energies

MDPI

Article

China's LNG Ex-Factory Prices by Integrating Large Language Models and Ensemble Learning: A Cutting-Edge Hybrid Approach

Jin Liang, Jiulong Wang, Xiaotian Luo and Hongqing Song *

School of Resource and Safety Engineering, University of Science and Technology Beijing, Beijing 100083, China
* Correspondence: songhongqing@ustb.edu.cn

Abstract

China's liquefied natural gas (LNG) ex-factory prices are affected by market-oriented reforms, supply structure changes, and diverse heterogeneous factors, presenting strong nonlinearity, nonstationarity, multiscale coupling, and high-frequency disturbances. Traditional models fail to capture such complicated dynamics, bringing difficulties to industrial capacity allocation, procurement, and inventory management. This paper proposes a CEEMDAN–LSTM–MiniMax hybrid framework combining signal decomposition, ensemble learning, and large language model semantic reasoning to realize closed-loop synergy of numerical forecasting and semantic decision-making. CEEMDAN decomposes raw price series, and LSTM models each component with validation set-based combinatorial optimization for ensemble prediction. The embedded MiniMax LLM extracts information from policies, market sentiment, and supply–demand texts to calibrate forecasts and support human–machine collaborative pricing, constructing a "numerical drive–semantic refinement–decision output" loop. Empirical results show MAE = 29.424, MAPE = 0.744%, RMSE = 38.557, and R^2 = 0.991. The model outperforms single LSTM and conventional CEEMDAN ensemble models in accuracy and adaptability to extreme price volatility.

Keywords: LNG price prediction; machine learning; ensemble models; LSTM; large language model; CEEMDAN–LSTM–MiniMax

1. Literature Review

In the context of the accelerated reform of the Chinese natural gas market and the profound adjustment in the supply-and-demand pattern, factory prices of LNG (liquefied natural gas) in China exhibit complex characteristics of "high volatility, strong uncertainty, and multi-factor coupling". Traditionally, based on patterns of seasonality and oil-price linkage, these rules have been broken. The supply structure has tilted toward unconventional gas sources and domestic production [1]. Although the fluctuation range has narrowed, the frequency of disturbances has increased, and policy reforms continue to reshape market expectations. These characteristics have led to the LNG factory price series exhibiting strong nonlinearity, nonstationarity, and multiscale coupling and being driven by multiple heterogeneous factors. From an industrial practice perspective, accurate price forecasting is the core basis for LNG production enterprises to optimize capacity allocation and control the cost of raw gas procurement; for downstream enterprises, precise forecasting can reduce inventory holding costs and procurement risks and provide scientific support for resource scheduling. From the perspective of industry governance, the price

Check for updates

Academic Editor: Nikolaos Koukouzas

Received: 17 August 2026
Revised: 16 October 2026
Accepted: 19 October 2026
Published: 25 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

forecasting results serve as important references for the government to improve the price regulation mechanism and prevent market abnormal fluctuations as well as key technical supports for promoting the marketisation reform of natural gas and establishing regional price benchmarks. Moreover, with the promotion of "carbon neutrality" LNG and the acceleration of new energy substitution, the future price formation mechanism will become more complex, and it is urgently necessary to conduct precise forecasting to provide a decision-making basis for the low-carbon transformation of the industrial chain. Therefore, researching the prediction of LNG factory prices in China is not only a practical need to solve the current market "increased volatility and difficult decision-making" predicament but also a strategic measure to enhance energy market governance capabilities and ensure the stable operation of the energy system.

In recent years, machine learning methods have been widely applied to natural gas price forecasting, demonstrating significant advantages in handling large-scale data and complex nonlinear relationships. Related research has experienced three stages of development.

1.1. Phase 1: Continuation and Optimization of Traditional Statistical Models

This phase focuses on the inheritance and optimization of machine learning built upon traditional statistical models. Time series models including ARIMA [1] and its seasonal variant SARIMA [2] are widely adopted as baseline tools for time series forecasting. They feature high interpretability and simplicity yet struggle with nonlinear relationships and multi-dimensional influencing factors.

To address this limitation, scholars combine deep learning with market variables for prediction. Machine learning approaches such as ANN [3], BP neural network [4], random forest (RF) [5], SVR [6], SHVAR [7], GARCH [8], ELM, LSTM [9], and XGBoost [10,11] achieve good performance under low-dimensional feature conditions.

Li Mengyi et al. carried out prediction and analysis of the PRTS Japan–South Korea benchmark index using time series models [12]. Huang Xin et al. built an EEMD–GRU hybrid model for natural gas price forecasting [13]. Wang Jianliang et al. improved existing data mining-driven price forecasting algorithms and proposed an enhanced data mining-based natural gas price prediction approach [14]. Taking Henry Hub futures as the research object, Jiarong Mu et al. verified the effective forecasting capability of LSTM for futures prices [15]. Dogukan Sahin compared XGBoost and LightGBM for daily spot natural gas price prediction and found that XGBoost yielded better results [16].

1.2. Precision Breakthrough for Hybrid Models

This stage focuses on innovative fusion and precision breakthroughs in hybrid models. It introduces a stacking strategy for multimodel integration using multiple benchmark model outputs as meta-model inputs. To address irrelevant features in original time series data, decomposition methods are introduced to preprocess data and convert nonlinear data into linear feature patterns, improving prediction accuracy. Mainstream methods include STL with local weighted least squares [17], EMD [18], and VMD [19]. For example, Du Xueping et al. proposed a BP–ARIMA combined model that improved prediction accuracy [20]. Pan Kai et al. used a CCA–ELM neural network with multi-factor data for prediction [21]; Sun et al. studied LNG prices using a VMD–CEEMDAN–ELM (variational mode decomposition–complete ensemble empirical mode decomposition with adaptive noise–extreme learning machine) combined model [22]. Zheng et al. developed an Informer–FFT framework for short-term dynamic prediction of daily natural gas prices, addressing high-frequency sequence modelling and periodic feature extraction [23]. Zhang Jinliang proposed a hybrid model based on CEEMD, ELM, and ARIMA [24]. Yu Lin et al.

proposed ICEEMDAN–STLSTM–TCN–CBAM, which performed well under different sliding windows [25]. Jun Wang et al. developed a CEEMDAN–SE and PSO–ALS–GRU hybrid method, effectively improving prediction accuracy [26]. Jinpei Liu et al. proposed an Ao–VMD–PCnet–SCE comprehensive prediction model using PCnet and network uncertainty to reveal potential features in historical data [27]. Jianliang Wang proposed a hybrid model combining SVR, LSTM, and improved pattern sequence similarity search (IPSS) for daily gas prices [28]. Duan et al. developed CEEMDAN–Bagging–HHO–SVR for Henry Hub monthly natural gas predictions, showing better performance than traditional models [29].

1.3. Multi-Source Heterogeneous Feature Engineering and Integration of Large Models (LLM)

Existing research has built a multi-dimensional "price–inventory–macro environment" feature system using Henry Hub spot price, CME futures, IEA inventory, and external factors such as crude oil prices, extreme weather indices, and BERT sentiment analysis of geopolitical events. Yang et al. selected eight variables including JKM, NBP, HH, and Brent crude oil and used LSTM, ANN, and SVM to predict JKM, finding that JKM, NBP, and Brent crude oil were the most influential [30]. Ceperic et al. used 12 variables with a feature selection (FS) algorithm for Henry Hub spot price, showing FS avoids variable redundancy [31]. Gao et al. used a time-varying parameter model, finding time-varying dynamics crucial for prediction accuracy [32]. Du et al. introduced 16 variables covering production, marketing, commodities, political/economic indicators, and temperature and proposed a multi-variable joint prediction model [33].

These methods work in stable markets but ignore unstructured text—sudden geopolitical events, policy shifts, and real-time sentiment—that often causes sharp price fluctuations. LLMs enable deep coupling of text intelligence and numerical prediction. In natural gas price prediction, research integrating LLMs and machine learning is growing, moving from simple text embedding to multimodal fusion.

Early work introduced open-domain event extraction. With mature pre-trained language representation, deeper integration became feasible. Guan et al. coupled BERT news embeddings with bidirectional LSTM, attention, and a Transformer encoder, confirming that financial news semantics significantly improve trend prediction [34]. Data-driven methods such as PLO–Transformer–CEEMDAN described by Wang et al. focus on price decomposition but remain largely numerical [35].

Basic models further optimized energy prediction. EF–LLM fine-tuned an LLM using prefix fine-tuning and low-rank adaptation and then combined it with a time series prediction head to handle sparse data and reduce hallucinations. Although these innovations matured first in power and carbon markets [36,37], their mechanisms transfer to natural gas. The model integrating LLaMA3 and RVM reported by Wang et al. used 3.3 million time series and 160,000 indicators, improving accuracy (MAPE 3.96%) and analyzing long- and short-term macro/micro factors [38].

2. Novelty of the Work

Despite progress, LLM–machine learning integration for natural gas prices has limitations: most large models are used for isolated text analysis, not deeply integrated with time series models, so no closed loop of "numerical prediction + semantic decision-making" is formed. To address this, this paper proposes a framework with innovations:

(1) A pioneering architecture deeply integrating LLM-driven semantic understanding with time series prediction, achieving the closed loop and building a CEEMDAN–LSTM–MiniMax model.
(2) In numerical prediction, a CEEMDAN–LSTM combined model is built, with multi-algorithm comparison and combination optimization across traditional machine learn-

ing and deep learning. A CEEMD with adaptive noise set (ANS) preprocessing decomposes prices into IMFs and residuals, reducing nonstationarity. After separate component prediction, combination optimization (e.g., genetic algorithms or ridge regression) reconstructs results. Experiments show that the CEEMD framework integrated with combinatorial optimization achieves significant improvements across multiple error indicators, particularly in terms of robustness amid extreme price fluctuations.

(3) In semantic decision-making, an intelligent pricing assistance framework integrates machine learning and MiniMax. It extracts nonlinear features and pricing rules from a "case pool" of historical market states, pricing schemes, and human experience; the LLM then semantically understands and reasons over multi-source heterogeneous information (market text, policy signals, supply–demand dynamics), correcting numerical predictions and generating preliminary pricing schemes. A "pricing decision group" (price setters and quantitative indicators) performs multi-criteria assessment and correction, achieving human–machine closed-loop optimization. This breaks traditional pricing's reliance on structured data and improves adaptability, interpretability, and decision efficiency.

3. Research Methods

3.1. Research Framework and Overall Technical Route

This study focuses on the high-precision prediction of the ex-factory price of liquefied natural gas (LNG) in China. Following the main line of "data preparation—signal decomposition—fusion modelling—performance evaluation—decision application", a combined prediction framework integrating complete ensemble empirical mode decomposition (CEEMDAN), multiple machine learning models, and large language models (LLM) was constructed.

During the data preparation stage, time series data on China's LNG ex-factory prices from January 2016 to February 2026 were collected. Seven exogenous variables were simultaneously introduced, covering macroeconomic indicators, policy and regulatory texts, public opinions on geopolitical and risk factors, major events, as well as expandable market, trade flow, and scenario data, to form a panel time series dataset. Preprocessing operations, such as filling missing values, detecting and handling outliers, standardizing, and normalizing, were applied to the original data to ensure data quality and the stability of model training.

In the signal decomposition stage, the CEEMDAN algorithm is used to decompose the preprocessed price sequence into several intrinsic mode functions (IMFs) and a trend term. By converting the original nonstationary, multiscale sequence into relatively stable, more regular subsequences, a foundation is laid for subsequent component-level modelling.

During the fusion modelling stage, first, individual machine learning models such as LSTM, BP neural network, XGBoost, random forest, and extreme learning machine (ELM) are separately constructed, and the decomposed and reconstructed model, namely CEEMDAN, is combined with the above five machine learning models to form CEEMDAN–LSTM, CEEMDAN–BP, etc. The initial price prediction values are obtained from these models. On this basis, a CEEMDAN–LSTM–MiniMax combination framework is further constructed. By integrating policy, public opinion, historical experience, and domain expert knowledge through a large language model, the initial predictions are corrected, and causal explanations and risk warnings are provided, yielding an intelligently priced solution with explainability.

In the model performance evaluation stage, the root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and mean absolute percentage error (MAPE) were used as evaluation metrics. The system compared the prediction accuracy and robustness of a single machine learning model, the decomposition-enhanced model (CEEMDAN–LSTM), and the proposed CEEMDAN–LSTM–MiniMax combined framework.

Finally, at the decision application level, the obtained prediction results are combined with knowledge bases such as historical pricing policies, historical pricing cases, and market environments to support the transition from auxiliary to active decision-making, providing a scientific basis for the pricing decision-making group and price setters. This research framework, through a closed-loop design of "decomposition–multimodel component learning—LLM knowledge enhancement—evaluation—application", effectively improves the accuracy, interpretability, and practical value of LNG plant price predictions. The overall framework of this study is presented in Figure 1.

3.2. Influencing Factors of Target Variables and Knowledge Base Selection

This study takes the China LNG factory price as the core prediction target. Considering the multiple driving factors of industrial production, market supply and demand, cost transmission, the international energy market, and environmental factors, seven types of key exogenous variables are selected as model inputs to construct a system of factors influencing the LNG factory price.

The data in this paper are derived from the Wind database, the Shanghai Petroleum and Natural Gas Exchange, and the China Meteorological Data Network. To ensure no information leakage, no overfitting, and the reproducibility of the prediction results, this study strictly divides the dataset in absolute chronological order without cross-division. The training set covers from 1 January 2016 to 4 February 2022 for model parameter initialization and basic learning. The validation set, from 5 February 2022 to 17 February 2024, was used for hyperparameter optimization, weight fusion, and model screening. The test set covers the period from 18 February 2024 to 28 February 2026 for simulating real-world future prediction scenarios. The influencing factors listed in Table 1 are highly relevant to the operational characteristics of China's LNG market. Details are shown in Table 1.

Table 1. Influencing factors of the LNG price forecast in China.

Influencing Factors	Notes
Production	It reflects the supply-side situation and affects the balance between supply and demand in the market.
Demand	It reflects the situation on the demand side and directly affects the price trend.
Stock	It reflects market inventory levels and influences price fluctuations.
Terminal price	The price of imported LNG terminal is in a competitive relationship with the LNG ex-factory price.
Gas source price	It reflects upstream costs.
Brent oil price	It reflects the impact of energy costs.
Temperature	It reflects the influence of seasonal factors.

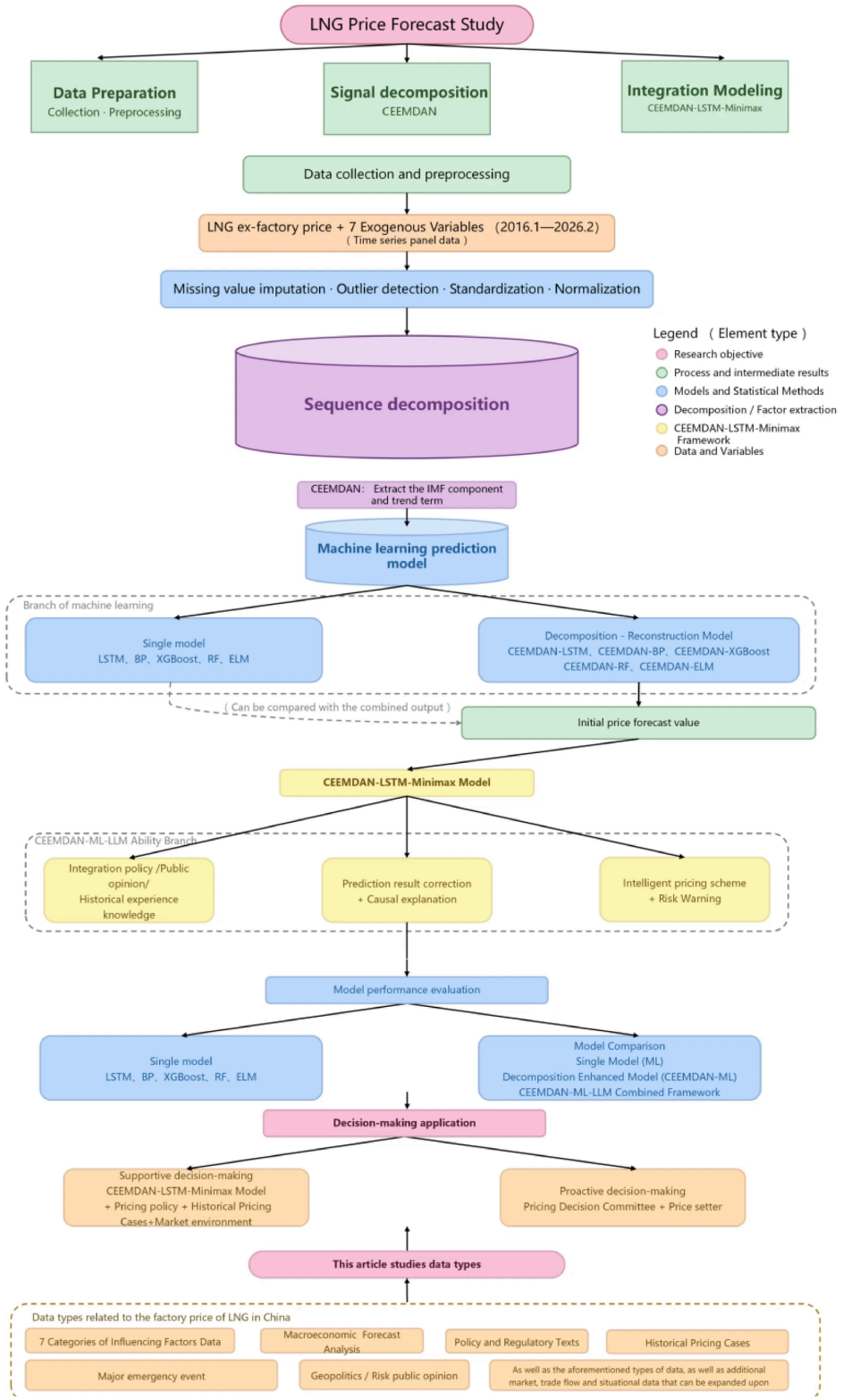


Figure 1. Framework of the study in this paper.

Missing value filling, outlier screening, and standardization were specified on the training set only; the resulting rules and the training set mean and scale were then applied unchanged to the validation and test sets. Later observations were not used to re-estimate these preprocessing parameters. Only after the split and the training-only preprocessing was CEEMDAN applied to the preprocessed training price series, producing intrinsic mode functions and one residual. LSTM, BP, XGBoost, random forest, and ELM were

then estimated on those training components. The validation set was used only to select hyperparameters, fusion weights, and the retained model. The test set was held out until the final evaluation and was not used to revise preprocessing, the decomposition, or the estimated parameters. The working order is therefore chronological split, training-only preprocessing, training set decomposition, training set estimation, validation-based selection, and test set evaluation.

1. Output and demand: They represent the regional supply and demand situation of LNG and are the fundamental factors determining the market equilibrium price. When supply fluctuations or demand shocks occur, LNG prices will respond sensitively [39]. Domestic studies have also shown that seasonal peak-to-valley changes in demand in North China can significantly affect LNG ex-factory prices [40].
2. Inventory: Inventory levels are also one of the main factors influencing natural gas prices [41] and act as a "buffer" in the LNG market, reflecting the ability of enterprises to cope with demand fluctuations. Low LNG inventories may have a psychological effect on the market price of [42]. Inventory changes are considered a core explanatory variable in international gas hubs such as Henry Hub and NBP and are also a common input in LNG spot pricing models [43].
3. LNG terminal prices, as a reference, reflect the transmission pressure from downstream markets to upstream prices. LNG terminal prices are mainly calculated based on international crude oil prices and are timely. They are an important yardstick for measuring international LNG prices and a barometer of price fluctuations [44]. Due to certain cost-plus and regional arbitrage mechanisms in China's LNG industry, LNG terminal prices often serve as a leading indicator of LNG ex-factory prices.
4. Gas source price: The production costs of LNG plants include gas source costs, purification and liquefaction costs, transportation costs, etc. Among them, the upstream pipeline gas source costs account for the largest proportion and directly affect the pricing space at the factory end of LNG plants [45]. For China's LNG plants, fluctuations in upstream pipeline gas prices will be transmitted to the factory end in the short term [46].
5. Brent crude oil price: Crude oil prices play a significant role in the global natural gas and LNG pricing system. Although the degree of marketization of natural gas has increased in recent years, the "oil price linkage" coefficient remains widespread in LNG import contracts in Asia and China [47]. Therefore, Brent prices are considered an important exogenous variable in international research for explaining fluctuations in LNG prices [48].
6. Temperature: As a seasonal variable, it directly affects heating load and urban gas demand and is an indispensable factor in China's natural gas demand forecasting and price model [49]. A large number of related studies have shown that low temperatures can significantly boost natural gas consumption and cause short-term price fluctuations [50], especially in North China, where heating demand is dominant.

Table 2 constructs the structured domain knowledge system for large language models (LLMs), comprising six major categories of professional knowledge bases supporting the China LNG price prediction system, all of which are used during the large model knowledge learning session. The prediction bias of the basic machine learning model can be calibrated and corrected by leveraging capabilities in text information mining, pattern extraction, and risk interpretation. Various knowledge bases supplement unstructured text information that is difficult for quantitative models to capture from multiple dimensions such as supply and demand fundamentals, policy systems, historical pricing experience, geopolitical risks, sudden disturbances, and real-time market trade, making up for the shortcomings of pure numerical machine learning models in analyzing qualitative factors

such as policy interpretation, geopolitical public opinion, and event shocks. The knowledge features digested and refined by the large model will be input into the prediction framework to optimize the accuracy of component predictions, thereby enhancing the accuracy and explainability of predictions and providing enhanced intelligent support for the final LNG pricing decision. Details are shown in Table 2.

Table 2. Knowledge base of the LNG price forecast in China.

Knowledge Base	Instructions
Macroeconomic forecast analysis	Reflecting macroeconomic prosperity and energy demand expectations, it affects industrial and power gas consumption, thereby shaping medium- and long-term LNG price trends.
Policy and regulatory texts	Reflecting national rules on gas pricing, tariffs, reserve peak shaving, and environmental/energy transition, it affects LNG circulation costs and supply structure.
Historical pricing cases	Reflecting historical LNG pricing methods and strategies, it reveals past pricing laws and adjustment mechanisms, thereby informing judgments on current and future price formation trends.
Geopolitics/risk public opinion	It reflects political, diplomatic, and supply chain risks in major exporters, transit countries, and China's import sources, which affect domestic prices via international LNG prices and import expectations.
Major Emergencies	It reflects how unexpected events disrupt LNG production and logistics, causing short-term supply–demand imbalance and demand-driven price fluctuations.
Market and trade flow intelligence information	Reflecting spot and long-term contract prices, vessel arrivals, inventories, and regional arbitrage, it gauges supply tightness and serves as a direct price discovery source.

3.3. Evaluation Index System

This study uses four types of complementary indicators to comprehensively evaluate model performance in terms of error level, relative deviation, degree of dispersion, and goodness of fit:

1. Mean absolute error (MAE): Measures the absolute error between the predicted value and the true value, with a smaller value indicating a better match between the two.
2. Mean absolute percentage error (MAPE): It represents the proportion of the absolute error between the predicted value and the true value to the true value, eliminating the influence of dimensions and orders of magnitude on error visualization. The smaller the value is, the higher the accuracy.
3. Root mean square error (RMSE): expresses a regression effect, with a smaller value indicating a better effect.

4. Coefficient of determination (R^2): It indicates how close the predicted value is to the true value. The larger the value is, the closer the two are, and the better the prediction effect [51,52].

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|Y_t - P_t\right| \quad (1)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Y_t - P_t)^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum_t^N (Y_t - P_t)^2}{\sum_{t=1}^{N}(Y_t - \overline{Y_t})^2} \quad (4)$$

In the formula, Yt represents the true value at time t, Pt represents the predicted value at time t, and N represents the sample size used for prediction.

4. Model Construction and Interpretation

4.1. CEEMDAN–LSTM Model

The LNG ex-factory price sequence exhibits complex characteristics, including nonstationarity, nonlinearity, multiscale coupling, and the superposition of high-frequency noise and low-frequency trends, which are difficult for traditional models to fit effectively. For this purpose, this study selects adaptive noise complete set empirical mode decomposition (CEEMDAN) as the multiscale decomposition tool to construct the CEEMDAN=-LSTM model, which captures temporal dependencies based on long short-term memory networks and uses linear readout for components, taking into account both efficiency and long-term dependency learning ability. It can adaptively decompose the original sequence into several intrinsic mode functions (IMFs) and one residual, effectively separating high-frequency random perturbations, mid-frequency periodic fluctuations, and low-frequency trend terms. Compared with EMD and EEMD, CEEMDAN can significantly suppress modal aliasing, achieve higher decomposition accuracy, and avoid the need for human parameter tuning, offering full adaptability. First, input the original time series of the LNG factory price into the CEEMDAN model. After adaptive decomposition, obtain multiple IMF components from high to low frequency (such as IMF1, IMF2, IMFk) and the residual term (Res). The components exhibit more stable, regular fluctuations, reducing the difficulty of subsequent modelling. The final prediction results are derived from the predicted values of each component using algebraic summation and reconstruction, following the decomposition–reconstruction principle of "component independent prediction, global summation restoration" to ensure the physical integrity of the original sequence.

The main model settings are described as follows. Every model in the comparison uses the chronological split and the preprocessing order stated in Section 3.2, so the differences in Table 3 come from the modelling choice rather than from a different sample. CEEMDAN is applied to the preprocessed training price series and returns several intrinsic mode functions, ordered from high frequency to low frequency, plus one residual. Each component is forecast on its own, and the component forecasts are added to restore the price level. The five base learners are LSTM, the BP neural network, XGBoost, random forest, and ELM. For the LSTM branch, information is updated with the gated recurrence in Equation (5): the forget gate and the input gate revise the cell state, a candidate state is written into the cell, and the output gate releases the hidden state. Each base learner is estimated on the training components only. Hyperparameters and fusion weights are selected using the validation set. The test set is used once for the metrics in Table 3. The hybrids CEEMDAN–LSTM,

CEEMDAN–BP, CEEMDAN–XGBoost, CEEMDAN–RF, and CEEMDAN–ELM differ from the corresponding single learners only by this decompose–predict–sum step. Each MiniMax variant adds the calibration layer in Section 4.2 on top of the corresponding hybrid and does not re-estimate the numerical component models on the test set.

Table 3. Comparison of the predictive performance of different CEEMDAN hybrid models.

Model	MAE	MAPE	RMSE	R^2
CEEMDAN–LSTM–MiniMax	29.424	0.744	38.557	0.991
CEEMDAN–LSTM	24.104	0.646	46.509	0.987
CEEMDAN–BP–MiniMax	25.397	0.673	47.091	0.986
CEEMDAN–XGBoost–MiniMax	26.213	0.692	47.318	0.986
CEEMDAN–RF–MiniMax	26.220	0.692	47.335	0.986
CEEMDAN–ELM–MiniMax	25.444	0.675	47.694	0.986
RF	30.861	0.806	54.799	0.982
XGBoost	40.424	1.050	64.741	0.974
CEEMDAN–ELM	55.152	1.435	70.597	0.969
CEEMDAN–BP	62.026	1.646	77.158	0.963
LSTM	61.353	1.573	90.408	0.950
CEEMDAN–RF	76.515	2.017	93.219	0.947
CEEMDAN–XGBoost	80.605	2.125	97.386	0.942
ELM	227.118	6.017	324.276	0.355
BP	312.345	8.377	372.129	0.151

A Long short-term memory (LSTM) network is a specialized recurrent neural network (RNN) proposed by Hochreiter and Schmidhuber in 1997, designed to address the Gradient vanishing or explosion problem encountered by traditional RNNs when processing long sequences. The core of LSTM lies in its sophisticated "gate mechanism" that introduces a cell state as a "thread" running through the entire sequence and utilizes three gate units—input gate, output gate, and forget gate—to precisely regulate information flow [53].

$$C_i = f_i \bigodot C_{i-1} + i_t \bigodot \widetilde{C}_t \tag{5}$$

Here, C_i and C_{i-1} are the current and previous cell states, respectively; f_i and i_t represent the forget and input gate activations, respectively; and $\widetilde{C}_t$ denotes the candidate cell state. The operator $\odot$ indicates element-wise multiplication. This mechanism enables selective memory retention and long-term dependency modelling. LSTM effectively captures long-term temporal dependencies, seasonal patterns, and nonlinearities in energy price and load forecasting.

4.2. CEEMDAN–LSTM–MiniMax Combined Model

This paper uses the open-source large model MiniMax, which, together with the decomposition and reconstruction model, forms the CEEMDAN–LSTM–MiniMax combined model. Relying on the powerful long time series data learning ability and multi-source information fusion advantage of the MiniMax large model, an information database of historical pricing cases, historical price policies, major market events, macroeconomic indicators, etc. is used to construct the LNG ex-factory price knowledge base, fully exploring the correlation between the historical price fluctuation patterns of LNG and multiple factors such as policy regulation impact and market supply and demand. Combined with the constructed, sequenced decomposition, and residual optimization strategy (CEEMDAN–LSTM), a combined prediction model that achieves high-precision price prediction and provides an explanation of driving factors can effectively enhance the practicability and scalability of the model in the LNG ex-factory price scenario. In this hybrid model, CEEMDAN–LSTM is responsible for precise numerical predictions, while the MiniMax large model equips

this precise "engine" with an intelligent "brain". Not only does it inject semantic understanding and logical reasoning into numerical predictions, but it also fuses information from multiple sources using multimodal capabilities, ultimately transforming numbers into decision recommendations with insights and evidence.

Different from conventional fixed fusion strategies, the MiniMax model adopts a dynamic cross-attention weighted calibration fusion framework to couple multi-source heterogeneous semantic information and numerical forecast outputs. The model avoids naive vector concatenation and static rule-based weighting and realizes adaptive dynamic fusion according to input data characteristics.

First, we obtain the numerical prediction feature vector $F_n \in R^d$ from the numerical forecasting module and extract the text semantic reasoning feature $F_s \in R^d$ from multi-source heterogeneous texts after encoding and semantic understanding. The two vectors are placed in the same dimension by the linear maps.

$$F_n = W_n h_n + b_n \tag{6}$$

$$F_s = W_s h_s + b_s \tag{7}$$

Here, h_n is the embedding of the CEEMDAN–LSTM forecast, and h_s is the MiniMax-01 encoding of the knowledge base in Table 2.

Subsequently, a cross-attention mechanism is introduced to calculate the interactive correlation between semantic features and numerical features and generate adaptive weight matrices W_s and W_n for the two feature branches as follows:

$$\text{Attention}(F_s, F_n) = \text{Softmax}\left(\frac{F_s \cdot F_n^T}{\sqrt{d}}\right) \tag{8}$$

The argument of softmax is the vector a. Define $Q_s = W_Q F_s$, $K_n = W_K F_n$, and $Q_n = W'_Q F_n$, $K_s = W'_K F_s$, where $F_s, F_n \in R^d$ and W_Q, W_K, W'_Q, W'_K are $d \times d$ matrices.

Set

$$a = \left[\frac{Q_s^T K_n}{\sqrt{d}}, \frac{Q_n^T K_s}{\sqrt{d}}\right]^T \tag{9}$$

Then

$$[\alpha_s, \alpha_n]^T = Softmax(a) = \frac{\exp(a)}{1^T \exp(a)} \tag{10}$$

with

$$\alpha_s \geq 0, \alpha_n \geq 0, \alpha_s + \alpha_n = 1 \tag{11}$$

The product of the two feature vectors inside the display above is a scalar. The softmax of one scalar is identically 1, so the score has to be the two-dimensional vector a.

$$W_s = \sigma(\text{Attention}(F_s, F_n) \cdot F_s), W_n = \sigma(\text{Attention}(F_n, F_s) \cdot F_n) \tag{12}$$

Here, σ denotes the Sigmoid activation function for weight normalization. Both branches use the same product. Let $V_s = W_V F_s$ and $V_n = W_V F_n$, and set $H_s = \alpha_s V_s$ and $H_n = \alpha_n V_n$. The adaptive weights are $W_s = \sigma(U_s H_s + c_s)$ and $W_n = \sigma(U_n H_n + c_n)$, where σ is applied element by element, U_s and U_n are $d \times d$ matrices, and $c_s, c_n \in R^d$. The product UH stands on both sides.

These are the weights in the fusion equation retained below:

$$F_{fusion} = W_s \odot F_s + W_n \odot F_n \tag{13}$$

Finally, the fused feature is input into the calibration regression layer to dynamically refine and calibrate the initial numerical forecast results and output the optimized final prediction value.

The calibration layer returns the adjusted price as follows:

$$\hat{y} = w_c^T F_{\text{fusion}} + b_c \tag{14}$$

where $\odot$ represents element-wise multiplication.

The core mechanisms of MiniMax include a hierarchical hybrid attention architecture, a hybrid expert (MoE) and parameter activation mechanism, and efficient reinforcement learning. The hierarchical hybrid attention architecture uses lightning attention in the first seven out of every eight layers of the model's 80 attention layers, while the eighth layer retains the traditional softmax attention. This 7:1 hybrid structure ultimately achieves near-linear complexity, solving the efficiency problem of processing very long sequences while retaining global modelling capabilities at critical levels. The hybrid expert (MoE) and parametric activation mechanism expert architecture form the core skeleton of model capacity scaling. To maximize model capacity with limited computing resources, MiniMax-01 uses the Mixture of Experts (MoE) architecture. The architecture consists of 32 expert modules, with each representing an independent feedforward neural network that dynamically selects the optimal combination of experts based on the semantic features of the input token using a gating network. This design enables the model to maintain a large total parameter capacity while activating only a small fraction of them per inference. Efficient reinforcement learning is the key engine for improving training quality. The CISPO algorithm's design concept of clipping importance sampling weights both accelerates convergence and significantly reduces training costs. The autonomous evolution mechanism further endows the model with the ability to self-optimize, setting up a ladder to the future of continuous evolution for achieving general artificial intelligence. The architecture of the model in this paper is presented in Figure 2.

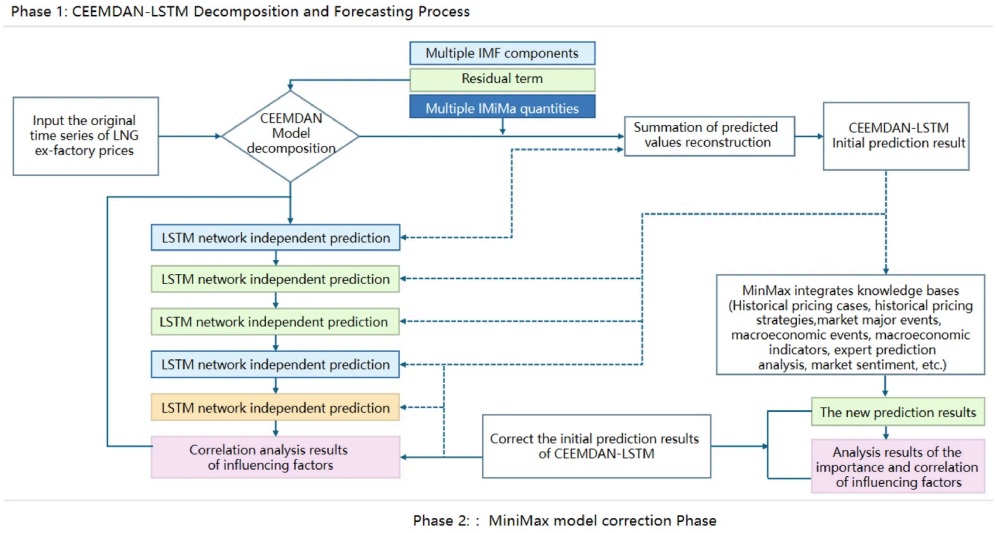


Figure 2. Architecture of the model in this paper.

The large language model used in this study is MiniMax-01. As described above, MiniMax-01 has 80 attention layers. In every group of eight layers, seven use lightning attention, and the eighth retains softmax attention. The backbone is a mixture of 32 experts. In the forecasting pipeline, its role is to calibrate the numerical forecast, not to replace the numerical model. CEEMDAN–LSTM, or the corresponding CEEMDAN–machine learning

hybrid, first produces the initial price forecast. MiniMax-01 then reads the knowledge base in Table 2, covering macroeconomic analysis, policy and regulatory texts, historical pricing cases, geopolitical and risk commentary, major emergencies, and market and trade flow information and encodes those texts as a semantic feature. The semantic feature and the numerical forecast feature are projected to a common dimension and combined using the cross-attention weights in Equations (6) and (7). The fused feature enters the calibration regression layer, which outputs the adjusted price. This adjustment is a learned mapping from the numerical forecast and the encoded text to a revised number, and it runs automatically. No analyst edits individual forecast points before the metrics in Table 3 are calculated. Human judgment stays outside that numerical loop. At the decision application stage, the pricing decision group may use the calibrated forecast together with the same knowledge base for a multi-criteria review and for a draft pricing scheme. That review does not change the test set errors reported in this paper.

5. Results and Discussion

5.1. Analysis of the Importance of Influencing Factors

An importance assessment of 15 predictive algorithms (including single models, CEEMDAN combined models, and variants combining MiniMax large models) for the seven influencing factors of the natural gas market was performed. The import price of LNG spot and gas source price have always been the two most dominant factors. In the single RF/XGBoost model, the importance of the gas source price is as high as 0.66. In the CEEMDAN–XGBoost–MiniMax and other models combined with MiniMax, the importance of the LNG spot import price can reach 0.40. It is particularly worth noting that the Brent oil price shows stable and non-negligible importance in all CEEMDAN combined models, such as CEEMDAN–ELM, CEEMDAN–BP, CEEMDAN–RF, and CEEMDAN–XGBoost, reaching 0.19–0.22 in the single ELM model. This is significantly higher than production, demand, and temperature (in most models lower than 0.12), indicating that the price linkage and substitution effect between crude oil and natural gas are clearly manifested after signal decomposition. In contrast, the single RF, XGBoost, LSTM, and BP models have lowered the relative contribution of Brent oil price due to the strong dominance of gas source price or spot price (0.03–0.05), which precisely reflects that CEEMDAN preprocessing can effectively remove strong interference and reveal the stable influence of Brent oil price as a medium- to long-term reference factor. The temperature factor only has a slightly stronger effect in a few models, such as BP (0.142), while the overall importance of production and demand is relatively low. The inventory level is generally low except for LSTM (0.401), which is extremely sensitive. Moreover, introducing the MiniMax large model significantly amplifies the importance of high-level features, such as the LNG spot import price. At the same time, CEEMDAN makes the distribution of each factor more balanced. Therefore, in actual prediction modelling, in addition to prioritizing the inclusion of international gas prices and gas source costs, the Brent oil price should always be retained as an important input feature, especially when using signal decomposition technology, as its indirect transmission effect on natural gas prices or loads cannot be ignored. Details are shown in Figure 3.

A comprehensive importance assessment of the influencing factors on LNG prices, based on 15 machine learning algorithms, shows that the driving effect of market competition factors is significantly stronger than that of supply and demand and fundamental factors. Among them, the price of imported LNG receiving stations ranks first with an average importance score of 0.306. Gas source prices (0.227) and Brent oil prices (0.154) follow closely. The cumulative contribution of these three factors exceeds 68.7%, indicating that the traditional pricing benchmark for the international LNG spot market, linked to oil prices, remains the primary driver of LNG price fluctuations. In contrast, the influence

weights for supply and demand and for seasonal factors (such as inventory levels, temperature, demand, and production) are significantly lower. This reflects that, in the current context of global trade and financialized pricing, the formation mechanism of LNG prices has shifted from being dominated by traditional supply-and-demand balance to being driven by international energy market linkage and financialized pricing. The marginal influence of fundamental factors has been significantly weakened. This conclusion provides an important basis for prioritizing the inclusion of market price variables in the subsequent construction of a high-precision LNG price prediction model. Details are shown in Figure 4.

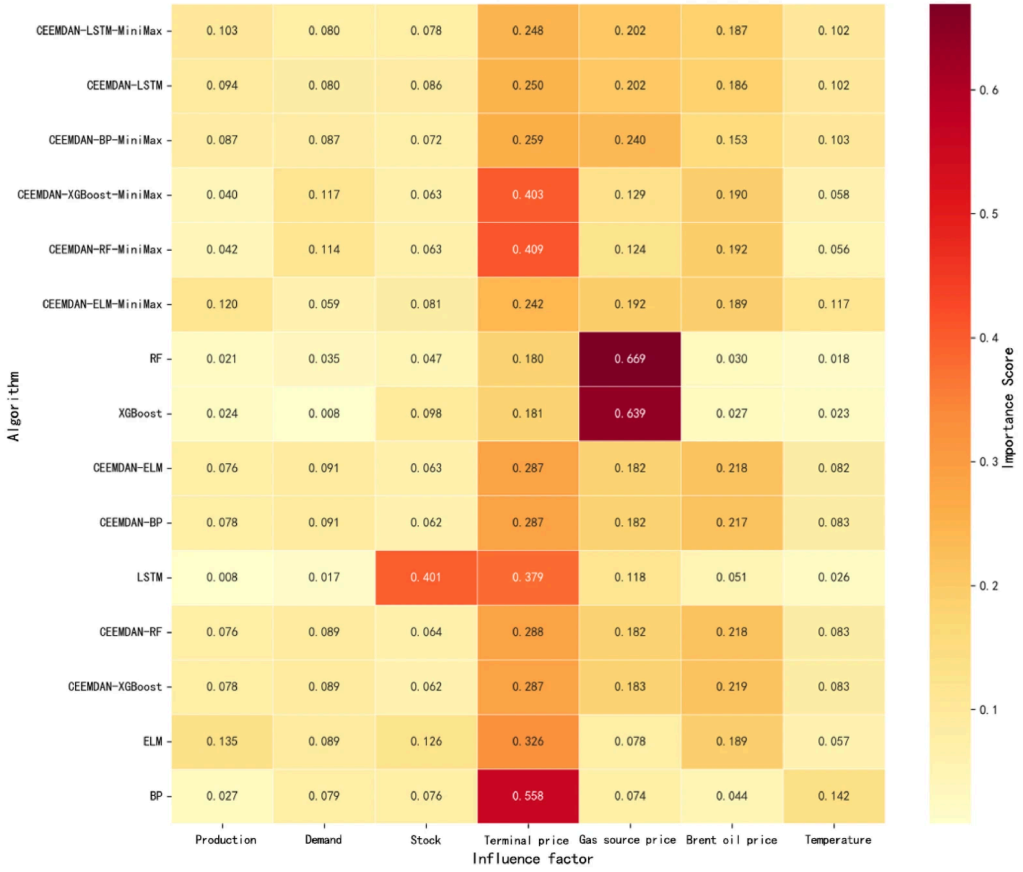


Figure 3. Heat map of seven types of impact factor importance analysis.

The scores in this subsection measure how much each input contributes to the predictions of a given algorithm. They describe predictive association inside the fitted models. They do not identify a structural cause, and they do not by themselves show that a variable drives the ex-factory price or that the pricing mechanism has moved from supply–demand balance to financialized pricing. A high score for the landed price of spot-import LNG, the gas-source price, or the Brent price means that, under the present feature set and sample, those series account for a large share of the models' fitted variation. Production, demand, inventory, and temperature can still matter in the market while receiving a lower score if their variation is smaller, if they move together with the price variables, or if they change more slowly. The ranking also changes with the learner: inventory is relatively important for LSTM and much less so for several tree models. These results are used to judge which series are worth keeping as predictors and to compare how different algorithms use the same inputs. A causal statement would require a design that separates external shocks from feedback between prices and quantities, which lies outside this forecasting study.

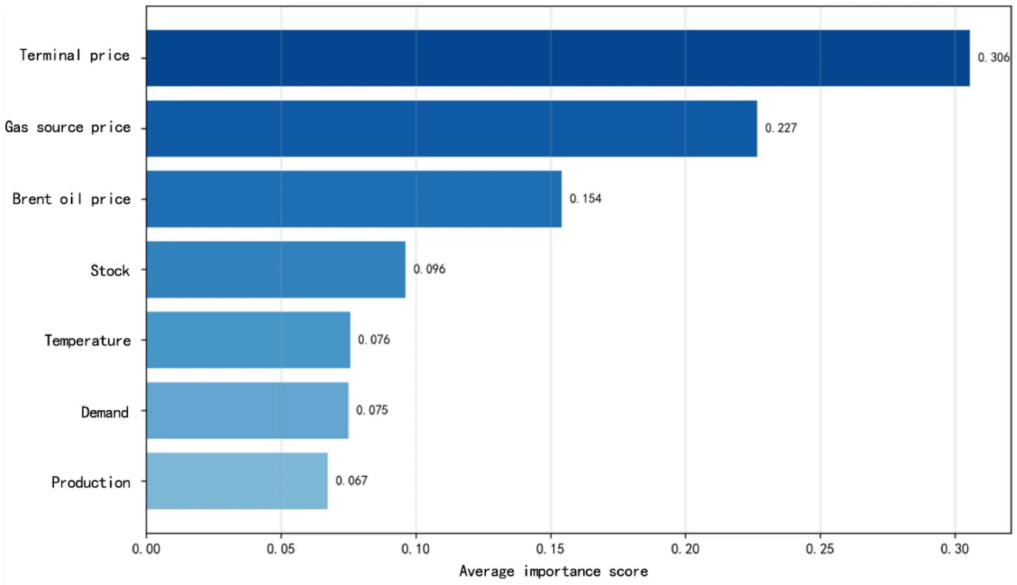


Figure 4. Comparison of the importance of seven types of influencing factors.

5.2. Correlation Analysis of Influencing Factors

The heat map analysis based on the correlation between the influencing factors of LNG factory prices and the prediction results using 15 machine learning algorithms reveals that the correlation patterns between different influencing factors and the price series exhibit significant heterogeneity: market and supply–demand factors such as the spot import LNG landed price, gas source price, Brent oil price, and temperature show strong positive correlations in most algorithms. Among them, the spot import LNG landed price shows a correlation coefficient of over 0.942 with LSTM and XGBoost models, indicating the dominant transmission effect of the international LNG spot market pricing mechanism on domestic factory prices. In contrast, fundamental factors such as production, demand, and inventory levels show negative or weak correlations in most algorithms, especially for inventory levels, which has a correlation coefficient as low as −0.838, indicating that inventory levels have an inverse regulatory effect on price fluctuations. At the same time, there are significant differences in the algorithms' ability to capture correlations. Traditional machine learning models (such as XGBoost and RF) tend to identify correlations more consistently, particularly between key variables, such as gas source price and spot import LNG landed price. In contrast, some decomposition–integration models are more robust in handling nonlinear correlations. This result not only verifies the core position of market price factors in LNG pricing but also reveals the limitations of a single model in correlation analysis, providing important empirical evidence for the subsequent construction of a multimodel fusion prediction framework. Details are shown in Figure 5.

The coefficients in this subsection are associations between each factor and the price series, as seen across the fifteen algorithms. A coefficient above 0.942 for the spot import LNG landed price in the LSTM and XGBoost views or a coefficient near −0.838 for inventory in some algorithms indicates a strong linear association in that fitted view. It does not establish that the landed price transmits international prices into the ex-factory price, and it does not establish that inventory exerts an inverse regulatory effect. Correlation does not identify direction, omitted variables, or reverse influence from price back to inventory or production. Differences across algorithms show that the measured association also depends on the model. Figure 5 therefore provides evidence about predictive relationships. It is not a test of a pricing mechanism. Its practical use is to check whether variables with high

importance also move with the price series and to avoid treating one model's correlation pattern as a market mechanism.

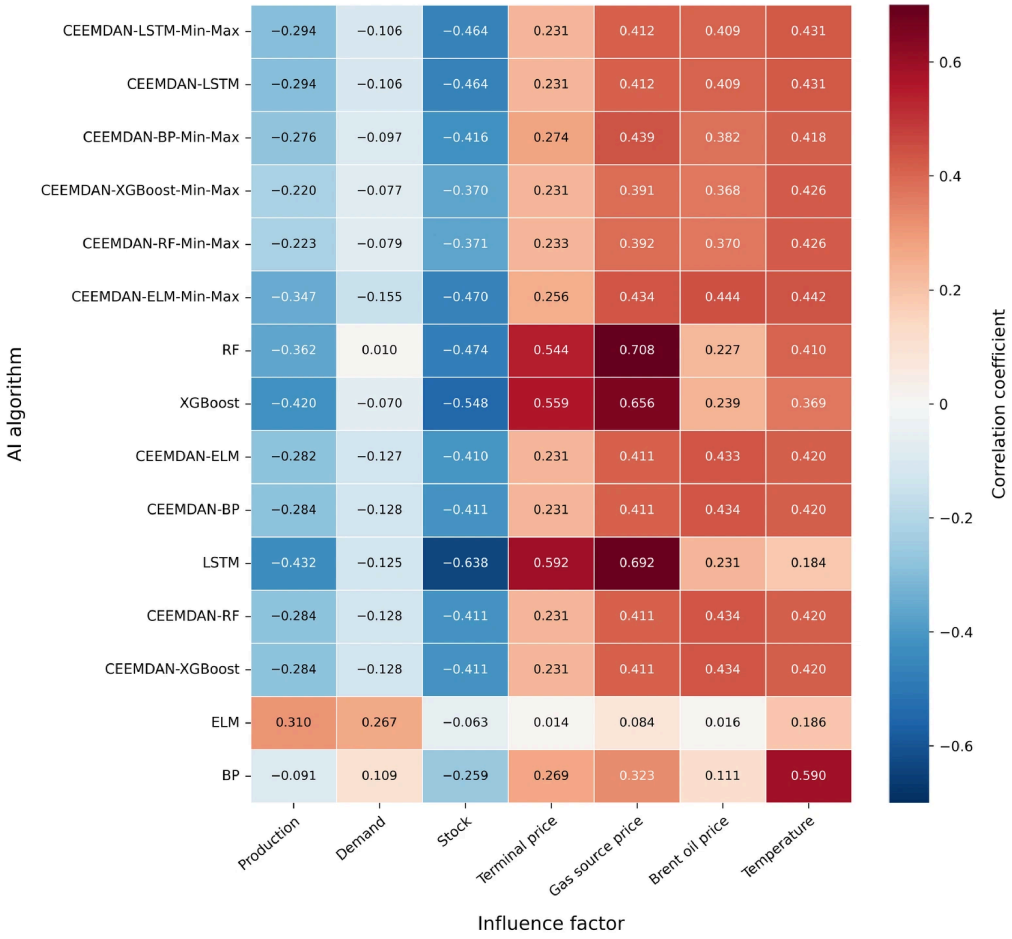


Figure 5. Heat map of the correlation between fifteen model influencing factors and predicted prices.

5.3. Comparison of Predictive Performance of Different Models

The CEEMDAN–LSTM–MiniMax model achieves the best comprehensive performance, with an MAE of 29.424 CNY/ton, an MAPE of 0.744%, a RMSE of 38.557 CNY/ton, and an R^2 of 0.991. This model's explanatory ability for LNG factory price fluctuations reaches 99.1%. Compared to the suboptimal model (CEEMDAN–LSTM), the RMSE is reduced by 17.1%, and R^2 increases by 0.4 percentage points. It demonstrates significant advantages in full-cycle price trend fitting and extreme turning point prediction. Compared to the enhanced decomposition of a single model, the CEEMDAN decomposition brings a significant improvement in accuracy, with MAE dropping to approximately 24–29 intervals and R^2 rising to 0.986.

Using a full-dimensional quantitative comparison of 15 models, this study confirmed the combined "CEEMDAN time series decomposition + MiniMax large model + time series-specific machine learning algorithm" framework as the optimal technical approach for LNG factory price prediction in China. The CEEMDAN–LSTM–MiniMax model serves as a high-precision benchmark for domestic LNG price prediction. The study clarified the performance hierarchy of LNG price prediction models: CEEMDAN + MiniMax combined model > traditional CEEMDAN combined model > single machine learning model. This pattern provides empirical evidence from the Chinese market for model selection in global natural gas price prediction. For energy price sequences with strong time series dependence and high volatility, such as LNG, the predictive capabilities of time series-specific

algorithms, such as LSTMs, are significantly superior to those of shallow neural networks and tree models, thereby verifying that "time series feature adaptation" is the core principle of natural gas price prediction model design. Details are shown in Table 3 and Figure 6 and Figure A1.

The four metrics do not rank the models in the same order, and no single model is best on every measure. CEEMDAN–LSTM–MiniMax records the lowest RMSE (38.557 CNY/ton) and the highest R^2 (0.991). Relative to CEEMDAN–LSTM, its RMSE is lower by 7.952 CNY/ton, about 17.1% of 46.509, and its R^2 is higher by 0.004. These two metrics indicate a closer fit in squared error terms and a slightly larger share of variance explained, which is consistent with smaller large deviations. CEEMDAN–LSTM, however, records the lower MAE (24.104 versus 29.424 CNY/ton) and the lower MAPE (0.646% versus 0.744%). The MiniMax calibration therefore improves the squared error and goodness-of-fit criteria while increasing the average absolute and percentage errors. One consistent reading is that the calibration reduces some large residuals and can at the same time shift a larger number of ordinary points. Where large misses are the main concern, CEEMDAN–LSTM–MiniMax is the stronger choice on this test set. Where typical absolute or percentage error is the main concern, CEEMDAN–LSTM is preferable. The other MiniMax hybrids, including CEEMDAN–BP–MiniMax, CEEMDAN–XGBoost–MiniMax, CEEMDAN–RF–MiniMax, and CEEMDAN–ELM–MiniMax, have RMSE values from 47.091 to 47.694 and an R^2 of 0.986. Each of them has a lower RMSE than the matching hybrid without MiniMax, and none of them matches CEEMDAN–LSTM in MAE or MAPE. Table 3 therefore supports a ranking that depends on the chosen indicator. It does not support a claim that one framework dominates all four measures.

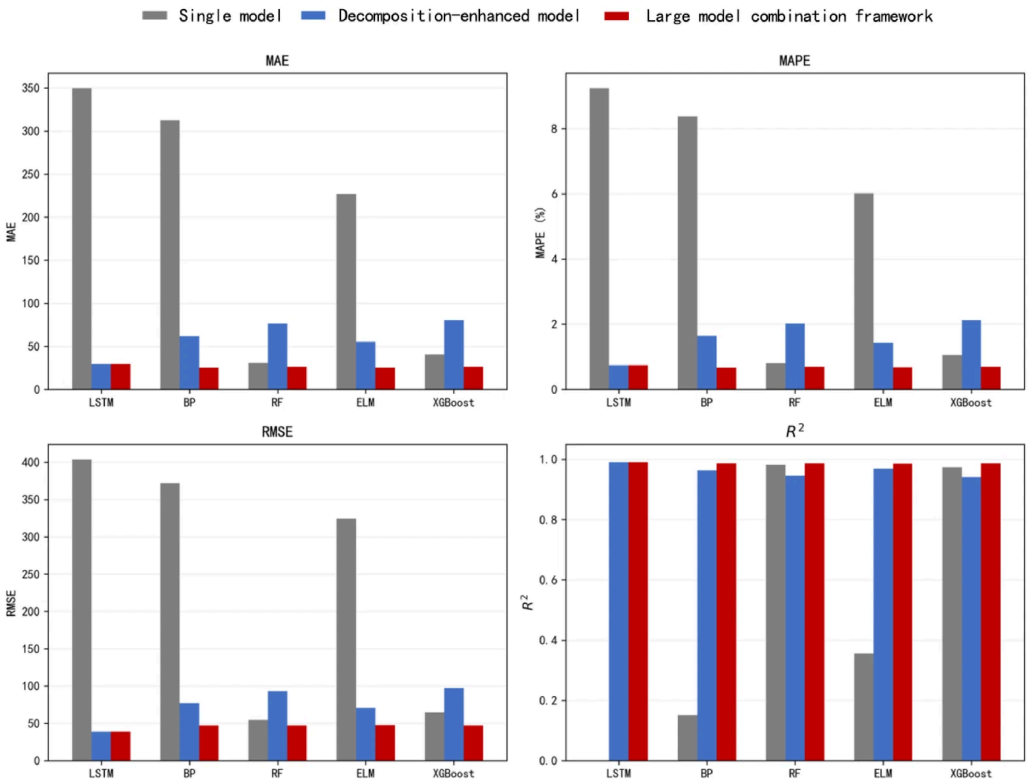


Figure 6. Bar charts of the prediction performance indicators of the three types of models.

5.4. LNG Ex-Factory Price CEEMDAN–LSTM–MiniMax Prediction Results

By comparing full-cycle LNG price prediction results from 15 models, the CEEMDAN–LSTM–MiniMax model achieved the best overall performance. This confirmed the superi-

ority of the hybrid prediction model driven by large language models in LNG ex-factory price prediction. Its prediction accuracy was significantly higher than that of the traditional CEEMDAN decomposition combination model and the single machine learning model. The MiniMax large model enhancement model effectively combined feature extraction, resulting in a significant improvement in overall fitting accuracy. Its prediction curve was the closest to the actual ex-factory price trend throughout all test periods. For example, at multiple price turning points from June 2024 to December 2025, the prediction direction and amplitude of this model were superior to those of other models. In the high-frequency fluctuation stage at the beginning of 2026, its deviation was the smallest, and there was no obvious divergence or phase error. The single model showed significant lag or deviation in local periods and a weak response to severe price fluctuations. The CEEMDAN decomposition model improved on the single model and better captured the true trend. However, there were still systematic deviations at some peaks. This demonstrated that, in scenarios with highly volatile energy commodity prices, introducing a hybrid modelling framework with large language models can effectively address the limitations of traditional models in fitting nonlinear and nonstationary time series data, providing empirical evidence from the Chinese market for model innovation in the global natural gas price prediction field. Details are shown in Figures 7–12.

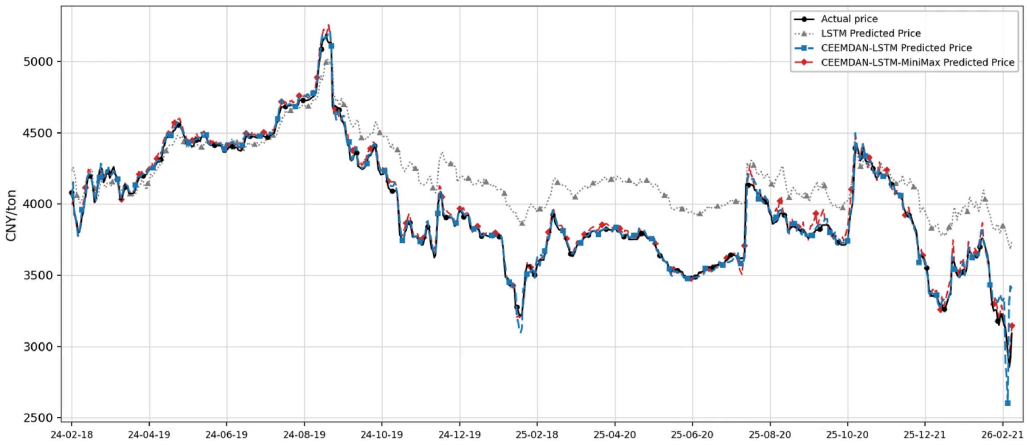


Figure 7. Prediction of LNG prices based on the CEEMDAN–LSTM–MiniMax model.

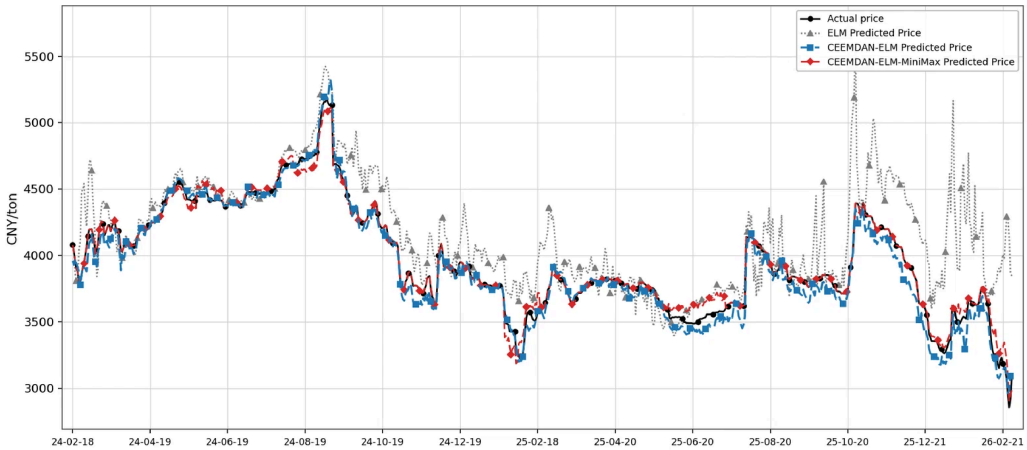


Figure 8. Prediction of LNG prices based on the CEEMDAN–ELM–MiniMax model.

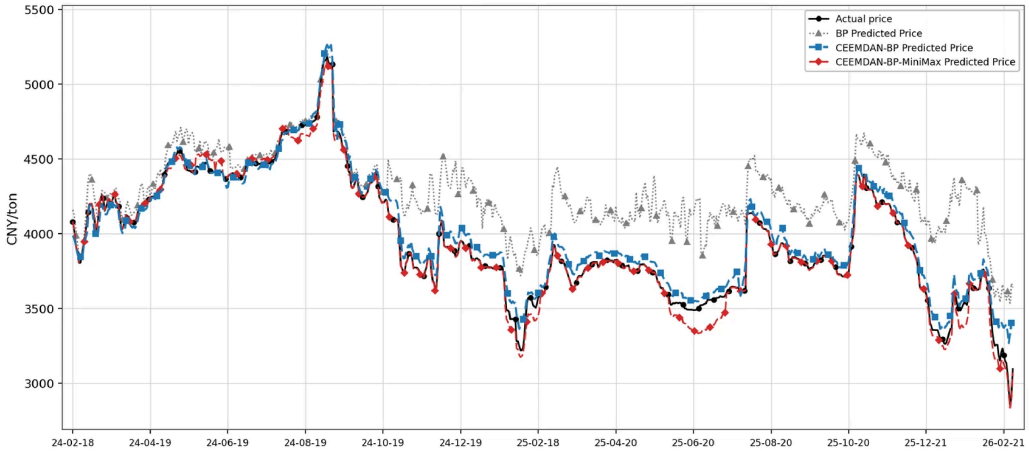


Figure 9. Prediction of LNG prices based on the CEEMDAN–BP–MiniMax model.

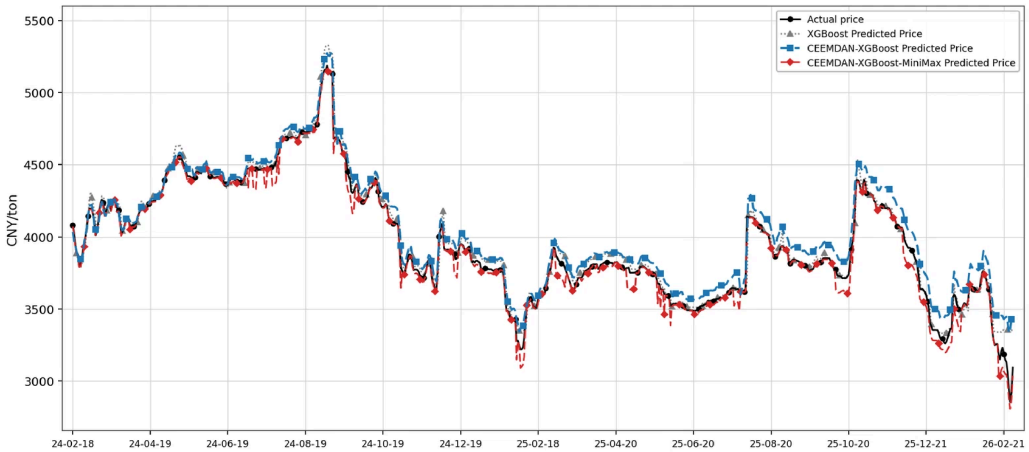


Figure 10. Prediction of LNG prices based on the CEEMDAN–XGBoost–MiniMax model.

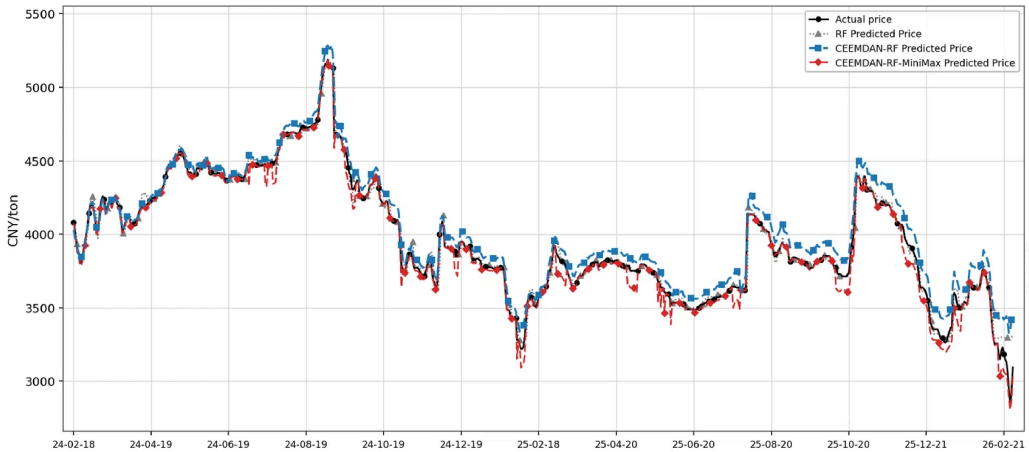


Figure 11. Prediction of LNG prices based on the CEEMDAN–RF–MiniMax model.

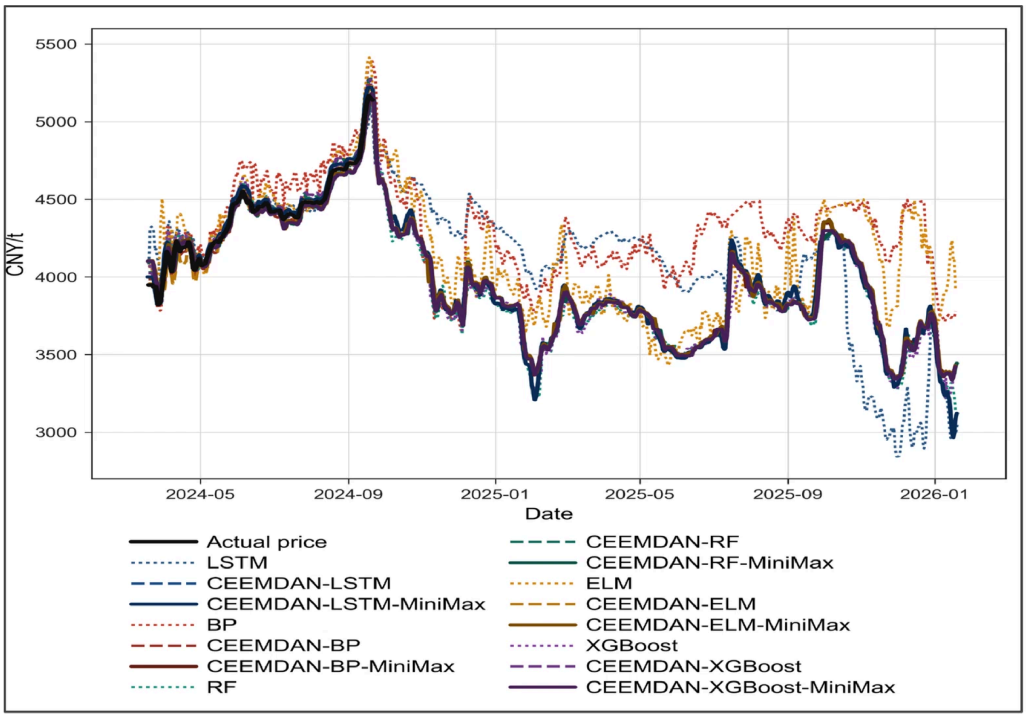


Figure 12. The prediction performance of fifteen models.

This paper uses the language model in a narrower role: MiniMax-01 does not replace the price model and does not produce the forecast from text alone. It calibrates a CEEMDAN–LSTM number that has already been formed. On this test set the calibration improves RMSE and R^2 and does not improve MAE or MAPE. The specific addition relative to the cited forecasting literature is therefore a CEEMDAN–LSTM numerical forecast of China's LNG ex-factory price, an automatic MiniMax-01 calibration layer, and an indicator-wise comparison rather than a single-metric ranking.

6. Conclusions

Considering the strong nonstationarity, multiscale fluctuations and complex coupling of external information within China's LNG production price sequence, this study develops a hybrid prediction framework combining signal decomposition, multimodel ensemble and LLM semantic reasoning. The empirical results demonstrate that the proposed CEEMDAN–LSTM–MiniMax hybrid model achieves the best prediction performance over all benchmark models on the test set, yielding MAE of 29.424, MAPE of 0.744%, RMSE of 38.557, and R^2 of 0.991. The introduction of MiniMax semantic calibration significantly improves prediction robustness during extreme market fluctuation periods, outperforming both standalone LSTM and CEEMDAN–LSTM variants without LLM correction.

The core innovation of this work lies in constructing a closed-loop "numerical prediction + semantic decision-making" paradigm that organically couples structured time series modelling and unstructured semantic reasoning. This framework mitigates traditional pricing models' over-reliance on historical numerical data and strengthens the model's ability to capture qualitative market drivers including emergency policy adjustments and market sentiment shifts. From an industrial perspective, the interpretable intelligent pricing outputs generated by our framework can provide quantitative decision support for LNG market participants: it helps optimize gas source procurement timing, supports dynamic inventory management, and facilitates rational capacity allocation to hedge against price volatility risks, thereby improving overall energy supply security.

Nevertheless, this study still has notable limitations. The incorporation of large language model semantic reasoning introduces non-negligible computational overhead, which poses challenges for low-latency real-time online inference in production environments. Future work will focus on model distillation and lightweight optimization of the LLM branch to reduce inference cost, while integrating more granular supply-side operational data to further enhance the reliability of semantic calibration. The proposed framework provides a robust and reproducible technical route for LNG price forecasting and offers a new pathway for integrating LLMs into energy time series prediction tasks. The framework is estimated for China's LNG ex-factory price from January 2016 to February 2026, using the seven structured inputs in Table 1 and the knowledge base in Table 2. The reported accuracy should be read inside that scope. Transfer to another hub, another sampling frequency, or a period governed by a new pricing rule is not established by the present test set. Four further limits qualify the findings. First, as clarified for Table 3, CEEMDAN–LSTM–MiniMax is preferable when RMSE and R^2 are the criteria, whereas CEEMDAN–LSTM is preferable when MAE and MAPE are the criteria. The hybrid is not uniformly superior. Second, the feature importance scores and the correlation coefficients describe predictive relationships inside the fitted models, and they are not evidence that a variable causes the ex-factory price. Third, the MiniMax-01 branch adds inference cost, which limits low-latency deployment. The calibration that produces the numbers in Table 3 is automatic, while a human pricing group remains necessary when those numbers are turned into a pricing decision. Fourth, the knowledge base texts and the 2024–2026 test window may not represent a later regime shift, so performance can deteriorate outside this sample. Useful next steps are to report prediction intervals as well as point metrics, to repeat the same pipeline on a later hold-out period or on another regional price, to compress the MiniMax-01 branch so that calibration remains usable under a tighter computing budget, and to add higher-frequency supply and logistics series only where they can be aligned without leakage. These extensions would show more clearly where the present ranking of indicators still holds.

Author Contributions: Conceptualization, J.L.; Methodology, Formal analysis, Writing—original draft, Review & editing, Project administration, J.W.; Validation, Software, X.L.; Supervision and Resources, Writing—review & editing, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China grant number No. 52274027.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A.

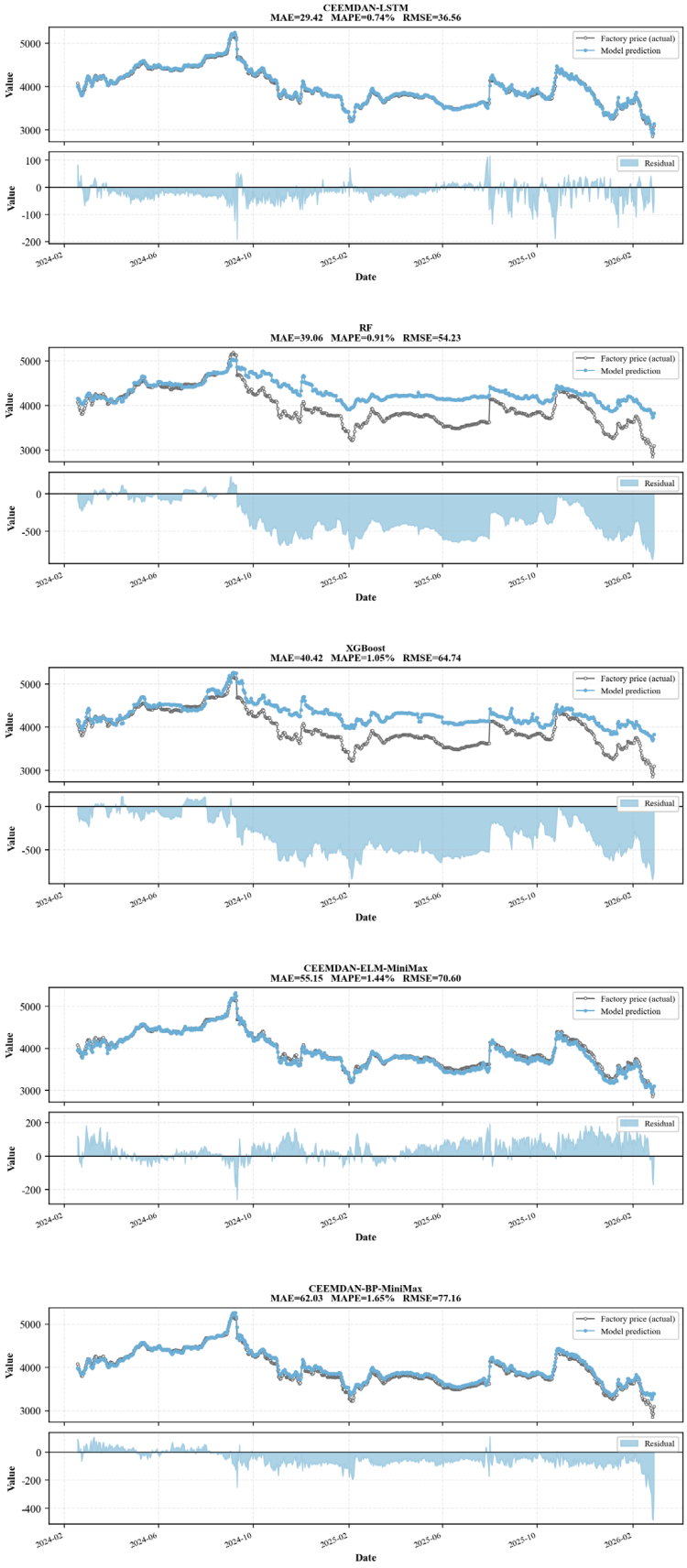


Figure A1. *Cont.*

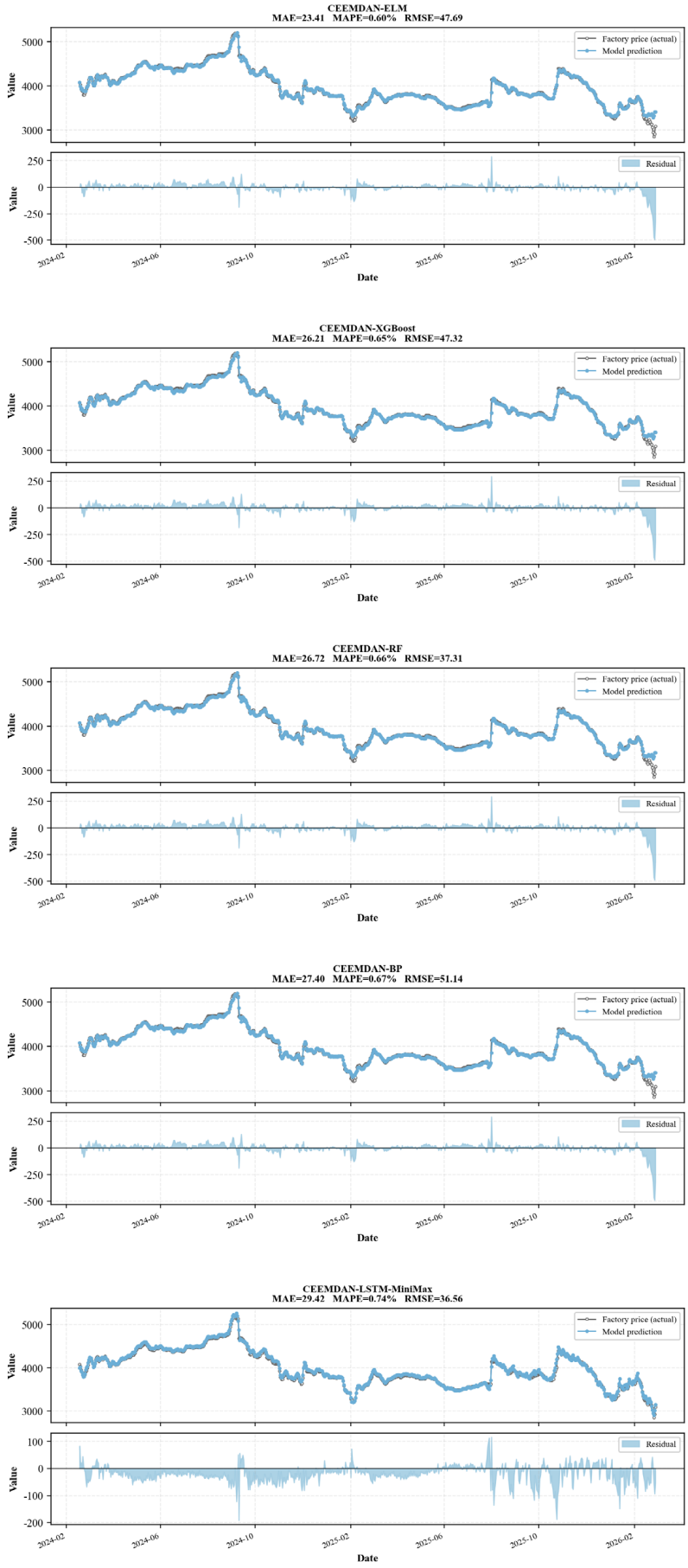


Figure A1. *Cont.*

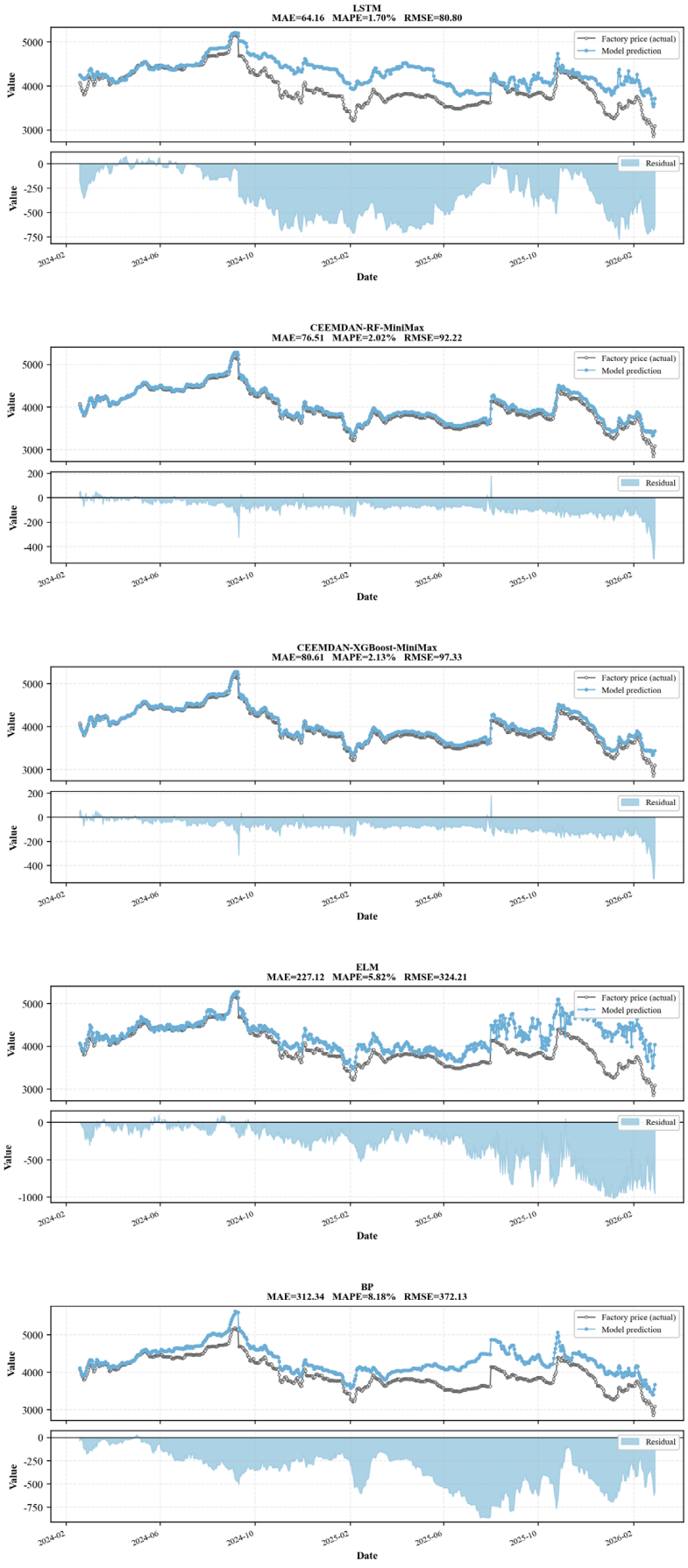


Figure A1. Prediction performance of fifteen models. Comparison of the prediction performance of the 15 models.

References

1. Jin, J.; Kim, J. Forecasting Natural GasPrices Using Wavelets, Time Series, and Artificial Neural Networks. *PLoS ONE* **2015**, *10*, e0142064. [CrossRef]
2. Bilgili, M.; Pinar, E. Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye. *Energy* **2023**, *284*, 128575. [CrossRef]
3. Salehnia, N.; Falahi, M.A.; Seifi, A.; Adeli, M.H.M. Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis. *J. Nat. Gas Sci. Eng.* **2013**, *14*, 238–249. [CrossRef]
4. Werbos, P.J. Generalization of backpropagation with application to a recurrent gas market model. *Neural Netw.* **1988**, *1*, 339356. Available online: https://www.sciencedirect.com/science/article/abs/pii/089360808890007X?via%3Dihub (accessed on 1 March 2026). [CrossRef]
5. Herrera, G.P.; Constantino, M.; Tabak, B.M.; Pistori, H.; Su, J.J.; Naranpanawa, A. Long-term forecast of energy commodities price using machine learning. *Energy* **2019**, *179*, 214–221. [CrossRef]
6. Nick, S.; Thoenes, S. What drives natural gas prices?—A structural VAR approach. *Energy Econ.* **2014**, *45*, 517–527. [CrossRef]
7. Hailemariam, A.; Smyth, R. What drives volatility in natural gas prices? *Energy Econ.* **2019**, *80*, 731–742. [CrossRef]
8. Van Goor, H.; Scholtens, B. Modeling natural gas price volatility: The case of the UK gas market. *Energy* **2014**, *72*, 126–134. [CrossRef]
9. Ioannis, E.L.; Emmanuel, P.; Niki, K.; Stavros, S. An Advanced Deep Learning Model for Short-Term Forecasting U.S. Natural Gas Price and Movement. In *Artificial Intelligence Applications and Innovations. AIAI 2020 IFIP WG 12.5 International Workshops*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 585, pp. 165–176. [CrossRef]
10. Mouchtaris, D.; Sofianos, E.; Gogas, P.; Papadimitriou, T. Forecasting Natural Gas Spot Prices with Machine Learning. *Energies* **2021**, *14*, 5782. [CrossRef]
11. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
12. Li, M.; Kong, Y. Prediction of Natural Gas Price Based on Time Series Model. *Pet. New Energy* **2023**, *35*, 61–66. [CrossRef]
13. Huang, X.; Zhao, M.; Huan, J.; Li, P.; Zhang, S. Construction of Natural Gas Price Prediction Model Based on EEMD-GRU Neural Network. *Microcomput. Appl.* **2024**, *40*, 13–17.
14. Wang, J.; Lei, C. Research on the forecasting Method for natural gas price based on the data Mining Technique. *China Min.* **2020**, *29*, 52–58.
15. Li, J.; Wu, Q.; Tian, Y.; Fan, L. Monthly Henry Hub natural gas spot prices forecasting using variational mode decomposition and deep belief network. *Energy* **2021**, *227*, 120478. [CrossRef]
16. Mu, J.; Liu, X.; Zang, P.; Li, S. Research on price prediction of natural gas futures using LSTM neural network. In Proceedings of the International Conference on Algorithms, High Performance Computing, and Artificial Intelligence (AHPCAI 2021), Guangzhou, China, 21 December 2021. [CrossRef]
17. Sahin, D. Comparative Analysis of XGBoost and Light GBM Methods for Day Ahead Spot Natural Gas Price Forecasting. Ph.D. Thesis, Stanbul Technical University, Istanbul, Türkiye, 2022; p. 301181010.
18. Zhang, L.; Wang, L.; Wang, X.; Zhang, Y.; Pan, Z. How macro-variables drive crude oil volatility? Perspective from the STL-based iterated combination method. *Resour. Policy* **2022**, *77*, 102656. [CrossRef]
19. Fang, T.; Zheng, C.; Wang, D. Forecasting the crude oil prices with an EMD-ISBM-FNN model. *Energy* **2023**, *263*, 125407. [CrossRef]
20. Chen, H.; Wu, H.; Kan, T.; Zhang, J.; Li, H. Low-carbon economic dispatch of integrated energy system containing electric hydrogen production based on VMDGRU short-term wind power prediction. *Int. J. Electr. Power Energy Syst.* **2023**, *154*, 109420. [CrossRef]
21. Du, X.; Zhao, Q.; Mi, L.; Lang, Z.; Liu, M. Prediction model for China's monthly LNG ex-factory prices based on BP-ARIMA. *Oil Gas Storage Transp.* **2024**, *43*, 1173–1179, 1188.
22. Pan, K.; Xie, X.; Zhang, X.; Liu, D.; Zhang, H.; Zhang, Y.; Deng, Y.; He, M.; Li, H.; Sun, R. Short-and medium-term forecast of domestic LNG ex-factory prices based on CCA-ELM model. *Int. Pet. Econ.* **2024**, *32*, 99–106.
23. Sun, R.; Deng, Y.; Li, H.; Liu, Z. A prediction method of China's LNG spot price based on sliding time window and combination model. *Nat. Gas Ind.* **2025**, *45*, 170–178.
24. Zheng, X.; Yuan, Y.; Guo, Y.; Sun, Y.; Li, M. An Informer-FFT method for predicting daily price of natural gas. *Nat. Gas Technol. Econ.* **2025**, *19*, 39–45.
25. Zhang, J.; Liu, Z.; Wang, M. Research on natural gas price forecasting model based on CEEMD-ELM-ARIMA. *Nat. Gas Oil* **2021**, *39*, 129–136.
26. Lin, Y.; Dai, D.; Yu, Y.; Li, Z.; Huang, W.; Zhao, L.; Xing, H. Forecasting natural gas prices using a novel hybrid model: Comparative study of different sliding windows. *Energy* **2025**, *329*, 136607. [CrossRef]

27. Wang, J.; Cao, J.; Yuan, S.; Cheng, M. Short-term forecasting of natural gas prices by using a novel hybrid method based on a combination of the CEEMDAN-SE-and the PSOALS-optimized GRU network. *Energy* **2021**, *233*, 121082. [CrossRef]
28. Liu, J.; Qiu, B.; Du, P.; Zhao, X.; Zhu, J. A novel probabilistic connectivity network link prediction modelfor natural gas price based on an improved K-shell algorithm. *Phys. A Stat. Mech. Its Appl.* **2025**, *671*, 130672. [CrossRef]
29. Wang, J.; Lei, C.; Guo, M. Daily natural gas price forecasting by a weighted hybrid data-driven model. *J. Pet. Sci. Eng.* **2020**, *192*, 107240. [CrossRef]
30. Duan, Y.; Zhang, J.; Wang, X. Henry Hub monthly natural gas price forecasting using CEEMDAN–Bagging–HHO–SVR. *Front. Energy Res.* **2023**, *11*, 1323073. [CrossRef]
31. Yang, S.-F.; Choi, S.-W.; Lee, E.-B. A Prediction Model for SpotLNG Prices Based on Machine Learning Algorithms to Reduce Fluctuation Risks in Purchasing Prices. *Energies* **2023**, *16*, 4271. [CrossRef]
32. Ceperic, E.; Zikovic, S.; Ceperic, V. Short-term forecasting of natural gas prices using machine learning and feature selection algorithms. *Energy* **2017**, *140*, 893–900. [CrossRef]
33. Gao, S.; Hou, C.; Nguyen, B.H. Forecasting natural gas prices using highly flexible time-varying parameter models. *Econ. Model.* **2021**, *105*, 105652. [CrossRef]
34. Du, P.; Zhang, X.-K.; Du, J.-T.; Wang, J.-Z. Multivariate natural gas price forecasting model with feature selection, machine learning and chernobyl disaster optimizer. *Pet. Sci.* **2025**, *22*, 4823–4837. [CrossRef]
35. Guan, R.; Wang, A.; Liang, Y.; Fu, J.; Han, X. International Natural Gas Price Trends Prediction with Historical Prices and Related News. *Energies* **2022**, *15*, 3573. [CrossRef]
36. Wang, X.; Wang, H.; Wang, Z.; Wu, D.; Xue, Y.; Ye, L.; Zhang, Y. A Hybrid PLO-Transformer-CEEMDAN Model for Natural Gas Price Forecasting. *IEEE Access* **2025**, *13*, 143489–143508. [CrossRef]
37. Zhou, M.; Du, P. Multivariate events enhanced pre-trained large language model for carbon price forecasting. *Energy* **2025**, *336*, 138377. [CrossRef]
38. Huang, Y.; Zhu, Y.; Lei, G.; Wang, A.; Zhu, J. LLM-Enhanced Short-Term Electricity Price Forecasting Method for Australian Electricity Market. *Appl. Sci.* **2026**, *16*, 200. [CrossRef]
39. Wang, W.; Luo, Y.; Ma, M.; Wang, J.; Sui, C. A novel forecasting framework leveraging large language model and machine learning for methanol price. *Energy* **2025**, *320*, 135123. [CrossRef]
40. Xia, X.; Li, A. Global natural gas market and changing trends of natural gas trade. In *Annual Report on China's Petroleum, Gas and New Energy Industry (2022–2023)*; China International United Petroleum & Chemicals Co., Ltd., Chinese Academy of Social Sciences, Peking University, Eds.; Springer: Singapore, 2024. [CrossRef]
41. Zhang, S.Q.; Sun, Z.; Liu, Y. Causes and countermeasures of seasonal gas supply shortage in 2017. *Nat. Gas Ind.* **2018**, *38*, 120–128.
42. Deng, Y.X.; Feng, X.L.; Sun, R.J.; Li, H.H. Research on influencing factors of China's LNG ex-factory price under the background of natural gas marketization reform. *Int. Pet. Econ.* **2024**, *32*, 80–86.
43. Zhang, W. Gas storage and peak shaving function of LNG projects—Also on the transformation of construction concepts of LNG projects. *Econ. Manag. Saf. Environ. Prot.* **2010**, *30*, 1–4.
44. Yarlagadda, B.; Iyer, G.; Binsted, M.; Patel, P.; Wise, M.; McLeod, J. The future evolution of global natural gas trade. *iScience* **2024**, *27*, 108902. [CrossRef]
45. Hua, B.; Luo, J.X. Analysis on price trend of international LNG market. *Nat. Gas Ind.* **2007**, *27*, 140–144.
46. Xiao, J.Z.; Wang, X. Transmission mechanism of China's spot price of liquefied natural gas. *Nat. Gas Ind.* **2019**, *39*, 117–125.
47. Ratan, R.; Ravi, S.; Samane, G.; Amit, K.; Manoj, K.T. Sustainable Energy Technologies and Assessments. *Sustain. Energy Technol. Assess.* **2016**, *18*, 140–152.
48. Zhang, D.; Shi, M.; Shi, X. Oil indexation, market fundamentals, and natural gas prices: An investigation of the Asian premium in natural gas trad. *Econ. Model.* **2018**, *69*, 33–41. [CrossRef]
49. Tiantian, W.; Fei, W.; Dayong, Z. Evolution of the oil and gas price-linkage with multiple uncertainties. *Econ. Model.* **2025**, *151*, 108893. [CrossRef]
50. Li, L.; Li, J.; Li, K.; Luo, X.; Jiao, J. Climatic impacts on residential natural gas consumption: Evidence from Hefei, China. *Energy Build.* **2022**, *275*, 112488. [CrossRef]
51. Manou, M.; Maria, C.A.; Gino, G.; Stefano, C. Can extreme weather forecasts lead to a risk premium? Evidence of a non-linear response in U.S. natural gas futures. *N. Am. J. Econ. Financ.* **2025**, *80*, 102494. [CrossRef]
52. Nicolosi, C.F.; Tina, G.M. Monte Carlo and bootstrap approaches for power system adequacy: A study on atmospheric variables correlations. *Int. J. Electr. Power Energy Syst.* **2025**, *169*, 110827. [CrossRef]
53. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.