Prediction and Analysis of Carbon Emissions from Transportation in Beijing Based on BP-LSTM-Attention

[1]Wang Jingyi and [2]Chang Chunyi,
[1,2]School of Statistics, Beijing Wuzi University, Beijing, China

***Abstract*—**Under the "dual carbon" goals, the low-carbon transformation of the transportation sector is urgently required. To accurately assess the future trends of carbon emissions from Beijing's transportation industry and support the formulation of emission reduction policies, this paper aims to construct a highly accurate and robust carbon emission prediction model.

Firstly, a "top-down" emission factor method was employed to calculate the CO_2 emissions from Beijing's transportation industry from 1997 to 2023. Secondly, 25 influencing factors were screened as input features through a literature review. Thirdly, the prediction performance of four individual models—BP neural network, GRU, LSTM, and Transformer—was compared. Finally, a combined model integrating temporal feature extraction (LSTM+Attention) and static feature extraction (BP) was constructed and its effectiveness was validated.

The results indicate that among the individual models, BP and LSTM demonstrated superior performance. The constructed combined model outperformed all individual models across the three metrics of MAPE, MAE, and RMSE. Specifically, its MAPE was 12.18% lower than that of the best individual model, significantly enhancing prediction accuracy and stability. This combined model effectively integrates temporal and non-temporal features, making it suitable for scenario-based forecasting of future carbon emissions from Beijing's transportation industry and capable of providing a scientific basis for relevant authorities to formulate differentiated emission reduction pathways.

Keywords*—Transportation industry; Carbon emissions; Deep learning*

I. INTRODUCTION

Achieving the goals of carbon peak and carbon neutrality represents a major strategic decision made by the central government and is also an inherent requirement for promoting high-quality economic and social development. As a fundamental, pioneering, and strategic industry of the national economy, transportation is also a significant area of energy consumption and carbon emissions. Its green and low-carbon transformation is of vital importance for implementing the guidance of the CPC Central Committee and State Council on Completely, Accurately, and Comprehensively Implementing the New Development Philosophy to Achieve Carbon Peak and Carbon Neutrality and for accelerating the construction of a country with strong transportation networks.

As the capital of China, Beijing has a large-scale and complex transportation system, facing particularly severe challenges related to carbon emissions. With the acceleration of urbanization and the continuous growth of vehicle ownership, the transportation industry has become one of the main contributors to carbon emissions in Beijing, exerting significant pressure on air quality improvement and public health protection. Against this backdrop, conducting in-depth research on the carbon emission characteristics of Beijing's transportation industry, accurately identifying key influencing factors, and scientifically predicting its future evolution trends can not only provide decision-making support for local governments to formulate differentiated and precise emission reduction policies but also address the practical needs of promoting the capital's ecological civilization construction and sustainable development.

A. Current Status of Carbon Emission Accounting

Carbon emission accounting serves as the foundation for trend prediction in the transportation sector. Currently, the mainstream accounting methods primarily include four approaches:

The "top-down" method, based on energy consumption statistics, calculates carbon emissions using the consumption volumes of various fuel types and their corresponding emission factors, making it suitable for macro-scale accounting. Wu et al. (2025) and Jiang et al. (2020) employed this method to calculate transportation carbon emissions in Guangzhou and the Yangtze River Economic Belt, respectively; Feng et al. (2025) applied it to empirical studies in Beijing.

The "bottom-up" method, based on vehicle kilometers traveled (VKT) and energy consumption per unit distance, aggregates data from the micro level of vehicle types and fuel categories, offering higher accounting accuracy but requiring extensive data. Xie (2025) disaggregated 13 vehicle types to construct a road carbon emission inventory for Fujian Province; Shen et al. (2021) used this method to estimate CO_2 emissions from motor vehicles in Beijing's urban areas.

The "total-structure" method estimates carbon emissions by analyzing structural parameters such as transportation turnover volume, modal share, and energy efficiency levels, making it suitable for scenario simulation. The life cycle assessment (LCA) method covers the entire process from vehicle production and use to recycling, and is applicable for technology evaluation; however, due to data complexity, it is mostly used for small-scale, refined studies.

Given that this paper aims to account for historical carbon emissions from Beijing's transportation industry and conduct trend prediction, while also considering data availability and methodological maturity, the "top-down" emission factor method is adopted. This approach can fully utilize official energy statistics data, ensuring the comparability and reliability of the accounting results, and laying a foundation for subsequent analysis of influencing factors and model construction.

B. *Current Status of Carbon Emission Prediction*

Carbon emission prediction in the transportation sector serves as a critical foundation for evaluating low-carbon transition pathways and formulating emission reduction policies. Existing prediction methodologies can be broadly categorized into several main paradigms: time series forecasting, machine learning prediction, the LEAP model, and hybrid frameworks.

Traditional statistical models primarily include regression predictions based on the STIRPAT model and its extended forms, as well as time series analysis methods such as ARIMA, VAR, and grey prediction models. Song Xiaowei et al. (2026) integrated the LMDI model with an extended STIRPAT model to forecast carbon emissions in Hebei Province from 2023 to 2060. Song Xiaohua et al. (2025) proposed a method combining ARIMA and the Markov matrix to simulate and predict green development in Gansu Province. Peng Jiaoting et al. (2026) employed a Vector Autoregression model to analyze the dynamic relationship between influencing factors and agricultural net carbon sinks and to predict their changes.

Machine learning models have been widely applied due to their advantages in nonlinear fitting. Yang Lishan et al. (2026) constructed a fuel consumption prediction framework based on LSTM and introduced Monte Carlo simulation to quantify uncertainty. Xu Dong et al. (2025) compared the predictive performance of multiple models for carbon prices and found that a weekly LSTM model based on national carbon price influencing factors performed best. Jin Yujie et al. (2026) incorporated four types of Lasso-Transformer neural network models to forecast carbon emissions in Hainan Province.

The LEAP model is an energy system simulation tool based on scenario analysis, used to predict carbon emission pathways by setting exogenous parameters, making it suitable for policy sensitivity analysis. Han Yan et al. (2025) built a carbon emission prediction model for industries in Lanzhou based on the LEAP model, systematically analyzing emission evolution paths under different policy scenarios.

In summary, traditional models offer strong interpretability but insufficient nonlinear fitting capability; machine learning models provide high accuracy but individual models have limitations; and the LEAP model is suitable for scenario simulation but relies heavily on parameter settings. This paper aims to compare multiple deep learning models and construct a combined model integrating temporal and static features to enhance prediction stability and accuracy.

II.CARBON EMISSION MEASUREMENT

A. *Carbon Emission Accounting Methods*

Based on the requirements of data availability and the precision of carbon emission accounting, this study selects the "top-down" emission factor method to calculate the CO_2 emissions from Beijing's transportation industry. The calculation formula for carbon dioxide emissions directly generated by energy consumption is presented in Equation (1):

$$E_{CO_2} = \sum FC_i \times EF_i \qquad (1)$$

Since the transportation sector is not listed as a separate category in China's current statistical, but is instead aggregated with the storage and postal services, the energy consumption data required for this study are sourced from the "Transport, Storage, and Postal Services" section of the historical Beijing Energy Balance Sheet (Physical Quantity). The average net calorific value for each energy source is derived from the "Reference Coefficients for Converting Various Energy Sources to Standard Coal" provided in the China Energy Statistical Yearbook. The carbon content per unit calorific value, the carbon oxidation rate, and the carbon dioxide emission factor for electricity are obtained from the Guidelines for the Preparation of Provincial Greenhouse Gas Inventories (Trial).

B. *Analysis of Current Status of Carbon Emissions*

Based on the estimation method described above, the carbon emissions from Beijing's transportation industry from 1997 to 2023 were calculated. Missing values for certain energy-related emissions were filled using linear trend interpolation based on neighboring points. The results are shown in the figure 1 below.

Figure 1: Trend of Total Carbon Emissions

The composition and trend of carbon emissions from Beijing's transportation sector are shown in Figure 1. Total carbon emissions increased from 3.5319 million tons in 1997 to 31.3434 million tons in 2023, representing an approximately 7.9-fold increase with an average annual growth rate of 8.6%. The period from 1997 to 2002 was characterized by a gentle growth phase, during which total emissions rose from 3.53 million tons to 6.71 million tons, with coal and diesel serving as the primary sources. The period from 2003 to 2016 witnessed accelerated growth; in 2003, with the inclusion of electricity consumption in the statistics, total emissions surged to 8.78 million tons. Amid Beijing's rapid development, emissions exceeded 29.94 million tons in 2016, with kerosene, diesel, and electricity making significant contributions. The period from 2017 to 2023 was marked by fluctuations: a peak of 33.97 million tons occurred in 2018, emissions dropped to 25.33 million tons in 2020 due to the pandemic, and subsequently rebounded to 31.3434 million tons in 2023.

III.CONSTRUCTION OF THE PREDICTION MODEL

A. *Model Introduction*

1. *Back Propagation Neural Network Model*

The Back Propagation Neural Network (BP) is a typical multi-layer feedforward network. Its core principle involves iteratively updating weights through the backpropagation of errors to achieve model optimization. The network structure comprises an input layer, hidden layers, and an output layer. Neurons between layers are connected via weights, while no direct connections exist within the same layer. The input layer receives data, the hidden layers perform nonlinear transformations and feature extraction, and the output layer produces the prediction results. During forward propagation, input signals are processed through weighted summation and activation functions, layer by layer, to the output layer. A loss function quantifies the error between the predicted values and the true values. Backpropagation, based on the chain rule, transmits this error backward and calculates the gradients of the weights. Subsequently, the weights are updated using a gradient descent algorithm combined with a learning rate. This process iterates repeatedly until the model converges. This network laid the foundation for deep learning, yet it has limitations such as a tendency to get trapped in local optima.

2. Long Short-Term Memory Neural Network Model

The Long Short-Term Memory neural network (LSTM) is a variant of the recurrent neural network (RNN) first proposed by Hochreiter and Schmidhuber. It overcomes the vanishing or exploding gradient problems of traditional RNNs through backpropagation training. Its core lies in a memory unit containing three gating mechanisms (forget gate, input gate, output gate) and a cell state. The forget gate controls the degree to which old information is discarded, the input gate regulates the proportion of new information to be stored, and the output gate determines the amount of information to be output. The cell state is responsible for storing and updating long-term memory, thereby capturing long-range dependencies in time series data and enhancing the rationality and accuracy of sequence modeling.

3. Gated Recurrent Unit Prediction Model

The Gated Recurrent Unit (GRU), proposed by Cho et al. in 2014, addresses the challenge of long-term dependencies in traditional RNNs through a simplified structure while maintaining performance comparable to LSTM. Its core components are the update gate and the reset gate. The update gate combines the functions of the LSTM's forget and input gates, controlling the balance between retaining old information and incorporating new information. The reset gate governs the extent to which past information should be forgotten and its impact on the current input. By calculating a candidate hidden state and fusing it with the previous hidden state, the current output is generated. The GRU offers advantages in terms of parameter count and computational efficiency, making it suitable for small datasets and high-frequency training scenarios, and it performs well in time series prediction.

4. Transformer Network

The Transformer network is a deep learning model that relies on a self-attention mechanism, discarding the sequential processing approach of RNNs and CNNs. Its core architecture consists of an encoder and a decoder. The input sequence, after being injected with positional information through word embedding and positional encoding, is fed into the encoder. The encoder, through stacked sub-layers including self-attention, feed-forward neural networks, residual connections, and layer normalization, extracts context-rich sequential features. The decoder, utilizing masked self-attention sub-layers, cross-attention sub-layers, and feed-forward neural network sub-layers, combines information from the encoder to generate the target sequence. Finally, the probability distribution of the target sequence is output through a linear transformation and a softmax function. This architecture has achieved remarkable success in the field of sequence transformation.

B. Data Preparation

The selection of indicators in this paper comprehensively considers factors reflecting the characteristics of economic activity intensity, including permanent population, GDP, and the number of patent grants. Factors reflecting energy consumption characteristics include transportation investment, transportation energy efficiency, and transportation energy structure. Indicators reflecting the characteristics of the transportation industry itself include transportation energy intensity, highway mileage, highway passenger volume, highway passenger turnover, highway freight volume, highway freight turnover, railway mileage, railway passenger volume, railway passenger turnover, railway freight volume, railway freight turnover, urban road mileage, the length of public transport operation lines, the number of public transport vehicles, public transport passenger volume, and the number of motor vehicles. Furthermore, factors reflecting urban development characteristics, such as the proportion of the tertiary industry, urbanization rate, and urban green coverage rate, are also incorporated.

C. Data preprocessing

1. Data preprocessing

Among the influencing factors selected in this article, there are obvious inconsistencies in the unit dimension. For example, the value of urbanization rate is between 0 and 1, while the value range of features such as permanent population, railway passenger volume, motor vehicle ownership, and highway passenger volume is in the tens of thousands. This will cause features with large values to dominate in model training. Through normalization, all features can be mapped to the same scale, ensuring that the influence of each feature on the model is balanced. In addition, during the algorithm training process, if the numerical values of each feature are very different, the convergence speed may be slowed down. Normalization processing helps improve the convergence speed and training efficiency of the algorithm. Therefore, before modeling, the collected data are normalized and the variables are processed into dimensionless data to eliminate the influence of dimensions and units, so that all input variables are at the same magnitude, which facilitates subsequent modeling analysis and improves the model prediction accuracy to a certain extent. The normalization method used in this article is shown in formula 2, that is, maximum value-minimum value normalization, which can scale all features to the interval of [0-1] and maintain the original data distribution state. The MinMaxScaler function in the Sklearn library provides an implementation of max-min normalization.

$$X_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \qquad (2)$$

In the formula:$X_i$represents the normalized feature values, represents the original data, and X_{max} represents the maximum value of the original data; X_{min} represents the minimum value of the original data.

2. Evaluation Indicators

This study collected the data of transportation carbon emissions and their influencing factors in Beijing from 1997 to 2020 as the training set for model learning, feature understanding and parameter optimization. The training set and the test set are divided in a ratio of 0.8:0.2 to test the generalization ability of the model. The above data are all from the Beijing Statistical Yearbook of previous years and the China Energy Statistical Yearbook.

To accurately evaluate the effectiveness of the prediction model adopted in this paper, and then adjust the model parameters and determine the optimal prediction model based on the model accuracy, this paper takes the mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute error (MAE) as the basis for judging the results of the experimental model. Three evaluation indicators are used to measure the difference between the predicted data and the actual data. The smaller the value, the higher the model accuracy and the better the prediction performance. Among them, MAPE is selected as the main evaluation index.see Equation (3-5)

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - y_i^*|}{y_i^*} \times 100\% \quad (3)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)^2} \quad (4)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i^*| \quad (5)$$

Among them,y_i^* is the actual value,y_i is the corresponding predicted value, and n is the number of indicator variables.

D. Single-item model Prediction and Result analysis

This paper selects four models, namely BP, LSTM, GRU and transformer, as the basic prediction models and makes predictions on the basic models respectively. The prediction process is as follows:

The first step is data preparation: Use the data on the influencing factors of carbon emissions in Beijing's transportation industry and the carbon emissions collected in the previous chapter as the input data for the prediction model. First, the missing values of the original data are processed. Then, the data is normalized to eliminate dimensional differences. Subsequently, the model is divided into a training set and a test set in an 8:2 ratio. The training set is used for learning model parameters, and the test set is used to verify the generalization ability of the model.

The second step is to build the optimal model: Take the original dataset of influencing factors as the input indicators for each model, and the carbon emissions of Beijing's transportation industry in the next year as the output indicator. Design appropriate input forms based on the characteristics of different models, and set core parameters based on the structural characteristics of the models for training experiments.

Step 3, model optimization: Conduct multiple rounds of adjustments and experiments on other hyperparameters such as the learning rate, batch size, dropout rate, and number of iteration rounds of the model. Introduce an early stop mechanism to prevent overfitting and continuous iteration, and obtain the optimal model.

The fourth step is to predict the carbon emissions of Beijing's transportation industry in future years: The relevant data that has been preprocessed in the first step is input into the optimal model in the third step. After the model outputs standardized predicted values, they are restored to the true carbon emission level through de-standardization, and finally the results of each model's prediction of Beijing's carbon emissions in future years are output.

After numerous experiments, the respective optimal models were obtained. The experimental results of each model are summarized as shown in Table 1.

Table 1 :Prediction Performance Metrics of Individual Models

	Bp	Gru	Lstm	Transformer
MAPE	11.91%	21.24%	17.37%	22.75%
MAE	321.48	604.56	502.26	642.11
RMSE	369.24	670.35	587.05	694.56

By integrating the three evaluation indicators, it can be found that the BP model demonstrates the best comprehensive performance in predicting the carbon emissions of the transportation industry in Beijing. Its relative error, mean absolute error and extreme error are all the smallest. This is mainly due to the simple structure and high training efficiency of the BP neural network, which can better fit the nonlinear variation law of the carbon emissions of the transportation industry in Beijing. Short-term or medium-term carbon emission prediction tasks applicable to this field.

As a recurrent neural network specifically for processing time series data, the LSTM model has a natural advantage in capturing time series features. However, its performance in this prediction is inferior to that of the BP model. It is speculated that the possible reasons for this are that the carbon emissions of the transportation industry in Beijing are affected by various factors such as policy regulation, changes in traffic flow, and energy structure adjustment. Temporal correlation is not the only core influencing factor, and the complex structure of LSTM models may lead to slight overfitting in small sample datasets, thereby reducing the prediction accuracy.

As a simplified version of the LSTM model, the GRU model reduces the number of gating units. Although it lowers the model complexity and training cost, it also loses some temporal feature capture capabilities. Therefore, its predictive performance is slightly inferior to that of the LSTM model. Although the Transformer model can capture long-distance temporal dependencies through its self-attention mechanism, it has a high requirement for data volume. However, this study focuses on the carbon emissions of the urban transportation industry, which leads to the inability of the self-attention mechanism to fully play its role. At the same time, the high complexity of the model is prone to overfitting, thereby resulting in the worst prediction performance. It is not applicable to this carbon emission prediction task.

E. Prediction and Result Analysis of Combined Models

Based on the comparison of the prediction performance of the four single-item models, BP, LSTM, GRU, and Transformer, in the previous section, the LSTM and BP models performed better in the prediction of carbon emissions in the transportation industry of Beijing. To fully leverage the advantages of both, this section constructs a combined model integrating temporal feature extraction and static feature extraction (LSTM+Attention and BP in parallel), with the aim of enhancing prediction accuracy and robustness through multi-dimensional information collaboration, providing a more reliable theoretical basis for the analysis of carbon reduction paths in Beijing's transportation.

1. Overall Framework of the Model

The core idea of the combined model lies in the collaborative modeling of temporal features and static features: the dynamic law of carbon emission data evolution over time is mined through the LSTM-Attention branch, the direct correlation between influencing factors and carbon emissions at a specific time point is captured through the BP neural network branch, and finally the prediction result is output by integrating the dual-branch information through the fusion layer. The model construction and prediction process is as follows:

Firstly, the original data set is preprocessed to separate the time dimension, dependent variable and independent variable. Min-Max standardization is adopted to eliminate the influence of dimensions, and the standardizer is retained for subsequent de-standardization. To meet the time series input requirements of LSTM, the data is reconstructed into a multi-input single-output time series sample structure: all the independent variable features of the previous N time steps are used as time series branch inputs, and the carbon emissions at the N+1 th time step are taken as the prediction target; Meanwhile, the independent variable features of the last time step of each time series sample are extracted and used as the static input of the BP branch.

Secondly, construct a dual-branch model. The time series branch adopts the LSTM network to extract the time series dependency of historical data, and superimposes the Attention mechanism to enhance the weight of key time step features. The static branch uses the BP neural network to process the independent variable features of the current time step. The dual-branch output is spliced at the fusion layer and integrated through a fully connected network, ultimately outputting the predicted carbon emission value.

Secondly, grid search is adopted to traverse and optimize the key hyperparameters, and the average absolute percentage error is used as the loss index to select the optimal parameter combination. The training process uses the Adam optimizer to minimize the mean square error and introduces an early stop mechanism to prevent overfitting.

Finally, three indicators, namely the root mean square error, the mean absolute error, and the mean absolute percentage error, are calculated on the test set to evaluate the model accuracy, and the prediction results are restored to the original dimension through de-normalization. Draw a comparison curve between the true value and the predicted value to visually display the model fitting effect.

2. Temporal Feature Extraction Branch: LSTM-Attention

The temporal branch uses the time series data of time window length × number of independent variables as inputs, and realizes the modeling of long-term temporal dependence through the gating mechanism and cell state of the LSTM network. The LSTM dynamically updates the cell state through the forgetting gate, input gate, and output gate, outputting the hidden state for each time step in the form of [number of samples, length of time window, number of LSTM hidden cells].

To enhance the contribution of key historical information to current predictions, an additive Attention mechanism is superimposed after the LSTM layer. Specifically, perform a linear transformation on the hidden states of all time steps output by the LSTM, calculate the similarity with the learnable query vector, and obtain the attention weights of each time step through Softmax normalization. Subsequently, the hidden states of each time step are weighted and summed, and the weighted time series feature vector is output, with the shape of [number of samples, number of LSTM hidden units]. This vector centrally reflects the comprehensive information of the historical time steps that contribute significantly to the prediction. Finally, a Dropout layer is added after the Attention layer (the dropout probability is determined by the grid search) to prevent overfitting of temporal branches.

3. Static Feature Extraction Branch: BP Neural Network

Static branches aim to capture the direct correlation between influencing factors and carbon emissions at the current time step, making up for the deficiency of temporal branches in modeling immediate features. This branch adopts a two-layer fully connected network: The first layer maps the features of the input independent variables to the higher dimension bp_units and introduces nonlinearity through the ReLU activation function; The second layer reduces the high-dimensional features to bp_units/2 and extracts the core static features; Then a Dropout layer is added to prevent overfitting. The final output is the core static feature vector, providing immediate information support for subsequent feature fusion.

4. Feature Fusion and Prediction Output

The fusion layer is the decision-making core of the composite model. First, the weighted temporal feature vectors output by the temporal branch and the core static feature vectors output by the static branch are dimensionally concatenated to form a fused feature vector, ensuring that the temporal dynamic information and the static real-time information are completely retained. Subsequently, a two-layer fully connected network is adopted to deeply integrate the fusion features: The first layer maps the fusion features to the fusion_units dimension, activates them by ReLU and applies Dropout; The second layer maps the integrated features to one dimension and outputs the final predicted value of carbon emissions. The nonlinear activation and dimension adaptation design of the fusion layer ensures the effective interaction between temporal and static features, avoiding the limitations of a single feature dimension.

Through the above structure, the combined model has achieved multi-dimensional and precise modeling of carbon emissions in Beijing's transportation industry, providing a reliable tool for subsequent scenario prediction.

F. Comparative Analysis of Prediction Results between single-item Models and combined models

Table 2:Comparison of Model Prediction Results

	BP	GRU	LSTM	Transformer	bp-lstm-attention
MAPE	11.91%	21.24%	17.37%	22.75%	10.46%
MAE	321.48	604.56	502.26	642.11	276.97
RMSE	369.24	670.35	587.05	694.56	351.64

To verify the improvement in the predictive performance of the combined model BP-LSTM-Attention, this paper takes MAPE, MAE, and RMSE as core indicators and conducts a horizontal comparison with four single-item models, namely BP, GRU, LSTM, and Transformer. The results are shown in Table 2.

It can be seen from Table 2 that the combined model outperforms all the individual models in all three indicators. In terms of the MAPE index, the combined model is 10.46%, which is 1.45 percentage points lower than the single-item optimal BP model. The relative error is further reduced, indicating that the combined model can more accurately capture the changing patterns of carbon emissions by integrating the advantages of multiple models and introducing an attention mechanism. In terms of the MAE index, the combined model is 44.51 lower than the BP model, 225.29 lower than the LSTM, 327.59 lower than the GRU, and 365.14 lower than the Transformer. The mean absolute deviation is significantly reduced, and the prediction results have a higher degree of fit. In terms of the RMSE indicator, the combined model decreased by 17.6 compared to the BP model, and the decline was more significant than that of other models, indicating that it has a stronger adaptability to extreme fluctuations in the data and better prediction stability.

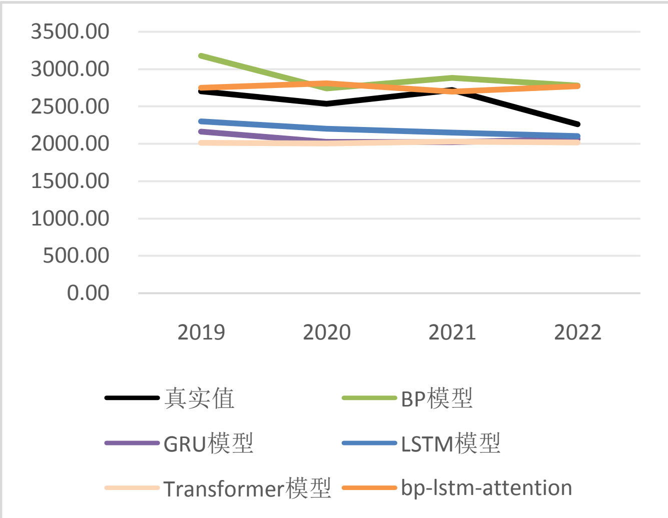

Figure 2 : Comparison of Predicted Values and Actual Values

The figures above present a comparison between the predicted values of each model and the actual values. Overall, the prediction performance of the four individual models ranks as follows: BP > LSTM > GRU > Transformer. The superior performance of the combined model is primarily attributed to its breakthrough in overcoming the feature extraction limitations of single models: the BP branch provides nonlinear fitting capability, the LSTM branch captures temporal dependencies, and the Attention mechanism enhances the weighting of key features. Among the individual models, the BP model, with its simple structure and strong nonlinear fitting ability, performs the best; the LSTM, while capable of capturing temporal sequences, does not integrate nonlinear advantages, resulting in slightly inferior performance; the GRU's simplified gating mechanism leads to some loss of temporal information; and the Transformer, which requires a relatively large sample size, struggles to fully leverage the advantages of its self-attention mechanism with limited data.

In summary, the combined model, through collaborative multi-dimensional feature modeling, significantly improves prediction accuracy and robustness, making it more suitable for accurately predicting carbon emissions from Beijing's transportation industry.

CONCLUSIONS AND PROSPECTS

A. *Main Conclusions*

This paper takes the transportation industry in Beijing as the research object. Based on a systematic review of domestic and international carbon emission accounting and prediction methods, the "top-down" emission factor method is adopted to calculate the carbon dioxide emissions of the transportation industry in Beijing from 1997 to 2023. Through the statistics of literature frequency, 25 influencing factors were screened out. Four deep learning models, namely BP, GRU, LSTM and Transformer, were selected for the comparison of prediction performance. Then, a combined prediction model integrating temporal feature extraction (LSTM+Attention) and static feature extraction (BP) was constructed. The main conclusions are as follows:

In the comparison of the predictive performance of individual models, the BP model and the LSTM model perform relatively better. The BP model, with its strong nonlinear fitting ability, can effectively capture the complex mapping relationship between carbon emissions and influencing factors. The LSTM model, through the design of the gated structure, has handled the long-term dependence characteristics of the carbon emission time series quite well. The GRU model loses some temporal information due to the simplification of the gating structure, and the Transformer model is difficult to fully leverage the advantages of the self-attention mechanism under the condition of limited samples. The prediction accuracy of both is relatively low.

(2) The BP-LSTM-Attention combined model constructed in this paper significantly outperforms all individual models in the three evaluation indicators. Among them, the MAPE value was 10.46%, which was 1.45 percentage points lower than that of the single-item optimal BP model. The MAE value was 2.7697 million t, which was 445,100 t lower than that of the BP model. The RMSE value was 3.5164 million t, which was 176,000 t lower than that of the BP model. This result indicates that the combined model achieves the collaborative modeling of temporal dynamic features and static real-time features through a dual-branch design. The introduction of the attention mechanism further strengthens the contribution weight of key historical information, effectively improving the prediction accuracy and robustness.

(3) The core reason why the combined model can achieve performance breakthroughs lies in breaking through the feature extraction limitations of a single model: the LSTM-Attention branch focuses on mining the dynamic laws of carbon emissions evolving over time, the BP branch captures the direct correlation between influencing factors and carbon emissions at a specific time point, and the fusion layer realizes the deep interaction of the two types of features. This architectural design provides an effective technical solution for the precise prediction of carbon emissions in Beijing's transportation industry.

B. *Research Prospects*

Although the combined model constructed in this paper has achieved good results in the carbon emission prediction of the transportation industry in Beijing, there are still the following directions worthy of in-depth exploration:

(1) Expansion of data dimensions and granularity. Due to the limitations of the statistical scope of energy consumption, this paper adopts the "top-down" method for calculation and fails to break down the carbon emission characteristics of different transportation modes such as roads, railways, and aviation, as well as different vehicle models. In the future, by integrating micro-data such as traffic flow monitoring data and the penetration rate of new energy vehicles, a more refined "bottom-up" accounting system can be constructed to further enhance the quality of data input into the model.

(2) Improvement of the influencing factor system. Based on the frequency of literature, this article has screened out 25 influencing factors, but some key variables may still be overlooked, such as the application level of intelligent transportation technology, changes in travel behavior patterns, and the impact of public emergencies. Subsequent research can introduce methods such as text mining and policy quantification to construct a more comprehensive system of influencing factors.

(3) Enhanced interpretability of the model. Although deep learning models have relatively high prediction accuracy, their "black box" feature limits the transparency of the decision-making process. In the future, interpretability analysis tools such as SHAP can be combined to quantify the marginal contribution of each influencing factor to the prediction results, providing a more targeted theoretical basis for the formulation of emission reduction policies.

(4) Multi-scenario prediction and policy simulation. This article focuses on model construction and performance verification, and has not yet conducted scenario predictions for future carbon emissions. Subsequently, multiple development paths such as benchmark scenarios, low-carbon scenarios, and enhanced low-carbon scenarios can be set in combination with Beijing's "14th Five-Year Plan" and the long-term goals for 2035. The combined model constructed by this research can be used to conduct scenario simulations, providing scientific support for the design of carbon peaking paths and emission reduction policies in Beijing's transportation industry.

References

[1] Wu Y, Yin Y M, Xiao Y, et al. Driving Factors and Projection of Carbon Emissions from Transportation in Guangzhou under the ‘Dual Carbon’ Goals[J]. Research of Environmental Sciences, 2025, 38(10): 2147-2154. DOI:10.13198/j.issn.1001-6929.2025.08.02.

[2] Jiang Z R, Jin H H, Wang C J, et al. Measurement and Efficiency Pattern of Transport Carbon Emissions in the Yangtze River Economic Belt (1985–2016)[J]. Environmental Science, 2020, 41(6): 2972-2980. DOI:10.13227/j.hjkx.201910221.

[3] Feng J, Zhang L Z, Han C, et al. Research on Carbon Emission Simulation and Reduction Pathways of Urban Transportation in Beijing Based on System Dynamics[J]. Journal of Shenyang University of Technology (Social Science Edition), 2025, 18(4): 407-418.

[4] Xie Z Y, Yang P, Wang Z X, et al. Identification of Driving Mechanisms of Urban Road Traffic Carbon Emissions Based on Interpretable Machine Learning[J/OL]. Environmental Science, 1-13 [2026-02-27]. https://doi.org/10.13227/j.hjkx.202505277.

[5] Shen Y, Wu T R, Yan J, et al. Research on Air Pollutants and Carbon Dioxide Emissions from Vehicles in Beijing Based on COPERT Model[J]. Journal of Environmental Engineering Technology, 2021, 11(6): 1075-1082.

[6] Song X W, Gao N X, Hao Y P, et al. Analysis of Influencing Factors and Trend Prediction of Carbon Emissions in Hebei Province Based on STIRPAT Model[J/OL]. Environmental Science, 1-26 [2026-03-04]. https://doi.org/10.13227/j.hjkx.202509030.

[7] Song X H, Zhang B J. Prediction of Regional Energy Green Development Based on ARIMA and Markov Matrix[J]. Computer Simulation, 2025, 42(11): 124-128, 220.

[8] Peng J T, Yuan Y T, Kang X, et al. Decomposition of Influencing Factors and Coupling Effects of Agricultural Net Carbon Sinks in Guizhou Province[J/OL]. Environmental Science, 1-14 [2026-03-04]. https://doi.org/10.13227/j.hjkx.202503354.

[9] Yang L S, Wang H L. Prediction of China's Sustainable Aviation Fuel Demand and Carbon Emissions Based on Deep Learning LSTM Model[J]. Coal Economic Research, 2026, 46(1): 130-139. DOI:10.13202/j.cnki.cer.2026.01.14.

[10] Xu D, Zhou X Y, Yao Z X, et al. Prediction of China's Medium- and Long-term Carbon Prices under a Multi-dimensional Coupling Model[J]. Price: Theory & Practice, 2025(11): 151-157. DOI:10.19851/j.cnki.CN11-1010/F.2025.11.425.

[11] Jin Y J, Jin X B, Hong X M, et al. Carbon Emission Prediction in Hainan Province Based on Lasso-Transformer Neural Network Model[J]. Environmental Science, 2026, 47(2): 781-792. DOI:10.13227/j.hjkx.202501263.

[12] Han Y, Zhou Y L, Wang S H. Scenario Simulation of Carbon Emissions and Emission Reduction Pathways in Industry-oriented Cities: A Case Study of Lanzhou[J/OL]. Environmental Science, 1-17 [2026-03-04]. https://doi.org/10.13227/j.hjkx.202506302.