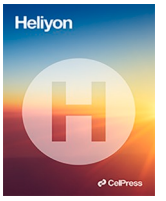
Heliyon 10 (2024) e33387

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Do multisource data matter for NGP prediction? Evidence from the G-LSTM model

Jun Hao [a,b], Shufan Shang [a,b], Jiaxin Yuan [a,b], Jianping Li [a,b,*]

[a] School of Economics and Management, University of Chinese Academy of Sciences, China
[b] MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Sciences, China

ARTICLE INFO

Keywords:
Multisource data
News text
Search index
Natural gas price forecast

ABSTRACT

Precisely predicting natural gas prices (NGPs) is important because it can provide the necessary decision-making basis for energy scheduling, planning and control. However, NGPs are affected by many factors and exhibit the characteristics of nonlinearity and randomness, which makes accurate predictions challenging. Therefore, in this paper, the information gain of multisource data and the global optimization ability of the gray wolf algorithm are used to build a multifactor-driven NGP hybrid forecasting model to improve the prediction performance. First, the emotional tendency and readability of news text are extracted and calculated by using VADER and textstat tools, respectively. Then the network search index is filtered and integrated by using the correlation coefficient method and the CRITIC method to form alternative variables of multisource data (news and search index). Second, the gray wolf optimization algorithm is used to find and determine the best key parameter group in long short-term memory model. Finally, the spot price of natural gas in Henry Hub from March 1, 2012 to February 28, 2022 is selected as the prediction object, and multi-scenario numerical experiments are carried out to verify the effectiveness of the proposed model. The ablation experiment results show that the information gain brought by multisource data can effectively improve the prediction effect of NGPs. Furthermore, the proposed model has the best prediction performance in different scenarios and can be regarded as a promising prediction tool.

1. Introduction

Due to global warming and rising energy demand, natural gas has become the best choice for different countries to adjust their energy structure and transform toward low-carbon development due to its high energy efficiency and economical, environmentally-friendly characteristics [1,2]. On a global scale, the tight supply of the natural gas market and the continuous increase in demand have led to a continuous rise in global natural gas prices. Changes in natural gas prices often have a profound impact on economic activities [3–6]. In this case, the accurate prediction of natural gas prices (NGPs) can effectively reduce energy risks, provide an important investment and decision-making basis for governments, investors and regulatory agencies, and promote the healthy and rapid development of the natural gas market. However, the NGP is often affected by a series of unconventional events or factors, such as market supply and demand levels, financial markets, foreign exchange markets, futures markets, geopolitics, and war, showing strong nonlinear and stochastic complex characteristics. There is no doubt that these complex data characteristics lead to severe challenges in

* Corresponding author. School of Economics and Management, University of Chinese Academy of Sciences, China.
E-mail address: ljp@ucas.ac.cn (J. Li).

https://doi.org/10.1016/j.heliyon.2024.e33387
Received 20 September 2023; Received in revised form 3 June 2024; Accepted 20 June 2024
Available online 21 June 2024

precisely predicting the NGP. Therefore, the establishment of an efficient and accurate prediction model to predict NGP has become a challenging issue in academic and industrial circles. As shown in Fig. 1, according to data released by the Henry Hub, natural gas spot price has fluctuated between $2 to $6 dollars per million btu in the past decade. It also has shown short-term downturn in 2015 and 2020 respectively, and shown a rebound trend after 2021. Due to the complex and diverse factors that determine energy prices, including supply-demand relationships, geopolitical factors, market expectations, etc., natural gas prices have not shown very regular cyclical fluctuations, Fig. 1 also reflects this point.

At present, NGP forecasting models can be divided into two categories: traditional statistical models and machine learning models. The former is mostly used in early price forecasting research, such as the gray model (GM), linear regression (LR) model, autoregressive integrated moving average (ARIMA), and generalized autoregressive conditional heteroskedasticity (GARCH). However, traditional statistical models need to meet the stationarity assumption, and the prediction effect is greatly affected by the law of data distribution [7–10]. Owing to the development of information technology, deep learning represented by long short term memory (LSTM) has good feature representation and fitting ability for complex price time series and is widely accepted and applied [11,12,13]. By introducing a gating mechanism to control the speed of information accumulation, LSTM can effectively solve the problem of gradient disappearance or explosion and reduce the risk of underfitting [14,15]. However, the setting and selection of hyperparameters in LSTM usually depend on the experience of users, which easily leads to the interference of subjective factors and reduces the prediction performance of the model. Therefore, it is necessary to use the optimization algorithm to optimize the hyperparameters of the model to weaken the influence of non-model factors on the prediction performance. The swarm intelligence algorithm connects single limited individuals through a certain mechanism to build swarm intelligence and uses the powerful overall ability to seek the optimal solution of complex problems through collaborative search in the solution space [16,17]. Among them, the gray wolf optimization algorithm (GWO) achieves the purpose of optimization by simulating the predation behavior and the cooperative mechanism of gray wolves. GWO has significant advantages in terms of solution accuracy and convergence speed by introducing the adjusted adaptive convergence factor and information feedback mechanism [18].

With the global marketization reform of natural gas, natural gas prices are driven by both market and unconventional factors, including financial market factors, exchange rate markets, commodity markets, and geopolitical factors. This makes the influencing factors and mechanisms of natural gas prices increasingly complex. In the era of big data, oil- and gas-related news reports are the mainstream media of current popular events in the energy field, with strong influence and representativeness, and can objectively record major social events in this field as well as relevant government decisions and market trends. Online news texts can be seen as an effective supplement to information, reflecting the impact of unconventional and unexpected events on natural gas prices in a timely manner [19]. The emotional tendency and readability of online news texts can not only measure the polarity and capacity of the information covered by the text, but also potentially affect the decisions and actions of readers. On the other hand, there may be a close correlation between online search data for social activity keywords and corresponding social activities [20,21]. Therefore, the search index can reflect the attention information of investors to a certain extent [22,23]. This raises two interesting research questions that we aim to answer in this paper: (1) Can multisource data mainly based on news texts and search indices effectively improve the effectiveness of NGP prediction? (2) How can multisource data be effectively utilized and efficient prediction models be established to achieve the accurate prediction of NGPs?

To this end, the following work is conducted in this study. First, macroeconomic fundamental indicators such as the Nasdaq Composite and the Global Economic Policy Uncertainty Index, as well as commodity prices such as oil futures prices and gold futures prices are selected as exogenous variables. On this basis, we embed the text indicators obtained after processing news texts and search indices and generate a total of 20 variables as explanatory variables of the model. Second, the gray wolf optimization algorithm is used to automatically search for and determine the optimal key parameter group in a deep learning model-LSTM to achieve the automatic

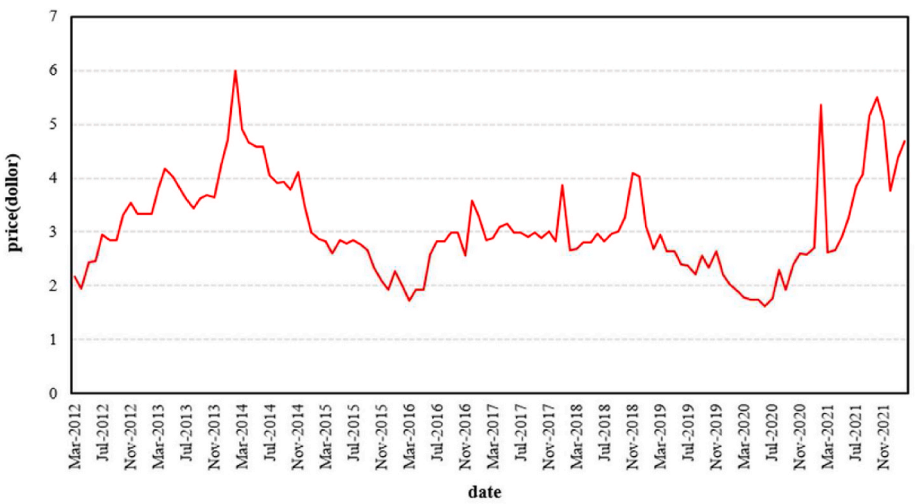


Fig. 1. The monthly spot price sequence of natural gas in Henry Hub.

optimization of the model's key parameters. Finally, the spot price of natural gas in Henry Hub is selected as the prediction object, and multi-scenario numerical calculation experiments are conducted to verify the effectiveness of the proposed model, attempting to answer whether the information gain of multisource data can improve the prediction accuracy of NGPs.

In short, the contributions of this paper to the literature on NGP prediction lie in the following three aspects: (1) Introducing text information such as online news and investor attention to predict NGPs, enriching the influencing factors of NGP prediction, and further verifying that the information gain of multisource data can effectively improve the NGP prediction performance. (2) Introducing the gray wolf optimization algorithm with global search capability to optimize the key parameters of the LSTM model, achieving the automatic optimization of key parameters. (3) Establishing a multisource data-driven hybrid forecasting model for NGP and design prediction experiments in different scenarios, emphasizing the applicability and effectiveness of the proposed model.

The remaining parts of this paper are arranged as follows. Section 2 presents a literature review. Section 3 provides an introduction to the main methods and prediction framework. Section 4 discusses empirical analysis, including descriptive analysis, parameter setting, predictive experiments, robustness analysis and DM testing. Finally, the conclusions of this paper and future outlooks are summarized in Section 5.

2. Literature review

Table 1 has shown the methods used by representative literature in the field of natural gas price prediction and related fields, as well as information on exogenous variables and specific research objects. Initially, scholars often used traditional statistical models to predict natural gas prices. However, as major countries around the world relaxed their control of natural gas prices, the influencing factors of natural gas prices became more diverse, and the fluctuation patterns of natural gas prices became more complex. The demand for improving the accuracy of time series prediction has led to a diversification of research methods in the field of NGP prediction.

Traditional statistical models are often based on the assumption of stationarity, emphasizing causal inference to uncover the laws behind prediction. Therefore, their empirical results have good interpretability, and they can be used to conduct in-depth research on driving factors while predicting NGPs. Econometric methods such as autoregressive integrated moving average (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH) and their variants were once favored by scholars. For example [25], established the GARCH-MIDAS-ES model to study the fluctuation law of the monthly futures price of natural gas and discussed the contribution of weather indicators to the prediction model [26]. constructed a state-space time series model to predict the daily closing price of natural gas and examined the impact of factors such as risk rollover on the price. However, traditional time series prediction requires solving parameters based on determining the model structure and using the solved model to complete the prediction task. Whether the parameters can be correctly selected largely determines the prediction effect. Therefore, the limitations of the parameter estimation process limit the prediction accuracy of the model.

In recent years, due to the rapid growth of data scale and increasingly complex data structures in price prediction tasks, scholars have found that deep learning models have relatively better fitting ability and prediction accuracy for complex time series. Therefore, their applications in the field of price prediction have become increasingly widespread. Deep learning is essentially a multilayer neural network, and its main models include convolutional neural networks (CNNs) and recurrent neural networks (RNNs). It is necessary to strictly define the input and output when using such models to predict prices. At present, many scholars use deep learning models and other technologies in the prediction process to achieve better prediction effects. For example [27], used VNN technology combined

Table 1
Prediction methods for NGPs.

Literature	Prediction object(s)	Exogenous variables	Method(s)	
[24]	Crude oil price	/	Traditional statistics	VTFM
[25]	Natural gas price	Temperature, precipitation, etc		GARCH-MIDAS-ES
[26]	Natural gas price	Crude oil price, coal price, etc		State-space time series model
[3]	Crude oil price, coal price, natural gas price	/		ARIMA, SES, KNN
[27]	Crude oil price, zinc price, natural gas price, gold price	/	Deep learning	VMD-ANN
[28]	Natural gas price	Natural gas consumption, crude oil price, etc		VMD-PSO-DBN
[29]	Natural gas price	/		CEEMDAN-SE and PSO-ALS-GRU
[30]	Natural gas price, carbon price	/		VMD and AR-IBiLSTMELMAN
[31]	Crude oil price, natural gas price	/		CEEMDan-SVM-ARMA
[32]	Crude oil price	US dollar index, Dow Jones Industrial Average, crude oil price, etc	Text assistance	News text
[33]	Carbon price	Coal price, crude oil price, electricity price, etc		Social media text and baidu index
[34]	Carbon price	Coal price, crude oil price, natural gas price		News text and google Index
[35]	Soybean price	/		News text
[36]	Crude oil price	Gold price, US dollar index, etc		News text

with an artificial neural network to predict four different price time series, including natural gas prices [30]. used AR, an Elman neural network and an IBiLSTM network to predict low-frequency components, high-frequency components and other sublayers after decomposing the time series, effectively improving the prediction effect of natural gas prices.

In addition, to incorporate some unconventional events or factors into the energy prediction system to optimize the prediction effect, some scholars have also used text information to assist in predicting energy time series. Many studies believe that this approach can effectively improve the prediction accuracy. Due to the similarity of time series predictions for different types of energy prices and the research approach of integrating text information for time series prediction having a certain universality, text-assisted prediction methods applied to energy prices such as crude oil and coal can also be used to predict natural gas prices.

In terms of news reporting, behavioral finance challenges the traditional efficient market hypothesis, proving that human emotions and behavior can play important roles in investment decision-making. Based on this theoretical basis, many scholars have confirmed the auxiliary role of sentiment analysis in forecasting. Extracting text sentiment for auxiliary prediction has been applied in the field of price prediction, e.g., see Refs. [32,35,36]. In addition, quantifying text readability is another effective means of processing text information. From research in marketing, it has been noted that text readability can affect the transmission and sharing of information, thereby guiding the information receiver and ultimately influencing the behavioral decisions of potential investors. Both [37,38] demonstrated the contribution of text readability to time series prediction. Although there are limited researches about using text information to assist natural gas price prediction, many scholars have confirmed the feasibility of this approach and proposed optimistic prospects for it [39]. introduced an economic index that cover text factors into the energy price prediction system, optimized the prediction results, and proposed that news text can be extracted to predict energy prices. In addition to news reports, there has been a strong interest in internet search data in academic communities in recent years. Previous studies have shown that search indices can be regarded not only as proxy variables for investors' attention but also as important factors affecting asset prices and therefore are also a type of important text information. Both [33,34] attempted to integrate search indices for energy price prediction and achieved certain results [40]. used Google search indices to assist natural gas price prediction and demonstrated that the introduction of text information such as search indices have significantly improved the accuracy of prediction.

Due to the unique structure of time series and certain correlations between their samples, some leading-edge neural network models often have better prediction performance. Considering that key parameters in neural network models usually need to be set based on the experience of users, parameter optimization can be used to improve the prediction accuracy. Among optimization algorithms, the swarm intelligence optimization algorithm can efficiently search for the optimal solution of complex problems through collaborative search in the solution space. Therefore, the combined method of using the swarm intelligence optimization algorithm to optimize key parameters of the neural network model can be used to predict the time series of energy prices. In this field [41], constructed a CEEMDAN-GWO CatBoost combination model to quickly and accurately capture the changing patterns of complex time series. They introduced the GWO algorithm into time series prediction model and achieved rapid parameter optimization [28]. proposed a mixed model based on variational mode decomposition (VMD), particle swarm optimization (PSO) and a deep belief network (DBN) [29]. proposed a new method combining complete ensemble empirical mode decomposition with adaptive noise and sample entropy (CEEMDAN-SE) with a gated recurrent unit optimized by particle swarm optimization and an adaptive learning strategy (PSO-ALS-GRU) [42]. demonstrated that a novel ICEEMDAN-R-AttGRU combination model can improve the prediction accuracy of various energy price time series, such as Brent crude oil, Dacheng steel coke, and NYMEX natural gas price sequences. All the above studies proved that combined models can solve the limitations of traditional methods and make more accurate predictions.

Meanwhile, previous studies have shown that neural network models have many key parameters, and changing these parameters will significantly change the prediction performance of the model: changing the number of hidden layer neurons will change the network structure and improve the optimized terrain; changing the batch size will change the loss reduction situation and affect the stability of training; and changing the learning rate will affect the rate of convergence and generalizability of the model. Therefore, the number of hidden layer neurons, batch size and learning rate can be selected for optimization to significantly improve the prediction accuracy [43–46].

In summary, on the one hand, integrating multisource data to predict natural gas prices can provide more information gain for the prediction system and ultimately improve the prediction accuracy. On the other hand, using a swarm intelligence algorithm to optimize neural network parameters for multivariate prediction of natural gas prices can theoretically effectively improve prediction performance. In this paper, a total of 16 exogenous variables are selected from two aspects: macroeconomic fundamental indicators and commodity futures prices. By integrating text information such as news reports and search indices, an NGP prediction system is established, and a GWO-LSTM combined model is constructed to predict natural gas spot prices. This is a relatively comprehensive and novel prediction approach.

3. Methodology equation

3.1. Long short term memory (LSTM)

The long short-term memory (LSTM) network is a special recurrent neural network. Compared to other deep learning models, LSTM has the following advantages in time series prediction: (1) It can effectively capture the long-term dependencies in sequences, so it can be more suitable for prediction tasks that rely on historical input data while ensuring a relatively simple structure. (2) It introduces the gating mechanism, so it can maintain the continuity of gradients and effectively solve the problem of gradient disappearance or explosion. (3) By relying on multiple structures and parameters to control the flow of information, it can not only reduce the risk of underfitting but also be robust to noise [47,48].

The unit structure of LSTM is shown in Fig. 2. The current hidden layer output h_t and the current internal state c_t are jointly determined by the previous hidden layer output h_{t-1}, the previous internal state c_{t-1}, and the current input x_t. LSTM uses three gates to control the flow of information: the forget gate f_t is used to set weights to select the proportion of information to be retained, the input gate i_t is used to filter new information to be input at the current time and the output gate o_t is used to extract effective information from the current internal state to generate a new hidden layer. The calculation method for the three gates is shown in Equations (1)–(3).

$$f_t = \sigma(W_t \cdot (h_{t-1}, x_t) + b_t) \tag{1}$$

$$i_t = \sigma(W_i \cdot (h_{t-1}, x_t) + b_i) \tag{2}$$

$$o_t = \sigma(W_o \cdot (h_{t-1}, x_t) + b_o) \tag{3}$$

where W_t and b_t are the weight matrix and corresponding bias term of the forget gate, respectively. W_i and b_i are the weight matrix and corresponding bias term of the input gate, respectively. And W_o and b_o are the weight matrix and corresponding bias term of the output gate, respectively.

An LSTM unit can process the current internal state using the tanh function to obtain the candidate state $\widetilde{c}_t$, as shown in Equation (4).

$$\widetilde{c}_t = \tanh(W_c \cdot (h_{t-1}, x_t) + b_c) \tag{4}$$

After obtaining the candidate state, the LSTM unit will update the internal state based on the information selection of the forget gate and the input gate and then combine with the output gate to generate a new hidden layer, as shown in Equations (5) and (6).

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \widetilde{c}_t \tag{5}$$

$$h_t = o_t \otimes \tanh(c_{t-1}) \tag{6}$$

3.2. Gray wolf optimization (GWO) algorithm

The gray wolf optimization (GWO) algorithm achieves the goal of optimization by simulating the predatory behavior of gray wolf populations and the mechanism of wolf group collaboration. Integrating the GWO algorithm with deep learning models to predict energy time series yields better results, as the combination model can leverage the following advantages of the GWO algorithm: (1) It has a convergence factor and information feedback mechanism that can be adaptively adjusted to achieve a balance between local optimization and global search. (2) Without constraints of centralized control, it has good universality and robustness. (3) It has a relatively simple structure, so it has significant advantages in improving both accuracy and convergence speed [18,49].

Fig. 3 depicts the general process of the GWO algorithm. The dominance level of gray wolves can be divided into four levels: α, β, δ and ω. Among them, the α wolf is the highest activity decision-making wolf and the leader in the population. This wolf is responsible for leading the wolves to hunt prey, which is the optimal solution in the algorithm. The levels of the β wolf, δ wolf and ω wolf decrease in sequence, with the lower-level wolves subordinate to the higher-level wolves. β wolves are responsible for assisting their superiors, namely, the suboptimal solutions in the algorithm; δ wolves obey the commands and decisions of their superiors, are responsible for tasks such as reconnaissance and sentry and are also the suboptimal solutions in the algorithm; and ω wolves will update their positions around their superiors, namely, the candidate solution in the algorithm. The GWO algorithm includes the processes of encirclement,

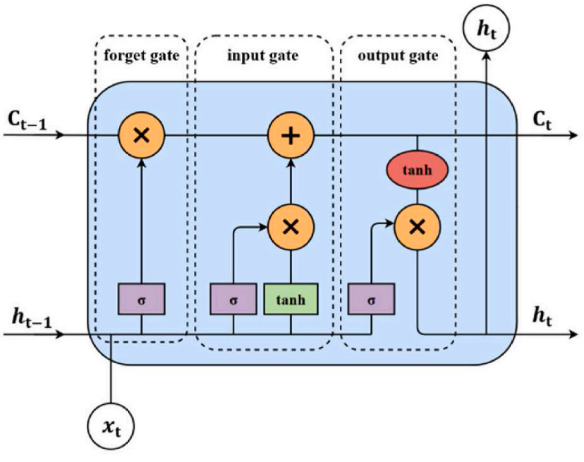


Fig. 2. The structure diagram of the LSTM unit.

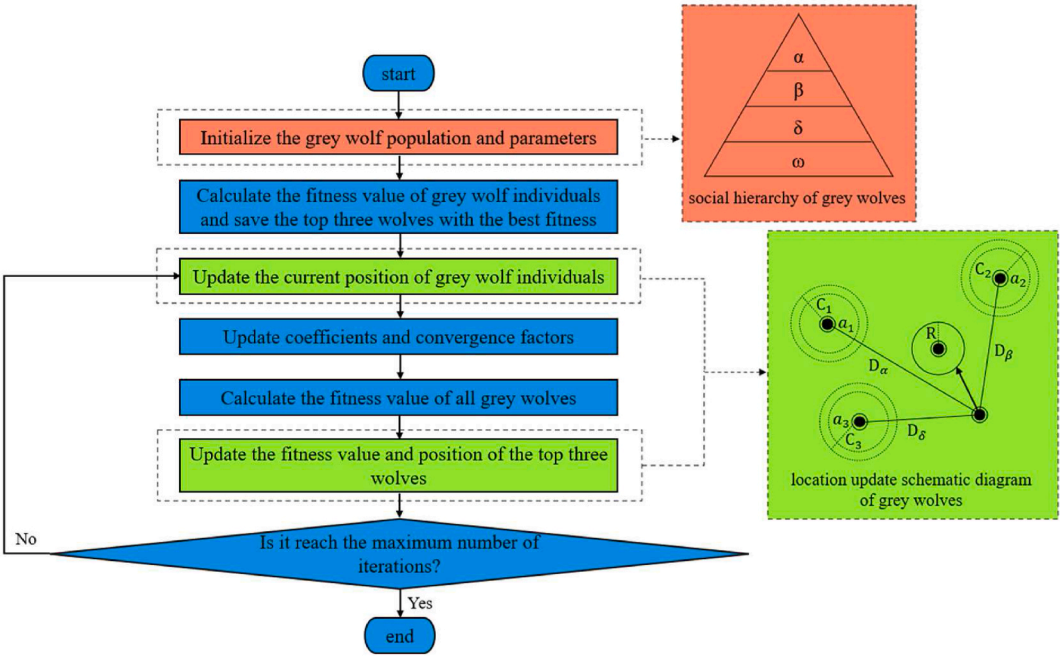


Fig. 3. Flowchart of the gray wolf optimization algorithm.

pursuit, and attack.

During the process of encirclement, the gray wolves update their positions, as shown in Equations (7)–(10).

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) \right| \tag{7}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \tag{8}$$

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{9}$$

$$\vec{C} = 2\vec{r}_2 \tag{10}$$

where $\vec{D}$ is the distance between the gray wolves and their prey, $\vec{X}$ and $\vec{X}_p$ are the position vectors of the gray wolves and their prey, respectively, and $\vec{A}$ and $\vec{C}$ are coefficients. In the calculation of coefficients, $\vec{r}_1$ and $\vec{r}_2$ are random vectors, and $\vec{a}$ is the convergence factor.

The process of hunting prey depends on the distance between different levels of wolves, as shown in Equations (11)–(13).

$$\vec{D}_\alpha = \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right| \tag{11}$$

$$\vec{D}_\beta = \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right| \tag{12}$$

$$\vec{D}_\delta = \left| \vec{C}_3 \cdot \vec{X}_\delta - \vec{X} \right| \tag{13}$$

where $\vec{D}_\alpha$, $\vec{D}_\beta$ and $\vec{D}_\delta$ are the distances between the ω wolf and the α wolf and between the β wolf and the δ wolf, respectively, while $\vec{X}_\alpha$, $\vec{X}_\beta$, and $\vec{X}_\delta$ are the positions of the α wolf, β wolf and δ wolf, respectively.

The position vector of the gray wolves is updated, as shown in Equations (14)–(17).

$$\vec{X}_1 = \left| \vec{X}_\alpha - A_1 \cdot \vec{D}_\alpha \right| \tag{14}$$

$$\vec{X}_2 = \left| \vec{X}_\beta - A_2 \cdot \vec{D}_\beta \right| \tag{15}$$

$$\vec{X}_3 = \left| \vec{X}_\delta - A_3 \cdot \vec{D}_\delta \right| \tag{16}$$

$$\vec{X}(t+1) = \left(\vec{X}_1 + \vec{X}_2 + \vec{X}_3 \right) \Big/ 3 \tag{17}$$

The process of attacking prey is influenced by controlling the value of the convergence factor, as shown in Equation (18).

$$a = 2 * (1 - t / T) \tag{18}$$

where T is the maximum number of iterations given.

3.3. Generation of text indicators

3.3.1. Online news reports

In this paper, the oil and gas news website Oil and Gas Daily was selected, and a web crawler was used to collect text data. The daily news from March 1, 2012, to February 28, 2022, were crawled, and the title, content, release time and other information of each news report were organized. The emotional tendency and readability of the news text were also analyzed. News sentiment analysis requires the use of sentiment analysis methods to provide the emotional scores and polarity of the text. Given that a method based on sentiment dictionaries can efficiently and comprehensively depict the nonstructured features of news texts, the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool was selected to perform sentiment analysis on news texts. When conducting emotional analysis on news texts, VADER accounts for factors such as negative expression, punctuation that can express emotional information and intensity, adverbs and phrases that express emotional intensity and obtains the sentiment score for text information (SS-TI). The analysis of news readability requires the integration of multiple readability indices to measure the readability of news texts from different perspectives. In this paper, the text readability calculation tool Textstat is used to analyze news texts. To measure the readability of the text by synthesizing multiple aspects such as words, sentences, letters, and syllables of the input text as much as possible, the Gunning Fog Index (GFI), Flesch Reading Ease (FRE) and Coleman–Liau Index (CLI) of the text were calculated and averaged to obtain the readability score for text information (RS-TI), as shown in Equations (19)–(22).

$$GFI = 0.4 \times \left(\frac{Num_{word}}{Num_{sentence}} + 100 \times \frac{Num_{long\ word}}{Num_{word}} \right) \tag{19}$$

$$FRE = 206.835 - 1.015 \times \frac{Num_{word}}{Num_{sentence}} - 84.6 \times \frac{Num_{syllable}}{Num_{word}} \tag{20}$$

$$CLI = 5.89 \times \frac{Num_{character}}{Num_{word}} - 30 \times \frac{Num_{sentence}}{Num_{word}} - 15.8 \tag{21}$$

$$RS - TI = (GFI + FRE + CLI) / 3 \tag{22}$$

where Num_{word}, $Num_{sentence}$, $Num_{long\ word}$, $Num_{syllable}$, and $Num_{character}$ represent the numbers of words, sentences, long words, syllables, and letters in the input text, respectively.

3.3.2. Network search index

In this paper, Google search indices from March 2012 to February 2022 were obtained by searching for keywords related to natural gas, and monthly data were converted into daily data through linear interpolation, as shown in Equation (23).

$$day_{mt+h} = \text{month}_t \times \frac{h}{m} + \text{month}_{t+1} \times \frac{m-h}{m} \tag{23}$$

where month_t is the index of the previous month and month_{t+1} is the index of the current month.

For search keywords, in this paper, the root words were customized, and then an association search was conducted. Based on the research results of previous scholars, root words, such as “natural gas”, representing natural gas, and “macroeconomy”, representing the macroeconomy, were selected. Multiple association searches were conducted for each root word to obtain a series of related words. Then, the Pearson correlation coefficient was used to quantify the correlation between related words and root words. Finally, the tested related words were selected as the final search keywords. The CRITIC method, which can objectively assign weights based on the data itself [50,51], was selected in this paper to determine the weights of keywords.

The calculation of conflict indicators and variation indicators is shown in Equations (24) and (25).

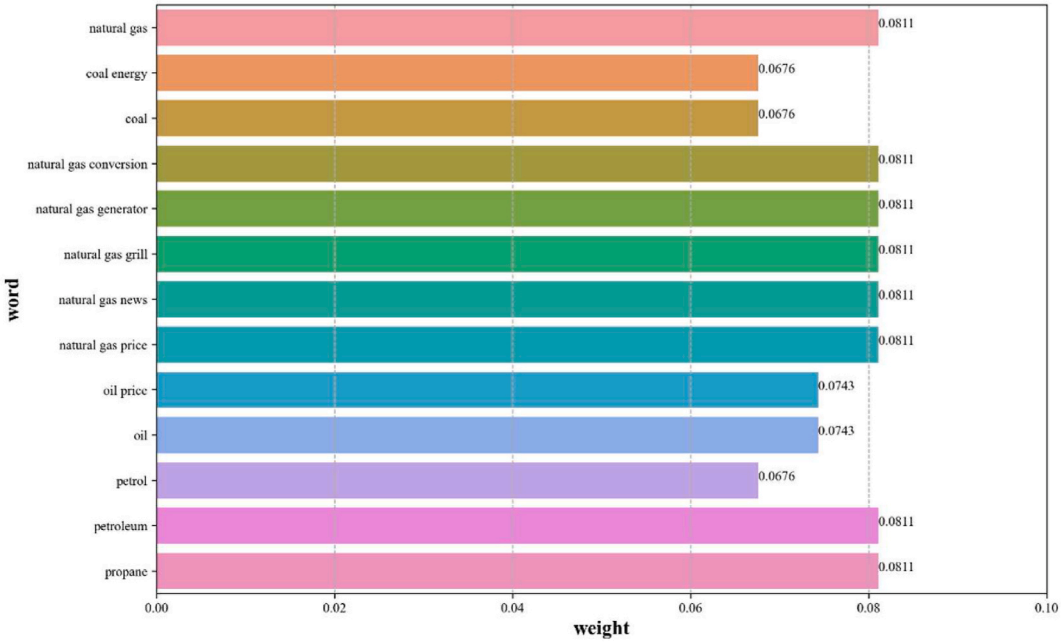
$$S_j = \sqrt{\frac{\sum_{i=1}^{n} \left(x_{ij} - \overline{x}_j \right)}{n-1}} \tag{24}$$

$$R_j = \sum_{i=1}^{p} (1 - r_{ij}) \tag{25}$$

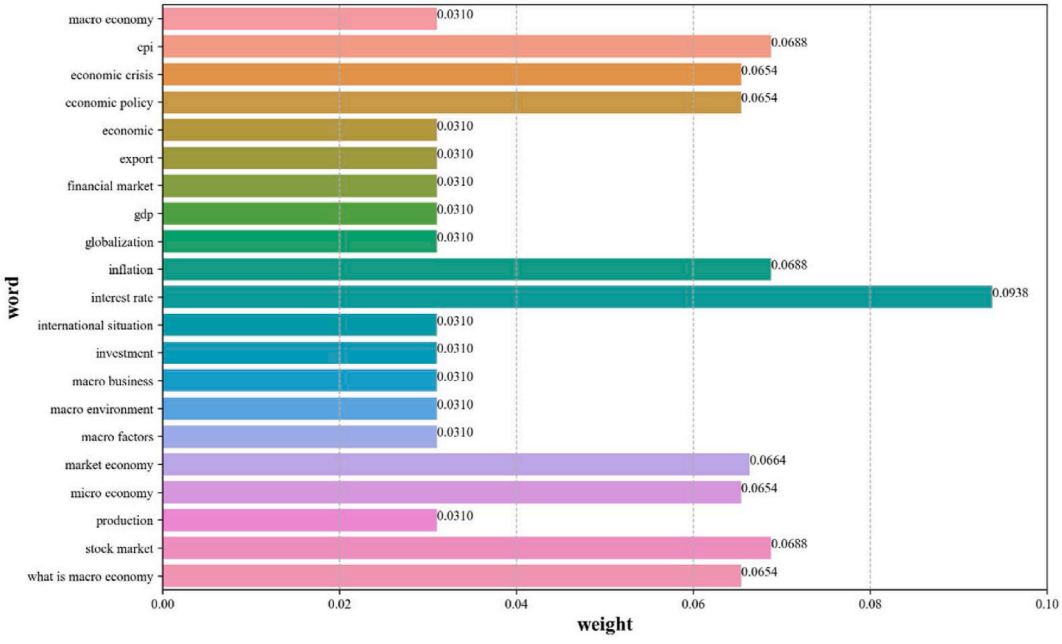
where $\bar{x}_j$ represents the mean of the j-th data and r_{ij} represents the correlation coefficient between the i-th and jth data.

The weight of the j-th element, w_j, is calculated based on the above indicators, as shown in Equations (26) and (27).

$$C_j = S_j \times R_j \tag{26}$$



(a) Weights of natural gas search terms



(b) Weights of macro economy search terms

Fig. 4. The chart of CRITIC weights.

$$w_j = \frac{C_j}{\sum_{j=1}^{m} C_j} \tag{27}$$

The significance test results of the search indices for root words and related words are shown in Appendix 1. The weight of each type of search term was calculated using the CRITIC method, as shown in Fig. 4. Through screening and weighting, two comprehensive indices were ultimately obtained, namely, "natural gas" in text information (NG-TI) (the weights are shown in Fig. 4 (a)) and "macroeconomy" in text information (ME-TI) (the weights are shown in Fig. 4 (b)).

3.4. Prediction framework

Given the information gain of multisource data and the global optimization ability of the gray wolf algorithm, a multisource data-driven combination prediction model was constructed for NGP, and its prediction framework is shown in Fig. 5.

Step 1. Data collection and preprocessing. First, online news texts, search indices, and macroeconomic variables were collected and obtained. Then, linear interpolation was used to fill in missing values (Equation (28)), outliers were removed using Equations (29) and (30), and the data were standardized using the min–max method (Equation (31)).

$$day_k = day_{k-i} \times \frac{i}{i+j} + day_{k+j} \times \frac{j}{i+j} \tag{28}$$

where day_k represents the data to be interpolated and day_{k-i} and day_{k+j} are the two data closest to day_k.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{29}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})^2} \tag{30}$$

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{31}$$

where x^* is the standardized data, x is the original data, $x_{\max}$ is the maximum value, and $x_{\min}$ is the minimum value.

Step 2. Proxy variables based on online news texts and search indices were constructed. In terms of news text, the VADER and Textstat tools were used to analyze emotional tendency and readability, respectively. In terms of the search index, the Pearson correlation test was used to screen the search terms. Then, the search index of similar keywords was weighted and synthesized through the CRITIC method to obtain the final comprehensive index.

Step 3. The GWO algorithm was used to optimize the parameters in the LSTM model, including the number of hidden layer neurons,

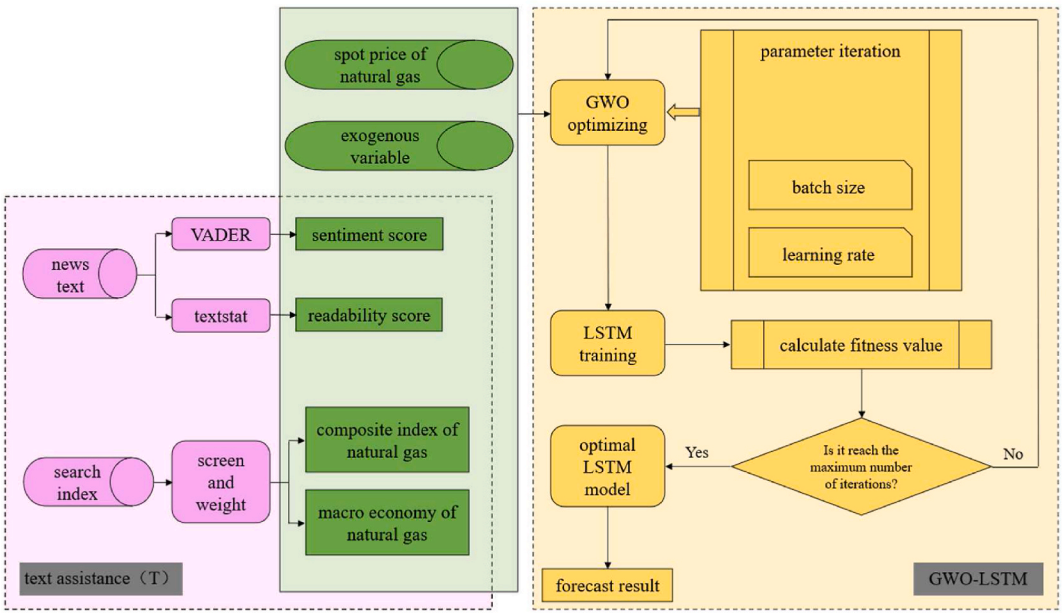


Fig. 5. Prediction framework.

batch size and learning rate. Building the LSTM model based on the optimal hyperparameters weakens the impact of non-model factors on prediction performance to obtain more accurate prediction results.

Step 4. The prediction results of the proposed model were compared with those of the benchmark model to verify the indispensability of the model components, the rationality of the combination process and the necessity of incorporating text information. In addition, the advantages of the proposed model were further demonstrated by setting up multistep prediction and DM testing.

4. Empirical analysis

4.1. Dataset descriptions

In this paper, the spot price of natural gas in Henry Hub in the United States was selected as the prediction object to verify the advantages of the proposed model. We provide a descriptive statistical analysis of the natural gas price and its related influencing factors, as shown in Table 2. Among them, the time span of natural gas prices is from March 1, 2012, to February 28, 2022, with a total of 2608 observations. The maximum value of the daily natural gas price is 23.86, the minimum value is 1.33, the average is 3.10, and the standard deviation is 1.01. According to the skewness result, the spot price of natural gas shows a right-skewed distribution, indicating that there may be maximum outliers in the spot price sequence of natural gas. The kurtosis result indicates that the distribution of the natural gas spot price sequence has a significant peak sharpness, indicating a clear trend toward central clustering.

In this paper, the dataset was divided into a training set and a testing set in an 8:2 ratio, with the data from March 1st, 2012, to February 27th, 2020, serving as the training set and the data from February 28th, 2022, to February 28th, 2022, serving as the testing set. The selection of relevant variables in this paper not only involves macroeconomic fundamentals and commodity prices but also introduces indicators representing news texts and search indices to characterize the impact of unexpected events and social factors on natural gas prices [52,53]. The correlation coefficient between variables is shown in Fig. 7. To visually display the experimental subject, we plotted the fluctuation curve of the NGP, as shown in Fig. 6. From Fig. 6, it can be observed that the NGP had significant volatility and structural changes during the research period, including a sharp increase on approximately February 10, 2014, and a sudden increase on February 17, 2021. The former may be due to the tense situation in Eastern Europe at the beginning of that year, which led to an increase in the demand for safe energy from neighboring countries. The latter is due to the decline in global crude oil

Table 2
Descriptive statistical results of variables.

Variable	Symbol	Mean	Std.	Minimum	Median	Maximum	Skewness	Kurtosis	Source
Spot price of natural gas	SPNG	3.10	1.01	1.33	2.91	23.86	4.36	71.29	Energy Information Administration
Sentiment score for text information	SS-TI	0.33	0.83	−1.00	0.90	1.00	−0.73	−1.31	Constructed in this paper
Readability score for text information	RS-TI	25.24	2.38	15.99	25.28	31.74	−0.47	0.41	
“natural gas” in text information	NG-TI	43.31	6.38	30.31	43.41	79.12	0.83	2.37	
“macro economy” in text information	ME-TI	49.50	4.38	41.03	48.80	64.76	0.99	1.21	
Dow Jones Industrial Average	DJIA-SMI	21796.45	6472.90	12101.46	20549.37	36799.65	0.55	−0.69	Tonghuashun Financial Database
Nasdaq Composite	NC-SMI	6880.21	3445.96	2747.48	5834.72	16057.44	1.08	0.23	
USD/EUR exchange rate	DE-ER	0.85	0.06	0.72	0.86	0.96	−0.46	−0.93	
USD/GBP exchange rate	DP-ER	0.71	0.07	0.58	0.72	0.87	−0.16	−1.32	
5-year break-even inflation rate	BIR-IR	1.78	0.43	0.14	1.75	3.17	0.29	0.90	Federal Reserve Bank of ST.Louis
10-year bond yield	BY-IR	2.04	0.61	0.52	2.07	3.24	−0.44	−0.28	Federal Reserve System
Discount rate	DR-IR	0.63	0.77	0.04	0.15	2.45	1.18	−0.13	
Consumer price index	CPI-O	246.46	13.06	227.72	243.59	283.63	0.74	−0.21	United States Department of Labor
Global economic policy uncertainty index	GEPUI-O	184.74	69.27	86.29	165.78	430.26	0.92	0.30	Tonghuashun Financial Database
Crude oil futures price	CO-FP	19.98	5.17	11.74	17.78	35.61	1.03	0.08	Energy Information Administration
Gold futures price	GO-FP	1425.73	405.76	820.00	1349.48	3150.00	1.69	3.75	New York Commodity
Silver futures price	SI-FP	1426.08	242.29	1050.80	1317.40	2051.50	0.66	−0.94	Exchange
Wheat futures price	WH-FP	14289.04	4426.05	9095.00	12970.00	29520.00	1.52	1.62	Chicago Board of Trade
Maize futures price	MA-FP	65.50	22.07	−37.63	59.63	110.53	0.34	−0.80	
Glass futures price	GL-FP	563.81	116.85	361.00	527.20	928.00	0.67	−0.44	Zhengzhou Commodity Exchange
Natural rubber futures price	NR-FP	444.95	120.83	301.50	383.00	760.20	1.15	−0.05	Shanghai Board of Trade
Training set		2012.03.01–2020.02.27							
Testing set		2020.02.28–2022.02.28							

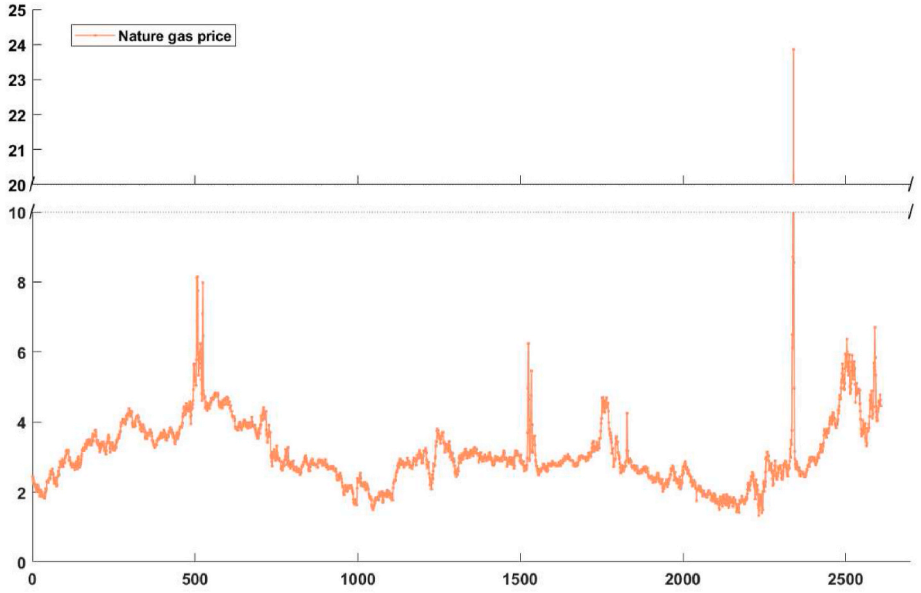


Fig. 6. The daily spot price sequence of natural gas in Henry Hub.

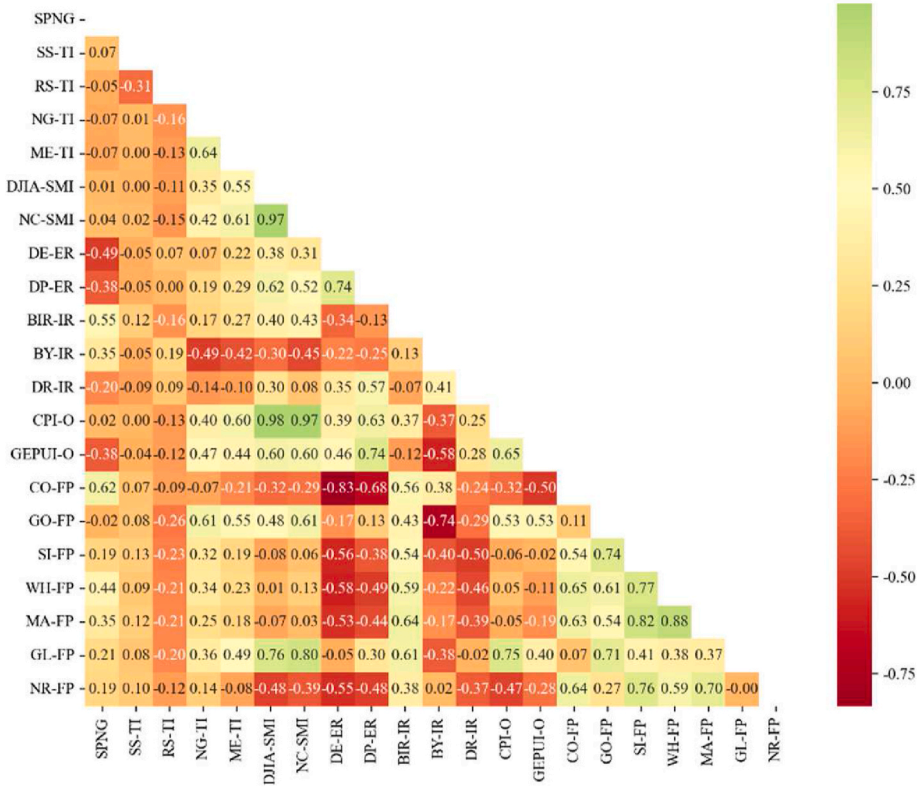


Fig. 7. Thermodynamic diagram of the variable correlation coefficient.

inventories during the pandemic and the strict implementation of production reduction agreements by OPEC countries.

4.2. Evaluation criteria

In this paper, the root mean square error (RMSE), mean square error (MSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) were selected to evaluate the accuracy of the model [54], as shown in Equations (32)–(35).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{32}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{33}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{34}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{35}$$

In addition, Diebold–Mariano (DM) statistical testing was introduced to determine whether the prediction accuracy of Model A is significantly better than that of Model B. The calculation method of DM statistics is shown in Equations (36)–(38).

$$DM = \frac{\frac{1}{n}\sum_{t=1}^{n} g_t}{\sqrt{(\gamma_0 + 2\sum_{l=1}^{\infty}\gamma_l)/n}} \sim N(0,1) \tag{36}$$

$$g_t = e_{A,t}^2 - e_{B,t}^2 \tag{37}$$

$$\gamma_l = \mathrm{cov}(g_t, g_{t-l}) \tag{38}$$

where $e_{A,t}$ and $e_{B,t}$ are the prediction errors of Model A and Model B, respectively.

4.3. Benchmark models and parameter settings

4.3.1. Benchmark model

To verify the advantages of the proposed model, 15 models were selected as benchmark models, including LSTM, GWO-LSTM, T-PSO-LSTM and T-GWO-GRU. By conducting multiple experiments on all benchmark models and calculating relevant evaluation indicators, the T-GWO-LSTM model was compared from different perspectives to demonstrate its predictive performance [55,56]. Information on the benchmark models is shown in Table 3.

In terms of neural networks, in this paper, RNNs and GRUs were selected for comparative analysis. A recurrent neural network (RNN) is a type of neural network that can efficiently utilize the characteristics of data sequences. However, as the network hierarchy deepens, RNNs are highly likely to experience gradient disappearance or explosion [57,58]. A gated recurrent unit (GRU) suppresses to some extent the characteristic of the RNN that cannot be relied on for a long time by introducing an update gate and reset gate. However, its relatively simple structure may lead to underfitting when there is a large amount of input information [59,60].

Table 3
Benchmark model explanation.

No.	Category		Model
1	Neural network	Using different neural network	RNN
2			GRU
3		Component of the proposed model	LSTM
4	Text assist + neural network	Using different neural network	T-RNN
5			T-GRU
6		Components of the proposed model	T-LSTM
7	Swarm intelligence optimize parameters + neural network	Changing the swarm intelligence part	ACO-LSTM
8			PSO-LSTM
9		Change the neural network part	GWO-RNN
10			GWO-GRU
11		Components of the proposed model	GWO-LSTM
12	Text assist + swarm intelligence optimize parameters + neural network	Changing the swarm intelligence part	T-ACO-LSTM
13			T-PSO-LSTM
14		Change the neural network part	T-GWO-RNN
15			T-GWO-GRU
16	The proposed model		T-GWO-LSTM

In terms of swarm intelligence, in this paper, ACO and PSO were selected for comparative analysis. Ant colony optimization (ACO) optimizes by simulating the behavior of ant colonies using pheromones to achieve indirect communication for foraging. This algorithm has strong robustness but slow convergence speed [61]. The core idea of particle swarm optimization (PSO) is that particles determine their next motion through their own experience and the experience of the best among their peers. Its advantage is that the steps are simple and relatively easy to converge, but it may also fall into a local optimal solution [62].

4.3.2. Parameter setting

The initial parameter settings for the combined model and its comparison model are shown in Table 4. To ensure fairness in the comparison, the structure, loss function and other parameters of different neural networks are set to be consistent. Similarly, parameters such as the number of individuals and the maximum number of iterations in each swarm intelligence optimization algorithm should also be set the same.

The GWO optimization process combines the hyperparameters to be optimized (number of hidden layer neurons, batch size, and learning rate) to form a multidimensional array for combinatorial optimization. Only when the input array reaches a stable minimum fitness value can the number of iterations in this state be determined as the optimal number of iterations. The convergence process of parameter optimization for the combined model is shown in Fig. 8. From Fig. 8(a), The fitness value stabilizes before 50 iterations, reaching 0.233657876, which also proves the rationality of the iteration number setting in this paper. In addition, as shown in Fig. 8 (b), (c) and (d), the three parameters to be optimized stabilize before 50 iterations. Therefore, we set the number of hidden layer neurons to 43, the batch size to 50, and the learning rate to 0.0117825032.

4.4. Prediction results

To verify the good performance of the T-GWO-LSTM combined model proposed in this paper, ARIMA, LSTM, T-LSTM, GWO-LSTM, T-PSO-LSTM, T-GWO-GRU and other models were selected as comparative models to calculate multiple evaluation indicators for each model. The results are shown in Table 5.

From Tables 5 and it can be observed that the T-GWO-LSTM model has the best prediction performance, and models that have good performance include T-RNN, GWO-LSTM, GWO-GRU, and T-GWO-RNN. Among them, the RMSE, MSE, MAPE and MAE values of the T-GWO-LSTM model decreased by 6.38 %, 12.26 %, 5.81 %, and 11.10 %, respectively, compared to those of the second-best model GWO-GRU. Meanwhile, compared with other comparative models, the T-GWO-LSTM model significantly improved in terms of evaluation indicators, which proves that the T-GWO-LSTM model has good predictive performance.

Note that the model proposed in this paper has a significant improvement in prediction accuracy compared to traditional models such as ARIMA and combination models such as ARIMA-SVR. This reflects the relatively better fitting ability of the combination model based on deep learning methods and swarm intelligence optimization for complex time series. This is not only due to the effective capture of patterns in a large amount of historical data by neural networks, but also due to the advantages of both universality and robustness of swarm intelligence optimization algorithms in determining parameters.

Multiple experiments were conducted in one step, and the results were compare with those of the reduced component model, as shown in Fig. 9, and the RMSE, MSE, MAPE, and MAE values of each model were shown in Fig. 9 (a)–(d), respectively. The evaluation index values of the T-GWO-LSTM model were significantly lower than those of the LSTM model, T-LSTM model, and GWO-LSTM model. At the same time, the data of each index value of the T-GWO-LSTM model were also relatively more concentrated, which proves the indispensability of the model components in this paper.

In addition, the results of the model that changed the swarm intelligence optimization algorithm are shown in Fig. 10, and the RMSE, MSE, MAPE, and MAE values of each model were shown in Fig. 10 (a)–(d), respectively. The evaluation index values of the T-GWO-LSTM model were significantly lower than those of ACO-LSTM, PSO-LSTM, GWO-LSTM, T-ACO-LSTM and T-PSO-LSTM. Meanwhile, the T-GWO-LSTM model had no abnormal data in multiple prediction processes, which proves that the GWO algorithm has a stable and significant advantage in the combination process.

Moreover, for the model with the modified neural network, the results are shown in Fig. 11, and the RMSE, MSE, MAPE, and MAE values of each model were shown in Fig. 11 (a)–(d), respectively. The evaluation index values of the T-GWO-LSTM model were significantly reduced compared to those of the comparison models such as GWO-RNN, GWO-GRU, GWO-LSTM, T-GWO-RNN, T-GWO-GRU and T-GWO-LSTM, which proves the rationality of choosing the GWO algorithm to optimize the LSTM network in this paper.

Table 4
Parameter setting.

Model		Parameters
Neural network	RNN	Structure: fully connected input layer + RNN layer + fully connected output layer, loss function: mae, iteration: 100
	GRU	Structure: fully connected input layer + GRU layer + fully connected output layer, loss function: mae, iteration: 100
	LSTM	Structure: fully connected input layer + LSTM layer + fully connected output layer, loss function: mae, iteration: 100
Swarm intelligence optimization	ACO	Number of ants: 10, iteration: 50, pheromone volatilization rate: 0.8, pheromone release amount: 1
	PSO	Number of particles: 10, iteration: 50, inertia weight: 2, acceleration coefficient: 1.5
	GWO	Number of gray wolves: 10, iteration: 50
Combined model		Search dimension: 3 Search scope (number of hidden layer neurons, batch size and learning rate): the lower limit is [10,50,0.001], the higher limit is [200,100,0.8]

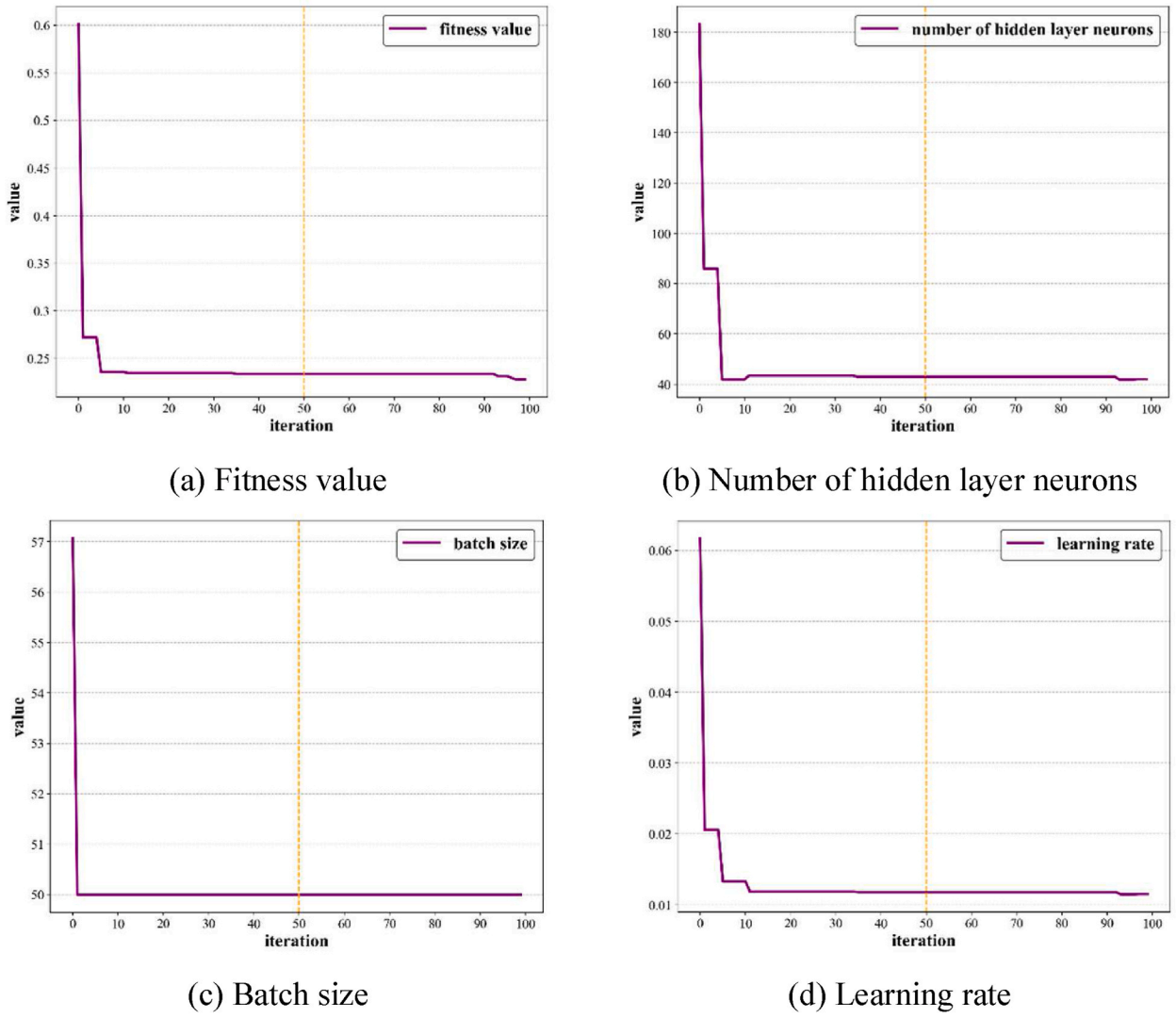


Fig. 8. The iterative convergence diagram of the parameter optimization process.

According to the results in Tables 5 and it can be seen that compared to not using swarm intelligence optimization algorithms to optimize neural network parameters, the evaluation index values of models using optimization algorithms are usually significantly reduced, and the indicator data obtained from multiple experiments are relatively more concentrated. This proves the advantage of constructing a combination model for prediction. At the same time, by replacing the optimization part of the T-GWO-LSTM model with other swarm intelligence optimization algorithms, the evaluation index values of the prediction results have not decreased. Compared with other combination models containing the GWO algorithm, the prediction results of the T-GWO-LSTM model are significantly optimal and relatively stable. This proves that choosing GWO algorithm to optimize LSTM network is the most reasonable choice in this paper.

Based on the analysis of Table 5 and Figs. 9–11, it was found that the prediction accuracy of the text-assisted prediction model was significantly better than that of the corresponding non-text-assisted prediction model by comparing the models predicted using only exogenous variables with the corresponding models with text-assisted prediction. This proves the necessity of adding text information and further verifies that the information gain of multisource data can effectively improve the prediction performance of the model.

Furthermore, in order to demonstrate the superiority of the prediction results, the evaluation index values of the empirical results in this paper are compared with those of relevant researches and summarized in Table 6. This paper compares the RMSE value and MAPE value of prediction results. RMSE preserves data units and is relatively insensitive to outliers, so it can be used to compare prediction results containing zero or data closed to zero. MAPE is calculated based on absolute percentage error and is not affected by data scale and range when comparing model performance on different datasets. Based on Tables 6 and it can be seen that the prediction results of the model proposed in this paper have certain advantages over other similar research topics in terms of RMSE and MAPE value, which further proves that the T-GWO-LSTM model has good prediction performance.

Table 5
Results of one-step prediction performance evaluation.

No.	Model	RMSE		MSE		MAPE		MAE	
		value	rank	value	rank	value	rank	value	rank
1	SVR	0.7583	10	0.5751	10	0.1517	9	0.5241	10
2	ARIMA	1.6355	19	2.6748	19	0.3294	18	1.2239	19
3	ARIMA-SVR	0.8352	13	0.6975	13	0.1687	10	0.5682	11
4	RNN	0.4554	8	0.2074	8	0.1114	8	0.3465	8
5	GRU	1.1289	18	1.2744	18	0.2119	13	0.7970	17
6	LSTM	0.9606	15	0.9227	15	0.2395	16	0.7330	14
7	T-RNN	0.3533	5	0.1248	5	0.0958	6	0.2642	5
8	T-GRU	0.8264	12	0.6830	12	0.2146	14	0.6574	13
9	T-LSTM	1.0246	17	1.0498	17	0.3559	19	0.9044	18
10	ACO-LSTM	0.9880	16	0.9762	16	0.2166	15	0.7742	16
11	PSO-LSTM	0.8090	11	0.6545	11	0.1848	11	0.6500	12
12	GWO-RNN	0.4249	7	0.1806	7	0.1048	7	0.3157	7
13	GWO-GRU	0.2587	2	0.0669	2	0.0688	3	0.1865	3
14	GWO-LSTM	0.3035	4	0.0921	4	0.0834	5	0.2264	4
15	T-ACO-LSTM	0.8856	14	0.7844	14	0.2587	17	0.7384	15
16	T-PSO-LSTM	0.5422	9	0.2940	9	0.2048	12	0.4736	9
17	T-GWO-RNN	0.2632	3	0.0693	3	0.0637	1	0.1808	2
18	T-GWO-GRU	0.3734	6	0.1395	6	0.0829	4	0.2653	6
19	T-GWO-LSTM	0.2422	1	0.0587	1	0.0648	2	0.1658	1

4.5. Robustness analysis

To further validate the significant advantages of the proposed model in prediction, multistep prediction experiments were conducted on T-GWO-LSTM and models that performed well in the one-step prediction empirical process (T-RNN, GWO-GRU, GWO-LSTM, and T-GWO-RNN). The results are shown in Table 7. It can be observed that as the step size increases, the prediction accuracy of each model decreases from Table 7. Although the T-GWO-LSTM model is not the optimal model in some steps, overall, the T-GWO-LSTM model still exhibits excellent prediction performance.

To provide a more intuitive test of the performance of the T-GWO-LSTM prediction model, prediction fitting and error analysis were conducted on the T-GWO-LSTM model with different step sizes. The fitting results and relative errors of six kinds of steps (1, 2, 3, 5, 10, and 20) are shown in Fig. 12 (a) - (f), respectively. As shown in Fig. 12, the one-step model has the best fitting effect. Compared to relative error of multistep prediction, the relative error of one-step prediction is lower and more stable, with a maximum relative error of no more than 0.6, and it mostly remains below 0.2. At the same time, the smaller the step size is, the closer the predictive value matches the true value, and the smaller the relative error, which is in line with objective logic.

From Fig. 12, the relative error values are relatively large in areas with dense sequence changes. On the one hand, intensive fluctuations may represent a high level of uncertainty in the market during the period, which is caused by multiple factors such as macroeconomic changes, policy adjustments, changes in market supply-demand, etc. The increase in uncertainty makes it difficult to predict, resulting in an increase in relative error values. On the other hand, intensive changes may indicate the presence of inefficient or incomplete information in the market during the period. Due to higher risks, companies and investors will make more cautious decisions, which will lead to delayed or reduced investments, resulting in more complex changes in natural gas prices. Market factors may also fail to reflect the real supply-demand dynamics in a timely manner, and the accuracy of prediction results may decrease.

In addition, the relative error values predicted at the outliers of the series are also relatively large. It may be due to extreme values typically indicate the occurrence of abnormal events, such as natural disasters, economic crises, political conflicts, etc. These situations often occur suddenly and uncontrollably, making them difficult to predict. Meanwhile, outliers may also be caused by intervention measures taken by the government and regulatory agencies, such as implementing new policies, changing the power of supervision, conducting market regulation, etc. These changes are structural, so they will reduce the accuracy of predictions using historical information. Therefore, the model needs more time to train appropriate parameters to achieve its original effect.

4.6. DM testing

The prediction performance of the T-GWO-LSTM model and other comparative models with different step sizes is tested, and the results are shown in Table 8. From Tables 8 and it can be seen that with one step, the p value of the DM test for the T-GWO-LSTM model proposed in this paper and the GWO-GRU model, which is the second-best model, is less than 0.05. Therefore, the original hypothesis can be rejected at a 5 % confidence level, indicating that the two models have different effects. The negative DM statistic value (−4.3227) of the test indicates that the proposed model is significantly superior to the second-best model. In addition, the majority of the results in Table 8 are consistent with p values less than 0.05 and negative DM statistic values, indicating that the two models tested have different effects at a 5 % confidence level, while the proposed model is significantly better (Li et al., 2021). Although the T-GWO-LSTM model has no significant advantage compared to the GWO-GRU model and T-GWO-RNN model in some steps, overall, it can be considered that the T-GWO-LSTM model is superior to the GWO-GRU model and T-GWO-RNN model. Therefore, it can be

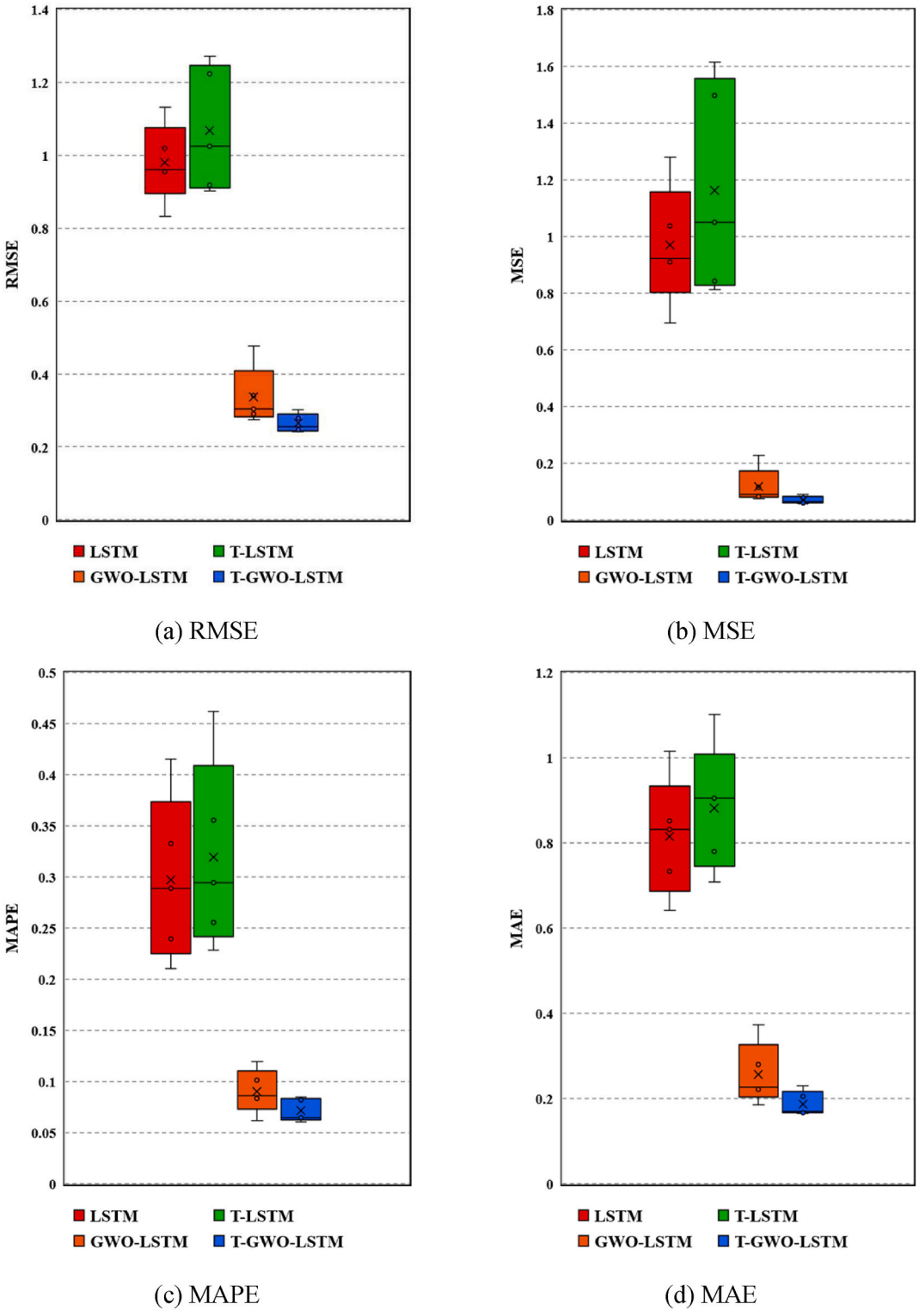


Fig. 9. Comparison of the prediction performance of different models.

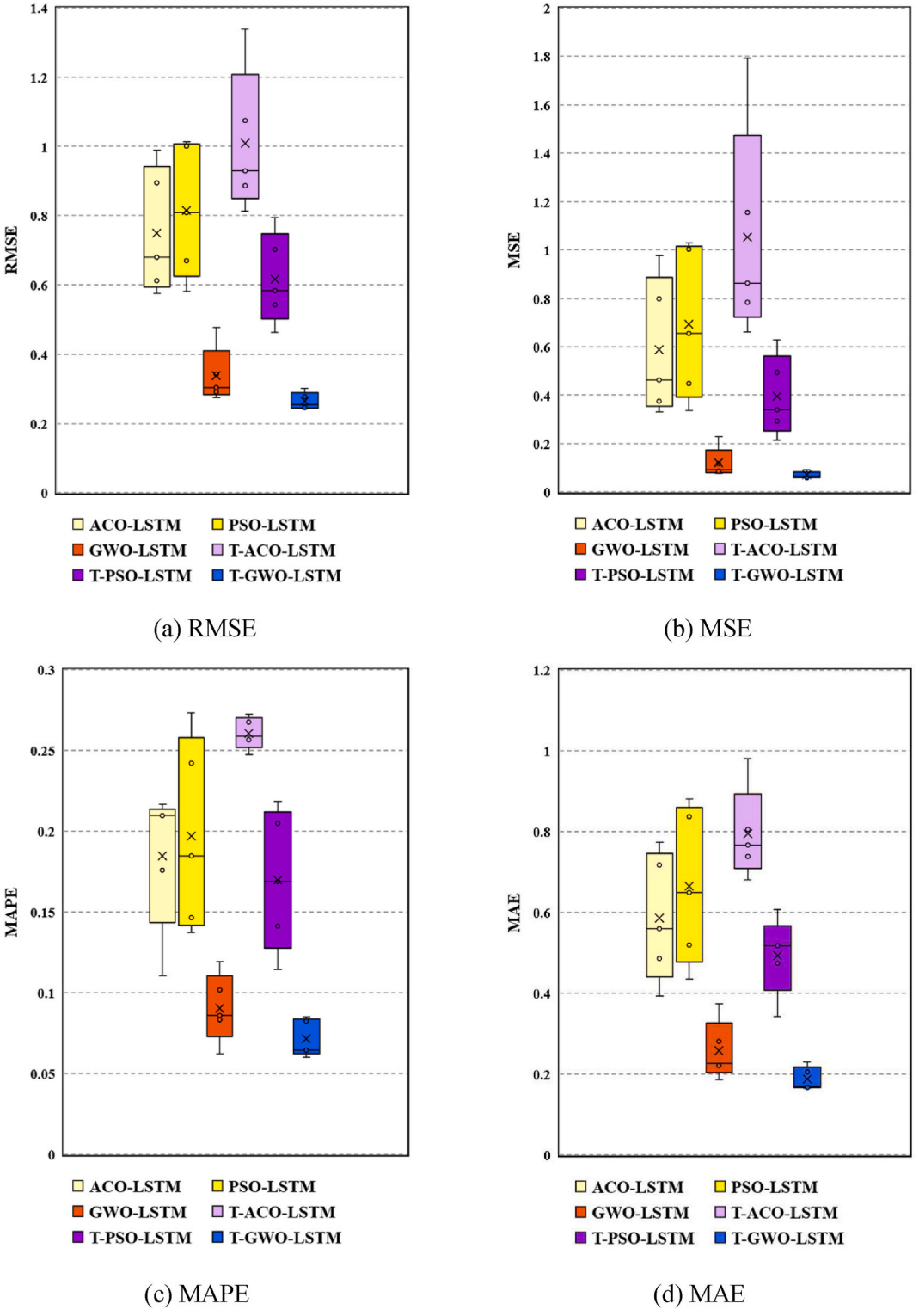


Fig. 10. Comparison of the prediction performance of different models.

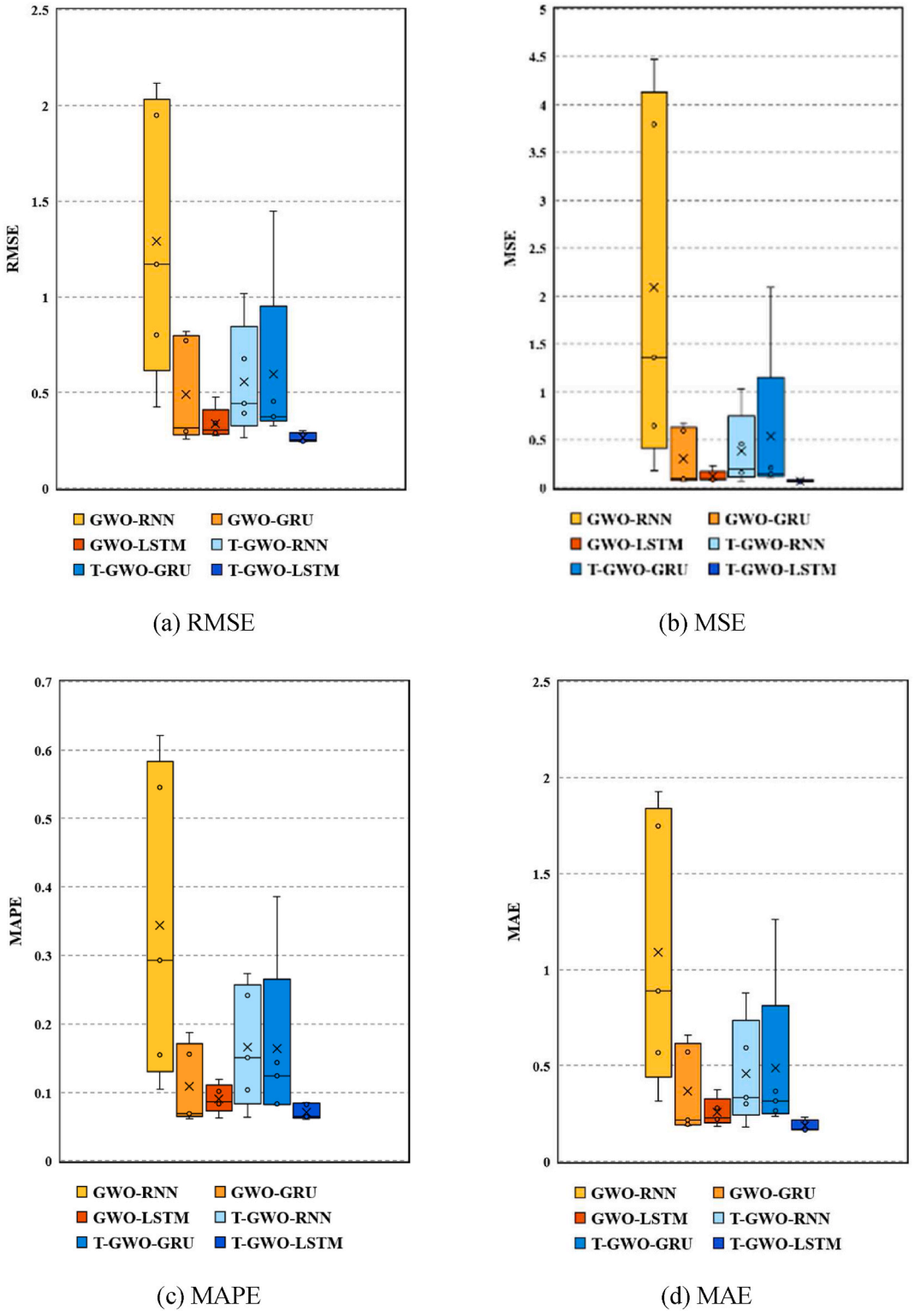


Fig. 11. Comparison of the prediction performance of different models.

Table 6
Evaluation index values of energy price prediction.

Literature	Prediction object(s)	Method(s)	RMSE	MAPE
This paper	Natural gas price	T-GWO-LSTM	0.242	0.065
[28]	Natural gas price	PSO-DBN	0.273	0.099
[30]	Natural gas price	VMD-LSTM	/	0.075
[3]	Natural gas price	ARIMA	5.073	2.565
[63]	Carbon price	ET-MVMD-LSTM	0.376	1.109
[64]	Carbon price	WT-GWA-LSTM	0.710	0.820

Table 7
Results of multistep prediction performance evaluation.

Number	Model	H = 1	H = 2	H = 3	H = 5	H = 10	H = 20
Panel A: RMSE							
4	T-RNN	0.3533	0.4118	0.4467	0.4930	0.5536	0.6604
10	GWO-GRU	0.2587	0.3395	0.3845	0.4490	0.5192	0.6255
11	GWO-LSTM	0.3035	0.4333	0.5210	0.6322	0.8327	0.9986
14	T-GWO-RNN	0.2632	0.3468	0.3868	0.4509	0.5213	0.6217
16	T-GWO-LSTM	0.2422	0.3445	0.3905	0.4301	0.5079	0.6607
Panel B: MAPE							
4	T-RNN	0.0958	0.1099	0.1203	0.1243	0.1508	0.1890
10	GWO-GRU	0.0688	0.0830	0.0933	0.0998	0.1104	0.1378
11	GWO-LSTM	0.0834	0.1091	0.1320	0.1432	0.1713	0.1986
14	T-GWO-RNN	0.0637	0.0823	0.0939	0.1044	0.1146	0.1353
16	T-GWO-LSTM	0.0648	0.0929	0.1059	0.1008	0.1075	0.1362

demonstrated that the T-GWO-LSTM model has statistically significant advantages under asynchronous steps.

5. Conclusions

Precisely predicting natural gas prices is crucial because it can provide the necessary decision-making basis for energy scheduling, planning, and control. However, natural gas prices are influenced by many factors and exhibit characteristics of nonlinearity and randomness, posing serious challenges to accurate prediction. This raises two interesting questions that we aimed to answer in this paper: (1) Can multisource data, mainly news texts and search indices, effectively improve the prediction effect of NGP? (2) How can multisource data be effectively utilized and efficient prediction models be established to achieve accurate prediction of NGPs? Therefore, we utilized the information gain of multisource data and the global optimization ability of the gray wolf algorithm to construct a multifactor driven hybrid forecasting model for NGPs. First, we extracted and calculated the emotional tendency and readability of news texts and then filtered and integrated network search indices to form alternative variables for multisource data (news texts and search indices). Second, the gray wolf optimization algorithm was used to find and determine the key parameter group in the deep learning model-LSTM. Finally, the spot price of natural gas in Henry Hub was selected as the prediction object, and multi-scenario numerical experiments were conducted to verify the effectiveness of the proposed model. The experimental results indicate that (1) the information gain brought by multisource data can indeed effectively improve the prediction effect of NGPs, indicating that using text information to assist NGP prediction is an effective approach. (2) The proposed model in this paper has the best prediction performance with different step sizes and can be regarded as a promising prediction tool. The RMSE, MSE, MAPE and MAE values of the T-GWO-LSTM model are reduced by 6.38 %, 12.26 %, 5.81 %, and 11.10 %, respectively, compared to those of the second-best model GWO-GRU. (3) By comparing the T-GWO-LSTM model with other models, such as GWO-LSTM, T-PSO-LSTM, and T-GWO-GRU, it can be proven that the components of the proposed model are indispensable, the combination process is reasonable, and the necessity of adding text information is essential.

The results of this paper can be served as the theoretical basis of feasible suggestions about stabilizing natural gas prices. Due to the various market factors involved in the natural gas price prediction system, governments of various countries should promote the implementation of natural gas price linkage policies to avoid the occurrence of natural gas purchase-sale inversion or difficulties on NGP stability, and alleviate the operational difficulties of gas enterprises due to political, climate and other reasons. Specific measures include: (1) Provide subsidies or incentives to natural gas producers and importers, reduce relevant taxes or lower tax rates, and encourage natural gas production and supply to stabilize price. (2) Establish a NGP regulation-reserve mechanism, purchase excess natural gas when there is excess supply, sell reserve natural gas when there is insufficient supply, balance supply and demand to ultimately stabilize price. (3) Strictly control the upper and lower limits of NGP, establish a corresponding adjustment mechanism, regularly monitor the market, evaluate and adjust the price, and ensure it fluctuate within the specified range. (4) Establish anti-monopoly agencies and price supervision agencies, promulgate relevant laws and regulations, publicly disclose market information on natural gas trading, and reduce the possibility of price manipulation and market monopoly.

In addition, this paper confirms that text information can assist in predicting NGP, providing important insights for NGP prediction. Relevant institutions can hire professional teams about news and public opinion collection and monitoring to regularly summarize text

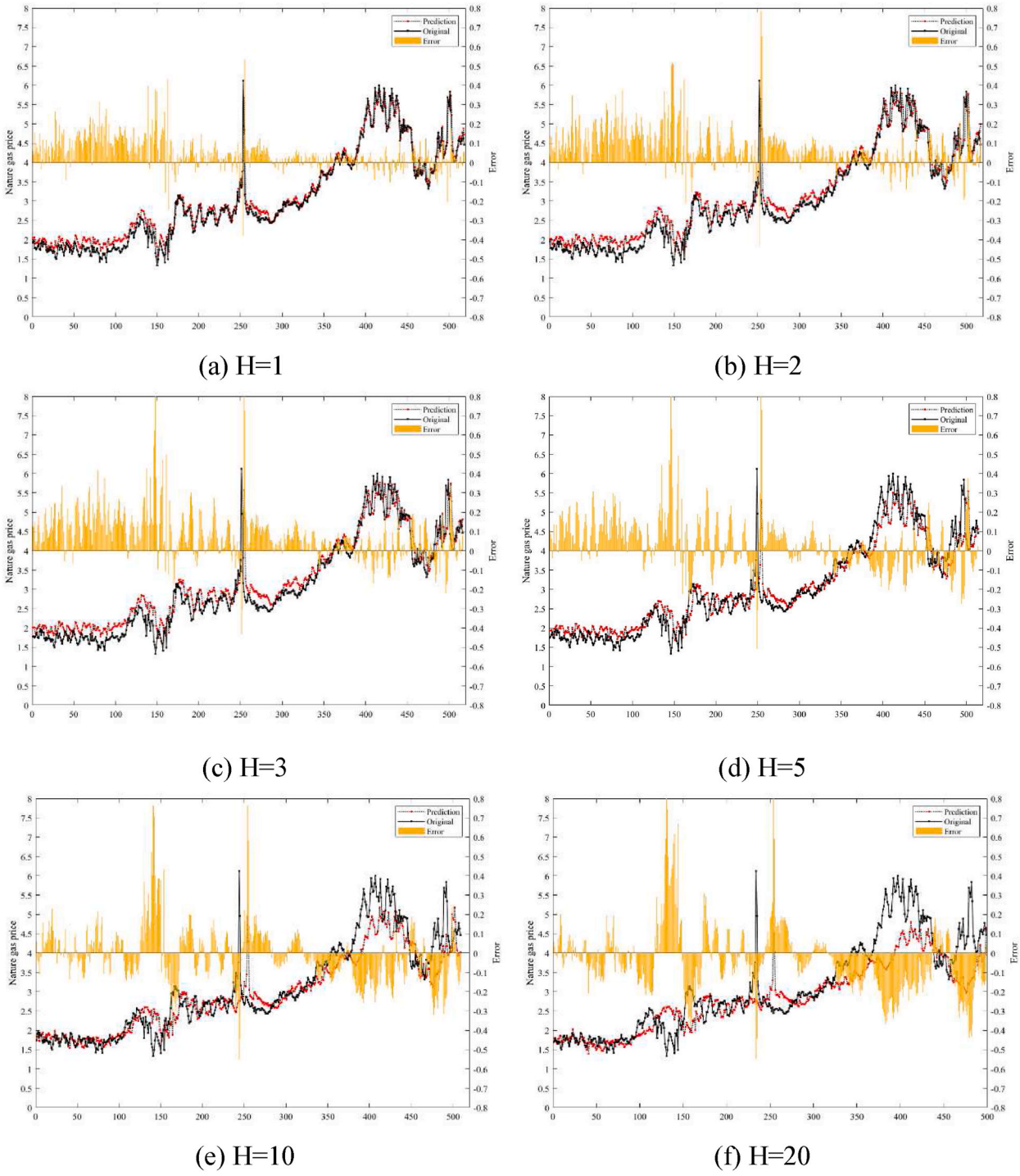


Fig. 12. Out-of-sample fitting curve and relative error value of the T-GWO-LSTM model.

data and timely incorporate the data into NGP prediction models, in order to better grasp the fluctuation of NGP. Specific measures include: (1) Establish a comprehensive data collection and analysis system to obtain data covering various market factors and energy production-marketing, including macroeconomic indicators and related energy prices. (2) Strengthen cooperation and information sharing with other countries and international organizations, establish cooperative relationships with energy departments of different countries and international energy agencies in order to obtain more accurate data and more comprehensive market information. (3) Establish a specialized text information collection system, combined with technologies such as web crawler and natural language

Table 8
DM statistical test results.

No.	Model	H = 1		H = 2		H = 3	
		DM	P value	DM	P value	DM	P value
1	RNN	−14.3146	(0.0000)	−8.8072	(0.0000)	−7.6430	(0.0000)
2	GRU	−17.9659	(0.0000)	−15.5375	(0.0000)	−14.8623	(0.0000)
3	LSTM	−21.3609	(0.0000)	−17.7380	(0.0000)	−16.5588	(0.0000)
4	T-RNN	−11.3478	(0.0000)	−6.3980	(0.0000)	−5.4036	(0.0000)
5	T-GRU	−23.3349	(0.0000)	−19.4137	(0.0000)	−17.9706	(0.0000)
6	T-LSTM	−37.4542	(0.0000)	−31.1464	(0.0000)	−29.3893	(0.0000)
7	ACO-LSTM	−22.1425	(0.0000)	−20.3835	(0.0000)	−18.8429	(0.0000)
8	PSO-LSTM	−21.9839	(0.0000)	−20.0846	(0.0000)	−11.4719	(0.0000)
9	GWO-RNN	−12.6712	(0.0000)	−7.5962	(0.0000)	−6.6700	(0.0000)
10	GWO-GRU	−4.3227	(0.0000)	1.2535	0.2106	1.4545	(0.1464)
11	GWO-LSTM	−10.4103	(0.0000)	−6.8034	(0.0000)	−8.1018	(0.0000)
12	T-ACO-LSTM	−27.9012	(0.0000)	−23.0191	(0.0000)	−21.5310	(0.0000)
13	T-PSO-LSTM	−30.2625	(0.0000)	−26.9673	(0.0000)	−13.7521	(0.0000)
14	T-GWO-RNN	−2.8765	(0.0042)	2.6138	(0.0092)	3.0773	(0.0022)
15	T-GWO-GRU	−8.8686	(0.0000)	−4.0557	(0.0001)	−3.0361	(0.0025)
No.	Model	H = 5		H = 10		H = 20	
		DM	P value	DM	P value	DM	P value
1	RNN	−12.6501	(0.0000)	−15.9707	(0.0000)	−18.1895	(0.0000)
2	GRU	−16.3953	(0.0000)	−17.5448	(0.0000)	−17.8886	(0.0000)
3	LSTM	−17.8674	(0.0000)	−17.6581	(0.0000)	−15.0153	(0.0000)
4	T-RNN	−7.5109	(0.0000)	−7.1790	(0.0000)	−3.8117	(0.0002)
5	T-GRU	−19.5828	(0.0000)	−19.6757	(0.0000)	−14.6946	(0.0000)
6	T-LSTM	−28.4431	(0.0000)	−24.9809	(0.0000)	−18.8525	(0.0000)
7	ACO-LSTM	−27.5971	(0.0000)	−26.3812	(0.0000)	−22.1296	(0.0000)
8	PSO-LSTM	−21.6371	(0.0000)	−26.9530	(0.0000)	−12.7673	(0.0000)
9	GWO-RNN	−13.2066	(0.0000)	−18.4513	(0.0000)	−19.7680	(0.0000)
10	GWO-GRU	−2.3140	(0.0211)	−1.8888	(0.0595)	2.1056	(0.0357)
11	GWO-LSTM	−12.5299	(0.0000)	−15.5484	(0.0000)	−15.4017	(0.0000)
12	T-ACO-LSTM	−23.0958	(0.0000)	−21.9583	(0.0000)	−16.5139	(0.0000)
13	T-PSO-LSTM	−18.2534	(0.0000)	−25.4698	(0.0000)	−21.6645	(0.0000)
14	T-GWO-RNN	−3.8207	(0.0001)	−4.2524	(0.0000)	3.8702	(0.0001)
15	T-GWO-GRU	−6.6922	(0.0000)	−9.9144	(0.0000)	−9.9034	(0.0000)

process, to collect relevant text data from multiple information sources and add the data into the NGP system. (4) Regularly evaluate the prediction results of NGP, continuously input new data and update parameters for prediction, continuously improve the accuracy and reliability of predictions.

In future work, multisource data can be used to provide richer information and achieve a systematic and comprehensive description of the object. Therefore, we will consider using multisource heterogeneous data to establish a prediction model to better improve the prediction performance of the model. Second, due to the numerous influencing factors in the natural gas market, it is necessary to establish dynamic prediction models to capture price trends. Finally, we will consider deploying a multisource data-driven combination prediction model into online prediction tools, which can benefit practitioners and better leverage their role in supporting investment decisions.

Data availability statement

Data will be made available on request.

CRediT authorship contribution statement

Jun Hao: Writing – review & editing, Writing – original draft, Project administration, Methodology, Conceptualization. **Shufan Shang:** Writing – original draft, Visualization, Software, Investigation, Data curation, Conceptualization. **Jiaxin Yuan:** Visualization, Software, Data curation. **Jianping Li:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

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

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (72201265, T2293774), China Postdoctoral Science Foundation funded project (2023T160635, 2022M723105), Fundamental Research Funds for the Central Universities, and MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation at the University of Chinese Academy of Sciences.

Appendix1

Table A1
Pearson correlation test results of root word and related word search index.

Root word	Related word	Correlation coefficient	P-value
natural gas	carbon	0.0685**	(0.4385)
	coal energy	0.5053	(0.0000)
	coal	0.3416	(0.0001)
	fuel	−0.0927**	(0.2941)
	natural gas conversion	0.5773	(0.0000)
	natural gas generator	0.2342	(0.0073)
	natural gas grill	−0.2660	(0.0022)
	natural gas heater	0.0777**	(0.3794)
	natural gas news	0.5038	(0.0000)
	natural gas price	0.2563	(0.0032)
	oil price	−0.2355	(0.0070)
	oil	−0.5143	(0.0000)
	petrol	−0.2507	(0.0040)
	petroleum	0.7905	(0.0000)
	propane	−0.3556	(0.0000)
	renewable energy	0.1626**	(0.0646)
	what is natural gas	0.1512**	(0.0859)
macro economy	cpi	0.3429	(0.0001)
	economic crisis	0.4857	(0.0000)
	economic policy	0.6736	(0.0000)
	economic	0.6972	(0.0000)
	export	0.2084*	(0.0173)
	financial market	0.8401	(0.0000)
	gdp	0.7061	(0.0000)
	globalization	0.7495	(0.0000)
	inflation	0.6589	(0.0000)
	interest rate	0.5737	(0.0000)
	international situation	0.7320	(0.0000)
	investment	0.5984	(0.0000)
	macro business	0.7817	(0.0000)
	macro environment	0.7835	(0.0000)
	macro factors	0.7930	(0.0000)
	market economy	0.6903	(0.0000)
	micro economy	0.7840	(0.0000)
	production	0.7325	(0.0000)
	stock market	0.3309	(0.0001)
	unemployment	0.0994**	(0.2604)
	what is macro economy	0.7618	(0.0000)
	world trade	0.0740**	(0.4025)

Note: * and ** represent insignificant at confidence levels of 1 % and 5 % respectively.

References

[1] M. Caporin, F. Fontini, R. Panzica, The systemic risk of US oil and natural gas companies, Energy Econ. 121 (2023) 106650, https://doi.org/10.1016/j.eneco.2023.106650.
[2] M. Greaker, E. Lund Sagen, Explaining experience curves for new energy technologies: a case study of liquefied natural gas, Energy Econ. 30 (2008) 2899–2911, https://doi.org/10.1016/j.eneco.2008.03.011.
[3] M.S. Alam, M. Murshed, P. Manigandan, D. Pachiyappan, S.Z. Abduvaxitovna, Forecasting oil, coal, and natural gas prices in the pre-and post-COVID scenarios: contextual evidence from India using time series forecasting tools, Resour. Pol. 81 (2023) 103342, https://doi.org/10.1016/j.resourpol.2023.103342.
[4] M. Kröger, M. Longmuir, K. Neuhoff, F. Schütze, The price of natural gas dependency: price shocks, inequality, and public policy, Energy Pol. 175 (2023) 113472, https://doi.org/10.1016/j.enpol.2023.113472.
[5] C.W. Su, M. Qin, H.-L. Chang, A.-M. Țăran, Which risks drive European natural gas bubbles? Novel evidence from geopolitics and climate, Resour. Pol. 81 (2023) 103381, https://doi.org/10.1016/j.resourpol.2023.103381.

[6] T. Wang, D. Zhang, D. Clive Broadstock, Financialization, fundamentals, and the time-varying determinants of US natural gas prices, Energy Econ. 80 (2019) 707–719, https://doi.org/10.1016/j.eneco.2019.01.026.
[7] N. Salehnia, M.A. Falahi, A. Seifi, M.H. Mahdavi Adeli, Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis, J. Nat. Gas Sci. Eng. 14 (2013) 238–249, https://doi.org/10.1016/j.jngse.2013.07.002.
[8] L. Wang, Y. Xia, Y. Lu, A novel forecasting approach by the GA-SVR-GRNN hybrid deep learning algorithm for oil future prices, Comput. Intell. Neurosci. 2022 (2022) 1–12, https://doi.org/10.1155/2022/4952215.
[9] M. Wang, L. Zhao, R. Du, C. Wang, L. Chen, L. Tian, H. Eugene Stanley, A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms, Appl. Energy 220 (2018) 480–495, https://doi.org/10.1016/j.apenergy.2018.03.148.
[10] Z. Xu, M. Mohsin, K. Ullah, X. Ma, Using econometric and machine learning models to forecast crude oil prices: insights from economic history, Resour. Pol. 83 (2023) 103614, https://doi.org/10.1016/j.resourpol.2023.103614.
[11] R. Ahmed, A. Shabri, Daily crude oil price forecasting model using arima, generalized autoregressive conditional heteroscedastic and Support Vector Machines, Am. J. Appl. Sci. 11 (2014) 425–432, https://doi.org/10.3844/ajassp.2014.425.432.
[12] F. Yu, X. Xu, A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network, Appl. Energy 134 (2014) 102–113, https://doi.org/10.1016/j.apenergy.2014.07.104.
[13] Y. Zhao, JP. Li, L. Yu, A deep learning ensemble approach for crude oil price forecasting, Energy Econ. 66 (2017) 9–16.
[14] V. Aryai, M. Goldsworthy, Day ahead carbon emission forecasting of the regional National Electricity Market using machine learning methods, Eng. Appl. Artif. Intell. 123 (2023) 106314, https://doi.org/10.1016/j.engappai.2023.106314.
[15] J. Wang, C. Lei, M. Guo, Daily natural gas price forecasting by a weighted hybrid data-driven model, J. Petrol. Sci. Eng. 192 (2020) 107240, https://doi.org/10.1016/j.petrol.2020.107240.
[16] M.K. Chegeni, A. Rashno, S. Fadaei, Convolution-layer parameters optimization in convolutional neural networks, Knowl. Base Syst. 261 (2023) 110210, https://doi.org/10.1016/j.knosys.2022.110210.
[17] P. Singh, S. Chaudhury, B.K. Panigrahi, Hybrid MPSO-CNN: multi-level particle swarm optimized hyperparameters of convolutional neural network, Swarm Evol. Comput. 63 (2021) 100863, https://doi.org/10.1016/j.swevo.2021.100863.
[18] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Software 69 (2014) 46–61, https://doi.org/10.1016/j.advengsoft.2013.12.007.
[19] A. Siganos, E. Vagenas-Nanos, P. Verwijmeren, Divergence of sentiment and stock market trading, J. Bank. Finance 78 (2017) 130–141, https://doi.org/10.1016/j.jbankfin.2017.02.005.
[20] W. Zhang, D. Shen, Y. Zhang, X. Xiong, Open source information, investor attention, and asset pricing, Econ. Modell. 33 (2013) 613–619, https://doi.org/10.1016/j.econmod.2013.03.018.
[21] X. Zhu, X. Chang, R. Li, H. Wang, Portal nodes screening for large scale social networks, J. Econom. 209 (2019) 145–157, https://doi.org/10.1016/j.jeconom.2018.12.021.
[22] Q. Ji, H.-Y. Zhang, J.-B. Geng, What drives natural gas prices in the United States? – a directed acyclic graph approach, Energy Econ. 69 (2018) 79–88, https://doi.org/10.1016/j.eneco.2017.11.002.
[23] S. Wiggins, X.L. Etienne, Turbulent times: uncovering the origins of US natural gas price fluctuations since deregulation, Energy Econ. 64 (2017) 196–205, https://doi.org/10.1016/j.eneco.2017.03.015.
[24] L.-T. Zhao, Y. Wang, S.-Q. Guo, G.-R. Zeng, A novel method based on numerical fitting for oil price trend forecasting, Appl. Energy 220 (2018) 154–163, https://doi.org/10.1016/j.apenergy.2018.03.060.
[25] C. Liang, Z. Xia, X. Lai, L. Wang, Natural gas volatility prediction: fresh evidence from extreme weather and extended GARCH-MIDAS-ES model, Energy Econ. 116 (2022) 106437, https://doi.org/10.1016/j.eneco.2022.106437.
[26] J. Berrisch, F. Ziel, Distributional modeling and forecasting of natural gas prices, J. Forecast. 41 (2022) 1065–1086, https://doi.org/10.1002/for.2853.
[27] W. Liu, C. Wang, Y. Li, Y. Liu, K. Huang, Ensemble forecasting for product futures prices using variational mode decomposition and artificial neural networks, Chaos, Solit. Fractals 146 (2021) 110822, https://doi.org/10.1016/j.chaos.2021.110822.
[28] J. Li, Q. Wu, Y. Tian, L. Fan, Monthly Henry Hub natural gas spot prices forecasting using variational mode decomposition and deep belief network, Energy 227 (2021) 120478, https://doi.org/10.1016/j.energy.2021.120478.
[29] J. Wang, J. Cao, S. Yuan, M. Cheng, Short-term forecasting of natural gas prices by using a novel hybrid method based on a combination of the CEEMDAN-SE-and the PSO-ALS-optimized GRU network, Energy 233 (2021) 121082, https://doi.org/10.1016/j.energy.2021.121082.
[30] Y. Lin, Q. Lu, B. Tan, Y. Yu, Forecasting energy prices using a novel hybrid model with variational mode decomposition, Energy 246 (2022) 123366, https://doi.org/10.1016/j.energy.2022.123366.
[31] R. Li, X. Song, A multi-scale model with feature recognition for the use of energy futures price forecasting, Expert Syst. Appl. 211 (2023) 118622, https://doi.org/10.1016/j.eswa.2022.118622.
[32] X. Li, W. Shang, S. Wang, Text-based crude oil price forecasting: a deep learning approach, Int. J. Forecast. 35 (2019) 1548–1560, https://doi.org/10.1016/j.ijforecast.2018.07.006.
[33] P. Wang, J. Liu, Z. Tao, H. Chen, A novel carbon price combination forecasting approach based on multi-source information fusion and hybrid multi-scale decomposition, Eng. Appl. Artif. Intell. 114 (2022) 105172, https://doi.org/10.1016/j.engappai.2022.105172.
[34] F. Zhang, Y. Xia, Carbon price prediction models based on online news information analytics, Finance Res. Lett. 46 (2022) 102809, https://doi.org/10.1016/j.frl.2022.102809.
[35] J. Li, G. Li, M. Liu, X. Zhu, L. Wei, A novel text-based framework for forecasting agricultural futures using massive online news headlines, Int. J. Forecast. 38 (2022) 35–50, https://doi.org/10.1016/j.ijforecast.2020.02.002.
[36] Y. Fang, W. Wang, P. Wu, Y. Zhao, A sentiment-enhanced hybrid model for crude oil price forecasting, Expert Syst. Appl. 215 (2023) 119329, https://doi.org/10.1016/j.eswa.2022.119329.
[37] A. Gosselin, J. Le Maux, N. Smaili, Readability of accounting disclosures: a comprehensive review and research agenda, Account. Perspect. 20 (2021) 543–581, https://doi.org/10.1111/1911-3838.12275.
[38] G. Carotta, M. Mello, J. Ponce, Monetary policy communication and inflation expectations: new evidence about tone and readability, Latin American Journal of Central Banking 4 (2023) 100088, https://doi.org/10.1016/j.latcb.2023.100088.
[39] J. Wang, F. Ma, E. Bouri, J. Zhong, Volatility of clean energy and natural gas, uncertainty indices, and global economic conditions, Energy Econ. 108 (2022) 105904, https://doi.org/10.1016/j.eneco.2022.105904.
[40] G. Xie, F. Jiang, C. Zhang, A secondary decomposition-ensemble methodology for forecasting natural gas prices using multisource data, Resour. Pol. 85 (2023) 104059, https://doi.org/10.1016/j.resourpol.2023.104059.
[41] X. Wang, Y. Mao, Y. Duan, Y. Guo, A Study on China coal Price forecasting based on CEEMDAN-GWO-CatBoost hybrid forecasting model under Carbon Neutral Target, Front. Environ. Sci. 10 (2022) 1014021, https://doi.org/10.3389/fenvs.2022.1014021.
[42] H. Niu, K. Xu, C. Liu, A decomposition-ensemble model with regrouping method and attention-based gated recurrent unit network for energy price prediction, Energy 231 (2021) 120941, https://doi.org/10.1016/j.energy.2021.120941.
[43] A. Altan, S. Karasu, E. Zio, A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer, Appl. Soft Comput. 100 (2021) 106996, https://doi.org/10.1016/j.asoc.2020.106996.
[44] X. Liu, J. Guo, H. Wang, F. Zhang, Prediction of stock market index based on ISSA-BP neural network, Expert Syst. Appl. 204 (2022) 117604, https://doi.org/10.1016/j.eswa.2022.117604.
[45] X. Ren, S. Liu, X. Yu, X. Dong, A method for state-of-charge estimation of lithium-ion batteries based on PSO-LSTM, Energy 234 (2021) 121236, https://doi.org/10.1016/j.energy.2021.121236.
[46] C. Sekhar, R. Dahiya, Robust framework based on hybrid deep learning approach for short term load forecasting of building electricity demand, Energy 268 (2023) 126660, https://doi.org/10.1016/j.energy.2023.126660.

[47] M.U. Ali, K.D. Kallu, H. Masood, K.A.K. Niazi, M.J. Alvi, U. Ghafoor, A. Zafar, Kernel recursive least square tracker and long-short term memory ensemble based battery health prognostic model, iScience 24 (2021) 103286, https://doi.org/10.1016/j.isci.2021.103286.
[48] Y. Wang, Z. Li, J. Rao, Y. Yang, Z. Dai, Gene based message passing for drug repurposing, iScience 26 (2023) 107663, https://doi.org/10.1016/j.isci.2023.107663.
[49] Y. Duan, X. Yu, A collaboration-based hybrid GWO-SCA optimizer for engineering optimization problems, Expert Syst. Appl. 213 (2023) 119017, https://doi.org/10.1016/j.eswa.2022.119017.
[50] A. Menekşe, H. Camgöz Akdağ, Medical waste disposal planning for healthcare units using spherical fuzzy CRITIC-WASPAS, Appl. Soft Comput. 144 (2023) 110480, https://doi.org/10.1016/j.asoc.2023.110480.
[51] N. Sharkasi, S. Rezakhah, A modified CRITIC with a reference point based on fuzzy logic and hamming distance, Knowl. Base Syst. 255 (2022) 109768, https://doi.org/10.1016/j.knosys.2022.109768.
[52] Q. Feng, X. Sun, J. Hao, J. Li, Predictability dynamics of multifactor-influenced installed capacity: a perspective of country clustering, Energy 214 (2021) 118831, https://doi.org/10.1016/j.energy.2020.118831.
[53] J. Guo, S. Long, W. Luo, Nonlinear effects of climate policy uncertainty and financial speculation on the global prices of oil and gas, Int. Rev. Financ. Anal. 83 (2022) 102286, https://doi.org/10.1016/j.irfa.2022.102286.
[54] J. Hao, Q. Feng, JP. Li, XL. Sun, A bi-level ensemble learning approach to complex time series forecasting: Taking exchange rates as an example, J. Forecast. 42 (2023) 1385–1406.
[55] J. Li, J. Hao, X. Sun, Q. Feng, Forecasting China's sovereign CDS with a decomposition reconstruction strategy, Appl. Soft Comput. 105 (2021) 107291, https://doi.org/10.1016/j.asoc.2021.107291.
[56] X. Sun, J. Hao, J. Li, Multi-objective optimization of crude oil-supply portfolio based on interval prediction data, Ann. Oper. Res. 309 (2022) 611–639, https://doi.org/10.1007/s10479-020-03701-w.
[57] Y.H. Liu, J. Zhu, C. Constantinidis, X. Zhou, Emergence of prefrontal neuron maturation properties by training recurrent neural networks in cognitive tasks, iScience 24 (2021) 103178, https://doi.org/10.1016/j.isci.2021.103178.
[58] L. Tao, J. Byrnes, V. Varshney, Y. Li, Machine learning strategies for the structure-property relationship of copolymers, iScience 25 (2022) 104585, https://doi.org/10.1016/j.isci.2022.104585.
[59] J. Chen, J. Zhang, H. Chen, Y. Zhao, H. Wang, A TDV attention-based BiGRU network for AIS-based vessel trajectory prediction, iScience 26 (2023) 106383, https://doi.org/10.1016/j.isci.2023.106383.
[60] Q. Xiong, W. Wang, M. Wang, C. Zhang, X. Zhang, C. Chen, M. Wang, Prediction of ground-level ozone by SOM-NARX hybrid neural network based on the correlation of predictors, iScience 25 (2022) 105658, https://doi.org/10.1016/j.isci.2022.105658.
[61] H. Xing, J. Zhu, R. Qu, P. Dai, S. Luo, M.A. Iqbal, An ACO for energy-efficient and traffic-aware virtual machine placement in cloud computing, Swarm Evol. Comput. 68 (2022) 101012, https://doi.org/10.1016/j.swevo.2021.101012.
[62] H. Moazen, S. Molaei, L. Farzinvash, M. Sabaei, PSO-ELPM: PSO with elite learning, enhanced parameter updating, and exponential mutation operator, Inf. Sci. 628 (2023) 70–91, https://doi.org/10.1016/j.ins.2023.01.103.
[63] K. Zhang, X. Yang, T. Wang, J. Thé, Z. Tan, H. Yu, Multi-step carbon price forecasting using a hybrid model based on multivariate decomposition strategy and deep learning algorithms, J. Clean. Prod. 405 (2023) 136959, https://doi.org/10.1016/j.jclepro.2023.136959.
[64] H. Wang, Z. Tan, A. Zhang, L. Pu, J. Zhang, Z. Zhang, Carbon market price prediction based on sequence decomposition-reconstruction-dimensionality reduction and improved deep learning model, J. Clean. Prod. 425 (2023) 139063, https://doi.org/10.1016/j.jclepro.2023.139063.