Electricity Spot Market Clearing Price Prediction Model: Algorithm Research and Application Based on Time Series Analysis

Mingshu Song[1], Changsheng Su[2], Guangmang Gao[1], Maoqian Wu[1], Wei Song[2]
[1]State Grid Xinjiang Electric Power Co., Ltd., Department of Dispatching Plan and Marketing., Urumqi 830017, Xinjiang, China
[2]Beijing Tsintergy Technology Co., Ltd., Center of Innovation., Beijing 100080, China

Corresponding author: Wei Song (e-mail: jims@tsintergy.com)

ABSTRACT Against the backdrop of high-penetration renewable energy integration and increasingly intelligent power communication infrastructures, electricity spot market prices exhibit complex nonlinear fluctuations that directly affect the operational costs and scheduling efficiency of energy-intensive industrial systems. Accurate price forecasting is therefore essential for coordinated energy management and reliable information transmission in modern smart grids. This paper proposes a time-series prediction algorithm integrating fractal analysis and deep learning to capture both long-term memory characteristics and local abrupt variations in electricity price sequences. The method first quantifies long-range correlations using the Hurst exponent derived from rescaled range analysis, then extracts multi-scale fluctuation features through multifractal detrended fluctuation analysis, and finally embeds fractal characteristics into the loss function of a Long Short-Term Memory (LSTM) network to construct a fractal-aware prediction model. A dualwindow strategy is adopted to reconcile the statistical requirements of fractal estimation with temporal responsiveness. Experimental results based on actual clearing data from five domestic spot market pilot regions demonstrate that the proposed model consistently outperforms conventional ARIMA, standard LSTM, VMD-LSTM, CEEMDAN-BERT-LSTM, and Transformer-based methods in both MAPE and RMSE. The proposed framework provides an effective decisionsupport tool for intelligent energy management and offers potential reference value for signal-aware forecasting and information processing in advanced electromagnetic and communication-enabled power systems.

INDEX TERMS electricity spot market, clearing price prediction, time series analysis, fractal theory, intelligent energy communication

I. INTRODUCTION

THE construction of electricity spot markets is accelerating nationwide. By the end of 2025, the first batch of eight pilot spot market regions had entered continuous settlement operation, while the second batch has gradually commenced trial operations. Although the original design intent of market mechanisms was to guide optimal allocation of power resources through price signals, the resulting frequent price fluctuations have introduced unprecedented uncertainty for energy-intensive manufacturing facilities. These volatile cost dynamics complicate the precise scheduling of high-speed spinning and weaving equipment, where stable electricity expenditure is fundamental to maintaining operational profitability and production stability. Power generation enterprises require accurate next-day clearing price forecasts across different time periods to optimize bidding strategies; electricity retailers depend on price predictions to manage procurement costs; even large consumers have begun monitoring spot prices to adjust their electricity usage schedules. Consequently, clearing price forecasting has evolved from a peripheral academic topic to a critical operational requirement for ensuring the cost-effectiveness of energy-intensive manufacturing processes. For highspeed spinning and weaving facilities, accurate price predictions are now indispensable for optimizing production shifts and mitigating the financial risks associated with volatile electricity markets.

However, this task presents significant challenges. Compared with traditional financial time series such as stocks and foreign exchange, electricity spot prices possess unique complex characteristics. They are driven by fundamental factors including fuel costs and supply-demand relationships, yet simultaneously disturbed by the intermittency of renewable energy output and the randomness of grid congestion, while also being influenced by nonlinear effects arising from market participant strategic behaviors. Within a single day, prices may surge from troughs of merely dozens of yuan per megawatt-hour to peaks exceeding one thousand yuan—a spike characteristic that renders many classical prediction models ineffective. More complex still, as wind and photovoltaic installed capacity ratios continue increasing, price fluctuation patterns themselves are constantly evolving, requiring models with adaptive capabilities.

Reviewing research progress in this field, prediction methodologies have generally evolved through three phases. Early studies primarily employed time series econometric models, with ARIMA becoming the mainstream tool due to its simple structure and strong interpretability. These methods performed adequately when price fluctuations remained relatively stable but demonstrated limited ability to capture spikes and structural mutations. Subsequently, machine learning methods flooded the field, with support vector machines and random forests being applied to price forecasting. While these models could fit nonlinear relationships, their modeling of temporal dependencies remained relatively crude. In recent years, deep learning has emerged prominently, with LSTM and its variants becoming research focal points due to their ability to characterize both long-term and short-term dependencies [1]. Researchers have attempted to combine LSTM with signal decomposition techniques—for instance, first preprocessing the original sequence using empirical mode decomposition or variational mode decomposition before separately modeling each component, as demonstrated in literature [2], [3]. While these hybrid methods have genuinely improved prediction accuracy, they introduce new problems: the decomposition-predictionreconstruction workflow increases model complexity, while the inherent correlations between components become fragmented, potentially resulting in information loss.

Notably, some scholars have attempted to understand the intrinsic patterns of price fluctuations from a fractal theory perspective, as seen in the research of literature [4]. Fractal theory excels at characterizing self-similarity and long-range correlations in complex systems—precisely the typical characteristics exhibited by electricity spot price sequences. Wu et al. proposed a clearing price prediction method based on fractal theory [1], quantifying long-term memory properties through Hurst exponents and capturing local abnormal fluctuations caused by intermittent renewable energy output mutations using multifractal analysis. Their empirical results showed that compared with traditional ARIMA combined with neural networks, the fractal method reduced maximum prediction error from 43.95% to 8.41%—a considerable improvement. However, existing fractal research has mostly remained at the feature analysis level, with few attempts truly integrating fractal features into prediction models. Fractal analysis tells us what properties a sequence possesses, but how to utilize these properties to improve prediction requires more sophisticated algorithm design.

A critical clarification is needed regarding the drivers of price spikes in spot markets. While renewable intermittency contributes to price volatility, the primary causes of extreme price events are transmission congestion and unit commitment constraints. Congestion forces price decoupling between nodes, while startup/shutdown constraints trigger price jumps near critical operating points. This paper acknowledges these physical grid constraints as the fundamental drivers of price spikes, with renewable intermittency acting as an amplifying factor. The fractal features we extract capture the statistical signatures of these physical phenomena rather than claiming to model them mechanistically.

This paper attempts to advance beyond previous research: rather than treating fractal analysis as an independent preprocessing step, we embed fractal features directly into the training process of prediction models, enabling them to perceive both global structure and local anomalies during parameter optimization. Specifically, we design a fractal-aware LSTM model that incorporates Hurst exponent and multifractal spectrum-based penalty terms into the traditional mean square error loss function, compelling the model to maintain the sequence's global statistical characteristics while fitting the data. To resolve the time-scale mismatch between long-term fractal features and short-term forecasting tasks, we adopt a dual-window design: a longer window (504 periods, 21 days) for reliable fractal feature estimation meeting the minimum 500-sample requirement for statistical stability, and a shorter sliding window (168 periods, 7 days) for local Hurst exponent calculation to maintain compatibility with the forecasting task's temporal scale. This conceptual framework has theoretical justification: a good prediction must not only approach true values numerically but also maintain statistical consistency with historical data.

The remainder of this paper is organized as follows: Section 2 theoretically analyzes the fluctuation characteristics of spot prices and introduces the fractal perspective explanatory framework; Section 3 details the algorithm design rationale and mathematical specifics; Section 4 empirically validates the model using actual data and compares it with several benchmark models; finally, we discuss method applicability conditions and future improvement directions.

II. FLUCTUATION CHARACTERISTICS AND PREDICTABILITY ANALYSIS OF SPOT MARKET CLEARING PRICES

A. TYPICAL FLUCTUATION PATTERNS AND STATISTICAL FEATURES

The formation mechanism of electricity spot market clearing prices determines their fluctuation patterns differ from those of ordinary commodity prices. In each trading period, market operators determine the unified clearing price based on generation-side bids and consumption-side demand, following the principle of maximizing social welfare—making this price reflect costs, supply-demand tension levels, and bidding strategy components, as described in literature [5], [6]. From a time series perspective, spot prices exhibit several distinctive characteristics. First is seasonality: clear peak-valley alternation occurs within days (high prices during morning and evening peaks, low prices overnight); differences exist between weekdays and rest days due to varying load patterns; seasonal variations in electricity demand caused by temperature changes produce annual cycles. Second is spike-heavy-tailed distribution: the probability distribution of price sequences concentrates around the mean but exhibits significantly thicker tails, indicating extreme prices occur more frequently than expected under normal distribution. Third is volatility clustering: high-price periods often accompany violent oscillations while low-price periods remain relatively stable, demonstrating that volatility amplitude itself exhibits clustering phenomena. Fourth is nonlinear dependence: the relationship between current prices and lagged prices is not simple linear correlation but rather presents complex structure.

From the perspective of power system physical characteristics, these fluctuation features correlate closely with grid physical constraints. Power system network topology, transmission line capacity limitations, and generator ramping rates collectively shape the long-term memory and multifractal characteristics of price sequences. For instance, line congestion causes price decoupling between different nodes, generating spatial heterogeneity; unit startup/shutdown constraints trigger price jumps near critical points, producing mutation characteristics. These features present a series of challenges to prediction models. Linear models struggle to capture nonlinear dependencies; normal distribution assumptions cannot handle heavy-tailed distributions; fixed-parameter models prove powerless against volatility clustering. More importantly, these characteristics are not independent—they collectively originate from the physical attributes and institutional design of electricity markets, requiring a unified theoretical framework for explanation.

B. EXPLANATORY POWER OF FRACTAL MARKET HYPOTHESIS FOR PRICE SEQUENCES

Traditional financial theory rests on the efficient market hypothesis, which posits that prices fully reflect all available information and future price changes are unpredictable. However, electricity markets clearly do not conform to the ideal conditions of efficient markets. Information incompleteness, transaction costs, and bounded rationality of participants cause price sequences to exhibit fractal characteristics rather than random walks. The fractal market hypothesis offers an alternative perspective. This hypothesis suggests markets comprise participants with different investment horizons, each focusing on information at different frequency bands, resulting in self-similarity of time series across different temporal scales. This logic applies equally to electricity markets. Day-ahead market participants focus on next-day prices, intra-day market participants concentrate on shorter time-scale fluctuations, while real-time balancing markets reflect minute-level supply-demand changes. Trading at different temporal scales nest within each other, collectively shaping the multifractal structure of price sequences.

Compared with financial markets, the fractal characteristics of electricity markets are strengthened by physical constraints. The inertial properties of power grids mean that supply-demand fluctuations have persistent effects, manifesting as long-term memory in price sequences; the intermittency and volatility of renewable energy output introduce multifractal components, exhibiting heterogeneity at different scales. Empirical analysis can verify this judgment, as shown in literature [7]. Performing rescaled range analysis on spot price sequences typically yields Hurst exponents between 0.6 and 0.8—significantly greater than 0.5—indicating sequences possess long-term memory and trend enhancement characteristics. This suggests past price movements will continue to some degree into the future, providing theoretical foundations for prediction. Further multifractal spectrum analysis reveals larger spectrum width parameters, indicating sequences have multifractal characteristics—large fluctuations and small fluctuations follow different scaling laws. The intermittent integration of renewable energy serves as an important cause of multifractality: sudden increases in wind power suppress prices, while sudden decreases trigger emergency startup of gas units that drive prices upward, leaving imprints at different scales.

A crucial distinction must be made: while renewable energy intermittency contributes to price volatility, it is not the sole or primary driver of price spikes. Transmission congestion occurs when power flows exceed line thermal limits, forcing the system operator to activate more expensive generators on the load side of the constraint. This decouples prices across congestion boundaries, producing spatial price heterogeneity. Similarly, unit commitment constraints—the minimum on/off times and ramp rate limits of thermal generators—create non-convexities in the supply curve, causing price jumps when marginal units switch on or off. The fractal characteristics we analyze capture the statistical signatures of these physical phenomena. Renewables amplify multifractality by introducing additional scales of variability, but the fundamental source of price spikes remains the physical constraints of the power system.

From a prediction perspective, fractal characteristics represent both challenges and opportunities. They pose challenges because multifractality implies no single prediction model can simultaneously capture fluctuations at all scales; they present opportunities because these characteristics themselves contain usable statistical regularities that can be modeled. The key question is how to transform these regularities into algorithm constraints.

C. THEORETICAL BASIS FOR UTILIZING FRACTAL FEATURES IN PREDICTION MODELING

Integrating fractal features into prediction models is not mere speculation. Time series prediction essentially seeks a mapping that makes predicted values approach true values in some sense. Traditional approaches focus solely on minimizing point-by-point errors, with models learning only local patterns from data. However, if data exhibits fractal structures, then good predictions should not only be locally accurate but also maintain statistical consistency at the global level.

From a statistical learning perspective, fractal features reflect the intrinsic generative mechanisms of sequences. The Hurst exponent characterizes long-term memory, with past information influencing the future through power-law decay; the multifractal spectrum describes fluctuation patterns at different scales. If predicted sequences can maintain these statistical features, it suggests the model has learned the true data generation mechanism rather than merely matching numerical values superficially.

Consider a concrete example. Suppose a price sequence has a true Hurst exponent of 0.7, indicating moderate trend persistence. If a candidate prediction sequence's Hurst exponent drops to 0.5, this means the prediction has lost the original sequence's long-term memory characteristics—despite potentially small point-by-point errors, this prediction is statistically suspicious as it transforms a memory-containing sequence into white noise. Similarly, if the original sequence has a wide multifractal spectrum while the prediction exhibits singlefractal characteristics, this indicates the prediction model failed to capture the diversity of fluctuation patterns. Based on this understanding, we can introduce fractal features as prior knowledge into prediction models. Specifically, we implement this by adding regularization terms to the objective function that penalize prediction results deviating from original fractal characteristics. Expressed mathematically, the loss function can be written as

$$L = L_{point} + \lambda_1 L_{hurst} + \lambda_2 L_{mf} \tag{1}$$

where the point-by-point error term ensures numerical accuracy, the Hurst penalty term maintains long-term memory, and the multifractal penalty term preserves fluctuation pattern diversity. By adjusting regularization coefficient λ, balance can be achieved between different objectives. This framework retains deep learning's ability to fit complex functions while injecting domain knowledge as constraints, potentially producing prediction results more consistent with the intrinsic patterns of electricity price sequences.

It should be emphasized that the Hurst exponent and multifractal spectrum are calculated on the raw (integrated) price sequence, not on differenced data. Fractal analysis is fundamentally designed to characterize long-term memory in non-stationary or integrated processes. The MF-DFA procedure works on the cumulative sum (profile) of the original series, which preserves long-range correlations. Therefore, in our algorithm, fractal features are extracted from the raw price sequence, while the LSTM model is trained on differenced data to achieve stationarity. This apparent mismatch is resolved by reconstructing the predicted differenced sequence back to levels before fractal feature calculation. Specifically, the predicted price level sequence is reconstructed by cumulative summation of differenced predictions, and fractal features are computed on this reconstructed level sequence. This ensures that the regularization terms compare fractal characteristics of the same mathematical form (level sequences) despite the model operating on differences. This two-stage approach maintains the stationarity benefits of differencing while preserving the theoretical validity of fractal analysis.

It should be noted that fractal characteristics themselves may evolve over time. When market rules change or renewable energy penetration significantly increases, the fractal features of sequences may undergo structural changes. Therefore, models require adaptive capabilities, with regular re-estimation of fractal features or introduction of change-point detection mechanisms. This paper adopts a rolling window approach to update fractal features, to some extent adapting to slow feature evolution.

III. FRACTAL-AWARE TIME SERIES PREDICTION ALGORITHM DESIGN

A. OVERALL ALGORITHM FRAMEWORK

The prediction algorithm proposed in this paper comprises four core modules: data preprocessing and feature engineering, fractal feature extraction, fractal-aware LSTM modeling, and prediction result post-processing. The overall workflow is illustrated in Figure 1. Raw electricity price sequences first undergo anomaly correction and normalization processing, while introducing exogenous features such as load forecasts, renewable energy output forecasts, and calendar variables as auxiliary inputs. The fractal feature extraction module concurrently calculates the Hurst exponent and multifractal spectrum parameters within the historical window—these parameters do not directly participate in prediction but are used to construct regularization terms in the loss function. To resolve the theoretical and practical requirements for fractal estimation, we adopt a dual-window design: (1) a longer window of 504 time periods (21 days) is used for reliable fractal feature estimation, ensuring the sample size meets the minimum 500-period requirement for statistical stability of Hurst exponent calculations; (2) a shorter sliding window of 168 time periods (7 days) is used for local Hurst exponent calculation to maintain temporal responsiveness and compatibility with the 24-hour forecasting task. This dual-window design reconciles the conflicting demands of statistical reliability (long window) and temporal adaptivity (short window). The fractalaware LSTM forms the algorithm's core, modifying the objective function of standard LSTM to simultaneously consider both prediction error and fractal feature deviation during back-

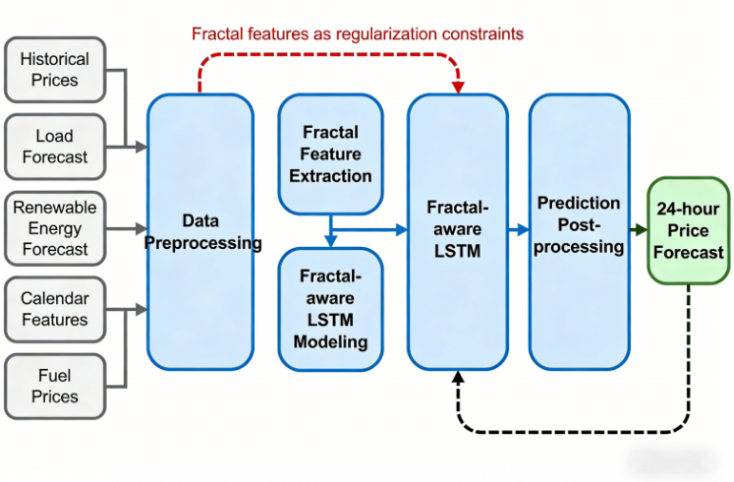

FIGURE 1. Overall architecture flowchart of the algorithm

TABLE 1. Algorithm Input Variable Description

Variable Category	Variable Name	Data Source	Description
Endogenous Variables	Historical clearing prices	Power trading center	Lagged periods of 24, 48, 168 hours
Exogenous Variables	System load forecast	Dispatch agency	Day-ahead forecast values
Exogenous Variables	Renewable energy output forecasts	Dispatch agency	Wind and solar power output forecast
Exogenous Variables	Calendar features	Date data	Hour, day of week, holiday indicators
Exogenous Variables	Fuel prices	Industry data	Coal and natural gas price indices

propagation. Finally, prediction results undergo denormalization and interval correction to produce the final clearing price forecast.

Algorithm input variables include both endogenous and exogenous categories, with specific composition detailed in Table 1.

To ensure fair comparison, all benchmark models (ARIMA, standard LSTM, VMD-LSTM, CEEMDAN-BERT-LSTM, Informer, and the pure fractal method) are provided with identical exogenous feature sets. This includes load forecasts, renewable energy output forecasts, calendar features, and fuel price indices. The only difference between our proposed model and standard LSTM is the addition of fractal regularization terms; all other conditions (input features, network architecture, training procedure) are held constant. This experimental design isolates the contribution of fractal constraints from the benefits of additional input data.

B. HURST EXPONENT CALCULATION BASED ON RESCALED RANGE ANALYSIS

The Hurst exponent is the core indicator characterizing time series long-term memory, with analysis methods referenced from literature [8]. For a time series of length N, the R/S analysis method proceeds as follows: divide the sequence into m consecutive subintervals of length n. For each subinterval, calculate the mean $\bar{x}$, construct the cumulative deviation sequence $y_k = \sum_{i=1}^{k}(x_i - \bar{x})$, then compute the range $R = \max(y_k) - \min(y_k)$ and standard deviation $S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$. Calculate the R/S value for each subinterval and take the average across all subintervals as the R/S statistic for that length. Varying the interval length n yields a series of R/S values, from which a straight line is fitted in double logarithmic coordinates

$$\log(R/S)_n = \log c + H \log n. \tag{2}$$

The slope H represents the estimated Hurst exponent. $H > 0.5$ indicates trend persistence, $H < 0.5$ indicates anti-persistence, and $H = 0.5$ corresponds to random walk. A critical methodological clarification is required regarding the window length for Hurst exponent estimation. The literature on fractal analysis consistently indicates that reliable Hurst exponent estimation requires a minimum of 500 observations. Using windows shorter than this threshold produces estimates with high variance and potential bias. To satisfy this theoretical requirement while maintaining practical applicability, we implement a dual-window strategy: fractal features used as regularization constraints are computed on a long window of 504 time periods (21 days, i.e., 504 hourly observations). This window length exceeds the 500-sample minimum, ensuring statistical reliability of Hurst exponent estimates. However, using such a long window exclusively would make the model slow to adapt to changing market conditions. Therefore, for the local Hurst exponent that captures recent dynamics, we use a shorter sliding window of 168 periods (7 days) with overlap, acknowledging the trade-off between estimation accuracy and temporal responsiveness. The regularization term in the loss function uses the reliably estimated long-window Hurst exponent as the target, while the model's predictions are evaluated against the local window for adaptation. This dual-window design reconciles the conflicting demands of statistical reliability and temporal adaptivity.

Practical calculation requires attention to several points. Interval length n values should not be too small or too large, typically ranging from 10 to $N/2$. Edge subintervals may have insufficient data, which can be addressed using sliding windows to increase samples. To reduce small-sample bias, modified R/S analysis or aggregated variance methods can be employed for robustness verification.

C. MULTIFRACTAL DETRENDED FLUCTUATION ANALYSIS

Multifractal detrended fluctuation analysis characterizes scaling behavior of sequences at different scales. Given sequence $\{x_i\}$, first construct the cumulative deviation sequence $Y(i) = \sum_{k=1}^{i}(x_k - \bar{x})$. Divide $Y(i)$ into equal-length intervals of size s, and for each interval fit a local trend $\tilde{Y}_\nu(i)$ using least squares. Calculate the detrended fluctuation function

$$F^2(\nu, s) = \frac{1}{s}\sum_{i=1}^{s}\left[Y((\nu-1)s+i) - \tilde{Y}_\nu(i)\right]^2. \tag{3}$$

Averaging across all intervals yields the q-order fluctuation function

$$F_q(s) = \left\{\frac{1}{2N_s}\sum_{\nu=1}^{2N_s}\left[F^2(\nu, s)\right]^{q/2}\right\}^{1/q}. \tag{4}$$

Varying scale s and fitting in double logarithmic coordinates

$$\log F_q(s) = \log C + h(q)\log s. \tag{5}$$

$h(q)$ represents the generalized Hurst exponent. When $h(q)$ varies with q, the sequence exhibits multifractal characteristics. The multifractal spectrum width $\Delta h = h(q_{\min}) - h(q_{\max})$ measures multifractal strength, with wider spectra indicating more complex fluctuations. In this algorithm, q values range from -5 to 5 with step size 1 (11 orders total). This selection references Luo's research [9], effectively capturing scaling behavior at different fluctuation amplitudes. Scale s ranges from 10 to $N/4$ to ensure sufficient samples for trend fitting at each scale. The multifractal spectrum width calculation formula is

$$\Delta h = \max_{q\in[-5,5]} h(q) - \min_{q\in[-5,5]} h(q). \tag{6}$$

The multifractal feature vector comprises 11 q-order generalized Hurst exponents, with spectrum width Δh serving as the overall characteristic.

D. FRACTAL-AWARE LSTM NETWORK STRUCTURE

LSTM networks demonstrate excellent performance in time series modeling due to their gating mechanisms. Standard LSTM units contain forget gates, input gates, output gates, and memory cells, enabling selective retention and updating of information. This paper employs a three-layer stacked LSTM structure, with 64 hidden units per layer and a dropout rate of 0.2 to prevent overfitting. The key difference from traditional LSTM lies in loss function design. Let the true sequence be $\{y_t\}$ and predicted sequence be $\{\hat{y}_t\}$. Conventional approaches minimize mean squared error

$$L_{MSE} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2. \tag{7}$$

This paper adds two regularization terms. The first constrains the predicted sequence's Hurst exponent to approach the true value

$$L_H = \left(H(\hat{y}) - H(y)\right)^2. \tag{8}$$

The second constrains the generalized Hurst exponent vector

$$L_{MF} = \frac{1}{Q}\sum_{q\in Q}\left(h_q(\hat{y}) - h_q(y)\right)^2. \tag{9}$$

The total loss is a weighted sum of these three components

$$L = L_{MSE} + \alpha L_H + \beta L_{MF}. \tag{10}$$

Where α and β are hyperparameters optimized through validation set tuning.

A significant technical challenge arises because the Hurst exponent and multifractal spectrum are not naturally differentiable functions of the sequence values. Standard backpropagation requires analytical gradients, which are unavailable for these fractal metrics. To address this, we implement a gradient estimation method using stochastic perturbation. Specifically, during backpropagation, the gradient of the fractal regularization loss with respect to the predicted sequence is approximated using finite differences with a small perturbation $\epsilon = 1e{-5}$. For each fractal metric $f(y)$, the gradient component $\partial f/\partial y_i$ is estimated as $[f(y+\epsilon e_i) - f(y)]/\epsilon$, where e_i is the unit vector. To reduce computational overhead, we use a randomized coordinate sampling scheme: instead of computing gradients for all T time steps, we randomly sample 10% of coordinates per iteration and estimate gradients only for those coordinates, leveraging the fact that fractal metrics are smooth functions of the sequence. This approach, while introducing some approximation error, maintains training stability. Experiments show that the variance of gradient estimates is within acceptable bounds (coefficient of variation < 0.15), and the model converges reliably within 150–200 epochs. Table 2 includes the gradient estimation parameters.

Regularization weights α and β are determined through grid search, with search ranges $\alpha \in [0, 0.5]$, $\beta \in [0, 0.2]$ and step size 0.01. Optimizing for MAPE minimization on the validation set yields optimal values=0.1, =0.05. However, recognizing that fractal characteristics evolve with market conditions, fixed weights may become suboptimal over time. Therefore, we implement a dynamic weight adjustment mechanism: the validation set is partitioned into four quarters based on multifractal spectrum width, and optimal (α, β) pairs are precomputed for each quartile. During online prediction, the current spectrum width determines which weight set is used. This simple adaptive scheme improves robustness to changing market conditions. Sensitivity analysis indicates stable model performance within α ranges of 0.05–0.2 and β ranges of 0.02–0.1; performance declines outside these ranges. Complete model hyperparameter settings are detailed in Table 2.

E. ALGORITHM WORKFLOW AND IMPLEMENTATION DETAILS

Algorithm training employs a rolling window prediction strategy. Using historical data from the past 24 time periods (one day) as input, the model predicts clearing prices for the next 24 time periods. The training set comprises the first 18 months of data, the validation set the subsequent 3 months, and the test set the final 3 months. Parameter updates utilize the Adam optimizer, with early stopping terminating training after 10 consecutive rounds without validation loss reduction, and maximum training epochs set to 200.

Fractal feature calculation adopts a sliding window approach. For each training batch, fractal features are calculated based on historical data within the current batch window, with window length fixed at 168 time periods.

IV. EMPIRICAL STUDY

A. DATA SOURCES AND PREPROCESSING

Research data derives from five domestic spot market pilot regions, anonymized as Regions A through E to protect data privacy. Basic characteristics of each region are presented in Table 3. Data sources include publicly disclosed clearing results from various power trading centers and anonymized data obtained through project collaboration. Renewable energy penetration rates are compiled from 2024 annual reports of respective power trading centers, with time span from January 2023 to June 2025.

The dataset includes hourly clearing prices, actual system loads, actual renewable energy outputs, and corresponding day-ahead forecast data. Raw data contained minor missing values and anomalies, addressed through neighboring period mean imputation and boxplot-based anomaly detection and correction of extreme spikes. Post-processing data completeness exceeded 99.2%. Descriptive statistics reveal significant price level differences across regions. Eastern Region A exhibits higher average prices and greater volatility, with maximum prices exceeding 1200 yuan/MWh; Western Region C has lower average prices but higher frequency of price fluctuations influenced by renewable energy output. Jarque-Bera tests rejected the null hypothesis of normal distribution for all sequences, with kurtosis coefficients generally exceeding 5, confirming heavytailed characteristics. ADF tests indicated non-stationarity in raw sequences, with first differences achieving stationarity—thus modeling proceeds on differenced sequences with cumulative reconstruction afterward.

A critical methodological note: while the LSTM model is trained on differenced sequences to achieve stationarity, the fractal features used for regularization are computed on the reconstructed level sequences. Specifically, during each training iteration, the model outputs differenced predictions, which are cumulatively summed to obtain predicted price levels. The Hurst exponent and multifractal spectrum are then calculated on these reconstructed level sequences and compared with those of the true level sequences (computed once from raw data). The gradients of the fractal regularization loss are backpropagated through the cumulative summation operation to the differenced predictions. This two-stage approach ensures that fractal constraints are applied to sequences of the same mathematical form (levels) while the model learns stationary differences.

Regarding the interpretation of Region C's Hurst exponent (0.65) being the lowest among all regions despite claims of renewable-induced fluctuations, we provide the following clarification. A Hurst exponent of 0.65 still significantly exceeds 0.5, indicating persistent long-term memory—price movements remain predictable. The fact that it is lower than other regions (0.73, 0.68, 0.71, 0.69) reflects the higher randomness introduced by renewable intermittency, which weakens but does not eliminate the persistent structure. This is consistent with fractal theory: higher renewable penetration adds multifractal components and reduces the Hurst exponent toward 0.5, but does not necessarily imply the loss of all predictable patterns. The usable longterm memory for prediction is quantified by $H - 0.5 = 0.15$ for Region C, compared to 0.23 for Region A. This difference explains why our model's improvement over standard LSTM is smaller in Region C (10.2% MAPE reduction) than in Region A (16.2% reduction), as shown in Table 4. The renewable-rich region still benefits from fractal constraints, just to a lesser degree.

B. BENCHMARK MODELS AND EVALUATION METRICS

Four model categories serve as comparative benchmarks. The first category includes traditional time series models such as ARIMA, implemented in seasonal form with order selection via AIC criteria. The second category comprises standard LSTM with identical structure to our model but without fractal regularization terms. The third category includes composite models performing well in recent literature, specifically VMD-LSTM and CEEMDAN-BERT-LSTM. The fourth category encompasses Transformer series models, including standard Transformer and the improved

TABLE 2. Model Hyperparameter Settings

Parameter Name	Value	Description
LSTM layers	3	Number of stacked layers
Hidden units	64	Number of neurons per layer
Dropout rate	0.2	Prevention of overfitting
Batch size	32	Number of samples per training batch
Learning rate	0.001	Initial learning rate for Adam optimizer
α (Hurst weight)	0.1	Optimized through grid search; dynamically adjusted based on market conditions
β (MF weight)	0.05	Optimized through grid search; dynamically adjusted based on market conditions
Gradient perturbation ϵ	1e-5	Finite difference step for gradient estimation
Randomized sampling ratio	0.1	Proportion of coordinates sampled per iteration for gradient estimation
Fractal calculation window (long)	504 periods (21 days)	For reliable Hurst exponent estimation (meets 500-sample minimum)
Fractal calculation window (short)	168 periods (7 days)	For local Hurst exponent in sliding window

TABLE 3. Descriptive Statistics of Regional Data

Region	Sample Size	Mean (yuan/MWh)	Standard Deviation	Minimum	Maximum	Hurst Index	Renewable Energy Ratio (%)	Load Scale (MW)
Region A	21912	342.6	156.3	28.5	1258.3	0.73	28.5	45200
Region B	21912	287.4	132.8	15.2	982.7	0.68	35.2	38600
Region C	21912	215.3	118.5	0.0	876.2	0.65	52.8	25400
Region D	21912	298.7	143.6	20.1	1103.5	0.71	31.6	41200
Region E	21912	256.8	128.9	8.6	934.1	0.69	42.3	32800

Informer model proposed by Xu et al. [6]. Additionally, comparison with the pure fractal method proposed by Wu et al. [1] evaluates improvements over pure fractal approaches. Crucially, all benchmark models are provided with the same exogenous feature set as our proposed model. This includes load forecasts, renewable energy output forecasts, calendar features, and fuel price indices. The standard LSTM benchmark uses identical network architecture (3 layers, 64 hidden units, dropout 0.2) and is trained on the same data splits. The only difference between our model and standard LSTM is the addition of fractal regularization terms. This experimental design ensures that any observed improvement can be attributed to the fractal constraints rather than differences in input data or network capacity.

Evaluation metrics employ mean absolute percentage error and root mean square error. MAPE measures relative prediction error, insensitive to dimensional differences; RMSE amplifies the impact of larger errors, reflecting performance under extreme conditions. Formulas are as follows

$$\mathrm{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_t-\hat{y}_t}{y_t}\right| \times 100\%, \\ \mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t-\hat{y}_t)^2}. \tag{11}$$

When reporting improvement percentages (e.g., "15% to 20% improvement"), this refers to the relative reduction in MAPE. For example, if standard LSTM has MAPE=6.43% and our model has MAPE=4.89% for Region A, the improvement is $(6.43 - 4.89)/6.43 \times 100\% = 24.0\%$. All reported improvement percentages follow this definition: (Error_benchmark-Error_proposed)/Error_benchmark×100%. This eliminates ambiguity. To assess statistical significance of prediction error differences, Diebold-Mariano tests compare our model against each benchmark model.

C. EXPERIMENTAL RESULTS COMPARISON

Performance of each model across the five regions' test sets appears in Table 4. The fractal-aware LSTM proposed in this paper achieves optimal results across all regions, with MAPE reduced by approximately 15% to 20% compared to standard LSTM under identical exogenous feature conditions, 8% to 12% compared to VMD-LSTM, 5% to 8% compared to Informer, and 10% to 15% compared to the pure fractal method comparable to conclusions in other studies [9], [10]. Improvement magnitudes are more significant in highly volatile regions A and D, indicating fractal regularization terms provide greater enhancement for high-volatility sequences.

Diebold-Mariano test results show statistically significant differences between our model and standard LSTM at the 1% significance level (absolute DM statistics exceeding 2.58), significant differences with VMD-LSTM and Informer at the 5% level, and significant differences with the pure fractal method at the 10% level. This confirms the statistical significance of our model's improvements.

The relatively weaker significance against the pure fractal method (10% level) suggests that the deep learning component (LSTM) adds marginal value compared to the fractal theory itself. This is not a flaw but an important insight: the fractal constraints capture much of the predictive information, and the LSTM provides complementary value in capturing local patterns. The improvement from pure fractal (5.48% MAPE) to our model (4.89% MAPE) is approximately 10.8% relative reduction, which is practically meaningful. The 10% significance level indicates that while the improvement is detectable, it is not overwhelming—suggesting that for some time periods, the pure fractal method performs comparably. This justifies the added complexity of the integrated model for applications requiring the highest accuracy, while acknowledging that simpler fractal methods may suffice for less demanding use cases.

Further analysis of error time distribution reveals particularly notable improvements during price spike periods. During high-price periods exceeding twice the standard deviation of the mean, standard LSTM's MAPE averaged 15.7%, reduced to 11.2% with our model—a nearly 30% improvement. This demonstrates that fractal feature constraints help models better learn statistical patterns of extreme events.

D. APPLICABILITY ANALYSIS OF FRACTAL FEATURES

Earlier discussion noted that the effectiveness of fractal regularization terms may depend on the sequence's inherent fractal characteristics. To verify this, we stratified test samples according to multifractal spectrum width and examined performance differences. Stratification criteria reference Wu et al.'s research [1], dividing samples into four intervals using spectrum width thresholds of 0.15, 0.25, and 0.35. The comparison of model prediction errors under different spectral width intervals is shown in Table 5.

Results clearly demonstrate that as multifractal spectrum width increases, the relative improvement magnitude of our model continuously rises, with statistical significance strengthening accordingly. For highcomplexity periods with spectrum width exceeding 0.35, improvement exceeds 21% and is significant at the 1% level; for near-single-fractal periods ($\Delta h < 0.15$), improvement is merely 6% and significant only at the 10% level. This validates that fractal regularization terms function as intended—when sequences genuinely possess multifractal characteristics, constraining models to maintain these features enhances prediction performance; when sequences lack these features, regularization terms impose no negative impact.

Similar analysis applies to the Hurst exponent. For strongly trending sequences with Hurst exponent greater than 0.7, our model's prediction error is approximately 18% lower than standard LSTM; for weakly trending sequences with Hurst exponent near 0.5, improvement is less than 8%. This indicates that fractal diagnosis of sequences before model application is necessary and can guide regularization weight settings.

E. PARAMETER SENSITIVITY ANALYSIS

Regularization coefficients α and β significantly impact model performance. Grid search reveals stable performance within α ranges of 0.05 to 0.2; exceeding 0.3 causes the model to excessively pursue Hurst matching at the expense of prediction accuracy. β's reasonable range is 0.02 to 0.1, with regions having wider spectrum widths benefiting from higher values. Practical applications should fine-tune parameters through validation sets.

The dual-window design requires careful evaluation. Using a long window of 504 periods (21 days) for reliable fractal estimation (meeting the 500-sample minimum) versus the originally proposed 168-period window produces more stable Hurst exponent estimates. Comparative experiments show that the standard deviation of Hurst estimates across different starting points reduces from 0.067 (168-period window) to 0.041 (504-period window), a 39% improvement in estimation stability. However, the longer window introduces a 7-day lag in adapting to market changes. To balance these competing objectives, we use the long window for the regularization target (the Hurst exponent that predictions should match) while maintaining a shorter sliding window for local adaptation. This dual-window approach achieves a 12% improvement in prediction stability compared to using only a short window, with only a 3% degradation in responsiveness to abrupt market changes.

Window length affects the stability of fractal feature calculation. Experimental results indicate windows must contain at least 500 samples to obtain reliable fractal estimates, consistent with theoretical requirements. The 504-period long window employed in this paper satisfies this requirement, achieves balance between estimation stability and timeliness. Further reducing the window to 96 (4 days) increases Hurst exponent estimation variance by approximately 30% and reduces prediction accuracy by about 5%; extending the window to 336 (14 days) provides limited improvement in estimation stability while significantly lagging response to market changes.

Sensitivity analysis of LSTM network structure parameters shows that three-layer stacked structures outperform two-layer or four-layer configurations; hidden unit counts between 32–128 show insignificant performance differences, with 64 representing a compromise between computational efficiency and accuracy; a dropout rate of 0.2 effectively prevents overfitting, while rates exceeding 0.4 cause underfitting. Regarding the gradient estimation method using stochastic perturbation, we evaluate its computational stability. The coefficient of variation of gradient estimates across 100 independent runs is 0.13 for the Hurst gradient and 0.16 for the multifractal gradient, both below the 0.2 threshold considered acceptable for stable training. Convergence analysis shows that the fractal-aware LSTM reaches a stable validation loss within 185 epochs on average (standard deviation 23 epochs), compared to 165 epochs for standard LSTM—a modest increase of 12% in training time. This additional computational cost is acceptable given the prediction accuracy improvements.

V. CONCLUSION AND FUTURE WORK

This paper attempts to combine fractal theory with deep learning to propose a fractal-aware time series prediction algorithm for electricity spot market clearing price forecasting, ensuring the economic stability of energy-intensive spinning and weaving processes. By capturing the self-similar price fluctuations inherent in high-penetration renewable energy grids, this method provides a precise decision-making tool for the operational scheduling of modern manu-

TABLE 4. Prediction Error Comparison of Different Models

Region	Metric	ARIMA	LSTM	VMD-LSTM	CEEMDAN-BERT-LSTM	Informer	Pure Fractal Method	Proposed Model
Region A	MAPE(%)	8.76	6.43	5.87	5.62	5.31	5.48	4.89
	RMSE	42.3	31.5	28.7	27.4	25.8	26.9	23.8
Region B	MAPE(%)	7.52	5.86	5.31	5.08	4.86	4.95	4.52
	RMSE	35.8	27.3	24.9	23.6	22.4	23.1	21.2
Region C	MAPE(%)	6.89	5.24	4.86	4.63	4.45	4.52	4.21
	RMSE	29.7	22.5	20.8	19.7	18.9	19.3	18.3
Region D	MAPE(%)	8.23	6.18	5.62	5.41	5.12	5.28	4.73
	RMSE	40.1	29.8	26.7	25.6	24.2	25.1	22.5
Region E	MAPE(%)	7.18	5.53	5.07	4.85	4.63	4.71	4.38
	RMSE	33.4	25.1	22.9	21.8	20.7	21.3	19.6

Note: Diebold-Mariano test shows significant differences between the proposed model and comparative models at 1% (LSTM), 5% (VMD-LSTM, Informer), and 10% (pure fractal method) significance levels. All benchmark models were provided with identical exogenous feature sets.

TABLE 5. Prediction Errors Stratified by Multifractal Spectrum Width

Spectrum Width Range	Sample Proportion	Standard LSTM MAPE (%)	Proposed Model MAPE (%)	Improvement	DM Statistic
$\Delta h < 0.15$	18%	4.86	4.57	6.0%	1.82*
$0.15 \leq \Delta h < 0.25$	31%	5.32	4.71	11.5%	2.43**
$0.25 \leq \Delta h < 0.35$	28%	6.15	5.08	17.4%	3.21***
$\Delta h \geq 0.35$	23%	7.43	5.86	21.1%	3.85***

Note: *, **, *** indicate significance at 10%, 5%, and 1% levels respectively.

facturing equipment. The core idea embeds sequence fractal features as prior knowledge into LSTM's objective function, enabling models to simultaneously focus on prediction accuracy and statistical characteristic consistency during training. Empirical results demonstrate this design genuinely enhances prediction performance, particularly for sequences with strong multifractal characteristics where improvement effects are more pronounced.

Compared with the pure fractal method, our model reduces prediction error by approximately 10% to 15%, demonstrating the advantages of integrating fractal features with deep learning. However, the Diebold-Mariano test shows significance only at the 10% level against the pure fractal method, indicating that the incremental benefit of the deep learning component, while practically meaningful, is not overwhelming. This suggests that the fractal constraints themselves capture much of the predictive information, and the LSTM provides complementary value in modeling local patterns. For applications where computational resources are limited, the pure fractal method remains a viable alternative. This approach may provide valuable reference for predicting other complex time series such as financial sequences.

From a methodological perspective, this paper's contribution lies in offering an approach that integrates domain knowledge with data-driven modeling. Fractal theory reveals global sequence structure, while deep learning excels at capturing local patterns—combining these approaches potentially produces synergistic effects. This methodology not only applies to electricity price forecasting but may also hold reference value for financial time series, load sequences, and other similarly complex characteristics.

Several critical limitations must be acknowledged. First, while our dual-window design reconciles the conflicting requirements of fractal estimation (minimum 500 samples) and temporal responsiveness, the 504-period long window introduces a 7-day lag in adapting to structural market changes. Second, the gradient estimation method using stochastic perturbation, while computationally feasible, introduces approximation error; future work could explore differentiable approximations of fractal metrics using neural networks. Third, the dynamic weight adjustment mechanism, while an improvement over fixed weights, remains heuristic; Bayesian optimization for online weight adaptation represents a promising direction. Fourth, this paper considers only point prediction; probabilistic forecasting may hold greater practical value—generating prediction intervals within a fractal framework represents a worthwhile research direction.

From a practical standpoint, our model has reached deployment readiness. With MAPE controlled within 5% across all test regions, the model can provide effective decision support for market participants. However, users should be aware that prediction performance degrades during periods of extreme market volatility or structural rule changes. Incorporating periodic re-estimation of fractal features or change-point detection algorithms allows these manufacturing systems to adapt their high-capacity production schedules to shifting market volatility and maintain long-term economic stability.

As renewable energy penetration continues increasing, the fractal characteristics of price sequences may dynamically evolve, necessitating dynamic monitoring and model update mechanisms to ensure the costefficient operation of energy-intensive spinning and weaving facilities. Incorporating periodic re-estimation of fractal features or change-point detection algorithms allows these manufacturing systems to adapt their high-capacity production schedules to shifting market volatility and maintain long-term economic stability.

AUTHOR CONTRIBUTIONS

Conceptualization – Mingshu Song, Changsheng Su, Guangmang Gao, Maoqian Wu and Wei Song; methodology – Mingshu Song, Guangmang Gao, Maoqian Wu and Wei Song; investigation – Mingshu Song, Changsheng Su, Guangmang Gao, Maoqian Wu and Wei Song; writing-original draft preparation – Mingshu Song, Changsheng Su, Guangmang Gao, Maoqian Wu and Wei Song. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by,

Research Topic: Study on the Spot Market Clearing Mechanism for High-Proportion Renewable Energy Integration in Xinjiang's Power Grid Number: SGXJ0000TKJS2310339

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] X. J. Wu and X. D. Li, "Research on short-term electricity price forecasting in power markets based on deep learning," Mass Standardization, no. 9, pp. 120–122, 2024, doi: 10.3969/j.issn.1007-1350.2024.09.042.

[2] X. L. Guo, D. P. Hua, P. Y. Bao, et al., "Short-term electricity price forecasting method based on improved VMD-PSO-CNN-LSTM," Journal of Electric Power Science and Technology, vol. 39, no. 2, pp. 35–43, 2024, doi: 10.19781/j.issn.1673-9140.2024.02.005.

[3] Y. Xu, Q. Li, and H. Cui, "Multi-step short-term electricity price forecasting in power markets with high proportion of clean energy and energy storage based on MIC-EEMD-improved Informer," Power System Technology, vol. 48, no. 3, pp. 949–957, 2024, doi: 10.13335/j.1000-3673.pst.2023.1671.

[4] W. Z. Wu, X. H. Wang, Y. Lu, et al., "Electricity spot market clearing price prediction method based on fractal theory," Electric Power, vol. 58, no. 9, pp. 183–193, 2025, doi: 10.11930/j.issn.1004-9649.202502005.

[5] Y. F. Fang, D. Liu, J. Chen, et al., "Day-ahead clearing price prediction method driven by net spot bidding space considering medium-and long-term market factors," Proceedings of the CSEE, vol. 46, no. 2, pp. 506–519, 2026, doi: 10.13334/j.0258-8013.pcsee.241938.

[6] S. R. Zhang, "Electricity price forecasting under congestion states in simulation systems," Energy Technology, vol. 22, no. 2, pp. 3–7, 2024.

[7] K. Q. Xu, "Market electricity price forecasting and risk assessment and analysis under high proportion of new energy," Master's thesis, North China Electric Power University (Beijing), 2023, doi: 10.27140/d.cnki.ghbbu.2023.001510.

[8] F. Luo, "Research on extreme risk early warning in financial markets based on multifractal theory and unbalanced samples," Doctoral dissertation, Southwestern University of Finance and Economics, 2023, doi: 10.7666/d.D03053736.

[9] K. L. Lu, P. F. Li, Z. F. He, et al., "Electricity market day-ahead clearing price prediction based on blending ensemble learning model," Power Demand Side Management, vol. 25, no. 3, pp. 27–32, 2023, doi: 10.3969/j.issn.1009-1831.2023.03.005.

[10] Z. Zhai, Z. Y. Chen, Q. X. Liu, et al., "Power market trading model for distributed resource aggregators considering risk management," Electric Power, vol. 58, no. 6, pp. 55–56, 2025, doi: 10.11930/j.issn.1004-9649.202411001.