Oil Demand Forecasting from The Perspective of Alternatives

Yi Huang [1, 3], Jing Yu [1], Wenhua Wu [2, 3, *]

[1] School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

[2] School of International Education, Nanjing University of Science and Technology, Nanjing, China

[3] Saudi Studies Center, Nanjing University of Science and Technology, Nanjing, China

* Corresponding Author Email: gswwh@njust.edu.cn

Abstract. In recent years, events such as oil sanctions, the Russia-Ukraine conflict, and the COVID-19 pandemic have led to global instability and extreme fluctuations in international oil prices. This has slowed the growth of oil consumption, while the development of new energy technologies and the introduction of alternative energy sources are gradually transforming China's energy demand structure. Accurately forecasting oil demand has thus become crucial for national energy security and economic development. This paper predicts China's future oil demand by examining commonly used energy sources and focusing on oil substitutes. Sixteen relevant factors were selected, and machine learning and neural network methods were applied to assess their impact on China's oil consumption demand. Experimental results demonstrate that considering oil substitutes in variable selection is feasible for predicting China's oil demand. The combined LSTM and attention mechanism approach achieved the highest prediction accuracy, stability, and overall performance. The findings of this empirical study offer valuable insights for future research utilizing deep learning techniques to forecast oil demand.

Keywords: Oil demand forecasting; Alternatives; Deep learning; Attention mechanism.

1. Introduction

In the past decade, China has become the focus of international attention due to sustained economic growth and increasing involvement in global financial markets. This persistent economic growth has primarily been driven by high energy consumption in the industrial and transportation sectors, while the expanding scale of the economy reflects the growing demand for energy. Among various energy sources, oil plays a particularly significant role. On one hand, China's rapid infrastructure development, industrialization, increased mobility, and urbanization have fueled this trend; on the other hand, the country's oil production has failed to keep pace with consumption. As a result, China's oil consumption exceeded its production for the first time in 1993, and today China is the world's largest net oil importer and the second largest oil consumer.

In recent years, the turbulent international environment—including oil sanctions and the Russia–Ukraine war has had a significant impact on global energy demand and the overall economy. Existing studies on oil sanctions consistently suggest that such sanctions affect oil demand and production [1]. Additionally, the COVID-19 pandemic, as an exogenous shock, has significantly impacted energy demand. Multiple studies [2][3][4] have shown that the pandemic, which caused a sharp rise in global mortality, led to reduced oil demand and increased uncertainty in oil supply and demand. Available data indicate that in 2020, global oil demand decreased by 8.76 million barrels per day compared to 2019.

These signs highlight the increasing strategic importance of oil in ensuring national energy security and economic development. As a fundamental strategic industry for the national economy and public welfare, the oil industry drives economic growth and ensures the stability of energy security and industrial chains. However, with the acceleration of the global energy transition and the development of new energy technologies and alternative energy sources, China's energy demand structure is gradually changing. According to statistics, in the first half of 2023, China's investment in new energy projects reached 5.2 trillion RMB. Alternative energy sources such as natural gas, nuclear energy, and electricity can partially substitute oil and reduce dependence on it. Therefore, to ensure an uninterrupted energy supply, it is necessary to analyze the future trend of oil demand in China from the perspective of substitutes.

Scholars at home and abroad have proposed various perspectives on forecasting oil demand. First, regarding the mechanism of oil demand, Chai et al. [5] analyzed the role of economic growth. Liu [6] studied the effect of motor vehicle ownership. Zhu [7] incorporated exogenous variables such as oil sanctions and the pandemic, and Huang et al. [8] examined the trend in oil demand from the perspective of integrated tourism development.

Second, in terms of data-driven models—both traditional econometric methods and artificial intelligence-based machine learning methods have been applied: (1) In traditional econometrics, Xuan et al. [9] used the least squares method to forecast oil consumption trends. Narayan and Smyth [10] developed a crude oil demand model for Middle Eastern countries using panel cointegration analysis. Dagoumas et al. [11] built a global crude oil demand model using the autoregressive moving average (ARMA) approach. Chen [12] applied the GM (1,1) model to predict oil demand. Xi et al. [13] extended this by using a dynamic GM (1,1) model. Fatima et al. [14] utilized a structural time series model to forecast China's oil demand. (2) In machine learning, Li et al. [15] applied a BP neural network to predict oil demand. Sun et al. [16] used generalized regression neural networks (GRNN) and backpropagation neural networks (BPNN). Tang and Tang [17] proposed a nonlinear prediction method using a chaotic particle swarm optimization support vector machine (CPSO-SVM). Behrang et al. [18] empirically demonstrated that the gravitational search algorithm (GSA) outperformed the particle swarm optimization (PSO) algorithm in estimating oil demand in Iran. Li et al. [19] proposed a combined model based on optimized AI algorithms to forecast China's oil consumption. Fattah and Aramco [20] used the GANNATS model to predict oil demand in Saudi Arabia and China.

From the perspective of microeconomics and general equilibrium theory, the demand for a good is influenced by the demand, supply, and prices of its substitutes [21]. Specifically, the mechanism of substitution can be explained by consumer choice behavior among different goods: when the price of the original good increases, consumers may switch to lower-priced or more cost-effective substitutes, reducing demand for the original good, and vice versa. Furthermore, the supply and demand of substitutes also impact the demand for the original good. For example, an increase in substitute supply, while demand remains constant, will lower substitute prices and expand their market share, thus reducing demand for the original product, and vice versa. However, existing oil demand forecasting research often focuses on single factors such as economy, industry, or price, while relatively little attention has been given to predictions based on the perspective of substitutes.

Based on this, this paper proposes to examine China's oil demand from the perspective of substitutes. By integrating deep learning methods with substitute factors, we construct a prediction model incorporating an attention mechanism and compare it with various machine learning models to comprehensively explore the key factors affecting oil demand. The findings aim to provide more valuable references for government agencies and oil enterprises in decision-making and contribute to the rational planning of China's future oil demand.

2. Methodology

2.1. Selection of Influencing Factors

The model's input variables consist of factors influencing China's oil demand, which can be broadly categorized into four dimensions: price factors, economic growth, population size, and supply-demand dynamics.

(1) Price plays a crucial role in shaping demand; thus, oil prices represent the primary determinant of oil demand. Additionally, prices of alternative energy sources, such as natural gas and coal, also significantly affect oil demand. (2) Economic development relies heavily on oil as a driving force. Consequently, oil demand is strongly influenced by the level of economic growth, as the advancement of the oil industry itself depends on the broader economy. Freight and passenger volumes, as key

indicators of economic activity, directly reflect the transportation sector's oil consumption. (3) Population size directly impacts total oil consumption and shapes per capita oil resource utilization patterns, making it an indispensable factor influencing oil demand. (4) The supply and consumption of other energy sources also affect oil demand by providing alternative options. Production and consumption of natural gas, coal, nuclear power, and hydropower directly influence the scale of oil demand, making supply-demand dynamics a critical consideration.

Drawing on existing literature in oil demand forecasting and guided by principles of comprehensiveness, comparability, and data availability, this study analyzes China's oil consumption from 2000 to 2023. The aforementioned variables are selected to predict oil demand, considering price factors, economic growth, population size, and supply-demand dynamics.

2.2. Deep Learning Forecasting Methods

(1) Recurrent Neural Networks

Recurrent Neural Networks (RNN) encompass two prominent variants: Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU). Each has its advantages and drawbacks: LSTM effectively addresses the long-term dependency problem inherent in vanilla RNNs, while GRU requires fewer parameters, reducing the risk of overfitting.

Therefore, this study integrates these two architectures to construct a hybrid recurrent neural network model for oil demand forecasting. This system extends traditional model structures and empirically validates its predictive performance for composite indicators involving oil substitute factors.

The LSTM model is governed by the following equations:

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

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{3}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{4}$$

$$h_t = o_t \odot tanh\,(C_t) \tag{5}$$

Here, x_t denotes the input vector at time, h_{t-1} is the output from the previous time step, σ is the sigmoid activation function, i_t the input gate, f_t the forget gate, C_t the updated cell state, o_t the output gate, h_t and the hidden state.

The GRU model is defined by:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{6}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{7}$$

$$\tilde{h}_t = tanh(W_c x_t + U_c(r_t \odot h_{t-1}) + b_c) \tag{8}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{9}$$

Here, x_t is the input at time, W are weight matrices, z_t the update gate, r_t the reset gate, $\tilde{h}_t$ the candidate hidden state, and h_t the hidden state.

(2) Attention Mechanism

The attention mechanism, inspired by studies of human vision, fundamentally shifts focus from processing all available information to selectively attending to key elements. Due to limited cognitive resources, humans selectively concentrate on a subset of information while disregarding other visible data. Analogously, the attention mechanism allocates limited computational resources to salient inputs, enabling efficient extraction of relevant information. This mechanism helps models to better identify and weigh critical components within input data, thereby enhancing overall performance. The mechanism operates by computing scores through Query (Q), Key (K), and Value (V) matrices to generate weighted outputs.

This study develops a hybrid model combining LSTM and attention mechanisms for oil demand forecasting, extending previous model architectures and empirically evaluating its predictive efficacy.

2.3. Data Preprocessing Methods

Raw data cannot be directly employed for model development; preprocessing is essential to enhance data mining quality. This study applies data normalization as the preprocessing technique. Normalization is performed using the following formula:

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{10}$$

Where x_i is the original value, $\min(x)$ and $\max(x)$ denote the minimum and maximum of feature x, respectively, and x_i' is the normalized value.

2.4. Evaluation Metrics for Forecasting Methods

The predictive accuracy of the models is assessed using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics validate the feasibility and effectiveness of the forecasting approaches under different model configurations.

Let n be the sample size, y_i the actual oil consumption in year i, $\hat{y}_i$ the forecasted oil demand for year i, and $\hat{y}$ the average actual oil consumption.

MAE is calculated as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{11}$$

MSE is calculated as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{12}$$

RMSE is calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{13}$$

MAPE is calculated as:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{14}$$

3. Empirical Analysis

3.1. Data Sources and Sample Data

This paper obtains the data related to oil demand forecasting from the CASMER database and the 2024 World Energy Statistical Yearbook.

From the perspective of oil energy substitutes, this study comprehensively considers 16 influencing factors, specifically as follows：China's Gross Domestic Product (100 million yuan), China's year-end population (10,000 persons), Total freight volume (10,000 tons), Total passenger volume (10,000 persons), China's oil production (thousand barrels per day), Dubai crude oil price per barrel (USD), China's natural gas production (exajoules), China's natural gas consumption (exajoules), Natural gas price (JKM, USD per MMBtu), China's coal production (million tons), China's coal consumption (exajoules), Coal price in South China (USD per ton), China's nuclear power production (TWh), China's nuclear energy consumption (exajoules), China's hydropower production (TWh), China's hydropower consumption (exajoules).

These influencing factors are used to forecast China's oil demand from four aspects: price factors, economic growth, population size, and supply-demand dynamics. Table 1 presents the descriptive statistics of the above 17 variables from 2000 to 2023.

Table 1. Descriptive statistics of variables

	N	Mean	Std	Min	25%	50%	75%	Max
Oil demand	24	10056.60	3625.95	4655.44	7195.86	9845.19	13162.27	16576.54
GDP	24	560542.17	379689.73	100280.14	211408.58	513260.07	853847.25	1260582.00
POP	24	135235.67	4901.99	126743.00	131275.00	135419.00	140143.50	141260.00
Freight vol	24	3432822.88	1440071.91	1358681.71	1993311.50	3897930.56	4716683.74	5570636.00
Passenger vol	24	1966954.57	824343.39	558737.63	1574153.25	1847819.06	2149183.97	3804034.90
Oil prod	24	3845.82	311.61	3256.84	3693.37	3847.09	4085.54	4308.84
Oil price	24	63.70	27.83	22.78	42.07	62.59	85.14	109.06
Natural gas prod	24	4.05	2.30	0.99	2.04	3.92	5.48	8.43
Natural gas cons	24	5.99	4.52	0.89	1.98	5.15	9.07	14.57
Natural gas price	24	9.53	4.04	4.27	6.49	9.21	12.48	16.98
Coal prod	24	3191.01	977.56	1384.18	2518.58	3476.00	3880.83	4710.00
Coal cons	24	68.89	19.17	29.56	59.54	78.46	82.18	91.94
Coal price	24	78.20	39.76	27.52	40.87	77.73	106.25	177.78
Nuclear prod	24	161.22	141.86	16.74	54.40	92.76	259.83	434.72
Nuclear cons	24	1.49	1.27	0.17	0.54	0.88	2.39	3.90
Hydropower prod	24	803.22	394.76	222.41	426.09	787.09	1173.52	1321.71
Hydropower cons	24	7.82	3.62	2.37	4.37	7.81	11.23	12.50

3.2. Characteristics of Factors Affecting Oil Consumption

This study uses the ANOVA (Analysis of Variance) method to evaluate the importance of the features. The ANOVA method can effectively distinguish the contribution of different features to the model's predictive capability. Through this method, the highest-scoring subset of features was selected. These high-scoring features not only significantly enhance the explanatory power of the model but also help prevent the issue of overfitting. During the feature selection process, the top 10 optimal features were chosen based on the ANOVA scores. The importance of these features has been validated through the ANOVA method, and the specific results are shown in Table 2.

According to the results of feature importance scores, natural gas production, gross domestic product, population size, natural gas consumption, and freight volume are the most influential factors.

Hydropower, as a representative of clean energy, is environmentally friendly and holds great potential, ranking next in importance. In addition, other substitute energy sources for oil also demonstrate relatively high importance scores, indicating that they have a significant impact on the demand for oil energy.

3.3. Evaluation and Analysis of Forecasting Results

(1) Testing Environment and Model Description

This study was implemented on a Windows 10 using Python 3.6, with TensorFlow 2.0 serving as the deep learning framework. Model performance was evaluated based on four key metrics outlined in Section 2.4.

A total of 16 variables—such as natural gas production, natural gas consumption, and nuclear energy consumption—were selected as input features, with oil consumption serving as the output target. To enhance prediction accuracy, the dataset was divided into training and testing subsets, with 70% allocated for training and the remaining 30% for testing.

Several models were constructed for comparison. Linear Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB) were trained using default parameters. The Fully Connected Neural Network (FC) was composed of four layers: the first three were dense layers, followed by a final output layer. The Long Short-Term Memory (LSTM) network consisted of three stacked LSTM layers and one fully connected layer for output. A hybrid recurrent network combined LSTM and Gated Recurrent Unit (GRU) layers in the first three layers, with a fully connected output layer. Additionally, an LSTM model with an Attention mechanism was developed, comprising an LSTM layer, an Attention layer, a Flatten layer, and a dense output layer. This model architecture represents an extension of prior oil demand forecasting studies by incorporating a diverse set of techniques, ranging from traditional linear models to advanced machine learning and deep learning frameworks—eight in total—offering a more comprehensive comparison. Given the relatively small sample size, the number of neurons in the hidden layers was tuned to 8 or 16 to mitigate overfitting. Each model employed 16 input neurons (corresponding to the 16 variables) and one output neuron. The MSE was chosen as the loss function, and the AdamW optimizer was used with default learning rate and convergence settings. Each model was trained for 100 epochs.

(2) Forecasting Results

The forecasting task was carried out across different time spans depending on the model type: machine learning algorithms generated predictions for the period 2016–2023, while deep learning models covered 2018–2023. After training, a strong alignment was observed between predicted and actual values, suggesting good model fit. A visual comparison is provided in Figure 1.

To assess predictive performance, this study calculated MAE, MSE, RMSE, and MAPE for all eight modeling approaches. Summary statistics are presented in Table 3.

These metrics offer comprehensive insights into forecast accuracy. MSE reflects the average squared difference between predicted and actual values, capturing both bias and variance, and thus serves as a robust indicator of model reliability. The MSE values obtained for the eight models were 0.110, 0.212, 0.069, 0.053, 0.028, 0.033, 0.027, and 0.005, respectively. Among them, the LSTM model incorporating Attention achieved the lowest MSE, demonstrating superior predictive precision. Similarly, the MAE and RMSE values provide information on the absolute and root-average errors, respectively. Lower values indicate stronger agreement between predictions and ground truth. Of all models, the linear regression model exhibited the highest RMSE, whereas the LSTM with Attention mechanism achieved the lowest, confirming the latter's advantage in reducing prediction error. Notably, the MAPE associated with the LSTM-Attention model was also significantly lower than that

Table 2. Feature importance score

Rank	Feature	Score
1	Natural gas prod	946.62
2	GDP	933.39
3	POP	660.08
4	Natural gas cons	496.88
5	Freight vol	468.72
6	Hydropower prod	430.86
7	Hydropower cons	356.76
8	Nuclear cons	195.41
9	Nuclear prod	185.77
10	Coal prod	111.78

of the others, reinforcing its robustness and generalization ability. Taken together, these results affirm that the LSTM with Attention mechanism outperforms the other seven models in all four-performance metrics. It demonstrates higher accuracy, better generalization, and is therefore the most suitable approach for forecasting oil demand among those evaluated.

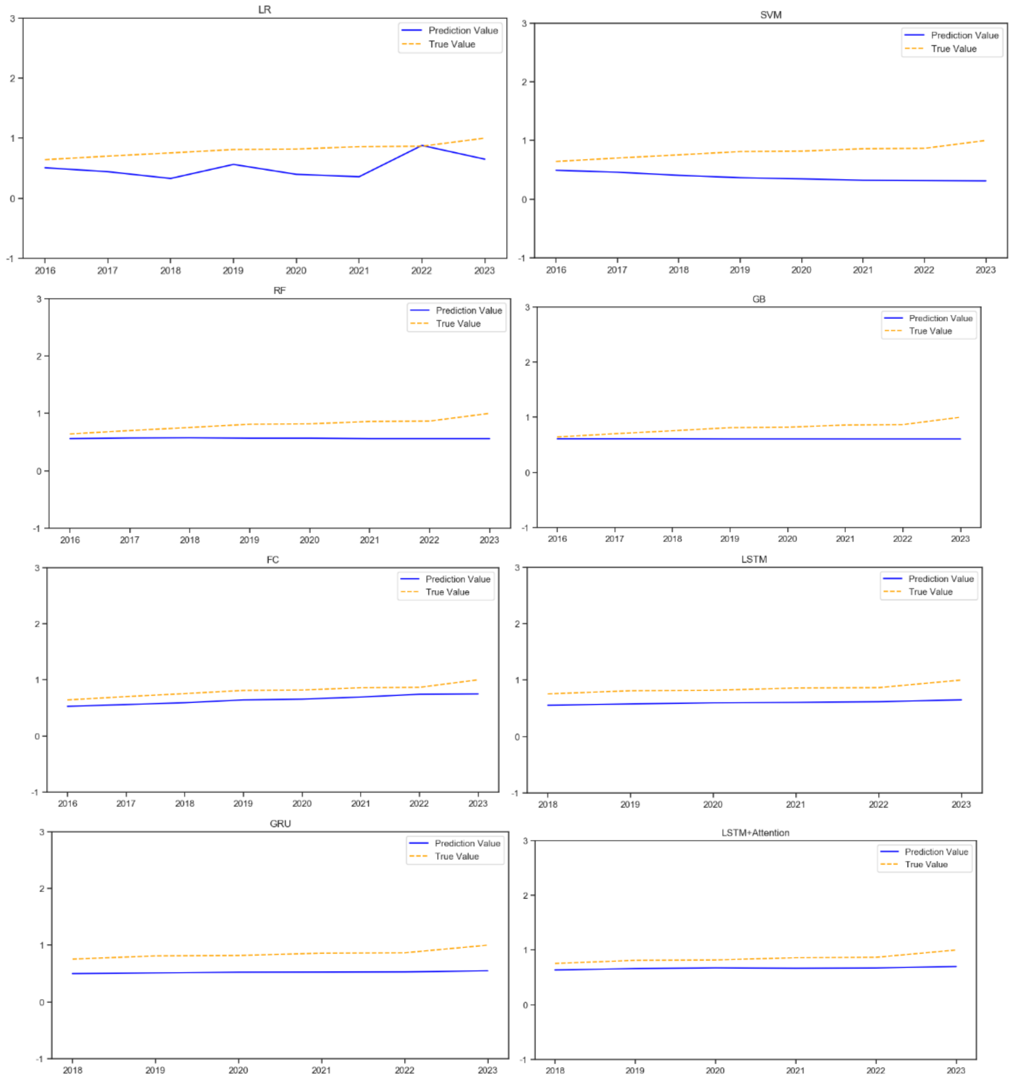

Fig. 1 Different model predictions

Table 3. Model error value

Model	MAE	MSE	RMSE	MAPE
LR	0.294	0.110	0.332	0.702
SVM	0.430	0.212	0.460	1.241
RF	0.239	0.069	0.262	0.423
GB	0.204	0.053	0.230	0.339
FC	0.161	0.028	0.166	0.250
LSTM	0.174	0.033	0.182	0.255
GRU	0.156	0.027	0.163	0.223
LSTM+Attention	0.066	0.005	0.074	0.072

4. Conclusion

Oil plays a vital role in ensuring national energy security and driving economic development. As a result, increasing attention has been paid globally to empirical research on oil demand. A wide array of models and estimation techniques has been developed for demand forecasting. This study adopts a novel perspective by predicting oil consumption based on the influence of oil substitutes. The ANOVA method was employed to evaluate the relative importance of various influencing factors, enabling the selection of variables with high relevance to China's oil demand. Utilizing empirical data from 2000 to 2023, this research constructed forecasting models using four machine learning and four deep learning approaches. Among these, a hybrid deep learning model combining LSTM with an attention mechanism was found to provide the best fit for the data, offering an optimal network structure and producing significantly lower forecasting errors than traditional models. Given the current era of profound global transformation, it is imperative to consider the impact of alternative energy sources when forecasting oil demand. Such a multidimensional approach aligns with the demands of big data analytics. The empirical findings of this study offer valuable insights for future applications of deep learning in energy demand forecasting, particularly in the selection of relevant indicators and the design of effective model architectures.

Acknowledgements

The authors gratefully acknowledge the financial support from Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX24_0729).

References

[1] Ou S, He X, Ji W, et al. Machine learning model to project the impact of COVID-19 on US motor gasoline demand[J]. Nature Energy, 2020, 5(9): 666-673.

[2] Wang Q, Zhang F. What does the China's economic recovery after COVID-19 pandemic mean for the economic growth and energy consumption of other countries?[J]. Journal of Cleaner Production, 2021, 295: 126265.

[3] Zhang X, Pellegrino F, Shen J, et al. A preliminary simulation study about the impact of COVID-19 crisis on energy demand of a building mix at a district in Sweden[J]. Applied Energy, 2020, 280: 115954.

[4] Nakhli S R, Rafat M, Dastjerdi R B, et al. Oil sanctions and their transmission channels in the Iranian economy: A DSGE model[J]. Resources Policy, 2021, 70: 101963.

[5] Jian C, Ting L, Youhong Z. Analysis of oil consumption under different regional industrialization—An empirical study based on Path-STR model[J]. Chinese Journal of Management Science, 2017, 25(11): 47-57.

[6] Liu G W, Yan Q, Yang J B. World oil demand based on S-Curve Model of the transport sector[J]. Resources Science, 2018, 40(3): 547-557.

[7] Zhu H. Oil Demand Forecasting in Importing and Exporting Countries: AI-Based Analysis of Endogenous and Exogenous Factors[J]. Sustainability, 2023, 15(18): 13592.

[8] Huang Y, Li S, Wang R, et al. Forecasting oil demand with the development of comprehensive tourism[J]. Chemistry and Technology of Fuels and Oils, 2021, 57(2): 299-310.

[9] Weiqi X, Xue L, Yuejun J. Analysis of future trend prediction of China's petroleum consumption[J]. Operations Research and Management Science, 2009, 18(3): 1-7.

[10] Narayan P K, Smyth R. A panel cointegration analysis of the demand for oil in the Middle East[J]. Energy Policy, 2007, 35(12): 6258-6265.

[11] Dagoumas A, Perifanis T, Polemis M. An econometric analysis of the Saudi Arabia's crude oil strategy[J]. Resources Policy, 2018, 59: 265-273.

[12] Haitao C. GM (1, 1) model predicts China's oil demand[J]. Ecological Economy, 2014, 30(2): 127-131.

[13] Xi W F, Lu Y J, Chen Q, et al. Forecast of China’s Oil Demand Based on the Dynamic Gm (1,1) Model[J]. Journal of Yunnan Normal University,2015,35(04):41-45.

[14] Fatima T, Xia E, Ahad M. Oil demand forecasting for China: fresh evidence from structural time series analysis[J]. Environment, development and sustainability, 2019, 21: 1205-1224.

[15] Li H, Li H, Wang H. Research on Prediction of China's Oil Demand Based on the BP Neural Network Model[J]. Journal of China University of Petroleum (Edition of Social Science), 2015: 8-11.

[16] SUN M, DENG X, DONG T. Application of Neural Networks and Kalman Algorithm in Oil Demand Prediction[J]. Journal of Beijing Institute of Technology (Social Sciences Edition), 2013 (6): 10-15.

[17] Tang L F. CPSO-SVM Based Petroleum Demand Prediction[J]. Applied Mechanics and Materials, 2013, 273: 91-96.

[18] Behrang M A, Assareh E, Ghalambaz M, et al. Forecasting future oil demand in Iran using GSA (Gravitational Search Algorithm) [J]. Energy, 2011, 36(9): 5649-5654.

[19] Li J, Wang R, Wang J, et al. Analysis and forecasting of the oil consumption in China based on combination models optimized by artificial intelligence algorithms[J]. Energy, 2018, 144: 243-264.

[20] Al-Fattah S M, Aramco S. Application of the artificial intelligence GANNATS model in forecasting crude oil demand for Saudi Arabia and China[J]. Journal of Petroleum Science and Engineering, 2021, 200: 108368.

[21] Pindyck, R.S., Rubinfeld, D.L. Microeconomics. 9th ed [M]. Pearson, 2020.