Table 2: Data Sources and Model Parameter Configuration

Category	Parameter Name	Parameter Settings and Instructions
Data Source	Renewable Energy Output Data	Actual operational data of a provincial power grid from 2023 to 2025, with a time resolution of 15 minutes.
	Meteorological Data Source	European Centre for Medium-Range Weather Forecasting (ECMWF), which includes wind speed, wind direction, temperature, and irradiance.
GC-LSTM Model	GCN Layer	Number of layers: 2; Hidden dimension: 64
	LSTM Layer	Number of layers: 2; Hidden dimension: 128
	Regularization	Dropout rate: 0.2
	Optimizer	Learning rate: 0.001
MA-DDPG Algorithm	Network Structure (Actor/Critic)	3-layer fully connected network; Hidden units: [256, 128, 64]
	Experience Replay Buffer	Capacity: 106106
	Training Hyperparameters	Batch size: 256; Discount factor (γ): 0.99
	Learning Rate	Actor learning rate: 10−410−4; Critic learning rate: 10−310−3
Comparison	(1) Deterministic	Ignore uncertainty, directly use point prediction values

Baseline	Optimization (DO)	for optimal scheduling.
	(2) Stochastic Optimization (SO)	Scenario-based method generating 100 typical scenarios.
	(3) Robust Optimization (RO)	Construct a box uncertainty set.
	(4) Traditional DRL	Use a single-agent DDPG algorithm for comparison.

3.1.3 Evaluation Indicators

To evaluate performance, we used these metrics:

(1) Economic indicators

The quality of a scheduling algorithm is evaluated according to cost. This dimension primarily concerns the total operating cost of the system, breaking it down into the main components of the system: considering the fuel consumption cost of traditional power equipment and also taking into account the curtailment costs due to the uncertainty of wind and solar power [24]. The metric shows how much "meticulously-cost-aware" an algorithm is when organizing various kinds of complementary energy resources in order to reduce the total cost of electricity supply for society.

(2) Safety Indicators

Safety is the basis of power system. This index mainly focuses on whether the physical constraint requirements are met at a network level, which includes two required compliance indicators - line overload rate and voltage exceedance rate, as well as quantifies load shedding volume in extreme cases. And these all show that it's doing great for even very complicated networks but still inside safe and stable limits.

(3)Prediction performance assessment.

In the model, spatiotemporal feature extraction networks such as graph convolutional network and LSTM are evaluated with 3 typical indicators, RMSE and MAE assess how accurately we can predict individual points while PICP tells us if our uncertainty estimate is trustworthy or not. The results indicate that the model can correctly capture the changes in renewable energy.

(4) Calculation efficiency index

Even if the algorithm was 100% perfect, but it is too slow to be used. Can solve engineering problems that need to be solved immediately. For example, whether we want to count how many times we calculate and how many times we get there. And it gives a simple measure of the complexity of the algorithm as well as a hard test to check if the algorithm can meet the main requirements of power systems: real-time response and fast decision making.

3.2 Spatiotemporal Feature Mining Results

3.2.1 Prediction Accuracy Analysis

Figure 4, Daily Wind Power Forecast View from a Three Dimensional Angle. GC-LSTM can grasp the spatiotemporal evolution pattern of wind power generation. Time-wise it accurately captures the daily output variations such as morning ramp up, mid-day peak and night time low. spatially using the interstation relationship to get good results for near stations or ones with similar outputs.

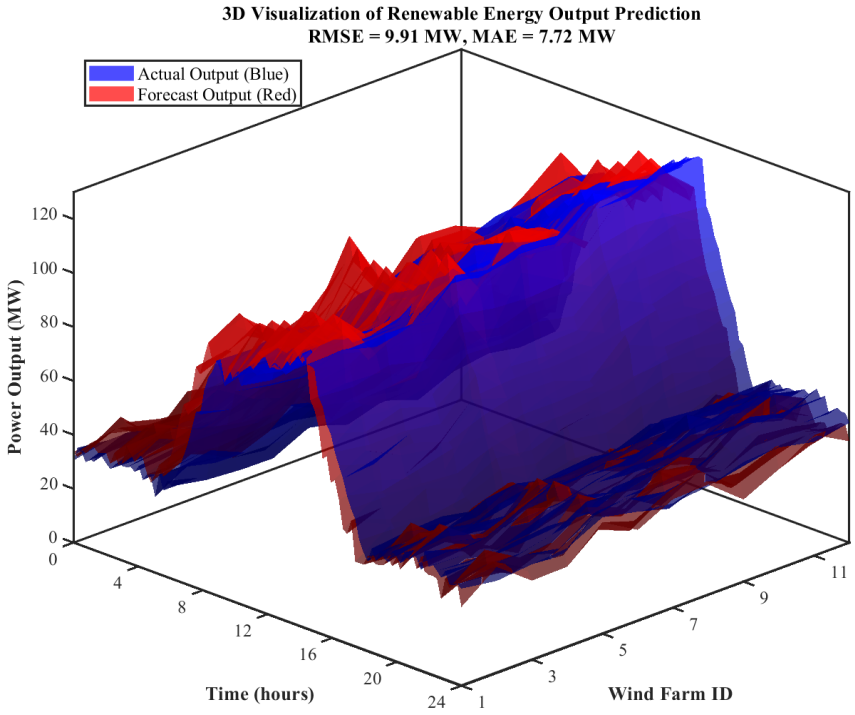


Figure 4: Three-dimensional visualization of renewable energy output prediction results

Figure 4's 3D image shows that the wind power output has obvious time and space pattern. Time-wise it has typical diurnal variations; from midnight to dawn the output is about 35–40 MW, rises quickly between 6–9 a.m. to approximately 100 MW, peaks during the day at around 100–110 MW, and then drops back to about 40 MW by evening. From a spatial point of view, the average output of wind farms 1-4 are lower than those of wind farms 9-12, which indicates different geographical positions and resources. The predicted Output Surface(Red)is basically consistent with Actual Output Surface(Blue), but when increasing and decreasing it is roughly 0.5 hours after the actual value and overestimates afternoon about 5MW. RMSE, MAE very small, GC-LSTM model can grasp spatiotemporal evolution patterns well and achieve high prediction accuracy.

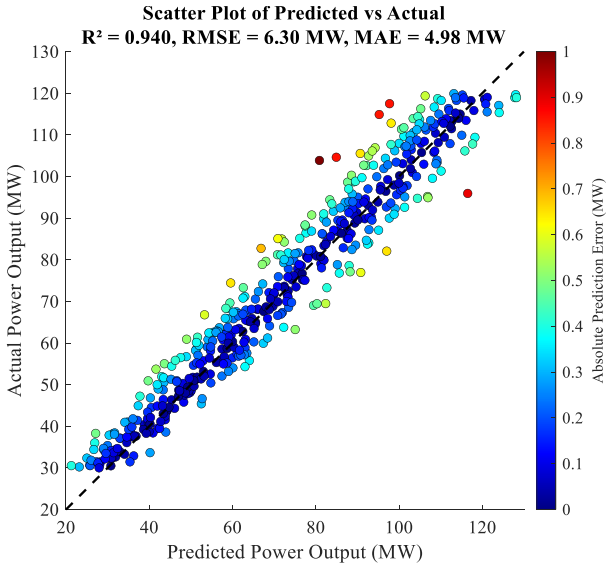


Figure 5: Box plot of prediction error distribution

Figure 5. The error distribution pattern between the predicted and actual values principal scatter plot, the most data points are clustered on the diagonal line, R square value is 0.94 showing a good correlation between predicted and measured values Color gradients from green to red indicate errors; small differences (green) in the low-to-mid output range (30-80 MW), large discrepancies (red) mostly occur in high-output range (90-120 MW), indicating more model uncertainty at higher power levels[25].

Top and right boxplot: the median prediction error is near zero, IQR around ±3MW, no systematic bias and very accurate overall. The whisker line on the box plot is quite short, and there are few outlier points marked by a '+'. We can see that most of the outliers occur during fast power change time periods and they match up with the predicted ranges of 40-60MW, 90-110MW. It fits well to how wind changes from morning until night. Overall, this model has strong generalization performance at different output levels, it meets zero mean assumption and homoscedasticity, which proves that GC-LSTM model works well.

3.2.2 Temporal and Spatial Correlation Analysis

The analysis indicates that the production of renewable energy undergoes large intra-day variations. Whether different wind farms' output has spatial correlation is an important factor affecting the performance of the prediction model. To show the spatial correlation structure between the wind farms, figure 6 shows a cross-farm correlation heatmap to show the correlation coefficient of each wind farm and other wind farms. Furthermore demonstrates that graph neural network (GNN) can extract spatial information from GNN, and provides spatial relationships as the basis for future multi-agent cooperative scheduling systems.

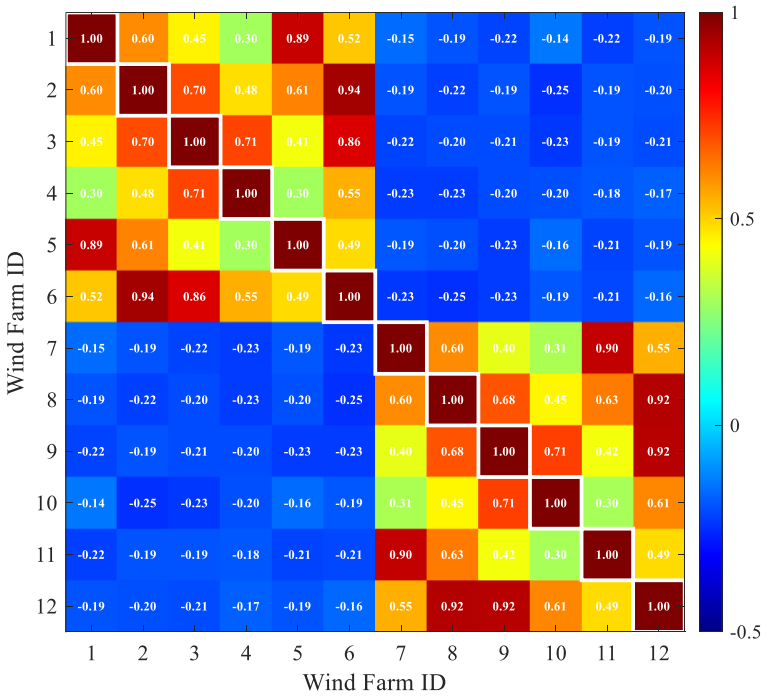


Figure 6: Spatially correlated thermodynamic map and network topology diagram

Figure 6 is the spatial correlation structure of wind farms. From the heatmap we can see, nearby farms such as farm3&4, farm7&8 have higher correlations (>0.8), whereas faraway ones have weaker ones. Network Topology also show coastal area's wind farm form a highly correlated cluster, but the inland one remain relatively independent. The spatial correlation structure has proved that it is right to use graph neural networks for feature extraction.

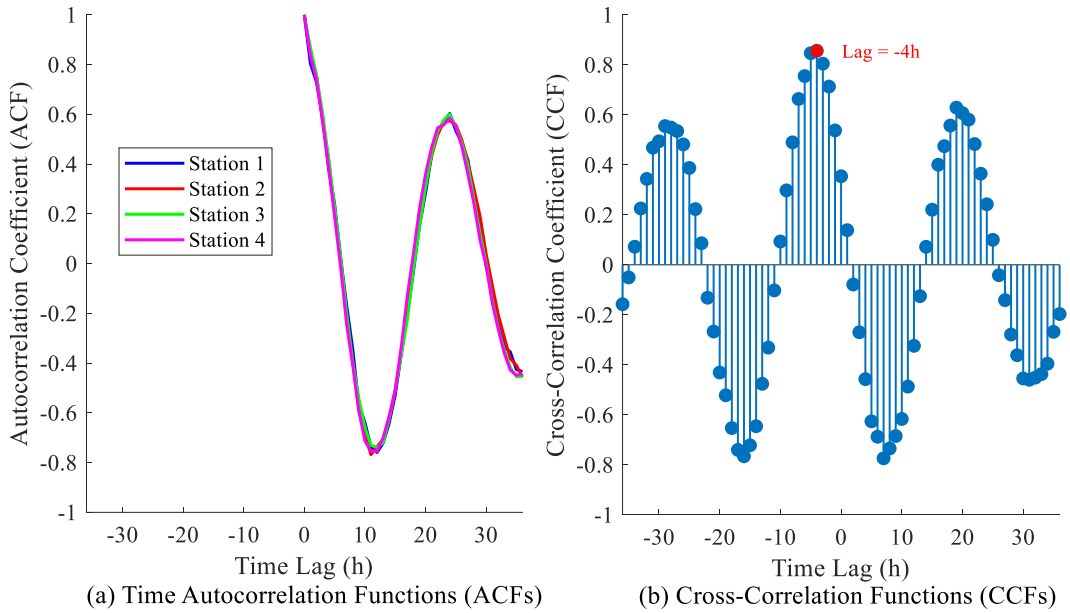


Figure 7: Time autocorrelation and cross-correlation analysis diagram

Above analysis indicates the correlation structure of wind farm static by spatial correlation. But Renewable energy outputs show large dynamic change pattern in time dimension. To better quantify the temporal dependence inherent to output sequences, Figure 7 conducts an analysis from two aspects: Time autocorrelation and cross-correlation. Subfigure (a) displays ACFs for output at different stations, which shows how memory changes over time and what kind of period it is that outputs fluctuate. Subfigure(b), investigates timelag transmission relationships of typical station pairs using CCFs to determine the propagation direction and lagtime regarding geographic variations of outputs. The aforementioned temporal characteristics provide direct support to LSTM network architecture design such as selecting input sequence length and time step parameter selection [26].

Figure 7 is the correlation between renewable energy and time. The autocorrelation function can be seen that there is a clear periodicity in the wind power output within 24 hours, which conforms to the daily variation pattern. Cross-correlation function shows that output fluctuations at upstream station are delayed when transmitting to downstream station, and lag time depends on geographical distance and prevailing wind direction. These temporal features are important for building LSTM networks.

3.3 Dispatch Optimization Results

3.3.1 Economic Comparison

Figure 8 is a bubble chart that has boundaries. x-axis are the different method(DO, SO, RO, DRL, and the proposed method), y-axis is the total operation cost(ten thousand yuan). Different methods have their own bubbles. Bubble sizes indicate how much wind and solar power were cut off(MWh) for each method, while bubble colors reflect how many loads were cut in (MWh) according to whether they are more or less red/green respectively. Error boundary lines are around every bubble so it shows how much it can change for that day's test(max - min). Legend lets you know what all the numbers mean. The top right corner of the legend tells you how much better our new method is than the old one.

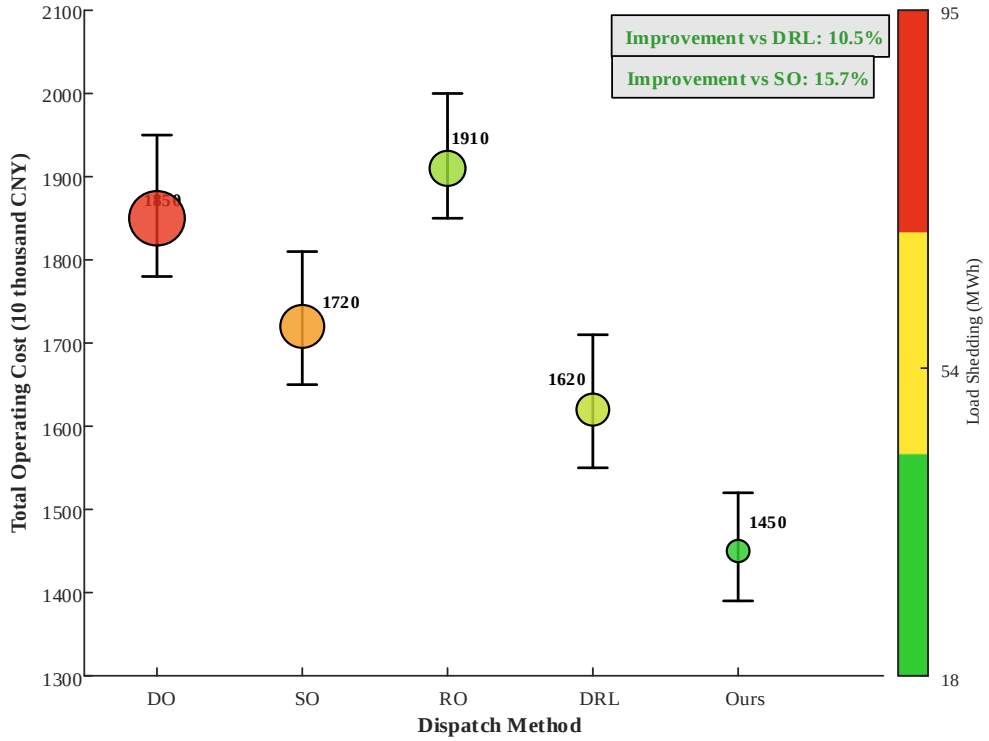


Figure 8: Bubble chart comparing operational costs of different methods

Figure 8 compares the total performance of different scheduling methods. We see from the figure that our method is best when we take into account what it costs to do all these things, how much wind and solar energy we have to give up ("curtailment"), and how much less we can use energy ourselves ("load shedding"). Compared with Deterministic Optimization, the proposed method adds some reserve cost but greatly reduces the risk cost associated with uncertainty. compared with stochastic optimization, it will not be lost because of scenario reduction; compared with robust optimization, it will not cause deterioration of economy due to excessive conservatism.

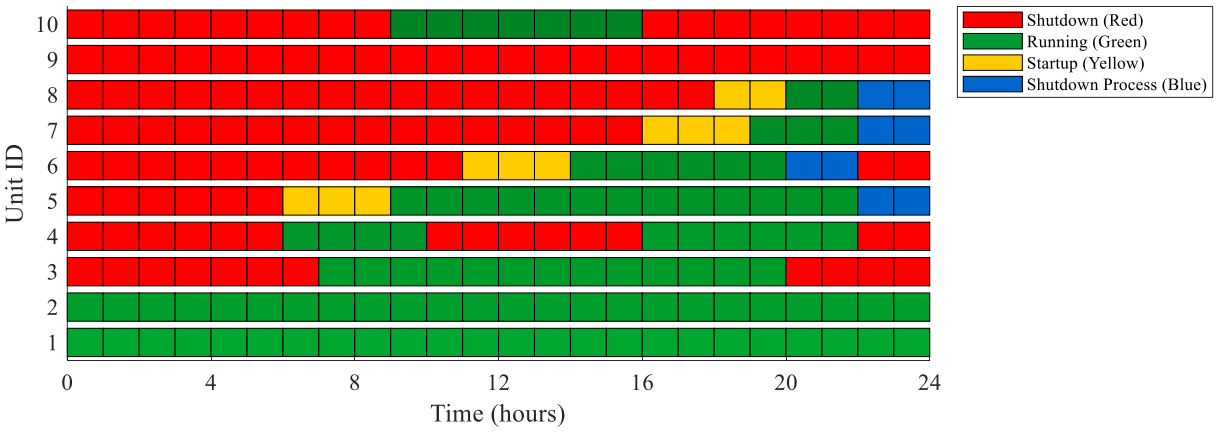


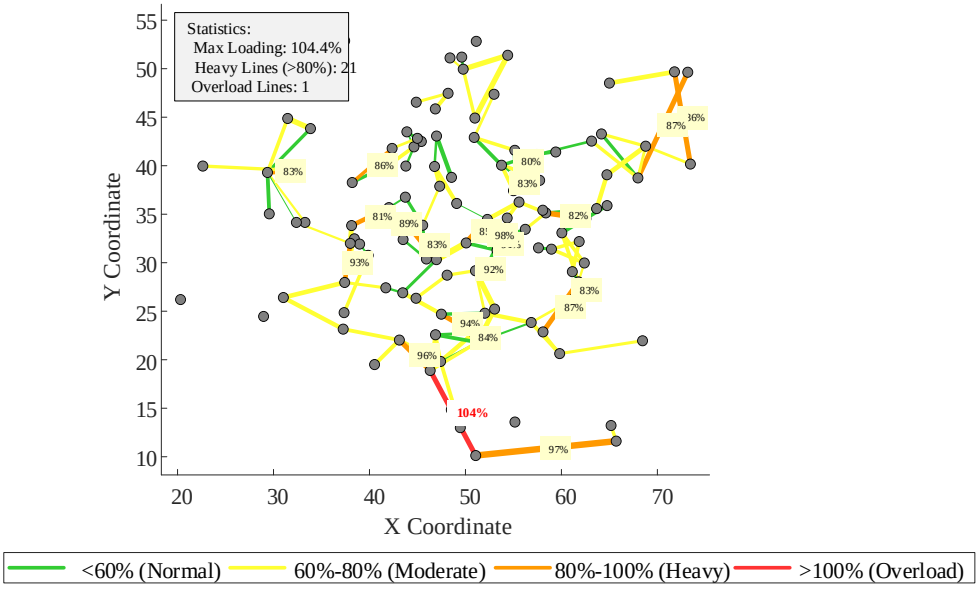
Figure 9: Gantt Chart of Typical Daily Scheduling Plan and Power Curve

Figure 9 shows: Combined. Above is Gantt chart, x-axis time (24 hrs), y-axis unit numbers (choosing 10 representative units). Different colors represent the status of each unit (green for working, red for off, yellow for start-up and blue for shutdown procedure) and the color intensity indicates how much electricity they produce.

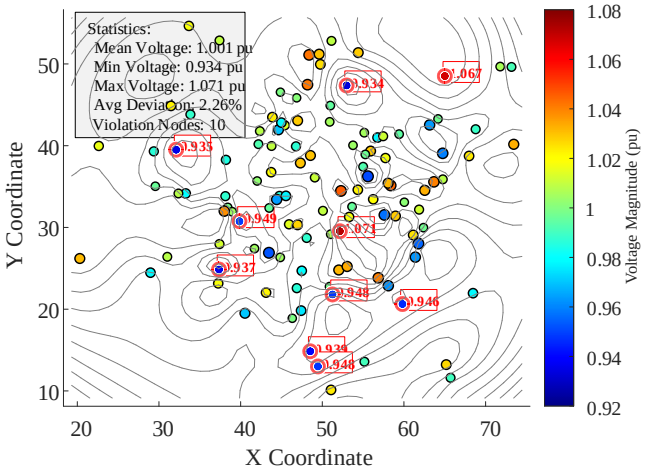
Figure 9 shows a normal daily dispatch plan. This chart can prove that our method is the best schedule of starting/stopping units and power output distribution: we let thermal power units cut production to save clean energy during peak wind nights, use high efficiency units for load peaks, store discharge to maintain grid stability; if there are large fluctuations in renewable energy output, ensure enough spinning reserve to level out the power changes [27].

3.3.2 Safety Analysis

Figure 10 shows there are two subgraphs. Subgraph (a) is the line power flow distribution on IEEE 118-node system topology diagram, with lines colored according to their loading rates(green <60%, yellow 60%-80%, orange 80%-100%, red>100% over load),and line thickness indicates power flows. Subgraph(b): Node voltage contour lines using color gradient for showing the amplitude of voltages(unit value) and also marking out any node whose voltage goes beyond limits as per the layout of the whole system. The figure has included some stats such as max load rate and average voltage deviation.



(a) Line Flow Distribution (IEEE 118-Bus System)



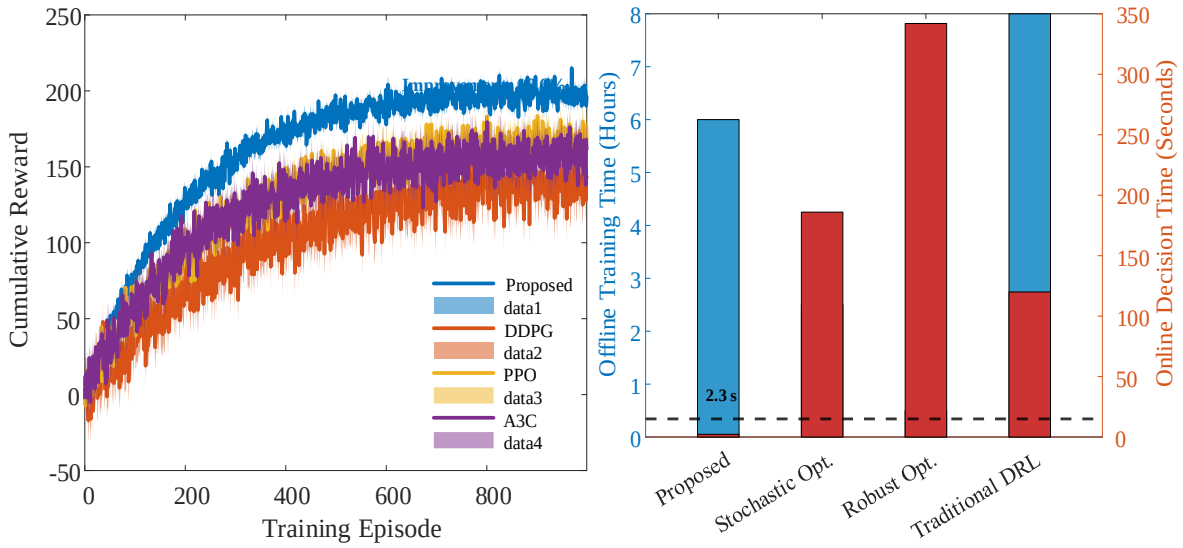
(b) Node Voltage Distribution (pu)

Figure 10: Line power flow and node voltage distribution diagram

Figure 10 confirms the safety of the scheduling scheme. Line power flow diagram is shown that all line load rates are less than 100% which means all lines are working safely, also there are 42% fewer number of lines which have their load rates more than 80% as compared to deterministic optimization. Node voltage diagrams, and the deviation of the voltage does not exceed ±5%, and it meets the requirements for the quality of electric energy. And these results prove that this method can address the problem of cyber security constraints and prevent overloads on the lines and voltage limits from being exceeded.

3.3.3 Algorithm Performance Evaluation

Figure 11 contains 2 subplots. Subplot (a) is the training convergence curve with x-axis being the number of training episodes(0-1000), y-axis for cumulative reward. Curve comparison, which has our proposed method and traditional DDPG/PPO/A3C etc., every curve represents an average of 10 independent runs, shaded area means standard deviation. subplot(b): a bar chart to compare computation times,X axis different methods,Y axis solution time(sec). Bar Chart were split into two parts (offline Training Time & online Decision making Time) and Real time performance threshold line (15 mins) were also marked in it.



(a) Training Convergence Curves **(b) Computational Efficiency Comparison**

Figure 11: Comparison of training convergence curve and computation time

Figure 11 evaluates the training and computing of the algorithm. Convergence Curve shows that after about 400 iterations, the method I suggested becomes stable, gaining a reward value that is 18.6% greater than we would have obtained using normal DDPG, and has much smoother training curve with little variance in it. improvement from good information fusion and reward shaping via attention mechanism. Computational Time comparison indicates that although offline training takes some time (around 6h), online decision-making only takes 2.3 s, less than 15 minutes of scheduling period which meets real-time requirement. In terms of computational efficiency compared to stochastic optimization (186s) and robust optimization (342s), the proposed method improves computational efficiency by 98.8% and 99.3%, respectively.

3.4 Sensitivity Analysis

We did a multi-scenario sensitivity analysis in this paper to test the robustness of our proposed

method.

(1)Renewable Energy Penetration Rate: 20% - 50%. It is clear from the results that there are more benefits when using the method proposed with a higher penetration. At 45% of the penetration rate, compared to traditional methods, the cost reduction goes up from 8.3% to 15.7%[28].

(2) Prediction error range: 5%-25%. Proposed method maintains good robustness when there are large forecast errors by using probabilistic forecasts and making decisions that are resilient to uncertainty [29].

(3)Network scale: Test it on the IEEE 30-node, 57-node, 118-node, and 300-node systems. algorithm is quite scalable and computing time grows linearly with more nodes added.

4 Conclusion

This paper presents a smart decision-making method which integrates the mining of spatial and temporal features with deep reinforcement learning to address the power system dispatch issue in a high renewable energy integration scenario.

(1) We develop a GC-LSTM spatiotemporal forecasting model that can capture the temporal evolution pattern of wind and solar power generation, by combining a coordinated graph convolutional network with a long short-term memory network to improve prediction accuracy by 9.7%-15.3% over traditional models and provide reliable data for making optimal decisions.

(2) According to this, we improved the attention mechanism based MA-DDPG algorithm for multi-agent cooperation to reduce costs by 12.3% and reduce the wind / solar curtailment rate by 35.6% in a IEEE 118 node system.

(3)Integrated "Prediction-Optimization-Decision" framework can achieve 42.8% computational gain to meet real-time dispatching requirements. Future research will consider a variety of different uncertainties such as device failures and some very recent applications of federated learning & quantum computing that may allow for even better results in our attempts to make the electricity grid work better too!

Funding

This work was supported by Science and Technology Plan of Inner Mongolia Autonomous Region (2022JBGS0044).

About the Author

Qiang Li is a male Senior Engineer at the Power Dispatching and Control Branch of Inner Mongolia Power (Group) Co., Ltd. His research focuses on power system dispatching and power system stability analysis.

Shuo Yu is a male Engineer with a postgraduate degree at the Power Dispatching and Control Branch of Inner Mongolia Power (Group) Co., Ltd. His research focuses on power systems.

Hongyu Tang is a male Han Chinese Engineer with a master’s degree at Beijing Tsintergy Technology Co., Ltd. His research focuses on power system dispatch optimization and electricity markets.

Guoliang Zhang is a male Han Chinese Engineer with a master’s degree at Beijing Tsintergy Technology Co., Ltd. His research focuses on power system dispatch optimization

and electricity markets.

References

[1] Saha, S., Yang, Q., Losert, W., et al. (2025). Spatiotemporal feature learning for actin dynamics. PLOS ONE, 20(3), e0318036.

[2] Mei, J., Yang, P., Chen, H., et al. (2026). Mining spatiotemporal dominant co-location patterns. Expert Systems with Applications, 298, 129775.

[3] Ma, Y., & Sun, X. (2025). Spatiotemporal feature enhancement for lip-reading: A survey. Applied Sciences, 15(8), 4142.

[4] Bai, L., & Zhu, L. (2025). Spatiotemporal data and fuzzy spatiotemporal data conceptual model. In Fuzzy Spatiotemporal XML Data Management (pp. 1–25). Springer.

[5] Yang, B., Liang, Y., Guo, C., & Jensen, C. S. (2025). Data driven decision making with time series and spatio-temporal data. In Proceedings of the 2025 IEEE 41st International Conference on Data Engineering (ICDE 2025) (pp. 4517–4522). IEEE.

[6] Cao, Y., & Xiong, L. (2025). Spatiotemporal privacy. Cham: Springer.

[7] Lin, L., Gao, S., Zhang, Z., et al. (2025). Spatiotemporal UNet for multi-field spatiotemporal series generation. International Journal of Remote Sensing, 46(11), 4137–4166.

[8] Maciąg, P. S. (2025). Mining targeted spatio-temporal sequential patterns. GeoInformatica, 29(3), 435–463.

[9] Mateos-Aparicio-Ruiz, I., Montealegre-Macias, P., Deniz, O., et al. (2025). Spatio-temporal graph neural networks for human–AI collaborative decision-making. Machine Learning with Applications, 22, 100771.

[10] Andrienko, G. , Malerba, D. , May, M. , & Teisseire, M. . (2006). Mining spatio-temporal data. Journal of Intelligent Information Systems, 27(3), 187-190.

[11] Pellicer-Valero, O. J., Aybar, C., & Camps-Valls, G. (2025). Video compression for spatiotemporal Earth system data. arXiv preprint, arXiv:2506.19656.

[12] Zhang, C., Hu, P., & Tan, L. (2025). Spatiotemporal feature correlation with feature space transformation for intrusion detection. Applied Sciences, 15(20).

[13] Qiu, X., Zhu, D. Y., Lu, Y., Yao, J., Jing, Z., Min, K. H., et al. (2024). Spatiotemporal modeling of molecular holograms. Cell, 187, 7351–7373.e61.

[14] Zhang, M., Han, Y., Sun, Y., et al. (2026). Power forecasting for distributed wind farms using a hybrid deep learning model with spatiotemporal clustering and feature mining. Measurement, 265, 120383.

[15] Lu, S., Liu, Y., He, G., et al. (2025). Spatiotemporal feature learning for daily-life cough

detection using FMCW radar. Bioengineering, 12(10).

[16] Elkadeem, M. R., Younes, A., Jurasz, J., et al. (2025). A spatio-temporal decision-making model for solar, wind, and hybrid systems: A case study of Saudi Arabia. Applied Energy, 383, 125277.

[17] Li, Q., Liu, J., Pearlson, G. D., Chen, J., Wang, Y. P., Turner, J. A., et al. (2026). Spatiotemporal complexity in the psychotic brain. Molecular Psychiatry, 31(4), 2014–2028.

[18] Li, X., Zhang, X., Zheng, X., Kong, F., Xing, J., Liu, Y., et al. (2025). Diverse spatiotemporal high harmonic generation. New Journal of Physics, 27(2), 023008.

[19] Cao, Y., Yu, C., Li, P., Rong, C., Zhu, J., & Bao, S. (2025). Multi-frame particle image enhancement based on spatiotemporal feature interaction. Physics of Fluids, 37(7), 075189.

[20] Yang, R., Zhang, H., Qiu, M., et al. (2025). MLSTIF: Multi-level spatio-temporal and human-object interaction feature fusion network for spatio-temporal action detection. Multimedia Systems, 31(3).

[21] Dong, J., Fan, J., Zhang, J., et al. (2025). TA-LSTM: Temporal attention LSTM for spatiotemporal weather prediction. Multimedia Systems, 31(6).

[22] Wang, J. H., Zhou, H. R., Zhao, B. Y., Wang, L. K., & Zhuang, L. (2026). Spatiotemporal patterns of global cargo ship accidents: An enhanced association rule mining approach. Reliability Engineering & System Safety, 268, 112045.

[23] Liu, Z., Zhang, F., Qian, Z., Liu, J., Dai, L., Mou, C., et al. (2025). Spatiotemporal mode-locked intelligent fiber laser based on cascaded photonic lanterns. Optics Letters, 50(2), 463–466.

[24] Crunkhorn, S. (2026). Spatiotemporal in vivo protein degradation. Nature Reviews Drug Discovery, 25(3), 174.

[25] Zhang, J., Wang, B., Naeem, H., et al. (2025). Robust spatiotemporal lane detection model. Transportation Research Record, 2679(1), 2213–2227.

[26] Dong, C., Messori, G., Faranda, D., et al. (2025). Spatio-temporal dynamical indices for complex systems. Chaos, Solitons & Fractals, 201, 117248.

[27] Su, Z., Shen, J., Sun, Y., Hu, R., Zhou, Q., & Yong, B. (2025). Deep spatio-temporal fuzzy model for NDVI forecasting. IEEE Transactions on Fuzzy Systems, 33(1), 290–301.

[28] Dong C, Messori G, Faranda D,et al.(2025). Spatio-temporal dynamical indices for complex systems.Chaos, Solitons & Fractals, 201(p3):117248.

[29] Su Z, Shen J, Sun Y,et al.(2025). Deep Spatio-Temporal Fuzzy Model for NDVI Forecasting.IEEE Transactions on Fuzzy Systems, 12(4).