ZIBELINE INTERNATIONAL™ PUBLISHING

ISSN: 2521-0858 (Print)
ISSN: 2521-0866 (Online)
CODEN: SHJCAS

Science Heritage Journal (GWS)

DOI: http://doi.org/10.26480/gws.01.2024.13.21

RESEARCH ARTICLE

PREDICTION OF CARBON EMISSIONS FROM TRANSPORTATION IN CHINA BASED ON THE ARIMA-LSTM-BP COMBINED MODEL

Sheng Kai*, Ye Shanli

School of Science, Zhejiang University of Science and Technology, Hangzhou 310023, China.
**Corresponding Author Email: 18858238294@163.com*

This is an open access article distributed under the Creative Commons Attribution License CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ARTICLE DETAILS

Article History:

Received 18 November 2023
Revised 20 December 2023
Accepted 28 January 2024
Available online 30 January 2024

ABSTRACT

Transportation is not only a significant force in promoting economic and social development but also one of the primary industries that consume energy and emit greenhouse gas emissions. In order to achieve China's overall goal of reaching the carbon peak by 2030, this paper selects six influencing factors, such as population, GDP and urbanization rate, and proposes a combined prediction model based on ARIMA-LSTM-BP, which predicts transportation carbon emissions in China from 2022 to 2050 under three scenarios of low carbon, benchmark and high carbon. The results show that the peak emissions of transportation in low-carbon, benchmark and high-carbon scenarios are 1624.7732 million tons, 1478.1694 million tons and 1367.5417 million tons, respectively, reaching the peak in 2031, 2034 and 2039. It can be seen that in China, the transportation industry alone cannot achieve the goal of reaching the peak by 2030, and more measures need to be taken to achieve the carbon peak of the transportation industry as soon as possible.

KEYWORDS

Carbon emission prediction; combination model; transportation industry; scenario analysis method

1. INTRODUCTION

The document of the 20th National Congress of the CPC Central Committee proposed that "actively and steadily promote carbon to peak and carbon neutrality" and achieve the goal of "double carbon" is not only the inevitable need for China to seek development at a new level of harmonious coexistence between man and nature but also an essential part of promoting Chinese style modernization. As one of the three significant areas of energy consumption and carbon dioxide emissions, transportation is an important source of carbon dioxide emissions (Yang, 2023). The transportation industry is developing rapidly and will become the key industry to achieve carbon emission reduction goals in the future. However, with the popularity and rapid increase of the number of transportation tools, the transportation industry's contribution to environmental pollution is also increasing, which will not only affect human health and the ecological environment but also hurt the development of the transportation industry. Therefore, it is of great practical significance to predict the peak carbon emissions of the transportation industry and formulate appropriate carbon emission reduction policies.

Scholars at home and abroad have carried out a series of research studies on traffic carbon emissions, among which research on the influencing factors and prediction of traffic carbon emissions is a hot spot in this field.

In the selection of influencing factors of traffic carbon emissions, scholars usually use logarithmic mean decomposition analysis (LMDI), Kaya identity, IPAT model, and extensible stochastic environmental impact assessment model (STIRPAT) to select influencing factors. Solaymani used the LMDI decomposition method to conduct in-depth research on seven major carbon emission countries and analyzed the driving factors of carbon emissions in the transportation sector (Solaymani, 2019). Jia et al. decomposed the factors of transportation carbon emissions in Hebei Province through the improved Kaya identity and calculated the correlation between each decomposition index and transportation carbon emissions by using the grey correlation model (Jia, 2020). Commoner proposed the classic IPAT model and used it to explain the relationship between population, wealth, and environmental pressure (Commoner, 1990). Because the IPAT model only explores the relationship between population, wealth and environment, it has certain limitations, so Dietz et al. proposed the STIRPAT model, which has also become an essential means for scholars to study the influencing factors of traffic carbon emissions in recent years (Dietz, 1997). Andrés et al. took energy intensity, per capita freight volume and other factors as influencing factors of traffic carbon emissions based on the STIRPAT model (Andrés et al., 2018). Liu et al. used the STIRPAT and GTWR model to reveal the impact of population size, urbanization rate, the number of private cars and other factors on traffic carbon emissions from the perspective of time and space (Liu et al., 2012). Zhang et al. pointed out through the expanded STIRPAT model that the population size and per capita GDP will promote the rise of traffic carbon emissions, while the energy intensity has a restraining effect (Zhang et al., 2020). Yan et al. selected urbanization rate, subway passenger volume and energy consumption per unit GDP as the main influencing factors of traffic carbon emissions by building the STIRPAT model (Yan et al., 2020). In the existing research, scholars usually choose factors such as population size, gross national product, energy intensity, and energy structure to study (Dong et al., 2020; Liu et al., 2023; Han et al., 2022; Wang et al., 2021). It can be found that the STIRPAT model has obvious advantages in selecting factors. Therefore, this paper uses the STIRPAT model to select the influencing factors of transportation carbon emissions.

In the field of transportation carbon emission prediction, scholars at home and abroad have conducted much research. The prediction models mainly include long-term energy alternative planning system (LEAP), autoregressive moving average model (ARIMA), STIRPAT model, long and short-term memory model (LSTM), BP neural network, etc. Fang et al. predicted the carbon emissions in the field of transportation in Hunan

Quick Response Code	Access this article online	
	Website: www.jscienceheritage.com	DOI: 10.26480/gws.01.2024.13.21

Cite The Article: Sheng Kai, Ye Shanli (2024). Prediction of Carbon Emissions from Transportation in China Based on The Arima-Lstm-Bp Combined Model. *Science Heritage Journal, 8(1): 13-21.*

Province from 2022 to 2035 through the leap model (Fang et al., 2023). The results showed that carbon emissions were expected to peak in 2033 and 2029 under the low-carbon and enhanced low-carbon scenarios. Hassouna et al. used the ARIMA model to predict China's aviation carbon emissions (Hassouna et al., 2020). Bian et al. estimated the future carbon emissions of the Qinghai Province's transportation system Based on the STIRPAT model simulation results. They believed improving the technical level was the key to reducing carbon emissions from transportation in Qinghai Province (Bian et al., 2019). In recent years, scholars have gradually applied machine learning methods to predict carbon emissions from transportation. Ağbulut used an artificial neural network model (ANN), support vector machine (SVM), and deep learning (DL) to predict the carbon emissions of transportation in Turkey. The research showed that the three methods had achieved ideal results (Ağbulut, 2022). Huang et al. used LSTM to predict China's carbon emissions (Huang et al., 2019). Pan et al. established a BP neural network model and predicted CO2 emissions in Gansu Province from 2021 to 2030, with a prediction error of 2 × 10-4, with high accuracy (Pan et al., 2023). Zhao et al. combined the lasso regression model with the BP neural network to predict the carbon emissions of Henan Province from 2021 to 2035 (Zhao et al., 2022). Hu et al. predicted China's carbon emission intensity based on the ARIMA-BP model, and the results showed that the prediction effect was excellent (Hu et al., 2022). Shao et al. constructed the grey neural network combination model of the grey prediction model BP neural network model combination. They verified that the combination model could improve the model's prediction accuracy (Shao et al., 2016).

There has been much research on predicting carbon emissions from transportation, and the research and improvement methods have their characteristics. In this paper, the ARIMA model is selected to better fit the linear characteristics in the prediction time series, LSTM has obvious advantages for long-term prediction, the BP neural network has self-adaptive solid ability and can deal with nonlinear characteristics well (Zhang et al., 2022).

Now, ARIMA, LSTM and BP models are combined, and a weighted combination model based on ARIMA-LSTM-BP is proposed. First, the ARIMA model is used to fit the linear characteristics in the sequence. Then, the LSTM neural network is used to correct the ARIMA model prediction residual to fit the nonlinear characteristics in the sequence, and then combined with the BP neural network prediction results, the error reciprocal method is used to determine the weight for combined prediction to reduce the error. Finally, combined with the scenario analysis method, this paper predicts the changing trend of China's transportation carbon emissions from 2022 to 2050 under the three baseline scenarios: low-carbon and high-carbon.

2. Data and Methods

2.1 Carbon Emission Measurement Of China's Transportation Industry

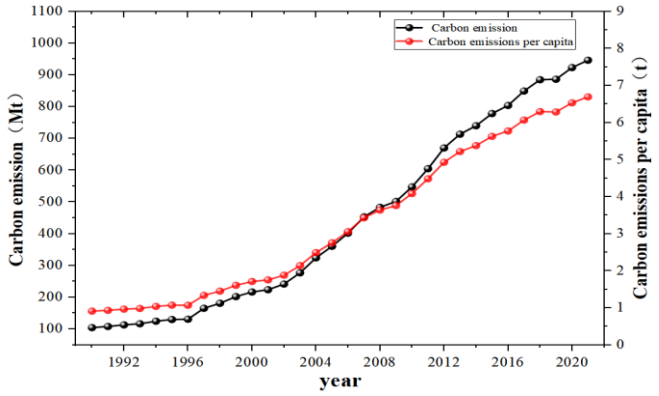
Regarding measuring carbon dioxide emissions from transportation, the IPCC 2006 guidelines for national greenhouse gas inventories provide measurement methods based on two perspectives. The first is the "top-down" method, which calculates the carbon emissions of the transportation industry according to the conversion coefficient of energy consumption and vehicles' energy carbon emission coefficient. The second is the "bottom-up" method, which calculates the energy consumption of the transportation industry according to the data of different types of vehicles and the fuel consumption per unit mileage to calculate the carbon dioxide emissions (Fang et al., 2017). This study simulates the future trend of carbon emissions in the transportation industry through prediction, so the "top-down" method is adopted for calculation. The calculation formula is shown in formula (1).

$$CO_2 = \sum_i C_i = \sum_i E_i \times F_i = \sum_i \left(E_i \times NCV_i \times CC_i \times COF_i \times \frac{44}{12} \right) \quad (1)$$

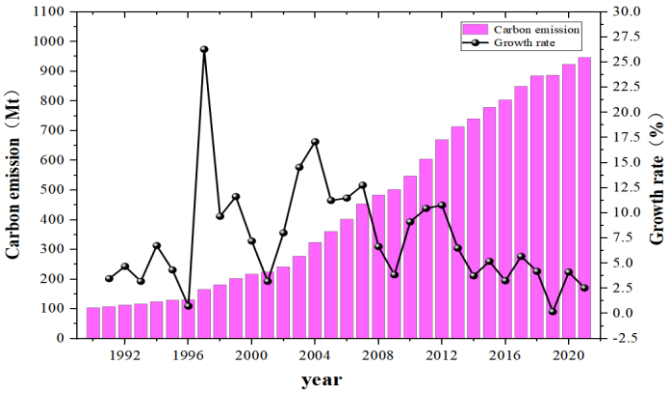
Where C is the carbon emission of the transportation industry; i refers to fuel type (such as gasoline, diesel, fuel oil, etc.); C_i is the amount of carbon dioxide produced by the ith energy consumption; E_i is the consumption of the ith energy; F_i is the carbon emission coefficient of the ith energy; Ncv_i is the average low calorific value of the ith energy source; CC_i is the carbon content per unit calorific value of the ith energy; COF_I is the carbon oxidation rate of the ith energy; 44/12 is the conversion coefficient between carbon and carbon dioxide.

At present, the energy consumption data of the transportation industry are not counted separately in the China Energy Statistical Yearbook and the China Statistical Yearbook, but are combined with the transportation, storage, and post and telecommunications industries. Because the latter two account for a relatively small proportion, and most of the energy consumption of the storage and post and Telecommunications industries is used for transportation, the consolidated data can also reflect the level of energy consumption in transportation to a large extent, so this study uses this consolidated data to measure the total carbon emissions generated by energy consumption in transportation. The carbon emission factors of different energy sources used in the accounting are determined by the correlation coefficients published by the IPCC (United Nations Intergovernmental Panel on Climate Change) and the National Development and Reform Commission. The calculation results of China's transportation carbon emissions and per capita transportation emissions from 1990 to 2021 are shown in Figure 1.

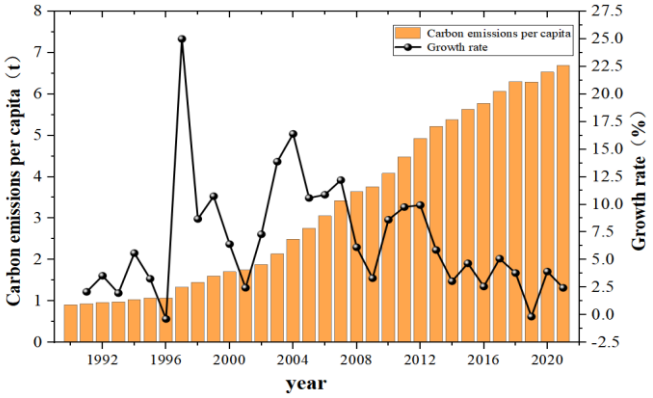
The results show that from 1990 to 2021, China's total carbon emissions from transportation showed an upward trend. By 2021, the carbon emissions from transportation were 946.667 million tons, with an average annual growth rate of 2.87%. At the same time, China's per capita transportation carbon emissions are also rising yearly. The annual increase in total and per capita carbon emissions shows that the transportation industry has a high potential for low-carbon emission reduction. Therefore, predicting transportation carbon emissions is significant to realizing China's dual carbon goal.



(a) Carbon emissions and per capita carbon emissions



(b) Carbon emissions



(c) Per capita carbon emissions

Figure 1: Emissions from China's transportation industry and per capita carbon emissions from transportation.

2.2 Selection Of Influencing Factors

The IPAT model first proposed comprehensively reflects the relationship among population, per capita wealth, technology and environment. It is a widely recognized model, and its general form is:

$$I = P \times A \times T \quad (2)$$

Where I refers to environmental pollution; P is the population; A is GDP per capita; T indicates technical status.

STIRPAT model is proposed based on the IPAT model, which can better explore the relationship between various influencing factors and carbon emissions. Its standard form is (Jiao et al., 2012):

$$I = aP^b A^c T^d e \quad (3)$$

Where, I represents environmental factors; a represents model coefficient; P、A and T respectively represent the population, financial status and scientific development technology; b、c and d respectively represent the changes of these reference values; e is random error.

Gu found that clean energy, transportation energy intensity, freight turnover, passenger turnover, urbanization level, per capita GDP and other factors significantly impact the transportation industry's carbon emissions (Gu, 2023). Zuo et al. summarized the existing research and found that many complex factors affect transportation's carbon emissions, which can be summarized into five aspects: transportation technology and management, economic development, residents' travel behavior, urban morphology and environmental factors (Zuo et al., 2023). Liu et al. believed that the per capita level of economic activity was crucial in improving carbon emissions in the transportation industry (Liu et al., 2015). Yang found that many variables, including energy intensity, transportation efficiency, travel intensity of urban residents and per capita economic activity level, will affect the carbon emissions of the transportation industry to varying degrees (Yang, 2023). Chen et al. selected seven indicators such as vehicle ownership, carbon emission intensity and urbanization rate as the influencing factors of regional traffic carbon emissions in China (Chen et al., 2018). Liu et al. found in the research on the influencing factors of carbon emissions from the transportation industry in a province in Southwest China that the population, per capita GDP and vehicle ownership are the main reasons for the increase of carbon emissions from transportation in the province (Liu et al., 2023).

After summarizing the existing literature, this paper expands the basic STIRPAT model and selects six influencing factors, including population, GDP, energy intensity, urbanization rate, vehicle ownership and passenger turnover, to establish the extended STIRPAT model of the transportation industry. The model expression is shown in formula (4):

$$C = a \times P^b \times A^c \times T^d \times U^e \times CP^f \times RP^g \quad (4)$$

Where C is the total carbon emission; P stands for population (ten thousand); A stands for GDP (trillion yuan); T stands for energy intensity (tons of standard coal/ ten thousand yuan); U represents urbanization rate (%); CP stands for vehicle ownership (billion); RP represents passenger turnover (billion person kilometers).

When measuring the level of the above six influencing factors, GDP adopts the constant price in 1990 to consider the actual GDP of each year to avoid the impact of price. The above-detailed data are mainly from each year's China Statistical Yearbook, and the descriptive statistical results of each variable are shown in Table 1.

Table 1: Descriptive Statistics of Variables

Variable	Unit	Minimum	Maximum	Mean	Standard deviation
population	ten thousand	114333	141389	130221.44	7978.01
GDP	Trillion yuan	1.89	54.16	20.05	15.99
urbanization rate	%	26.41	64.19	44.19	12.28
Energy intensity	Tons of standard coal/10000 yuan	0.51	1.56	0.88	0.27
Vehicle ownership	Billion	0.15	3.79	1.55	1.17
Passenger turnover	billion person kilometers	5628.40	35349.24	18946.32	9554.26

Pearson correlation is used to analyze the correlation between various influencing factors and carbon emissions from transportation, and the correlation coefficient is calculated. The analysis results are shown in Table 2. It can be seen from the table that there is a significant positive correlation between population, GDP, urbanization rate, vehicle ownership, passenger turnover and transportation carbon emissions, and a significant negative correlation between energy intensity and transportation carbon emissions. These six factors are strongly correlated with transportation carbon emissions.

Table 2: Correlation Analysis Of Influencing Factors

Influencing factors	Transportation carbon emissions
population	0.988***
GDP	0.991***
urbanization rate	0.988***
energy intensity	-0.864***
vehicle ownership	0.998***
passenger turnover	0.843***

[1] Note: * * *, ***, * respectively represent 1%, 5% and 10% significance levels

2.3 ARIMA Model

The autoregressive moving average model is a time series prediction and analysis method widely used in time series data analysis (Wang et al., 2022). The model is expressed as ARIMA (p, d, q), where the parameters p, d and q represent the structure of the prediction model, AR represents autoregression, and p is the number of autoregressive terms; MA stands for moving average, q is the number of moving average items; d represents the difference order of the stationary sequence, and the calculation formula is shown in formula (5):

$$Y_t = \mu + \sum_{i=1}^{p} \alpha_i Y_{t-1} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-1} + \varepsilon_t \quad (5)$$

Where Y_t is the data involved in the calculation at time t; μ is a constant; p is the number of autoregressive terms; q is the number of moving average items; α_i is the data autocorrelation coefficient; θ_i is the error autocorrelation coefficient; ε_t is the white noise sequence.

The modeling steps of the ARIMA model are as follows: 1) check the stationarity of the time series, and if the series fails the test, carry out the difference operation. In general, the first-order difference can get the stationary series, and the high order of the difference will lead to the loss of data information, which is not conducive to the establishment of the model; 2) Draw the autocorrelation function diagram and partial autocorrelation function diagram to find the possible value range of parameters p and q, and then determine the optimal combination of p and q in the model through Akaike information criterion (AIC) and Bayesian information criterion (BIC). The more pronounced the AIC and BIC values are, the better the model effect is; 3) Check the residual error to see if it conforms to the normal distribution. If it does not conform to the normal distribution, determine the p and q values and the order again; 4) The prediction and error analysis are carried out after the test.

2.4. Long Short-Term Memory

Long short-term memory networks (LSTM) is a variant of recurrent neural networks (RNN) proposed by Hochreiter et al., which can remember long and short-term information (Hochreiter et al., 1997). The model solves the problem that RNN has limited processing ability for too long or too short sequences, overcomes the influence of gradient disappearance in the process of RNN training, and has good processing ability for time series data. Compared with RNN, each neuron in the LSTM model contains three basic cell gates, which are divided into input gate i_t, output gate O_t and forgetting gate f_t according to different functions. The cell structure of LSTM is shown in Figure 2.

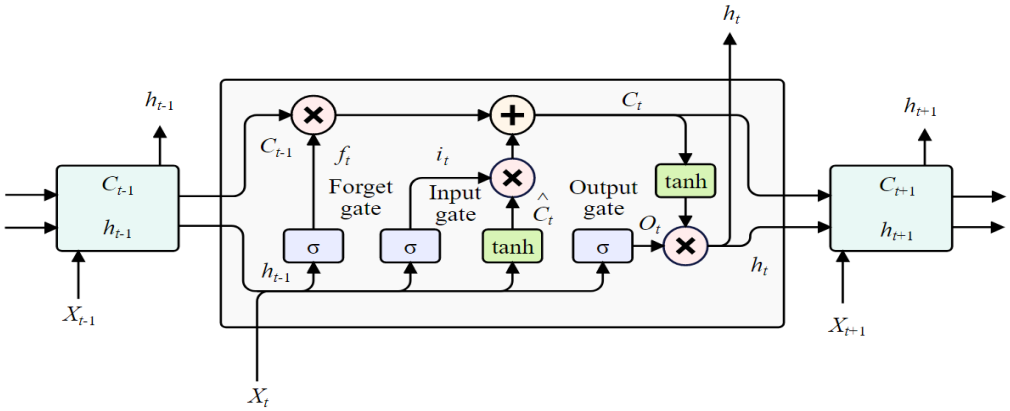


Figure 2: LSTM Hidden Layer Structure Diagram.

The steps of LSTM network prediction are as follows:

Step 1: Calculate the forgetting gate. The forgetting gate controls the information deleted or forgotten in the memory unit. h_{t-1} represents the hidden layer information of the previous time, and X_t represents the data input at the current time. The amount of information to be forgotten and retained through the forgetting gate can be determined. The calculation formula is as follows:

$$f_t = \sigma\left(W_f \times [h_{t-1}, X_t] + b_f\right) \quad (6)$$

Where f_t is the output value of the forgetting gate at time t, σ is the sigmoid activation function, W_f is the weight matrix of the forgetting gate, and b_f is the bias coefficient of the forgetting gate.

Step 2: Calculate the input gate. The input gate controls the amount of new information flowing into the control unit. First, the sigmoid activation function takes the input data and the output of the previous time step as the input and then creates a candidate value C through the tanh layer. Combine these two parts to update the cell state. The calculation formula is as follows:

$$i_t = \sigma(W_i \times [h_{t-1}, X_t] + b_i) \quad (7)$$

$$\tilde{C}_t = \tanh(W_c \times [h_{t-1}, X_t] + b_c) \quad (8)$$

Where i_t is the output information of the input gate at time t; W_i is the weight; b_i is the input gate offset coefficient; σ is the sigmoid activation function; b_C is the bias coefficient of the cell state gating unit.

The input gate will continuously update the state and update C_{t-1} to C_t. The memory information C_{t-1} at the previous time multiplied by f_t indicates the information to be discarded and retained. Then, add the candidate value $\tilde{C}_t$ to obtain the new cell state C_t, and the calculation formula is as follows:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (9)$$

Step 3: Calculate the output gate. The output gate determines how much information is extracted from the memory unit and uses the following step to predict. First, control how many memory units will be output according to the sigmoid function, then use the tanh function to scale the value, calculate the value to be output, and finally multiply the two results to determine the output information h_t. The calculation formula is as follows:

$$O_t = \sigma(W_o \times [h_{t-1}, X_t] + b_o) \quad (10)$$

$$h_t = O_t \times \tanh(C_t) \quad (11)$$

Where O_t is the output information of the output gate at time t; W_o is the weight; bo is the offset coefficient of the output gate; σ activates the function for sigmoid.

2.5 Back-Propagation Neural

As a widely used depth method, the BP neural network is famous for effectively storing and understanding many data association characteristics. It is a multilayer feedforward neural network trained according to the error backpropagation algorithm. Firstly, the features of the learning samples are input from the input layer through forward propagation, then processed by each hidden layer, and finally output from the output layer. For the error between the actual output and the expected output of the network, the weight and threshold of the model need to be continuously adjusted by backpropagation, and the gradient descent method is used to optimize the model parameters until the final fitting value error is the minimum, so as to end the training. A neural network generally comprises input, hidden, and output layers. The number of neurons in each layer varies according to different situations.

Taking the historical value of carbon emissions of China's transportation industry every three years as the input, and through continuous training, a BP neural network prediction model with a topology of 3-6-1, which is influenced by the historical value, is finally constructed, as shown in Figure 3. Where X_{t-1}、X_{t-2}、X_{t-3} is the carbon emission at time t-1、t-2、t-3 , and Y_t is the predicted carbon emission at time t.

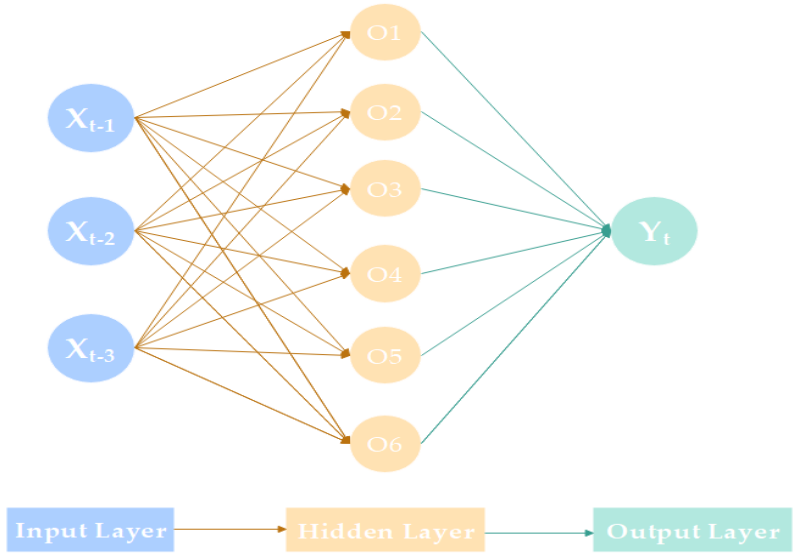


Figure 3: BP Neural Network Structure.

2.6 Combined Forecasting Model

In order to solve the problem of the accuracy of a single model not being high in the prediction and improve the prediction accuracy and generalization ability of the model, the single model is combined in different combinations. The advantages of each model are used to make up for the defects of the single model to improve the prediction's stability and accuracy. This study mainly adopts the combination method based on residual sequence and error weight. The structure of the combination model is shown in Figure 4.

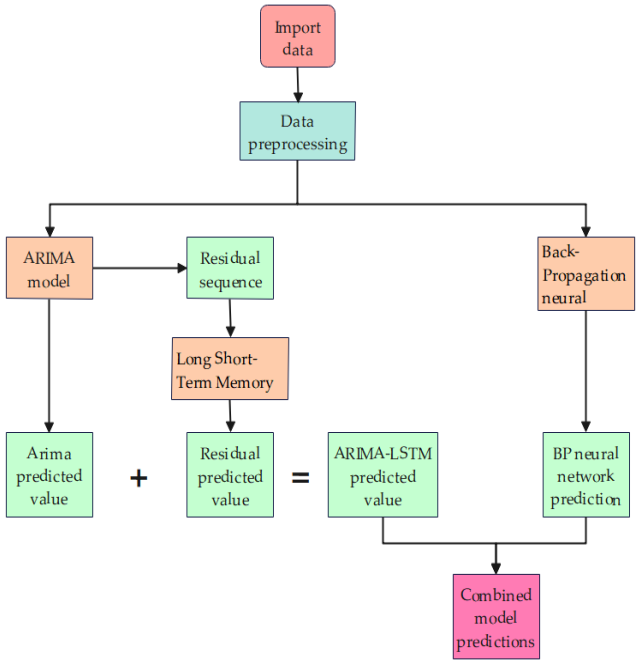


Figure 4: Structure diagram of composite model.

Combination method based on residual sequence: To compensate for the disadvantage that the ARIMA model is challenging to extract nonlinear components of time series, the ARIMA-LSTM combination model is established by combining the advantages of the ARIMA model and LSTM neural network model. Firstly, the ARIMA model is used to predict the carbon emissions of the transportation industry, and the predicted value X_1 of the ARIMA model is obtained, and the error sequence e is obtained. Then, taking the error sequence e as the input, the LSTM neural network is used to predict the error, and the error prediction value e_1 is obtained. Finally, the ARIMA-LSTM combined model prediction value Y_1 is obtained by adding the ARIMA model prediction value and the error prediction value;

$$Y_1 = X_1 + e_1 \tag{12}$$

Combination method based on error weight: To reduce the prediction error of the combination model, the model with more minor errors is given an enormous weight. Firstly, the prediction value Y_2 of the BP neural network model is obtained, and the prediction error e_2 of the model is calculated. Similarly, the prediction error e_3 of the ARIMA-LSTM model is obtained. Secondly, the weight l of the prediction error of the two models is calculated respectively l_1、l_2. The predicted value y of the ARIMA-LSTM -BP combined model is obtained by weighting the predicted value according to the error weight.

$$l_1 = \frac{e_2}{e_2+e_3} \tag{13}$$

$$l_2 = \frac{e_3}{e_2+e_3} \tag{14}$$

$$Y = l_1 \times Y_1 + l_2 \times Y_2 \tag{15}$$

2.7 Model Evaluating Indicator

This study mainly uses the average absolute percentage error (MAPE), root mean square error (RMSE), and average absolute error (MAE) to evaluate the model's prediction results. The smaller the indicators, the more accurate the prediction results are. The specific formula is as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \times 100\% \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{17}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(y_i-\hat{y}_i)| \tag{18}$$

Where Y_i is the real value at time i; $\hat{y}_i$ is the predicted value at time i; n is the number of samples.

3. Results and Discussion

3.1 ARIMA Model Construction

ADF test is used to judge whether the time series data is stable, and the result shows that p = 0.152 > 0.05, indicating that the original series is a non-stationary time series. The difference operation of d = 1 on the original sequence shows that p = 0.039 < 0.05. The significance test shows that the first-order difference sequence is stable. Therefore, the first-order difference sequence can be modeled.

The autocorrelation function (ACF) and partial autocorrelation function (PACF) are used to determine the order of the ARIMA model, and the autocorrelation function diagram and partial correlation function diagram are drawn. It can be seen from Figures 5 and 6 that the first-order difference sequence ACF and PACF diagrams fall within the confidence interval after the second order, so p and q can be taken as 0, 1 and 2.

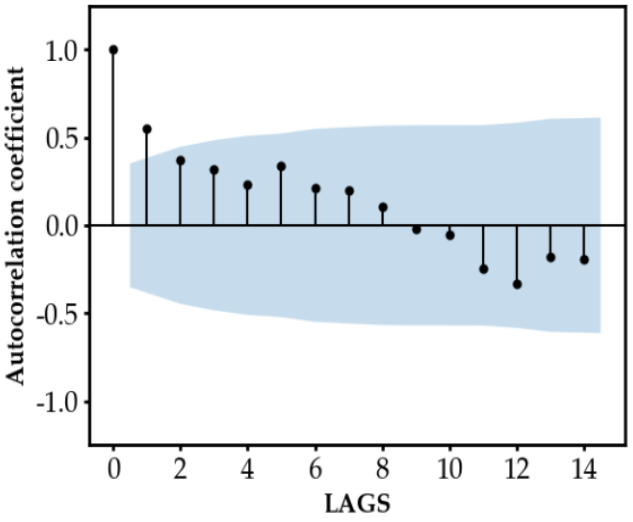


Figure 5: Autocorrelation function diagram.

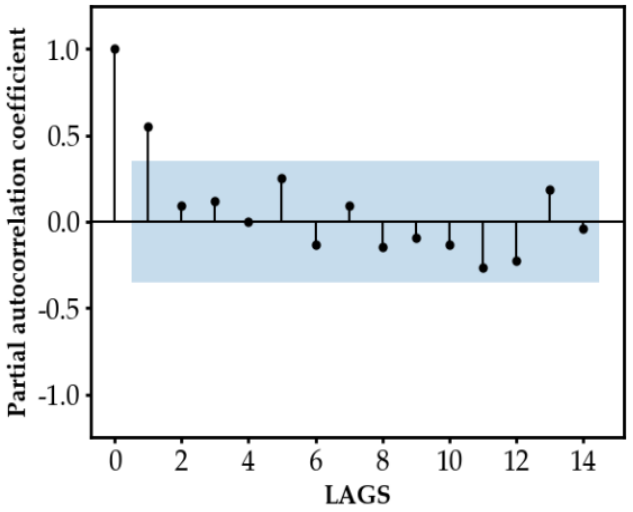


Figure 6: Partial autocorrelation function diagram.

The available values of ARIMA (p, d, q) obtained from ACF and PACF charts in the previous step are selected as the optimal model using AIC, BIC and HQIC criteria. The results are shown in Table 3. It can be seen from Table 3 that when p=1, d=1 and q=1, AIC, BIC and HQIC are the smallest, indicating that ARIMA (1, 1, 1) has the best fitting effect on the data. After the ARIMA (1, 1, 1) model is established, the first 23 years of data are used as training data to predict the carbon emissions in the next nine years (2013-2021).

Table 3: AIC、BIC and HQIC Results.

(p, d, q)	AIC	BIC	HQIC
(0, 1, 0)	305.66	307.09	306.12
(0, 1, 1)	282.04	284.90	282.97
(0, 1, 2)	280.63	284.93	282.03
(1, 1, 0)	264.09	266.96	265.02
(1, 1, 1)	261.46	265.76	262.86
(1, 1, 2)	262.75	268.48	264.62
(2, 1, 0)	264.06	268.36	265.46
(2, 1, 1)	262.96	268.69	264.83
(2, 1, 2)	262.03	269.20	264.37

3.2 ARIMA-LSTM Combined Model

The residual sequence of 23 years from 1990 to 2012 is selected as the training sample, and the residual sequence of 9 years from 2013 to 2021 is selected as the test sample. The maximum minimum method is used to normalize it, and the residual predicted by the ARIMA model is converted into a value between 0 and 1. Set the data set time step to 3 (that is, use the carbon emissions of the previous three years to predict the carbon emissions of the following year), set the number of hidden layers of the LSTM neural network to 1, and the number of training rounds is epoch=600. After repeated training, we can get the best number of training times and the number of hidden nodes. Among them, Batch_Size=9, Hidden_Size=9. The loss function is the mean square error function MAE, and the model optimizer selects Adam.

Then, the LSTM output data is de-normalized to obtain the residual prediction value of the LSTM model. Add the ARIMA model prediction value obtained in 3.1 and the LSTM model residual prediction value to obtain the ARIMA-LSTM combined model prediction value.

3.3 Back-Propagation Neural

First, normalize the data, and select the first 23 years of the data set as the training set of the model, and the last 9 years of the data set as the test set. The normalized historical carbon emissions every three years are used as input, and the carbon emissions affected by historical values are used as output. Levenberg-Marquardt backpropagation algorithm is selected as the BP neural network training algorithm; Set the maximum number of training times as 1000, determine the learning rate as 0.01; Set the number of hidden layer neurons to 1-20, and finally get the BP neural network model with the optimal number of hidden layer neurons of 6.

3.4 ARIMA-LSTM-BP Combined Model

According to the predicted values of the ARIMA-LSTM combined model and BP neural network model obtained in 3.1, 3.2 and 3.3, the model is recombined using the combination method based on error weight to obtain the predicted results of the ARIMA-LSTM-BP combined model.

3.5 Analysis of Prediction Results

In order to verify the prediction accuracy and effectiveness of the combined model for carbon emissions from the transportation industry, the ARIMA model, LSTM model, BP neural network model, ARIMA-LSTM model and ARIMA-LSTM-BP combined model were respectively constructed for comparison. The comparison between the predicted value and the actual value of each model is shown in Figure 7, and the comparison results of the evaluation indicators of each model are shown in Table 4:

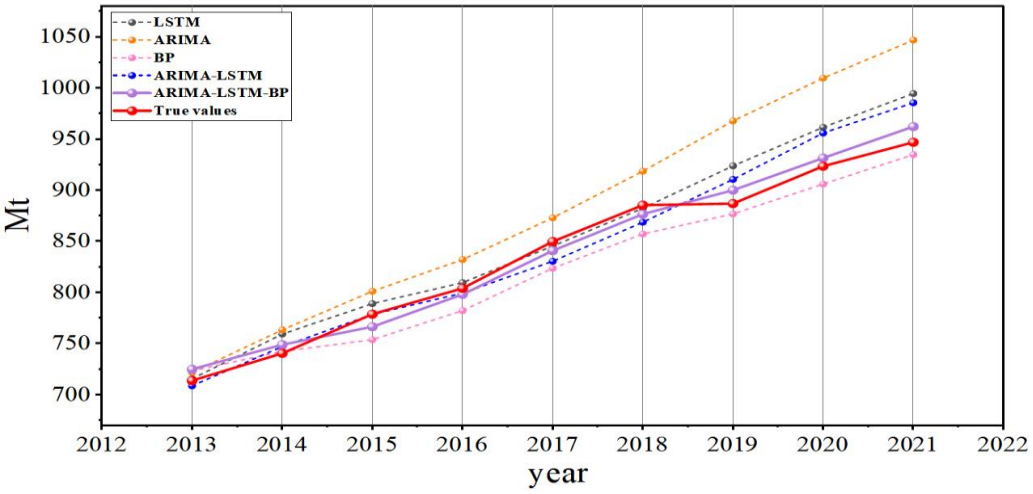


Figure 7: Comparison between predicted and actual values of various models.

Table 4: Comparison Of Evaluation Indicators Of Various Models.

Model	RMSE	MAE	MAPE
ARIMA	52.75	41.11	4.65%
LSTM	24.99	18.53	2.11%
BP	24.41	21.36	2.53%
ARIMA-LSTM	18.92	16.90	2.01%
ARIMA-LSTM-BP	13.04	12.12	1.42%

It can be seen from the chart that the ARIMA-LSTM-BP combined model has the best prediction accuracy and the most stable prediction result. It shows that the combined model can thoroughly combine the advantages of a single model and learn from each other. Therefore, the ARIMA-LSTM-BP combined model can be used to predict the future carbon emissions of China's transportation industry.

3.6 Multi Scenario Prediction

3.6.1 Scenario Settings

Through the development of the transportation industry in China and referring to a large number of policy documents, the six influencing factors in the prediction model are taken as the research object, and three scenarios of high carbon, benchmark, and low carbon are set. The prediction period of scenario analysis is 2022-2050. Considering the long prediction period, in order to ensure the rationality and scientificity of the prediction results, three prediction frequency bands will be set for each influencing factor, as shown in Table 5.

Population. Historical data shows that China's population growth has slowed significantly in the past six years. From 2016 to 2021, the population increased by only 21.57 million, with an average annual growth rate of about 0.25%. The seventh national census bulletin shows that China's total fertility rate is only 1.3, and the trend of negative population growth is visible to the naked eye. According to the Research Report on the national population development strategy, China's population will peak at 1.442 billion in 2028, showing a negative growth trend after that. Therefore, the growth rate of the population in different periods under the benchmark scenario is set to be 0.15% from 2022 to 2028, -0.1% from 2029 to 2038 and -0.25 from 2039 to 2050.

GDP. In 2021, facing the double test of the ups and downs of the epidemic situation in the century and the complex and severe external environment, China effectively coordinated the epidemic prevention and control and economic and social development, strengthened the cross-cycle adjustment of macro policies, and the national economy continued to

recover, with an increase of 7.81% over the previous year at constant prices. The research on China's long-term low-carbon development strategy and transformation path mentioned that during the "14th five-year plan" period, the average annual GDP growth rate will exceed 5.00%. The "14th five-year plan" and the proposal of the long-term goal of 2035 also clearly put forward the goal of doubling the economic aggregate by 2035 compared with 2020. Using the quantitative economic model, Xiao estimated that the average annual growth rate of China's GDP will reach about 5.3% from 2023 to 2025 (Xiao, 2023). Therefore, the growth rate of GDP in different periods under the benchmark scenario is set to be 5% from 2022 to 2028, 4% from 2029 to 2038, and 3.5% from 2039 to 2050.

Urbanization rate. In recent years, the level of urbanization in China has developed rapidly. In 2021, the urbanization rate reached 64.19%, and the urbanization rate of developed countries in the world has stabilized at about 85%. Ou et al. have studied China's urbanization rate, which is expected to reach about 67.45% by 2025 (Ou et al., 2021). Hu predicted that China's urbanization rate would reach 73.41% -74.53% in 2035 through multiple model screening. It is expected to reach 80% in 2050 (Hu, 2023). Therefore, the growth rate of urbanization rate in different periods under the benchmark scenario is set to be 1.1% from 2022 to 2028, 0.7% from 2029 to 2038, and 0.35% from 2039 to 2050.

Energy intensity. In recent years, China's energy intensity has shown a downward trend, decreasing by 17.65% from 2016 to 2021, with an average annual decline of about 3%. The comprehensive work plan for energy conservation and emission reduction during the 14th five-year plan puts forward the goal of reducing energy intensity by 10% in 2025 compared with 2021. From 2021 to 2050, energy intensity decreased by about 3.5% annually. Hu et al. pointed out that with the increasing difficulty of technical emission reduction, the energy consumption growth rate per unit transportation turnover will slow down (Hu et al., 2022). Based on this view and the analysis of the actual situation, the growth rate of energy intensity in different periods under the benchmark scenario is set as -3% from 2022 to 2028, -4% from 2029 to 2038 and -4.5% from 2039 to 2050.

Vehicle ownership. With the continuous development and expansion of the overall urban scale in China, the demand for automobiles continues to grow. By 2021, the number of motor vehicles was 394 million, an increase of 104 million compared with 2016, with an average annual growth of about 4.4%. Zhou et al. predicted that by 2028, the number of motor vehicles would be about 440 million (Zhou et al., 2023). Therefore, the growth rate of vehicle ownership in different periods under the benchmark scenario is set to be 4% from 2022 to 2028, 3% from 2029 to 2038, and 1.5% from 2039 to 2050.

Passenger turnover. In the ten years before the outbreak, from 2010 to 2019, the average annual increase in passenger turnover was 2.11%. Affected by the epidemic, compared with 2019, the passenger turnover in 2020 and 2021 decreased significantly to 1.29 trillion person kilometers. However, with the epidemic's end, passenger turnover will gradually warm up, and it is expected to return to its average level in 2028. Therefore, the growth rate of passenger turnover in different periods under the benchmark scenario is set to be 10% from 2022 to 2028, 2.5% from 2029 to 2038, and 1.5% from 2039 to 2050.

Table 5: Scenario Settings For Various Parameters In The Year.

Scene	Time	Growth rate of influencing factors					
		Population	GDP	Urbanization rate	Energy intensity	Vehicle ownership	Passenger turnover
high carbon	2022-2028	0.25%	5.5%	1.3%	-2%	5%	11.5%
	2029-2038	0.05%	4.5%	0.9%	-3%	4%	3.5%
	2039-2050	-0.15%	4%	0.55%	-3.5%	2.5%	2.5%
benchmark	2022-2028	0.15%	5%	1.1%	-3%	4%	10%
	2029-2038	-0.1%	4%	0.7%	-4%	3%	2.5%
	2039-2050	-0.25%	3.5%	0.35%	-4.5%	1.5	1.5%
low carbon	2022-2028	0.05%	4.5%	0.9%	-4%	3%	8.5%
	2029-2038	-0.2%	3.5%	0.5%	-5%	2%	1.5%
	2039-2050	-0.35%	3%	0.15%	-5.5%	0.5%	0.5%

3.6.2 Multi Scenario Prediction and Analysis

Based on good fitting and validity test results of the ARIMA-LSTM-BP combined model, carbon emissions from 2022 to 2050 are predicted. The prediction results are shown in Figure 8.

The prediction results show that the peak value of carbon emissions from transportation in China is different under the three scenarios of high carbon, benchmark and low carbon, and the time to reach the peak value is also different, which is 2031, 2034 and 2039 respectively. The transportation industry alone cannot achieve the goal of reaching the peak carbon by 2030. Under the high carbon scenario, the transportation industry will reach a peak of 1624.7732 million tons in 2039, an increase of 678.1062 million tons compared with 2021, and 125.7585 million tons in 2050, 1.32 times higher than 2021. Under the benchmark scenario, the transportation industry will reach a peak of 1478169400 tons in 2034, an increase of 531.524 million tons compared with 2021, and will drop to 113.7572 million tons in 2050, 1.20 times higher than 2021. Under the low-carbon scenario, the transportation industry will peak at 1367.5417 million tons in 2031, an increase of 420.8747 million tons compared with 2021, and drop to 980.7545 million tons in 2050, 1.03 times higher than 2021. By comparing the above three scenarios, in order to better keep up with the country's goal of reaching the carbon peak by 2030, we should choose the low-carbon development scenario. Compared with the baseline and high-carbon scenarios, the time of reaching the carbon peak has advanced by 3 and 8 years, respectively.

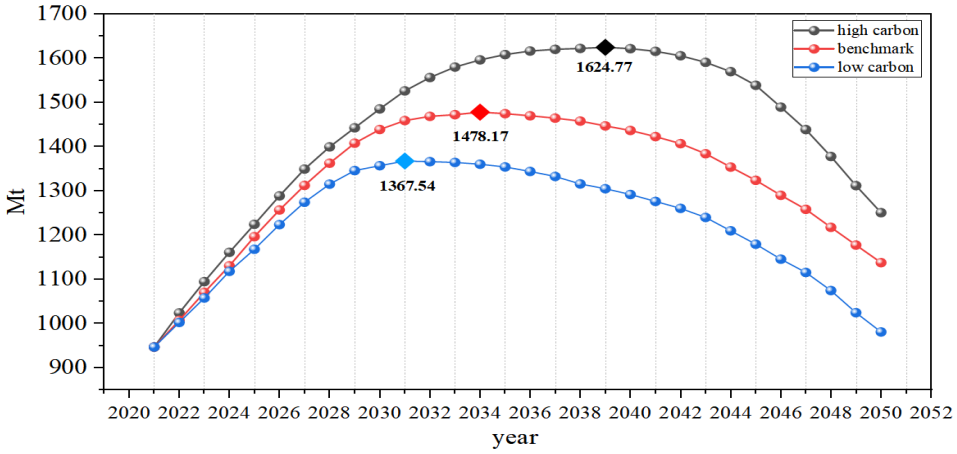


Figure 8: Prediction Results of Carbon Emissions from Transportation in China.

4. Conclusions and Suggestions

4.1 Conclusions

This study follows the logic of "Measurement Research - influencing factors - prediction model selection - scenario analysis" to conduct a specific study of China's transportation carbon emissions and draws the following conclusions:

Firstly, from 1990 to 2021, China's total emissions from transportation showed a gradual upward trend. By 2021, the total carbon emissions were 946.6672 million tons, with an average annual growth rate of 7.37%.

Secondly, compared with ARIMA, LSTM, BP, and ARIMA-LSTM, the ARIMA-LSTM-BP combined model shows better prediction ability, higher accuracy, and more robust stability.

Thirdly, the ARIMA-LSTM-BP combined model is used to predict the carbon emissions of transportation in the three scenarios. It can be seen that the carbon peak time is 2031, 2034, and 2039, respectively, and the peak value is 1624773200 tons, 1478169400 tons, and 1367541700 tons.

Fourthly, if the transportation industry wants to achieve the goal of a carbon peak by 2030, it also needs to strengthen low-carbon measures and promote low-carbon emission reduction.

4.2 Suggestions

Based on the above research and scenario analysis, in order to achieve the carbon peak of transportation in China as soon as possible, the following suggestions are put forward:

Firstly, optimize the energy structure. China's transportation industry is still an oil-based energy structure, but if we want to achieve a carbon peak as soon as possible, optimizing the energy structure is the only way. Therefore, the government should increase financial subsidies to support the development of new clean energy for the transportation industry. Actively promote new energy vehicles, reduce energy consumption, and vigorously develop "zero emission" and low-energy trams.

Secondly, improve the quality of the population. To improve the population's quality and promote social technology development, we must strengthen energy management, vigorously promote a conservation-oriented economy, strengthen the publicity and education of the environment, and popularize a conservation-oriented culture to create a healthy, safe, and comfortable ecological civilization. In order to protect our environment, we must actively promote resource conservation and encourage people to take adequate measures in daily life to reduce dependence on resources to achieve sustainable development.

Thirdly, improve the level of technological innovation. In order to achieve green and sustainable development, we should strengthen technological innovation, improve energy efficiency, vigorously promote the low-carbon circular economy, and achieve green and sustainable economic development to save energy, reduce pollution, and reduce emissions.

Through the above suggestions, it is hoped that the transportation carbon peak can be achieved as soon as possible and that the national goal of reaching the carbon peak by 2030 can be achieved.

Author Contributions

S.K.: methodology, software, formal analysis, writing—original draft. Y.S.: conceptualization, investigation, validation, supervision. The statement that all authors have approved the final article should be true and included in the disclosure.

Funding

This research was funded by the Natural Science Foundation of Zhejiang Province (Grant No. LY23A010003).

Declaration of Interest

The authors declare no conflict of interest.

References

Andrés, L., Padilla, E., 2018. Driving factors of GHG emissions in the EU transport activity. Transport Policy, 61, Pp. 60-74.

Ağbulut, Ü., 2022. Forecasting of transportation-related energy demand and CO2 emissions in Turkey with different machine learning algorithms. Sustainable Production and Consumption, 29(1), Pp. 141-157.

Bian, L.; Ji, M. 2019. Research on Influencing Factors and Prediction of Traffic Carbon Emissions in Qinghai. Ecological Economy, 35 (2), Pp. 35–39 (in chinese).

Commoner, B., 1990. Making Peace with the Planet. Pantheon Books, New York.

Chen, L., Wang, J., He, T., et al. 2018. Research on Prediction of Regional Traffic Carbon Emissions Based on SVR. Transp. Syst. Eng. Inform., 18(2), Pp. 14–19 (in chinese).

Dietz, T., Rosa, E.A., 1997. Effects of Population and Affluence on CO2 Emissions. PNAS, 94(1), Pp. 175–179.

Dong, M., Han, Z., Guo, J., 2020. Carbon Emission Efficiency Measurement and Influencing Factors Analysis of China's Marine Transportation Industry. Ocean Bull., 39, Pp. 169–177 (in chinese).

Fang, X., Luo, Y., 2017. Comparative Study on Calculation Methods of Urban Traffic Carbon Emissions. Transportation Energy Conservation and Environmental Protection, 13(4), Pp. 81–83 (in chinese).

Fang, H., Liu, C., Jiang, K., et al. 2023. Research on the Peak Path of Carbon Emissions in the Field of Transportation in Hunan Province. Transp. Syst. Eng. Inf., 23 (04), Pp. 61–69 (in chinese).

Gu, D., 2023. Regional Correlation Analysis of Carbon Emissions from China's Transportation Industry. J. Shanghai Marit. Univ., 44(03), Pp. 64–70+118 (in chinese).

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation, 9, Pp. 1735–1780.

Huang, Y.S., Shen, L., Liu, H., 2019. Grey Relational Analysis, Principal Component Analysis and Forecasting of Carbon Emissions Based on Long Short-Term Memory in China. Journal of Cleaner Production, 209 (10), Pp. 415–423 (in chinese).

Hassouna, F., Alsahili, K., 2020. Environmental Impact Assessment of the Transportation Sector and Hybrid Vehicle Implications in Palestine. Sustainability, Pp.12(19).

Han, R.; Li, L.; Zhang, X.; et al. 2022. Spatial-Temporal Evolution Characteristics and Decoupling Analysis of Influencing Factors of China's Aviation Carbon Emissions. Chinese Geographical Science, 32(2), Pp. 218–236 (in chinese).

Hu, J., Luo, Z., Li, F., 2022. Prediction of China's Carbon Emission Intensity under the "Carbon Peak" Target: Analysis Based on LSTM and ARIMA-BP Model. Financial Science, (2), Pp. 89–101 (in chinese).

Hu, M., Zhen, Y., Li, Y., 2022. Research on Prediction of Peak Carbon Emissions from Transportation in Hubei Province Under Multi Scenario. J. Environ. Sci., 42, Pp. 464–472 (in chinese).

Hu, A., 2023. China's Urbanization Rate and the Main Spatial Form of Urban Agglomeration in 2035. Technical Economy, 42(05), Pp. 174-188 (in chinese).

Jiao, W., Chen, X., 2012. Environmental Impact Analysis of Gansu Province Based on STIRPAT Model -- Taking Energy Consumption from 1991 to 2009 as an Example. Resources and Environment in the Yangtze River Basin, 21 (01), Pp. 105–110 (in chinese).

Jia, J., 2020. Research on Influencing Factors and Countermeasures of Transportation Carbon Emissions in Hebei Province Based on GRA. Hebei Enterprise, (9), Pp. 72–73 (in chinese).

Liu, Z., Guan, D., Wei, W., et al., 2015. Reduced Carbon Emission Estimates from Fossil Fuel Combustion and Cement Production in China. Nature, 524, Pp. 335–338 (in chinese).

Liu, J., Li, S., Ji, Q., 2021. Regional Differences and Driving Factors Analysis of Carbon Emission Intensity from Transport Sector in China. Energy, 224 (in chinese).

Liu, C., Qu, J., Ge, Y., 2023. Carbon Emission Prediction of China's Transportation Industry Based on LSTM Model. China Environmental Science, 43 (05), Pp. 2574–2582 (in chinese).

Liu, J., Huang, F., Chen, B., 2023. Research on Influencing Factors and Emission Reduction Strategies of Carbon Emissions in Transportation Industry. Highway, 68(05), Pp. 252–259 (in chinese).

Ou, Y.H., Li, Z., Li, P.L., 2021. The change trend and policy implications of China's urbanization rate during the 14th Five Year Plan period. Urban Development Research, 28(6), Pp. 1-9 (in chinese).

Pan, S.Y., Zhang, M.L., 2023. Prediction and Influencing Factors of Carbon Dioxide Emissions in Gansu Province Based on BP Neural Network. Environmental Engineering, 41(07), Pp. 61-68+85 (in chinese).

Shao, M., Cheng, T., Ma, X., 2016. Combined Forecasting of Railway Freight Volume Based on Grey Neural Network. J. Transp. Eng. Inf., 14 (3), Pp. 129–135 (in chinese).

Solaymani, S., 2019. CO2 emissions patterns in 7 top carbon emitter economies: The case of transport sector. Energy, 168, Pp. 989-1001.

Wang, Y., Li, H., Guo, X., et al., 2021. Research on Influencing Factors of Carbon Dioxide Emission from Railway Operation in China. Journal of Railways, 43 (6), Pp. 189-195 (in chinese).

Wang, X., Li, A., Li, Y., et al., 2022. Integrated Energy System Load and Wind and Solar Resources Prediction Based on ARIMA-LSTM Model. J. Xi'an Univ. Archit. Technol. (Nat. Sci. Ed.), 54(05), Pp. 762-769 (in chinese).

Xiao, H., 2023. Progress Evaluation and Prospect of Main Indicators in the Outline of the 14th Five-Year Plan. Economic Aspect, (07), Pp. 27-33 (in chinese).

Yan, S., Chen, L., 2020. Analysis on Influencing Factors of Traffic Carbon Emissions: A Case Study of Xi'an. Statistics and Decision Making, 36(04), Pp. 62–66 (in chinese).

Yang, J., 2023. Research on Carbon Emission Measurement and Emission Reduction Path of China's Transportation Industry. Ph.D. thesis, Jiangxi University of Finance and Economics, Jiangxi, China.

Zhang, G., Su, Z., 2020. Decomposition of Influencing Factors and Scenario Prediction of Transportation Carbon Emissions in the Yellow River Basin. Management Review, 32(12), Pp. 283–294 (in chinese).

Zhang L., Li P., Pang W., et al. 2022. Precipitation Prediction Based on Seasonal Autoregressive Integral Moving Average and Deep Learning Long-Term and Short-Term Memory Neural Network. Sci. Technol. Eng., 22(9), Pp. 3453-3463 (in chinese).

Zhao, J.H., Li, J.S., Wang, P.L., et al., 2022. Research on the Path of Carbon Peak in Henan Province Based on Lasso BP Neural Network Model. Environmental Engineering, 40 (12), Pp. 151-156,164 (in chinese).

Zhou, J., Li, N., Feng, W., et al., 2023. Estimation of Carbon Dioxide Emissions from Road Sources and Prediction of Future Scenarios in China. J. Environ. Sci., Pp. 1-12 (in chinese).

Zuo, D., Zhao, L., Xiong, Q., et al., 2023. Review of Research on Traffic Carbon Emissions: Accounting Methods, Influencing Factors and Mechanism. J. Transp. Eng. Inf., Pp. 1-22 (in chinese).