http://dx.doi.org/10.1016/j.jclepro.2024.142053

Forecasting Carbon Peaking in China Using Data-Driven Rule-Base Model: An In-depth Analysis Across Regional and Economic Scenarios

Long-Hao Yang [a], Yu-Qiong Lei [a], Fei-Fei Ye [b, *], Haibo Hu [c], Haitian Lu [d], Ying-Ming Wang [a]

[a] Decision Sciences Institute, Fuzhou University, Fuzhou, 350116, PR China

[b] School of Cultural Tourism and Public Administration, Fujian Normal University, Fuzhou, 350117, PR China

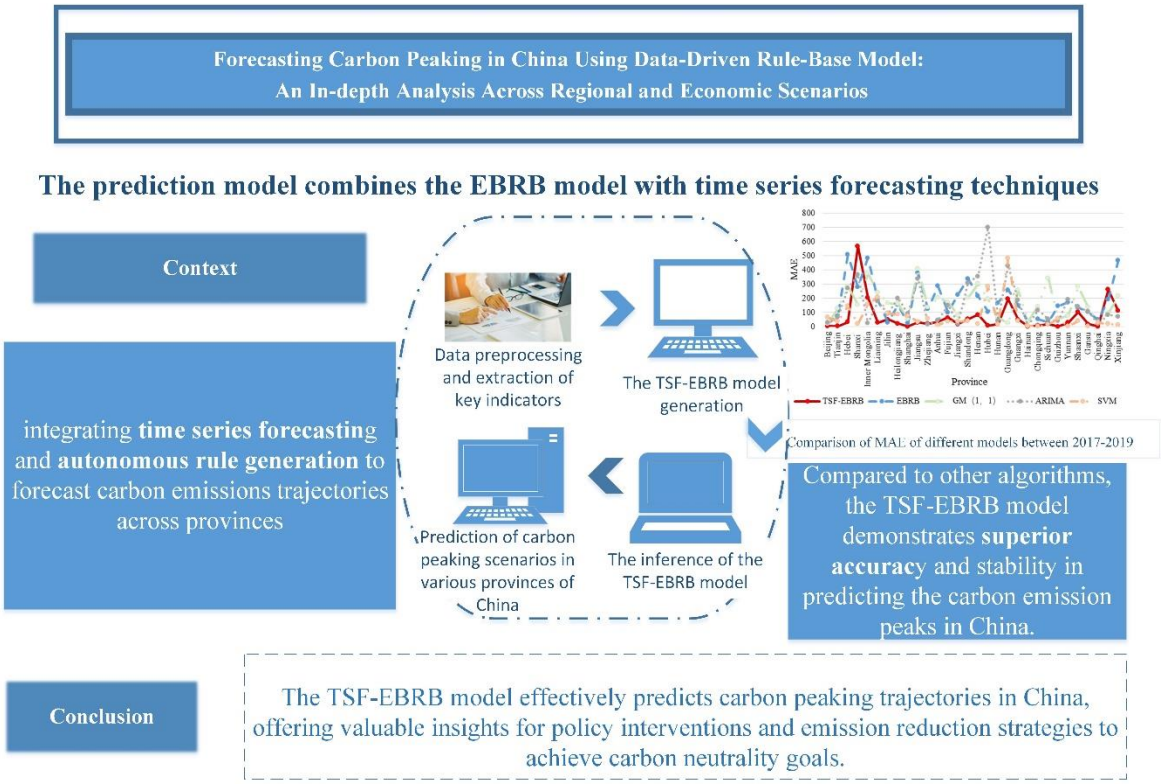
[c] Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, PR China

[d] School of Accounting and Finance, The Hong Kong Polytechnic University, Hong Kong, PR China

[*]Corresponding Author: 13075810934@163.com

Abstract: At the 2020 United Nations Climate Summit, China officially announced its goal to achieve carbon peaking by 2030. Exploring whether it is possible to reach the peak of carbon emissions earlier necessitates an urgent and imperative need for precise long-term forecasting of China's carbon emissions dynamics. However, the current predictions of carbon peaking mostly depend on mechanical or mathematical models, which failed to consider the interdependence between carbon emissions and the time series-based patterns existed in carbon emission data. Therefore, this study presents a novel carbon peaking prediction method based on time series forecasting with autonomous rule generation, which is implemented by the adaptation of the extended belief rule base (EBRB) model designed for handling time series data and thus it is referred to as the time series forecasting EBRB (TSF-EBRB) model. The TSF-EBRB model not only captures and measures the temporal correlations within the data throughout the processes of modeling and inference, but also consists of a novel parameter optimization model based on the temporal correlations. The study collected carbon emission data from 30 provinces in China for empirical analysis. It computed and predicted the carbon peaking trajectories of each province under three different scenarios from 2022 to 2030, validating the effectiveness and superiority of the TSF-EBRB model. The results indicate that, with policy interventions, the majority of provinces are projected to reach carbon peaking before 2030. Finally, based on the predicted outcomes under different scenarios, recommendations for carbon emission reduction are proposed.

Graphical abstract



Keywords: Carbon peaking prediction; Autonomous rule generation; Time series forecasting; Extended belief rule base

1. Introduction

Time series forecasting involves the utilization of historical time series data and domain-specific prior knowledge to explore and analyze various time series patterns, including periodicity, volatility, and trends. Subsequently, these identified patterns are employed to predict outcomes in future periods. In the field of environmental science, time series forecasting is widely employed to investigate aspects, such as climate change, the frequency of natural disasters, and ecosystem variations. However, in today's industrialized society, carbon emissions have become an exceedingly critical subject due to their direct relevance to climate change(Jian-Bo et al., 2006; Sippel et al., 2020), environmental conservation(Wen et al., 2021), and sustainable development(González-Torres et al., 2021).

The issue of global climate change is becoming increasingly severe, with carbon dioxide emissions being one of the main sources of greenhouse gas emissions. Controlling carbon dioxide emissions is of paramount importance for mitigating climate change. In 2019, global carbon dioxide emissions reached 36.4 billion tons, with China's total carbon dioxide emissions increasing from 9.085 billion tons in 2015 to 10.2 billion tons, accounting for 27.9% of global emissions, making it the world's largest emitter of greenhouse gases. In the face of the challenges posed by global climate change, the Chinese government has recognized the urgency of emissions reduction. In accordance with the Paris Agreement, China updated and strengthened its nationally determined contributions in 2020, committing to peak carbon emissions by 2030 and striving to achieve carbon neutrality by 2060. This goal entails a comprehensive transition to a green economy through various energy-saving and emission reduction policies. In this context, accurately predicting China's future carbon emissions trends becomes particularly important. Accurate predictions of future carbon emissions can assist governments(Qing et al., 2024) in discerning the anticipated trajectory of carbon emissions. This foresight empowers governments to adjust policy directions, promote green technology innovation, and guide industrial restructuring(Shen et al., 2023) to achieve carbon emissions goals. Moreover, businesses can strategically invest in green technologies(Qing et al., 2023) and sustainable development projects based on carbon emission forecasting data, thereby enhancing their environmental performance(Dagestani & Qing, 2022) and strengthening their competitiveness.

Despite the widespread application of time series forecasting in environmental science, there are still some limitations in the decomposition of carbon emission factors and the forecasting of carbon emissions: 1) The factors considered in the decomposition of carbon emission factors are overly rigid. Because of the evolving dynamics of economy, society, and technology, the primary factors, such as population and economy, influencing carbon emissions are no longer confined to conventional factors. Carbon emissions are usually influenced by a range of factors, including regional energy resource endowments and the stage of forestry development, among others. 2) In the forecasting of carbon emissions, the majority of studies relied on classical regression models. However, due to the intercorrelation between carbon emission factors, there exists a high degree of covariance among variables. In other words, a functional relationship exists between the variables themselves, making it challenging to accurately analyze the relationship between variables and outputs. As a consequence, this impacts the accuracy of regression prediction models. 3) Traditional time series forecasting models, including ARIMA and the gray prediction model, have relatively low data requirements and can quickly uncover simple nonlinear patterns within data. However, these methods struggle to accurately capture complex nonlinear characteristics.

In view of the limitation of time series prediction in the environmental science field, this study proposes a novel time series forecasting model based on the extended belief rule base (EBRB) model(Calzada et al., 2015) . Presently, conventional forecasting methods tend to be overly rigid in decomposing carbon emission factors, failing to adapt dynamically to changes in economic, social, and technological factors. Additionally, these traditional models struggle to effectively manage the

uncertainties inherent in carbon emission data and accurately capture nonlinear relationships, resulting in suboptimal prediction accuracy. In contrast, EBRB model, grounded in rule-based reasoning, exhibits robust flexibility and adaptability, effectively managing uncertainties and variations among factors, thereby enhancing model stability and reliability. Leveraging its rule-based reasoning mechanism enables the EBRB model to capture nonlinear characteristics and intricate relationships within carbon emission data, further improving prediction accuracy. Thus, selecting EBRB as the foundational model is pivotal in overcoming the challenges inherent in current time series forecasting, elevating the accuracy and stability of carbon emission predictions, and contributing novel theoretical and methodological advancements to the field of environmental science forecasting.

The proposed model is therefore referred as time series forecasting EBRB (TSF-EBRB) model. In the modeling process, the TSF-EBRB model firstly considers the significance of data variables to propose an effective method for identifying key variables; Secondly, the temporal correlation is defined as an additional parameter of TSF-EBRB model to improve the efficacy of handling time series data; Thirdly, a corresponding parameter optimization method is proposed based on time-related correlations to further enhance the accuracy of TSF-EBRB model; Finally, the TSF-EBRB model with the parameter optimization method are applied to handle carbon peaking prediction.

In the empirical analysis, the environmental indicator data and carbon emission data of 30 Chinese provinces from 2004 to 2021 are collected as the benchmark dataset to verify the effectiveness of the TSF-EBRB model. These provinces cover a vast geographical area and exhibit diverse levels of economic development, which are crucial for understanding China's carbon emission trends. Provinces in the eastern coastal region, such as Beijing, Shanghai, and Guangdong, typically have high levels of industrialization and population density, resulting in higher carbon emissions that significantly impact the national total. In contrast, provinces in the central and western regions, such as Sichuan, Guizhou, and Inner Mongolia, primarily rely on resource-based industries, and their carbon emissions are greatly influenced by energy structure and regional development levels. Therefore, in-depth research into the carbon emissions of these provinces can provide scientific evidence for formulating regional carbon reduction policies and measures, as well as crucial references for China's carbon peak prediction and emission reduction goals. By applying the TSF-EBRB model to predict carbon emissions in these provinces, we can gain a comprehensive understanding of regional disparities and development trends in China's carbon emissions, thereby offering more targeted strategic and policy recommendations to address climate change.

Consequently, a total of 30 independent TSF-EBRB models are construct to predict the carbon peaking of 30 Chinese provinces, where these 30 models are specifically designed to account for the distinct features of each province and the indicated models are used to assess the efficacy and precision of tackling carbon emissions prediction. Additionally, three kinds of scenarios entitled baseline, low-carbon, and technology development scenarios, four economic areas, and six geographic regions are considered to assess the feasibility of using the TSF-EBRB model to predict the carbon peaking of 30 Chinese provinces at 2030. The outcomes of the three scenarios provide valuable insights into early carbon peaking realization in China.

Compared with the existing carbon peaking prediction models, the TSF-EBRB model has the following advantages:

(1) The TSF-EBRB model is able to identify key carbon emission indicators while mitigating information redundancy. The use of all carbon emission indicators not only have negative impact modeling, but also results in substantial redundancy within the dataset. For this reason, the TSF-EBRB model employs an analytical approach to evaluate the correlations among indicators to effectively select the most essential carbon emission indicators.

(2) The TSF-EBRB model incorporates the time correlation of data in the process of rule generation and rule inference, thereby significantly enhancing the precision accuracy of time series data. Additionally, the TSF-EBRB model incorporates a data-driven method for parameter optimization, which aims to mitigate the problem of human subjective effect on parameter values. This further bolsters the predictive accuracy of the TSF-EBRB model.

(3) The core of the TSF-EBRB model is an interpretative expert system, which employs extended belief rules for the representation and storage of diverse information, and utilizes evidential reasoning (ER) method to produce final prediction outcomes. Notably, the TSF-EBRB model not only provides a comprehensive explanation of its operational principles, but also excels in attaining high prediction accuracy, particularly dealing with nonlinear data.

(4) The TSF-EBRB model exhibits significant theoretical innovation. By incorporating the temporal correlation of time series data and employing data-driven parameter optimization methods, it accurately captures the dynamic patterns inherent in the data during the prediction process. Compared to traditional models, the TSF-EBRB model can more precisely forecast carbon emission peaks, thereby enhancing the accuracy and stability of the prediction results.

The method employed in this study exhibits superiority over traditional approaches for several reasons. By integrating time series analysis, it captures temporal data dynamics more effectively, thereby enhancing prediction accuracy. Additionally, a data-driven parameter optimization method is utilized, automatically adjusting model parameters for improved adaptability to diverse datasets, thereby enhancing prediction accuracy and stability. Furthermore, the introduction of an interpretative expert system and extended belief rules enhances the model's interpretability and the credibility and comprehensibility of prediction results. Moreover, the method comprehensively analyzes data from different regions or provinces and tailors predictions based on their unique characteristics, providing more accurate decision-making guidance. Through these enhancements, the method proves to be more adept at forecasting future data trends and offering reliable decision-making guidance.

The main theoretical contribution of this study lies in the introduction of the TSF-EBRB model, which combines with the traditional EBRB model to improve the accuracy and stability of time series prediction. While the traditional EBRB model can handle multidimensional data, it overlooks the temporal relationships among data, potentially leading to inaccuracies in time series prediction. In contrast, the TSF-EBRB model incorporates time series analysis to better capture the temporal dynamics of the data, thus enhancing prediction accuracy.

In the context of carbon emission prediction, the traditional EBRB model may fail to accurately capture trends in carbon emissions due to its disregard for temporal dynamics, resulting in significant prediction errors. However, the TSF-EBRB model, by considering the temporal correlation characteristics, can more accurately predict future trends in carbon emissions. For example, if a province's carbon emissions have been increasing annually in recent years, the TSF-EBRB model can better capture this trend and make more accurate predictions for the future. Therefore, the TSF-EBRB model makes a significant theoretical contribution by addressing the influence of temporal dynamics on data, thereby enhancing prediction accuracy and reliability compared to the traditional EBRB model.

The rest of this paper is organized below: Section 2 provides literature review and challenges; Section 3 introduces the EBRB model and the existing problems; Section 4 proposes an improved EBRB model and its application on carbon peaking prediction. Section 5 is the empirical analysis and the conclusions are summarized in Section 6.

2. Literature Review and Challenges

Carbon emission prediction has attracted the attentions of many scholars in the past decades, and thus various kinds of prediction models have been developed for accurately and efficiently predicting carbon emissions. From the previous studies, the existing prediction models can be roughly divided into four categories: econometric models, time series forecasting models, scenario analysis methods, and machine learning models.

In the area of carbon projections, econometric models have made remarkable progress, leveraging statistical methods to explore the relationships and effects of different factors. For example, the Kaya (González-Torres et al., 2021) provided a basic framework focusing on the relationship among population, economic growth and energy consumption and carbon emissions. The STIRPAT extended the Kaya (Huang et al., 2021) by introducing technology level and urbanization to comprehensively explain the changes in carbon emissions. Thereafter, due to the potential limitations of STIRPAT, namely the challenges in capturing complex carbon emission changes and considering lots of factors, an improved STIRPAT was proposed in order to explain the influencing factors of carbon emissions(Wu et al., 2021).

Thanks to the fact that the econometrics models are usually difficult to capture seasonal changes and trends in data, the time series forecasting models were introduced to predict carbon emissions. For example, the autoregressive moving average model (Yang & O'Connell, 2020) was suitable for processing carbon emission time series data with complex seasonal and stochastic components; the Grey model (Nie & Duan, 2023) was used to deal with the situation of limited data quality or lack of resources by transforming the original data into grey data. The exponential smoothing method (Wang et al., 2020), which relies on a weighted average of past observations, was used to capture trends and cyclical fluctuations in different types of data and was capable of processing carbon emissions data with trend and seasonal variations.

However, the time series forecasting models may not be able to capture the diversity of carbon emissions under different scenarios. the scenario analysis methods (Ma et al., 2019) were introduced to predict carbon emissions. The recent trends of applying the scenario analysis methods on carbon emission prediction include the consideration of uncertainties and the adoption of standardized methods (Niu et al., 2023). These developments enable scenario analysis to help governments, businesses and international organizations to better understand the future trends of carbon emissions, so as to effectively respond to climate change challenges and develop effective carbon emission policies and strategies.

Considering that the scenario analysis methods have some limitations in dealing with uncertainty and large-scale data, the machine learning models were introduced to capture non-linear relations and deal with high-dimensional data. Some common machine learning models include artificial neural networks and support vector machine. For example, the artificial neural network (Du et al., 2022; Zhang et al., 2022) can effectively capture the complex nonlinear relationships in the data of carbon emissions. the support vector machine (Xu et al., 2007) emphasis on finding optimal decision boundaries in high-dimensional space makes it powerful for dealing with Galway data and nonlinear problems.

Despite the increasing attentions on carbon emission prediction, there are several challenges needed to be overcome:

(1) The econometrics models are mainly based on straightforward mathematical formulas to estimate the coefficients of different factors, and then a regression analysis method is used to predict carbon emissions. Although models with simple structures and algorithmic implementations fall short in accounting for potential cross-impacts among various carbon emission factors. The weighting of coefficients usually fails to accurately capture the intricate relationships between variables, thereby compromising the final prediction's accuracy.

(2) Comparing with econometrics models, the machine learning models have better performance in handling carbon emission prediction. However, its modeling relies on manual features, and only considers the current time feature, which fails

to capture the time dependence coexisted in time series data. The carbon emission data, which is a time series dataset, demonstrates robust interconnections between current and past data, displaying notable temporal correlation. Consequently, it is unattainable to obtain a desired prediction result when the models are directly used to predict carbon emissions.

(3) With increasing prevalence of machine learning models and complex ensemble models, carbon emission prediction has transcended the confines of traditional prediction models. While these models can offer superior prediction accuracy, the 'black box' nature is still a significant drawback and needs to be overcome. Moreover, as the increase of model complexity, it is becoming a difficult target that there needs a clear explanation about the core modeling and inference processes of the carbon emission prediction models.

3. Introduction to the EBRB Model and Hypothesis Description

This study adopts EBRB as the foundational model to address the challenges of time series forecasting in the field of environmental science. Based on the theory of belief functions, the EBRB model effectively manages uncertainty, thereby enhancing model stability and reliability. Furthermore, the EBRB model features a flexible rule-based reasoning mechanism, enabling it to capture complex nonlinear relationships and data patterns, consequently improving prediction accuracy. Additionally, the EBRB model exhibits strong adaptability, capable of flexibly handling multi-source data and adjusting the model adaptively to suit different forecasting scenarios.

To effectively utilize the EBRB model for carbon peaking prediction, this section briefly reviews its rule representation, rule generation, and rule inference processes, as shown in Fig. 1. Furthermore, the challenges of applying the EBRB model to carbon peaking prediction are discussed.

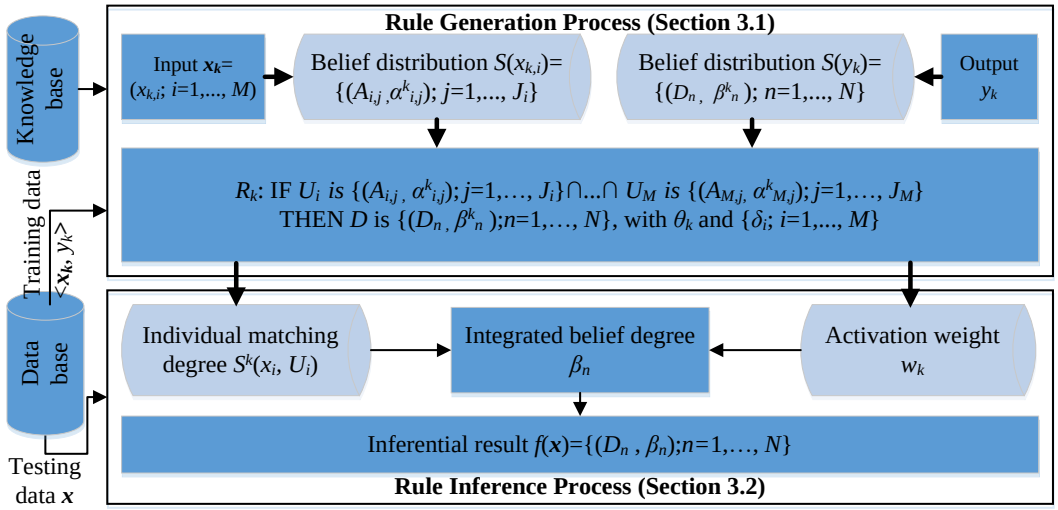


Fig. 1 The basic process of EBRB model.

3.1 Rule representation and rule generation process for the EBRB model

In an EBRB model, the EBRB serves as the rule base consisting of IF-THEN rules with belief distributions to save qualitative knowledge and quantitative data. Assuming that there are M antecedent attributes U_i (i=1,…, M) and one consequent attribute D, where the reference value of the ith antecedent attribute is $A_{i,j}$ (j=1,…, J_i) and the reference value of the consequent attribute is D_n (n=1,…, N), the kth (k=1,…, L) extended belief rule is denoted as follows:

$$\begin{aligned} &R_k : IF\ U_i\ is\ \{(A_{1,j}, \alpha^k_{1,j}); j=1,...,J_1\} \wedge ... \wedge U_M\ is\ \{(A_{M,j}, \alpha^k_{M,j}); j=1,...,J_M\} \\ &\text{THEN}\ D\ is\ \{(D_n, \beta^k_n); n=1,...,N\},\ \text{with}\ \theta_k, \{\delta_1,...,\delta_M\} \end{aligned} \tag{1}$$

where α^k_{ij} ($0 \leqslant \alpha^k_{i,j} \leqslant 1$) and β^k_n ($0 \leqslant \beta^k_n \leqslant 1$) denote the belief degree associated with the reference values $A_{i,j}$ and D_n in the kth rule, and these parameters satisfy $\sum_{i=1}^{J_i} \alpha^k_{i,j} \leq 1$ and $\sum_{n=1}^{N} \beta^k_n \leq 1$. If $\sum_{n=1}^{N} \beta^k_n = 1$, then the kth rule has complete information;

Otherwise, the kth rule contains uncertain information; Additionally, θ_k ($0<\theta_k\leqslant 1$) and δ_i ($0<\delta_i\leqslant 1$) denote the weights of the kth rule and the ith antecedent attribute, respectively.

The generation of extended belief rules is a vital component of the EBRB model and it relies on expert knowledge to initialize utility values $u(A_{i,j})$ for reference values $A_{i,j}$ (j=1,..., J_i), $u(D_n)$ for reference values D_n (n=1,…, N), and antecedent attribute weights δ_i (i=1,…, M). The specific steps are as follows:

***Step* 1**: To initialize utility values and attribute weights using expert knowledge. Before transforming the sample data into extended belief rules, it need to pre-determine the basic parameters of the EBRB model using expert knowledge. These parameters include utility value $u(A_{i,j})$ (j=1,..., J_i), $u(D_n)$ (n=1,…, N), and antecedent attribute weights δ_i (i=1,…, M).

***Step* 2**: To calculate the belief distribution of attributes. In this step, utility values are employed to convert sample data into belief distribution. Assuming that $x_{k,i}$ represents the input value of the ith (i=1,…, M) antecedent attribute in the kth (k=1,…, L) data, the belief distribution of the antecedent attribute is represented as follows:

$$S(x_{k,i})=\{(A_{1,j},\alpha_{1,j}^k);\ j=1,...,J_i\} \tag{2}$$

$$\alpha_{i,j}^k=\frac{u(A_{i,j+1})-x_{k,i}}{u(A_{i,j+1})-u(A_{i,j})}\ and\ \alpha_{i,j+1}^k=1-\alpha_{i,j}^k, if\ u(A_{i,j})\leq x_{k,i}\leq u(A_{i,j+1}) \tag{3}$$

where $\alpha_{i,j}^k$ denotes the belief degree of input data $x_{k,i}$ in reference value $A_{i,j}$. Noting that $\alpha_{i,j}^k$ is equal to 0 when t=1,…, J_i and $t\neq j$, j+1; Similarly, the belief distribution of consequent attributes can be obtained as:

$$S(y_k)=\{(D_n,\beta_n^k);n=1,...,N\} \tag{4}$$

***Step* 3**: To calculate rule weights. After transforming the sample data into belief distributions, the rules R_a and R_k are selected to calculate the similarity of rule antecedent (SRA) and the similarity of rule consequent (SRC) firstly:

$$SRA(R_a,R_k)=1-\max_{i=1,...,M}\left\{\sqrt{\sum_{j=1}^{J_i}(\alpha_{i,j}^a-\alpha_{i,j}^k)^2}\right\};a=1,...,L;a\neq k \tag{5}$$

$$SRC(R_a,R_k)=1-\sqrt{\sum_{j=1}^{J_i}(\beta_{\mathrm{n}}^a-\beta_n^k)^2};a=1,...,L;a\neq k \tag{6}$$

Next, the consistency of rules R_a and R_k is calculated by

$$Cons(R_a,R_k)=\exp\left(\frac{(\frac{SRA(R_a,R_k)}{SRC(R_a,R_k)}-1)^2}{(\frac{1}{SRA(R_a,R_k)})^2}\right) \tag{7}$$

Then, the inconsistency of the kth rule R_k with other rules is calculated by

$$ID(R_k)=\sum_{a=1,a\neq k}^{L}(1-Cons(R_a,R_k)) \tag{8}$$

Finally, the weight of the kth rule R_k is calculated by

$$\theta_k=1-\frac{ID(R_k)}{\sum_{j=1}^{L}ID(R_j)} \tag{9}$$

3.2 Rule inference process for the EBRB model

The rule inference process of the EBRB model is based on the Evidential Reasoning (ER) algorithm and it consists of calculating activation weights for all rules; integrating the belief distribution of all activated rules; and calculating final prediction results from the integrated belief distribution. The corresponding steps are as follows:

***Step* 1**: To calculate individual matching degrees. Assuming that the input data is $\boldsymbol{x}$=(x_1,…,x_M), the input data is first

converted into a belief distribution according to Eq. (3):

$$S(x_i) = \{(A_{i,j}, \alpha_{i,j}); J = 1,..., J_i\} \tag{10}$$

Next, the Euclidean distance is used to calculate the individual matching degree of the *i*th antecedent attribute:

$$S^k(x_i, U_i) = 1 - \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2} \tag{11}$$

where $a_{i,j}$ is the matching degree of x_i belonging to $A_{i,j}$, and $\alpha_{i,j}^k$ denotes the belief degree of the *k*th rule on the reference value $A_{i,j}$ of the *i*th antecedent attribute U_i.

***Step* 2**: To calculate activation weights. Based on the belief distribution $S^k(x_i, U_i)$, the antecedent attribute weight δ_i and rule weight θ_k are used to calculate the activation weight of the *k*th rule:

$$w_k = \frac{\theta_k \prod_{i=1}^{M} S^k(x_i, U_i)^{\overline{\delta_i}}}{\sum_{a=1}^{L} \theta_k \prod_{i=1}^{M} S^k(x_i, U_i)^{\overline{\delta_i}}}, \overline{\delta_i} = \frac{\delta_i}{\max\limits_{j=1,...,M} \delta_j} \tag{12}$$

***Step* 3**: To integrate the activated rules using the ER algorithm. If the activation weight of a rule is greater than 0, it needs to be integrated using the analytical ER algorithm below:

$$\beta_n = \frac{\prod_{k=1}^{L} (w_k \beta_n^k + 1 - w_k \sum_{i=1}^{N} \beta_i^k) - \prod_{k=1}^{L} (1 - w_k \sum_{i=1}^{N} \beta_i^k)}{\sum_{i=1}^{N} \prod_{k=1}^{L} (w_k \beta_n^k + 1 - w_k \sum_{j=1}^{N} \beta_j^k) - (N-1) \prod_{k=1}^{L} (1 - w_k \sum_{j=1}^{N} \beta_j^k) - \prod_{k=1}^{L} (1 - w_k \sum_{i=1}^{N} \beta_i^k)} \tag{13}$$

Finally, the inferential results of the EBRB model can be shown as follows:

$$f(\boldsymbol{x}) = \{(D_n, \beta_n); n = 1,..., N\} \tag{14}$$

3.3 Hypothesis toward carbon peaking prediction

To date, the EBRB model has shown significant achievements and also demonstrated superior performance in terms of prediction accuracy and model interpretability comparing to traditional machine-learning algorithms on many domains, such as pipeline leakage detection (Zhou *et al*., 2009), environmental management cost planning (Ye *et al*., 2020), and behavior prediction (Zhang *et al*., 2024). However, the EBRB model is still rarely used for forecasting carbon peaking. Moreover, due to the fact that forecasting carbon peaking involves diverse environmental indicators, the application of time series forecasting, and the requirement of high prediction accuracy, the following hypothesis should be defined to clarify how the accurate carbon peaking prediction is achieved by the EBRB model.

***Hypothesis* 1:** The use of key environmental indicators could improve the performance of the EBRB model for carbon emission prediction

Environmental indicators typically encompass diverse data dimensions, including economic factors, regional attributes, and governance investments. Using all these environmental indicators as inputs of the EBRB model may lead to a decreased computational efficiency due to data redundancy. Additionally, there exists the linear correlation among environmental indicators, with the truly critical information typically residing in a small subset of these indicators. Hence, this study proposes ***Hypothesis* 1** and aims to verify ***Hypothesis* 1** by extracting the most informative indicators from multidimensional data for constructing the EBRB model.

***Hypothesis* 2**: The combination of time series forecasting and the EBRB model could improve its performance for carbon emission prediction

The carbon emission data displays distinctive temporal patterns, whereby the present values are influenced by both the current data and past indicator data. Nonetheless, the inference process of the EBRB model fails to take into account the

temporal relationships among the data. When the EBRB model is directly applied to predict carbon emission data, it may disregard the temporal correlations within the data, thus impacting the model's predictive performance. Therefore, this study proposes ***Hypothesis* 2** and aims to verify ***Hypothesis* 2** by embedding the time-associated parameters into the EBRB model for carbon emission prediction.

***Hypothesis* 3**: Optimizing time-associated parameters could improve the performance of the EBRB model for carbon emission prediction

The EBRB model encompasses numerous parameters and their initial values have significant influence on prediction accuracy. This is reason why parameter optimization are typically used to improve the performance of the EBRB model. The time-associated parameters are the special parameters used to achieve the combination of the EBRB model and time series forecasting, which means that none of existing parameter optimization methods is available to optimize time-associated parameters for the EBRB model. Hence, this study proposes ***Hypothesis* 3** and aims to verify ***Hypothesis* 3** by optimizing the time-associated parameters of the EBRB model for carbon emission prediction.

4. Carbon Peaking Prediction based on Data-Driven Rule-Base Model

To address the issues outlined in Section 3.3, this section first develops an improved EBRB model and its parameter optimization method to achieve better time-series forecasting. Subsequently, a new carbon peaking prediction method is proposed based on the improved EBRB model.

4.1 An improved EBRB model for time series forecasting

Time series data usually exhibits the characteristics of autocorrelation and mutual correlation, which means that a given data is inherently linked to its historical data when the given data belongs to time series data(Koopmans, 2021). For the aim of enhancing the accuracy of time series forecasting, the time-associated parameters are designed to improve traditional machine learning algorithms by capturing and reflecting the hidden information of these characteristics, such as recurrent neural networks (RNN) (Sherstinsky, 2020) and long short-term memory (LSTM) (Bukhari et al., 2020). From the view of this point, this section proposes an improved EBRB model that contains the time-associated parameter entitled correlation probability for time series forecasting, where the improved EBRB model is therefore referred to as TSF-EBRB model. Based on the above description, the specific modeling steps for the TSF-EBRB model are shown below:

***Step* 1**: To select key variables using time series data. Assuming that the time series data are collected from duration T and the data at time t (t=1,…, T) is the $<\boldsymbol{x}^t=(x_s^t; s=1,\ldots, S), y^t>$, where x_s^t denotes the value of the sth variable at time t. Thus, the key variables of time series data can be selected by the following sub-steps:

***Step* 1.1**: To generate standardized matrix from time series data. In this step, the decentralization operation is performed on time series data $\{(x_s^t; s=1,\ldots, S); t=1,\ldots, T\}$ to generate standardized matrix $\boldsymbol{Z}=(z_s^t)_{T\times S}$ below:

$$z_s^t = \frac{x_s^t - \bar{x}_s}{\sigma_s}; s = 1,...,S; t = 1,...,T \tag{15}$$

where $\bar{x}_s$ denotes the mean of the sth variable; σ_s denotes the standard deviation of the sth variable.

***Step* 1.2**: To generate correlation coefficient matrix from the standardized matrix. Based on the standardized matrix $\boldsymbol{Z}$, the correlation coefficient matrix $\boldsymbol{R}=(r_{ij})_{S\times S}$ can be generated below:

$$r_{ij} = \frac{1}{n-1}\sum_{t=1}^{T} z_i^t z_j^t, (i, j = 1,..,S) \tag{16}$$

where r_{ij} ($r_{ij} = r_{ji}$) denotes the correlation coefficient of the ith variable related to the jth variable.

***Step* 1.3**: To calculate eigenvalues and eigenvectors from correlation coefficient matrix. From the correlation coefficient

matrix $\boldsymbol{R}$, S eigenvalues and S eigenvectors can be calculated by using singular value decomposition method, denoted as $\boldsymbol{\lambda}=\{\lambda_i; i=1,\ldots, S\}$ and $\boldsymbol{v}=\{\boldsymbol{v_i}=(v_{i,1},.., v_{i,S}); i=1,\ldots, S\}$, respectively。

***Step* 1.4**: To identify key variables from eigenvalues and eigenvectors. First, the order of the S eigenvalues and S eigenvectors are sorted from biggest to smallest according to the eigenvalues, denoted as $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_S$; Subsequently, the number of key variables is determined by the cumulative contribution rate below:

$$\frac{\sum_{j=1}^{M} \lambda_j}{\sum_{j=1}^{S} \lambda_j} \geq \sigma \tag{17}$$

where σ denotes the minimum cumulative contribution rate based on prior knowledge, and M denotes the number of key variables that satisfy the minimum cumulative contribution rate. Afterward, the key variables are selected by considering the corresponding eigenvectors associated with M key variables below:

$$I(\boldsymbol{v_i}) = \arg\max_{s=1,\ldots,S}\{v_{i,s}\}; i = 1,\ldots,M \tag{18}$$

where $I(\boldsymbol{v_i})$ denotes the index of the key variable selected from the ith eigenvectors; for the sake of convenient representation, assuming that the time series data at time t (t=1,…, T) related to the M key variables is denoted as $<\boldsymbol{x^t}=(x_i^t; i=1,\ldots, M), y^t>$.

***Step* 2**: To construct TSF-EBRB. Firstly, the M key variables are used as antecedent attributes and the variable of y^t as consequent attribute to form the basic framework of EBRB; Secondly, according to expert knowledge, the utility values $\{u(A_{i,j}); i=1,\ldots, M; j=1,\ldots, J_i\}$ of M antecedent attributes and the utility values $\{u(D_n); n=1,\ldots, N\}$ of consequent attribute are provided, where J_i denotes the number of reference values at the ith antecedent attribute; N denotes the number of consequents at the consequent attribute; Thirdly, an EBRB can be constructed from T time series data $\{<\boldsymbol{x^t}, y^t>; t=1,\ldots, T\}$ based on the rule generation process in Section 3.1; Finally, P time-associated parameters entitled correlation probability v_p (p=1,…, P) are designed to upgrade EBRB as TSF-EBRB, where v_p denotes the probability of the current time series data related to the time series data at previous pth moment.

***Step* 3:** To generate prediction results by considering time-associated parameters. Assuming that the data needed to be predicted is denoted as $\boldsymbol{x^{T+1}}=(x_i^{T+1}; i=1,\ldots, M)$. Based on the TSF-EBRB constructed in Step 2, the corresponding prediction result considering P correlation probabilities $\{v_p; p=1,\ldots, P\}$ can be generated as follows:

***Step* 3.1:** To calculate time-associated activation weights. Taking into account the influence of P correlation probabilities on the rule activation process, the activation weight of the kth rule at the current moment should be calculated by the activation weights at the previous P moments below:

$$g_k = \sum_{p=1}^{P} w_k^p v_p \tag{19}$$

where w_k^p denotes the activation weights of the kth rule at the previous pth moment; g_k denotes the time-associated activation weight of the kth rule. The process of calculating g_k is shown in Fig. 2.

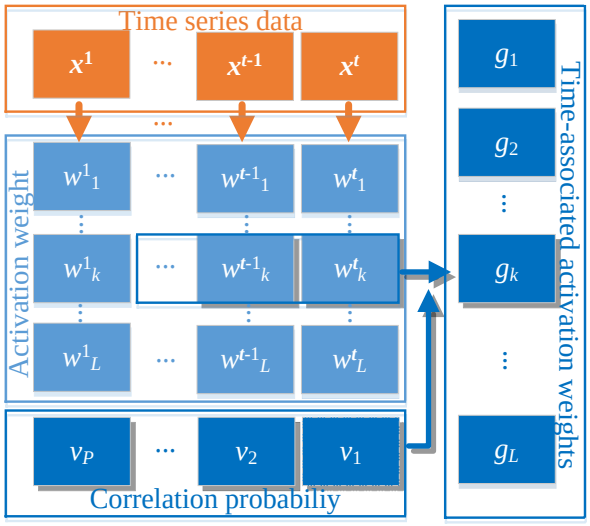


Fig.2 Illustration of calculating time-associated activation weights

***Step* 3.2**: To calculate time-associated belief degrees. Firstly, all rules with time association activation weights greater than 0 are regarded as activated rules. Without loss of generality, assume that there are L activated rules, the time-associated belief degrees can be calculated by using the analytical ER algorithm:

$$\beta_n = \frac{\prod_{k=1}^{L}(g_k\beta_n^k + 1 - g_k\sum_{i=1}^{N}\beta_i^k) - \prod_{k=1}^{L}(1 - g_k\sum_{i=1}^{N}\beta_i^k)}{\sum_{i=1}^{N}\prod_{k=1}^{L}(g_k\beta_n^k + 1 - g_k\sum_{i=1}^{N}\beta_i^k) - N\prod_{k=1}^{L}(1 - g_k\sum_{i=1}^{N}\beta_i^k)} \tag{20}$$

where β_n (n=1,…, N) denotes the time-associated belief degree of the nth consequent.

***Step* 3.3**: To calculate time-associated prediction results. Based on the utility value of N consequents $\{u(D_n); n=1,\ldots,N\}$, the time-associated prediction result generated by the TSF-EBRB model can be obtained below:

$$f(\boldsymbol{x}^T) = \sum_{n=1}^{N} u(D_n)\beta_n \tag{21}$$

4.2 A parameter optimization method for TSF-EBRB model

As with the existing rule-base models, such as EBRB model (Yang et al., 2019), belief rule-base (BRB) model(Jian-Bo et al., 2006), and fuzzy rule base (FRB) model(Chen & Linkens, 2004), the performance of the TSF-EBRB model has to depend on the value of basic parameters. However, all these basic parameters usually cannot be predetermined by expert knowledge because it may be really difficult to accurately determine the basic parameters subjectively, particularly, for a complicated time series forecasting problem. For this reason, it is more than necessary that the basic parameters should be determined by a data-driven approach and thus a parameter optimization method is proposed to optimize the basic parameters of the TSF-EBRB model.

The dynamic change of time series must be taken into account when optimizing the parameters of time series. An effective way to deal with this problem is to assume that time series data grow or decrease at a constant rate. Thus, a fixed percentage growth and decline rate can be set based on the rate of change in current data and then estimate the upper and lower limits of future data based on the trend of historical data. By using this method, the time series data dynamics can be better captured , ensuring that the model is more robust and can adapt to future data trends.

According to the introduction of the TSF-EBRB model in Section 4.1, its basic parameters include utility values $\{u(A_{i,j}); i=1,\ldots,M; j=1,\ldots,J_i\}$ for M antecedent attributes, utility values $\{u(D_n); n=1,\ldots,N\}$ for consequent attribute, attribute weights $\{\delta_i; i=1,\ldots,M\}$ for M antecedent attributes, and P correlation probabilities $\{v_p; p=1,\ldots,P\}$. Firstly, the following constraints are given:

(1) The utility values of antecedent attributes, which must satisfy the increasing relationship between adjacent utility values in the same antecedent attribute. The specific constraints are as follows:

$$u(A_{i,j}) \le u(A_{i,j+1}); i = 1,...,M; j = 1,...,J_i - 1 \tag{22}$$

$$u(A_{i,1}) = \min\{\min_{t=1,...,T}\{x_i^t\}, x_i^T(1+\Delta_i^-)^H; i = 1,...,M \tag{23}$$

$$u(A_{i,J_i}) = \max\{\max_{t=1,...,T}\{x_i^t\}, x_i^T(1+\Delta_i^+)^H\}; i = 1,...,M \tag{24}$$

where x_i^t denote the data of the ith key variable at time t; T denotes the value of duration to collect time series data. Δ_i^- and Δ_i^+ denote the negative and positive growth rate of the ith key variable; H denotes the number of future times. Here, it is worth noting that the negative and positive growth rate can be obtained from the historical data.

(2) The utility values of the consequent attribute, which must satisfy the increasing relationship between adjacent utility values. The specific constraints are as follows:

$$u(D_n) \le u(D_{n+1}); n = 1,...,N-1 \tag{25}$$

$$u(D_1) = \min\{\min_{t=1,...,T}\{y^t\}, y^T(1+\Delta^-)^H\} \tag{26}$$

$$u(D_N) = \max\{\max_{t=1,...,T}\{y^t\}, y^T(1+\Delta^+)^H\} \tag{27}$$

where y^t denote the data of the variable at time t; T denotes the value of duration to collect time series data. Δ^- and Δ^+ denote the negative and positive growth rate; H denotes the number of future times.

(3) The attribute weights of antecedent attributes, which must lie in the interval 0 to 1 below:

$$0 \le \delta_i \le 1; i = 1,...,M \tag{28}$$

(4) The correlation probabilities, which must lie in the interval 0 to 1 and the sum of all correlation probabilities equals 1. The specific constraints are as follows:

$$0 \le v_p \le 1;\ p = 1,...,P \tag{29}$$

$$\sum_{p=1}^{P} v_p = 1 \tag{30}$$

Based on the above constraints, the objective function should be further defined to provide a parameter optimization model. Assuming that there are T time series data $\{<\boldsymbol{x}^t, y^t>;\ t=1,\ldots,\ T\}$, the objective function is therefore expressed below when the mean absolute error (MAE) is used as the evaluation criterion:

$$\min MAE\{u(A_{i,j}), u(D_n), \delta_i, v_p\} = \frac{1}{T}\sum_{t=1}^{T} | f(\boldsymbol{x}^t) - y^t | \tag{31}$$

where $f(\boldsymbol{x}^t)$ is the time-associated prediction result of the TSF-EBRB model to reply time series data $\boldsymbol{x}^t$; y_t is the actual result of time series data $\boldsymbol{x}^t$.

***Remark* 1**: For the above-mentioned parameter optimization model, which only provides a mathematical model to show how to obtain the optimal value of the basic parameters used to construct TSF-EBRB model, the corresponding parameter optimization algorithms should be further introduced to finally optimize the basic parameters of TSF-EBRB model. From the previous studies on optimizing rule-base models (Chang et al., 2018; Jiao et al., 2020; Ren & Long, 2021), the optimization toolbox in MATLAB and swarm intelligence algorithms are available to achieve the goal of parameter optimization for the TSF-EBRB model.

4.3 Carbon peaking prediction method based on TSF-EBRB model

From Sections 4.1 and 4.2, an improved EBRB model, namely the TSF-EBRB model, and the corresponding parameter optimization method were introduced to handle time series forecasting problems. As one of time series forecasting problems,

this section aims to propose a novel carbon peaking prediction method based on the TSF-EBRB model. The fundamental framework is illustrated in Fig. 3.

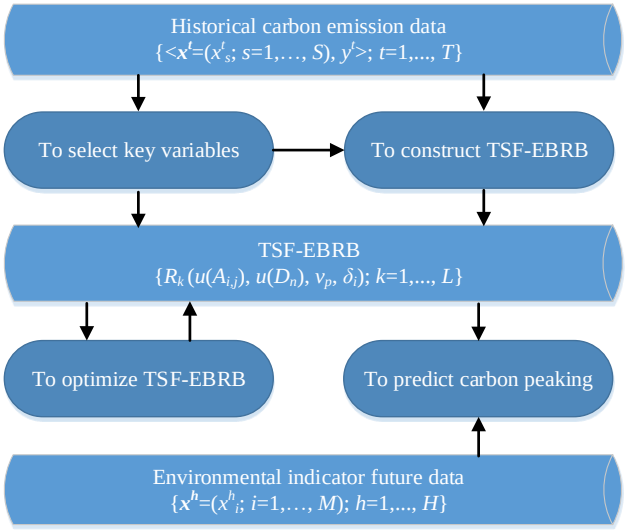


Fig.3 Carbon peaking prediction method based on TSF-EBRB model

From Fig. 3, the specific steps of the carbon peaking prediction method proposed are as follows:

***Step* 1**: To construct TSF-EBRB model using historical carbon emission data. Assuming that the carbon emission is related to S environmental indicators and their data are collected from T years. Thus, the carbon emission-related data at the tth (t=1,…, T) year are denoted as $<\boldsymbol{x}^t=(x_s^t; s=1,\ldots, S), y^t>$, where x_s^t denotes the data of the sth environmental indicator at the tth year, y^t denotes the data of carbon emission at the tth year. According to the construction steps detailed in Section 4.1, a TSF-EBRB model with initial basic parameters can be constructed from historical emission data.

***Step* 2**: To optimize TSF-EBRB model using the parameter optimization method. Assuming that the M key environmental indicators are selected from S environmental indicators and the corresponding carbon emission-related data for T years are denoted as $\{<\boldsymbol{x}^t=(x_i^t; i= 1,\ldots, M), y^t>; t=1,\ldots, T\}$. According to the objective function and constraints detailed in Section 4.2, the basic parameters of the TSF-EBRB model can be optimized to enhance the prediction accuracy of the TSF-EBRB model when handling carbon peaking prediction problem.

***Step* 3**: When employing the TSF-EBRB model for carbon peaking prediction, the assumption is made that the data for M key environmental indicators for the subsequent H years is represented as $\{\boldsymbol{x}^h=(x_i^h; i=1,\ldots, M); h=1,\ldots, H\}$. Based on an analysis of policy information and relevant data, the average annual growth rates of all indicators for the forthcoming H years are estimated, with the carbon emission data from the last year serving as the baseline for calculating the projected values $\boldsymbol{x}^h$ for each h-th year. Following the EBRB generation steps outlined in Section 3.1, the $\{\boldsymbol{x}^h=(x_i^h; i=1,\ldots, M); h=1,\ldots, H\}$ dataset, along with the optimized optimal parameters, is employed to generate input data for the TSF-EBRB model. This input data is then combined with the enhanced inference steps detailed in Section 4.1 to produce time-correlated carbon emission data for the subsequent H years, denoted as $\{f(\boldsymbol{x}^h); h=1,\ldots, H\}$.. It's imperative to note that if these H carbon emission data points exhibit a decrease followed by an increase, it indicates that carbon peaking has been achieved.

***Remark* 2**: Compared with the traditional EBRB model in Section 3, the TSF-EBRB model not only has an improved parameter optimization model to obtain the optimal basic parameters in the modeling process but also introduces the time-associated parameters that aim to take into account the characteristics of autocorrelation and mutual correlation in carbon peaking prediction, so as to greatly improve the performance of the proposed carbon peaking prediction method.

5. Empirical Analysis

To verify the validity of the proposed carbon peaking prediction method, the carbon emission data of 30 Chinese provinces, except for Tibet, Hong Kong, Macao, and Taiwan which is missing data, are used to conduct empirical analysis.

5.1 Data description and method setup

In the empirical analysis, the carbon emission and environmental data are collected from the China Statistical Yearbook and the China Environmental Statistics Yearbook, which are the most comprehensive and authoritative data sources and authoritative data sources are organized by the National Bureau of Statistics to publish Chinese statistical data each year. Based on the two data sources and the existing studies (Pan er al., 2022; Hao et al.,2022; Liu et al., 2022) the dataset with fourteen indicators is used in this study, namely the carbon emission is related to thirteen environmental indicators. Table 1 provides a description of these indicators and Table 2 shows the statistical analysis of these indicators in the terms of four major economic regions, namely, Eastern, Central, Western, and Northeastern regions, where the Eastern region includes Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan; the Central region includes Shaanxi, Anhui, Jiangxi, Henan, Hubei, and Hunan; the Western region includes Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang; and the Northeastern region includes Liaoning, Jilin, and Heilongjiang.

Table 1 Statistical analysis of input and output indicators.

Indicator type	Specific interpretation	Abbr.
Input indicator	Gross Domestic Product	GDP
Input indicator	Rate of forest cover	RFC
Input indicator	Afforestation area	AA
Input indicator	Area Forestry land	AFL
Input indicator	Area of planted forests	APF
Input indicator	Completion of Investment in industrial Pollution Control	IPC
Input indicator	Completion of Air Pollution control Project Investment	APC
Input indicator	Local Environmental Spending	LEP
Input indicator	Number of industrial enterprises above designated size	NIE
Input indicator	Resident population at year-end	RP
Input indicator	Electricity consumption	EC
Input indicator	Urban natural gas supply	TUA
Input indicator	Coal consumption	CC
Output indicator	Carbon emission	CE

Table 2 Statistical analysis of indicators in the four major economic regions

Abbr.	Eastern China		Central China		Western China		Northeast China	
	Min	Max	Min	Max	Min	Max	Min	Max
GDP	802.70	124369.70	2942.40	58887.40	443.70	53850.80	2455.20	27584.10
RFC	8.20	66.80	14.10	61.20	4.00	55.00	35.10	43.80
AA	0.71	600.96	3.40	861.90	5.38	713.50	19.90	246.77
AFL	7.46	1080.29	439.40	4499.17	179.03	2599.44	1297.58	2453.77
APF	5.97	615.51	102.74	733.53	4.44	507.68	148.94	315.32
IPC	476.00	2085343.00	25678.00	941275.00	2612.00	466884.00	8063.00	520470.00
APC	140.00	1568744.00	8072.00	708403.00	1677.00	321029.00	3513.00	326098.00
LEP	1.21	747.40	11.21	358.70	7.16	267.01	14.04	220.27
NIE	335.00	66307.00	2275.00	23679.00	386.00	16453.00	2774.00	23832.00

RP	818.00	12684.00	2393.00	9941.00	539.00	8372.00	2375.00	4379.00
EC	68.66	7867.00	389.20	3957.00	199.64	3460.00	378.23	2576.00
TUA	0.01	288.06	0.01	109.34	0.01	254.38	1.69	75.01
CC	106.70	43132.99	3366.71	113548.72	680.05	29848.06	5715.02	18710.78
CE	17.27	1094.20	93.93	2021.90	19.38	594.93	117.22	517.35

From Table 2, it can be found that the eastern region boasts a significant economic size, with its GDP ranging from 802.7 billion to 124369.7 billion, indicating its substantial economic strength and notable disparities. Following the eastern region, the central region ranks second in economic scale, reaching up to 58887.4 billion. Conversely, the western and northeastern regions exhibit relatively lower economic output, possibly reflecting their comparatively lagging economic development and weaker infrastructure. These disparities not only underscore the uneven economic development among regions but also provide crucial insights for in-depth analysis of regional development status and formulation of corresponding policies. The graph highlights that the central region possesses the widest range of afforestation and forest land areas, with afforestation reaching up to 861.90 million hectares and forest land ranging from 439.40 to 4499.17 million hectares. The maximum afforestation area even surpasses that of the eastern and northeastern regions, indicating more comprehensive policies and measures in forestry resource utilization and protection in this region.

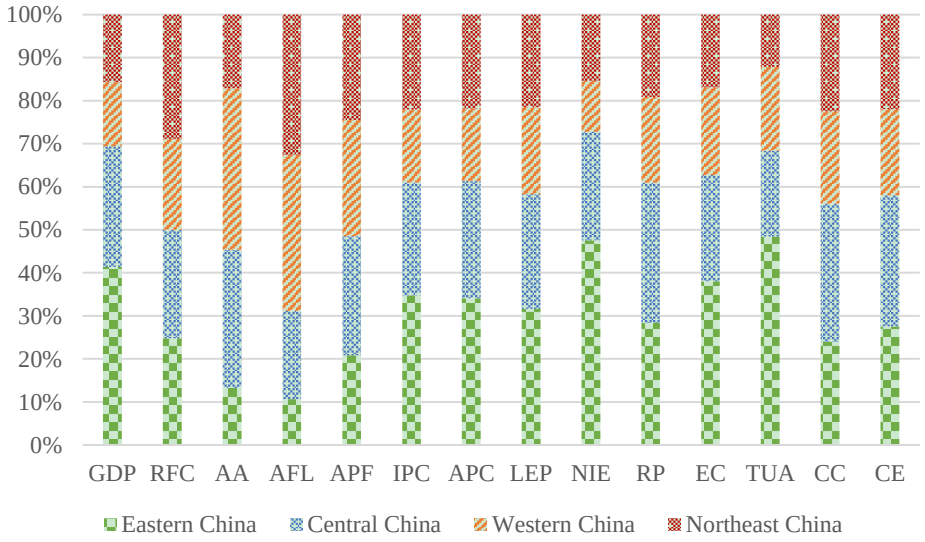


Fig.4 Average metrics across four major regions

Given the significant disparities in the average values of indicators among the four major economic regions, this paper utilizes a stacked percentage bar chart to present the average values of indicators for each region, as shown in Figure 4. Through the analysis of various indicators across the four major economic regions, significant differences are evident. The Eastern region stands out with its highly developed economy and relatively high forest coverage, boasting a notably higher GDP (30806.41 billion yuan) and RFC (34.76%) compared to other regions. Additionally, the Eastern region's investment in IPC and APC far exceeds that of other regions, reflecting substantial efforts in environmental protection. Central China excels in AA (227.61 million hectares) and RFC (35.45%). However, despite its abundant resources, the Western region exhibits a relatively lower GDP (11084.48 billion yuan) and RFC (29.62%). The Northeast region, characterized by a lower number of industrial enterprises and total GDP, consequently displays a relatively lower CE . Overall, these data underscore the diverse developmental characteristics of each region, spanning economic, environmental, and industrial structural aspects.

For the construction of TSF-EBRB models using the indicators shown in Table 1, it should be noted that using all these indicators directly as antecedent attributes may lead to a low computing efficiency in the process of transforming data into

rules. Moreover, there are different correlation levels among these indicators, suggesting that these indicators possess a certain degree of interchangeability. Therefore, it is imperative to conduct a selection process to identify key indicators in the unit of provinces, because a total of 30 TSF-EBRB models should be constructed independently for the 30 provinces.

For the optimization of the TSF-EBRB models according to Section 4.2, the negative and positive growth rate of the selected key variable should be calculated firstly. This is because time series data usually show consistent fluctuations throughout time, including periodic increments, decrements, or unpredictable fluctuations. The GDP data shown in Fig. 5 exemplifies a discernible rising trajectory over several years. However, traditional EBRB models typically rely on historical data for each indicator to calculate utility values. The magnitude of these utility values partially determines the potential range of forecast outcomes. When historical data are directly used as references to establish utility values, the predicted values will be constrained by historical data thresholds. This approach fails to capture the evolving trend of data changes over time, potentially leading to prediction biases. Additionally, considering the developmental disparities among different provinces, this study calculates the historical rate of change based on historical indicator data for each province after aggregating key indicators at the provincial level. If an indicator shows a growth trend, the maximum growth rate is selected as the fixed rate of change to estimate the upper limit of future data, and vice versa.

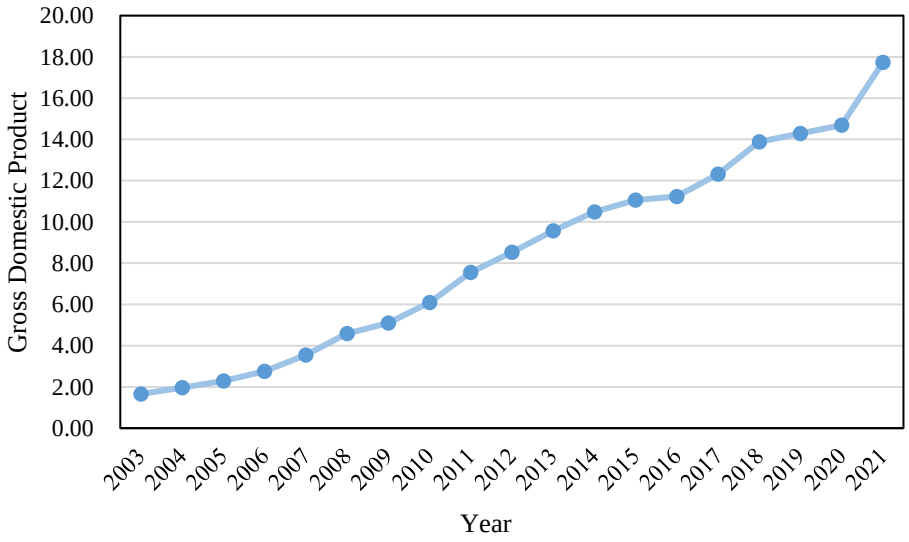


Fig. 5 GDP of China (2003-2021)

Additionally, the other basic settings are defined as follows: 1) the cumulative contribution rate of determining number of key environmental indicators is σ=95%; 2) the number of reference values for antecedent attributes is J_i=5 and the number of consequents for consequent attribute is N=3; 3) the attribute weight of all antecedent attributes is δ_i=0.6; 4) the number of correlation probabilities is P=2; 5) the carbon emission-related data from 2004 to 2018 are used as training dataset and the data from 2019 to 2021 as testing dataset; 6) the parameter optimization is performed by using Fmincon function in MATLAB toolbox with 300 iterations.

5.2 Carbon emission prediction based on the TSF-EBRB model

In this section, the TSF-EBRB model is used to predict carbon emissions of 30 provinces in China. Due to the fact that each provinces has different characteristics in carbon emissions, it is necessary to independently construct TSF-EBRB model for each province. In the aim of exemplifying the construction of TSF-EBRB model, the carbon emission prediction in Beijing is used as example to show each step detailed in Section 4.

Firstly, key environmental indicators are selected by utilizing carbon emission data from 2004 to 2018. Following the

Step 1 in Section 4.1, the data of all environmental indicators should be used to generate standardized matrix, correlation coefficient matrix, eigenvalues and eigenvectors, so that the key environmental indicators can be identified by using the cumulative contribution rate and eigenvectors, where the corresponding results are shown in Table 3 and Table 4.

Table 3 Cumulative contribution ratio of Beijing.

No	Eigenvalue	Contribution rate	Cumulative contribution rate
1	9.3673	0.7206	0.7206
2	2.1999	0.1692	0.8898
3	0.7309	0.0562	0.9460
4	0.2776	0.0214	0.9674
5	0.1726	0.0133	0.9806
6	0.1263	0.0097	0.9904
7	0.0750	0.0058	0.9961
8	0.0333	0.0026	0.9987
9	0.0081	0.0006	0.9993
10	0.0067	0.0005	0.9998
11	0.0020	0.0002	1.0000
12	0.0004	0.0000	1.0000
13	0.0000	0.0000	1.0000

Table 4 Eigenvectors of key indicators in Beijing.

Indicator name	Eigenvector 1	Eigenvector 2	Eigenvector 3	Eigenvector 4
GDP	0.3174	-0.0971	-0.0735	-0.1142
RFC	0.3172	0.0517	-0.1688	0.0267
AA	0.1834	-0.2332	**0.8551**	-0.2649
AFL	0.3005	0.1608	-0.2137	-0.3376
APF	0.3134	0.1018	-0.1912	-0.1381
IPA	0.0172	0.6415	0.2712	-0.0377
APC	-0.0242	**0.6531**	0.1011	0.1549
LEP	0.3045	0.1991	0.0959	0.0142
NIE	-0.2955	0.0534	-0.2222	-0.4731
RP	0.2989	-0.0883	0.0019	**0.6425**
EC	0.3170	-0.0934	-0.0719	0.0949
NGS	**0.3218**	-0.0095	-0.0666	-0.1325
CC	-0.3197	0.0126	0.0289	0.3069

From Table 3, it reveals that the number of key environmental indicators is 4 because the cumulative contribution rate regarding the 4th and the 5th eigenvalues is 94.60% and 96.74%, respectively, where the former one is smaller than 95% and the latter one is greater than 95%. Furthermore, from Table 3, the specific environmental indicators can be identified by using the corresponding eigenvectors, which are AA, APC, RP, and NGS, because their value is the maximum one comparing to other values in the same eigenvector. The detailed descriptions of these indicators are provided in Table 5.

Table 5 Selected input indicators of Beijing

Indicator	NGS	APC	AA	RP	CE
Min	27.02	1832	12.2	1493	85
Max	191.6	94599	45.81	2195	103
Average	108.16	44938.72	25.16	1979.78	94.91
standard deviation	60.39	33498.45	10.88	255.27	5.57

Secondly, the identified key indicators are employed as antecedent attributes and the carbon emissions as consequent attribute to construct TSF-EBRB model. Based on the parameter setting given in Section 5.1, the initial and trained values of basic parameters are provided in Table 6 and Table 7, respectively. This establishes the framework for the optimization process, taking into account the constraints and objective functions elucidated in Section 4.2. Following many rounds of training, the

parameters of the TSF-EBRB model have been optimized, resulting in the identification of the best parameter values as shown in Table 7. The impacts of this optimization initiative are shown in Table 8. By comparing the initial parameters with the optimized parameters, it can be observed that most of the parameters have been improved. This indicates that the optimization process has successfully enhanced the model's performance.

Table 6 Initial basic parameters of TSF-EBRB model for Beijing

Indicator type	Indicator name	Utility value					Indicator weights	Probability
		$u(VL)$	$u(L)$	$u(M)$	$u(H)$	$u(VH)$	δ_i	v_p
Input Indicator	NGS	27.02	99.25	171.49	243.72	330.24	0.6	0.37
	APC	2731.00	25697.87	48664.91	71631.96	94599.00	0.6	0.63
	AA	12.20	20.60	29.01	37.41	45.81	0.6	/
	RP	666.30	3689.74	6712.95	9736.99	12760.16	0.6	/
output indicator	CE	/	75.22	89.97	104.71	/	/	/

Table 7 Trained basic parameters of TSF-EBRB model for Beijing

Indicator type	Indicator name	Utility value					Indicator weights	Probability
		$u(VL)$	$u(L)$	$u(M)$	$u(H)$	$u(VH)$	δ_i	v_p
Input Indicator	NGS	27.02	106.92	181.22	261.04	338.99	0.9725	0.8189
	APC	2731.00	25023.74	48216.90	71407.69	94599.00.	0.2647	0.1811
	AA	12.2	20.39	29.98	40.33	45.81	0.9725	/
	RP	666.30	3689.74	6712.95	9736.99	12760.16	0.9570	/
output indicator	CE	/	75.22	86.68	104.71	/	/	/

Then, the optimized optimal parameters are used as input data to re-generate the TSF-EBRB model, which combined with the inference step to predict the carbon emission data of Beijing from 2019 to 2021. The mean absolute error (MAE) is calculated by comparing the inference output with the actual output, and the comparison results before and after parameter optimization are shown in Table 8. As can be seen from Table 8, the accuracy is greatly improved after parameter optimization.

Table 8 Comparison of parameter optimization in Beijing from 2019-2021.

Year	MAE before optimization	MAE after optimization	Improvement
2019	1.4717	0.4412	70.02%
2020	15.3207	9.0538	40.90%
2021	15.5408	13.4120	13.70%

Considering the different economic conditions and energy consumption of each province, key indicators are identified for each province through the same indicator screening process, and the relevant results are shown in Table 9.

Table 9 Key input indicators for 30 provinces in China

Provinces	Key environmental indicator	Provinces	Key environmental indicators
Beijing	{NGS, APC, AA, RP}	Henan	{GDP, CC, AA, NGS}
Tianjin	{GDP, IPC, CC, NIE, NGS}	Hubei	{EC, APC, CC, IPC}
Hebei	{GDP, APC, CC, NGS}	Hunan	{EC, IPC, CC, RP}
Shanxi	{GDP, IPC, NIE, AA}	Guangdong	{RP, IPC, NIE, CC}
Inner Mongolia	{GDP, RP, APC, AA}	Guangxi	{GDP, IPC, RP, NGS, NIE}
Liaoning	{EC, NIE, APC, AA, IPC}	Hainan	{NGS, AA}
Jilin	{GDP, CC, IPC, AA}	Chongqing	{EC, IPC, CC, NGS}
Heilongjiang	{NIE, IPC, AFL, GDP}	Sichuan	{GDP, CC, AA, NIE, NGS}
Shanghai	{GDP, AA, APC, NIE}	Guizhou	{EC, IPC, NGS, RP}
Jiangsu	{NGS, EC, NIE}	Yunnan	{EC, CC, APC, NGS}
Zhejiang	{NGS, RP, NIE}	Shaanxi	{IPC, AA, GDP}
Anhui	{AA, RP, IPC}	Gansu	{GDP, IPC, NIE, RFC, APC}
Fujian	{EC, APC, AFL, NIE, CC}	Qinghai	{AFL, APC, NGS, NIE}

Jiangxi	{AA, IPC}	Ningxia	{RP, IPC, AFL, AA}
Shandong	{EC, AFL, AA}	Xinjiang	{CC, AFL, AA}

To assess the efficacy of the TSF-EBRB model, four different types of models are selected for error comparison, including the GM(1,1) model, the ARIMA model, the standard EBRB model, and SVM model. Carbon emissions forecasts for each province were made for the years 2017 to 2019 using data from 2004 to 2016. The predictions generated by the diverse models were compared against the actual values to gauge their reliability. Fig. 6 displays the mean absolute errors associated with the different algorithms.

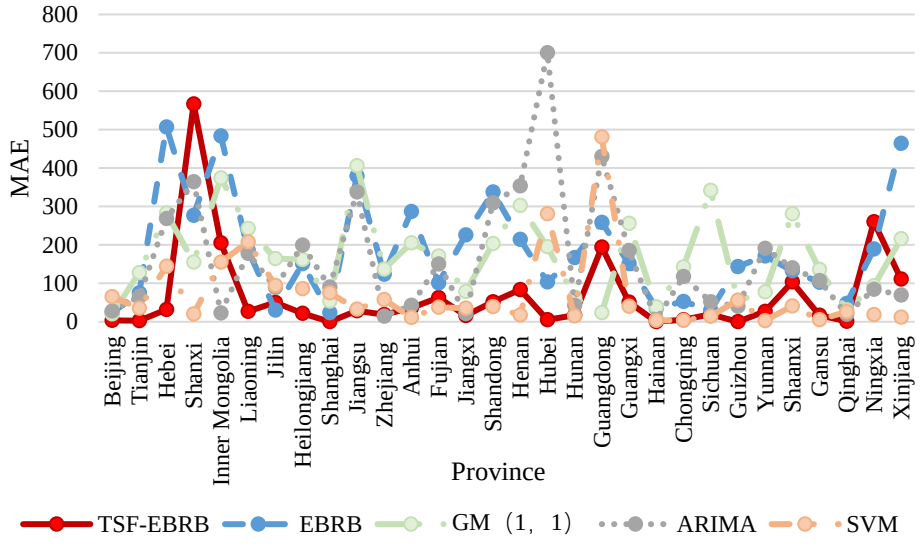


Fig.6 Comparison of MAE of different models between 2017-2019.

Fig. 6 reveals that the TSF-EBRB model exhibits superior overall prediction accuracy compared to the other models. It is worth mentioning that the conventional EBRB model has a lower accuracy than the TSF-EBRB model. This is because the conventional EBRB model fails to take into account specific parameters for temporal data correlation. Conversely, GM(1,1) and ARIMA models are particularly suitable for datasets with few observations and demonstrate their efficacy in making predictions for shorter and intermediate periods. However, these models encounter difficulties in handling nonlinear data, leading to relatively greater prediction errors compared to the TSF-EBRB model. The SVM performs exceptionally well in some provinces. Nevertheless, its relatively high MAE may be attributed to its higher model complexity. In summary, the TSF-EBRB model outperforms other carbon emission prediction models in terms of overall accuracy. The lower MAE, stability, consideration of time series data, and enhanced business interpretability make it the preferred choice for achieving more precise predictions of carbon emissions across various provinces.

To validate the three hypotheses, three control model experimental plans were devised. Firstly, in the experiment for ***Hypothesis* 1**, the same data and parameter settings as the TSF-EBRB model will be employed, but without the screening of key indicators, to evaluate the model performance without the screening of key indicators. The control model for this experiment is named NO-PCA-EBRB. Secondly, in the experiment for ***Hypothesis* 2**, a control model named NO-TIME-EBRB will be set up. The basic process remains the same as the TSF-EBRB model, but without considering temporal correlation, i.e., excluding time series analysis from the model. Finally, in the experiment for ***Hypothesis* 3**, a control model named NO-OPTI-EBRB will be established. Parameter optimization will not be conducted for this model, while other parameters remain unchanged from the modeling process. The predictive performance of the model without parameter optimization will be evaluated. Subsequently, the Mean Absolute Error (MAE) and other metrics will be used to assess the

prediction errors of the TSF-EBRB model and the control models for the years 2017-2019, as illustrated in Fig. 7. Through the above experimental procedures, the three hypotheses proposed in Section 3.3 can be validated:

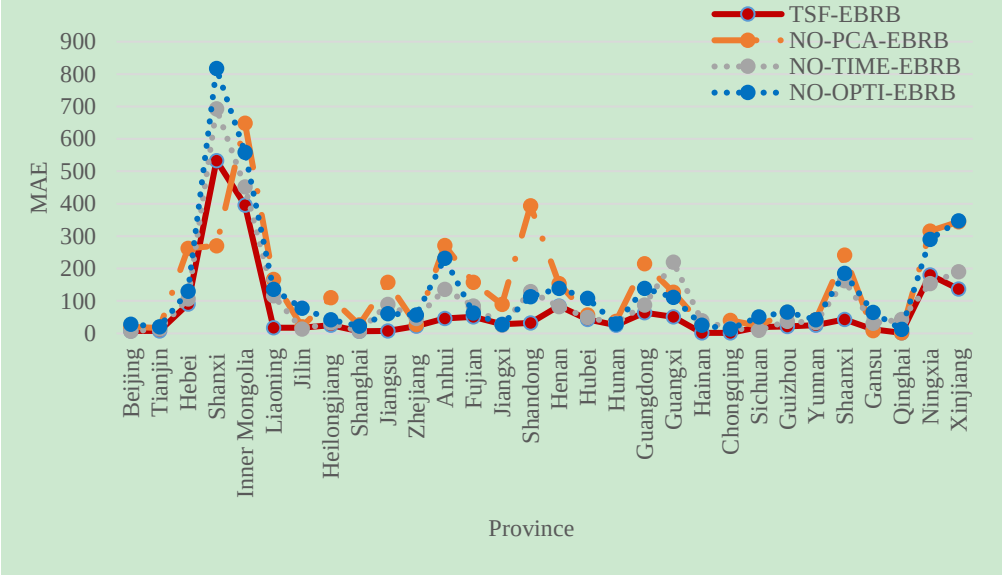


Fig.7 Comparison of MAE between control model and TSF-EBRB model

Firstly, the ***Hypothesis* 1** is validated by the fact that the accuracy of the TSF-EBRB model is improved after selecting key environmental indicators, where the key indicators for 30 provinces can be found in Table 9 and the comparison of TSF-EBRB model and the conventional EBRB model is shown in Fig. 6.

To verify this hypothesis, key environmental indicators were initially selected from carbon emission data spanning from 2004 to 2019. Utilizing methods such as correlation analysis and principal component analysis, environmental indicators closely associated with carbon emissions were identified, forming the basis for constructing the TSF-EBRB model. The model was trained on the available dataset, allowing it to learn the relationships between environmental indicators and carbon emissions. Subsequently, the predictive performance of the TSF-EBRB model was evaluated by forecasting with the testing dataset and calculating the MAE. Furthermore, comparison with the NO-PCA-EBRB that did not involve the selection of key environmental indicators revealed a significant enhancement in prediction accuracy when using the TSF-EBRB model with consideration of these indicators. This finding further supports the validity of ***Hypothesis* 1**, indicating the notable advantage of the TSF-EBRB model in carbon emission prediction. Consequently, it can provide more precise reference points for relevant decision-making processes.

***Hypothesis* 2**: To validate this hypothesis, the TSF-EBRB model was constructed by integrating time series forecasting techniques with the EBRB model. Time series forecasting considers the correlation between current data and historical data by adding parameters to connect them, enabling the model to better capture the temporal relationships between data points. In the experiment, the EBRB model without considering time correlations served as the control model and was compared with the TSF-EBRB model. By calculating evaluation metrics such as MAE, the predictive performance of the TSF-EBRB model was compared with NO-TIME-EBRB. This comparative analysis further confirmed the superior predictive performance of the TSF-EBRB model in carbon emission prediction, thereby validating ***Hypothesis* 2**.

***Hypothesis* 3**: The study employed FMINCON function to optimize the time-related parameters. Initially, the parameters requiring optimization were used as inputs to generate model prediction errors, aiming to minimize these errors as the optimization objective. Subsequently, the FMINCON function was invoked to optimize the parameters under specified constraints, seeking the parameter values that minimize the objective function. Upon completion of parameter optimization,

the TSF-EBRB model was updated with the optimized parameter values. Following this, the EBRB model without time-related parameter optimization served as the control model. By comparing the predictive performance of the optimized model and NO-OPTI-EBRB on the testing dataset, the effectiveness of the ***Hypothesis* 3** was further validated.

In the resistance test of TSF-EBRB model, this paper adopts several methods to evaluate its robustness. Firstly, the model is tested by adding random noise and interfering input data. The random noise is used to simulate the uncertainty and noise interference of real data, and the interference input data is used to evaluate the response of the model to the change of input data and interference. Targeted attack methods are also employed, such as inserting fake data and deleting important information, to simulate the robustness of the model in the face of malicious attacks. The MAE of different antagonism samples was calculated using the data from Beijing 2019-2021. It can effectively deal with additional noise, interference input and specific attacks, and maintain stable predictive performance. This shows that TSF-EBRB model has good generalization ability and robustness, and is suitable for carbon emission prediction in complex environment.

Table 10 Comparison of Mean Absolute Error (MAE) under different methods of adversarial training

Add noise	Interference input	TSF-EBRB
8.99085854	9.46836667	7.51902

5.3 Carbon peaking prediction based on the TSF-EBRB model

In this section, the carbon peaking of 30 Chinese provinces is predicted based on the TSF-EBRB model under baseline, low-carbon, and technology development scenarios. The purpose of these scenarios is to predict the maximum levels of carbon emissions in the provinces of China during the period of 2022 to 2030.

5.3.1 Process of carbon peaking prediction under the three scenarios

In order to apply the TSF-EBRB model to carbon peaking prediction, the baseline scenario, low-carbon scenario and technology development scenarios are designed, as described below:

(1) The baseline scenario, commonly used as a reference scenario, represents the current development path or trends. In this scenario, various factors continue to evolve at their current growth rates and trends. Key environmental indicators follow their current developmental trends without the implementation of additional policies or intervention measures to alter these trends. Therefore, it serves as a baseline for assessing the impacts of other scenarios since it does not account for additional low-carbon or technological development measures.

(2) The low-carbon scenario is an environment friendly scenario aimed at reducing carbon emissions and mitigating the impacts of climate change. In this scenario, various policies, technologies, and practices are adopted to reduce carbon emissions, enhance energy efficiency, utilize clean energy sources, and implement other measures to decrease the carbon footprint. To better simulate the low-carbon scenario, the average values of various indicators in the current data are used as baseline values. The preset rates of indicator growth under different policies are used as the indicator growth rates.

(3) The technology development scenario emphasizes technological innovation and rapid progress to enhance productivity, economic growth, and societal well-being. In this scenario, various new technologies and innovations are widely adopted, leading to increased productivity and improved resource utilization efficiency. In general, the technology development scenario is typically considered as an advanced evolution of the low-carbon scenario, where all indicator growth rates are higher than those in the low-carbon scenario.

To observe the carbon emission peak of each province before 2030, this study predicts the carbon emissions of 30

provinces from 2022 to 2035. Firstly, historical data from 2004 to 2021 is used, with the period from 2022 to 2035 designated as the forecast period. After setting the forecast period, the study divides it into two equal-length intervals based on current sustainable development policies and economic growth conditions: 2022-2028 and 2029-2035. Additionally, the study analyzes current policies to determine the growth rates of various indicators under different scenarios. It is assumed that the growth rate of each indicator in the later period is lower than that in the former period, reflecting the reality of sustainable development policies and economic growth slowdown. Subsequently, using 2021 data as a baseline, it is assumed that all indicators grow or decline at the set rates during the forecast period. Furthermore, trained optimal parameters are utilized as inputs to predict carbon emissions using the TSF-EBRB model. Through this scenario prediction method, the study obtains carbon emission forecasts under different scenarios, enabling analysis and evaluation of carbon emission peaks.

Table 11 Growth rate of thirteen environmental indicators under different scenarios.

Indicators	Baseline scenario		Low-carbon scenario		Technological development scenario	
	2022-2028	2029-2035	2022-2028	2029-2035	2022-2028	2029-2035
GDP	0.0650	0.0600	0.0600	0.0550	0.0550	0.0500
RFC	0.0070	0.0065	0.0078	0.0075	0.0080	0.0075
AA	0.0129	0.0122	0.0135	0.0126	0.0138	0.0131
AFL	0.0180	0.0130	0.0200	0.0150	0.0230	0.0180
APF	0.1243	0.1240	0.1241	0.1238	0.1243	0.1240
IPC	0.0345	0.0340	0.0350	0.0345	0.0350	0.0345
APC	0.0313	0.0263	0.0357	0.0307	0.0377	0.0357
LEP	0.0880	0.0850	0.0910	0.0880	0.0920	0.0890
NIE	0.0800	0.0355	0.1200	0.0700	0.1350	0.0750
RP	0.0025	0.0021	0.0021	0.0018	0.0027	0.0022
EC	0.0263	0.0213	0.0310	0.0260	0.0325	0.0275
TUA	0.0396	0.0346	0.0476	0.0426	0.0482	0.0432
CC	0.0001	-0.0031	-0.0282	-0.0252	-0.0288	-0.0258

To calculate the year of carbon peaking under baseline scenario, the ARIMA model is employed to predict the future values of carbon emission-related indicators. This enables the carbon emission prediction model to make prediction using the future trends of the indicator data. In order to illustrate the year of carbon emission in the unit of economic regions, the 30 provinces are divided into Eastern, Central, Northeastern and Western regions, as detailed in Fig 8.

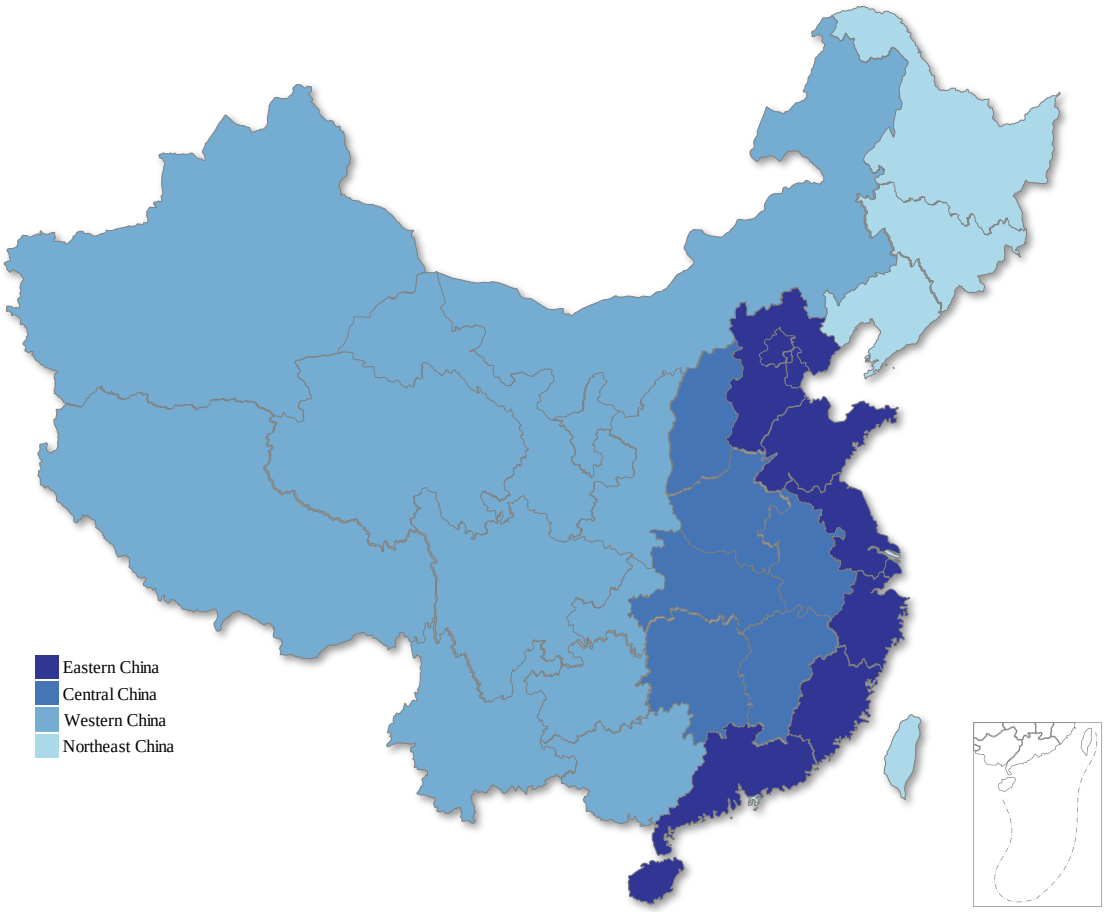


Fig.8 Division of four major four economic regions

Due to the complexity of economy and industrial structure, relying solely on the division into four major economic regions often fails to accurately capture the carbon emissions situation in each region. Additionally, the division into four economic regions overlooks the influence of geographical factors such as climate, topography, and resource conditions on carbon emissions. To address these limitations, the four major economic regions are further subdivided into six geographical regions: Northeast, North, Northwest, East, South and Southwest, as shown in Fig. 9. This subdivision into six geographical regions provides a more detailed and accurate basis for data analysis, facilitating a deeper understanding of carbon emissions and mitigation potentials under different geographical environments. Furthermore, geographical region division enables more effective resource allocation, promoting the implementation and effectiveness of carbon reduction efforts. It also aids in the formulation of more precise carbon reduction policies and measures tailored to the specific needs and challenges of each region. The refinement of geographical region division facilitates more efficient resource allocation and promotes the implementation and effectiveness of carbon reduction efforts. Fig. 10 - 15 illustrate the carbon emissions of the four economic regions and the six further economic regions.

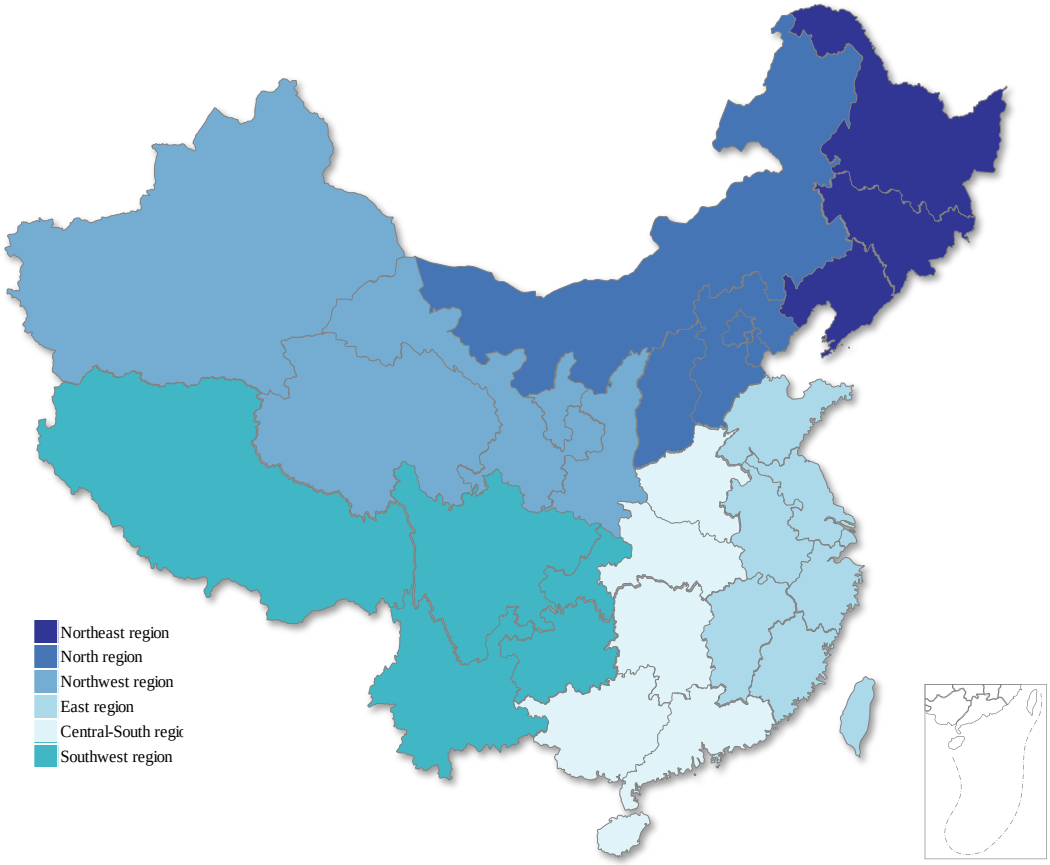


Fig.9 Division of six major geographical regions

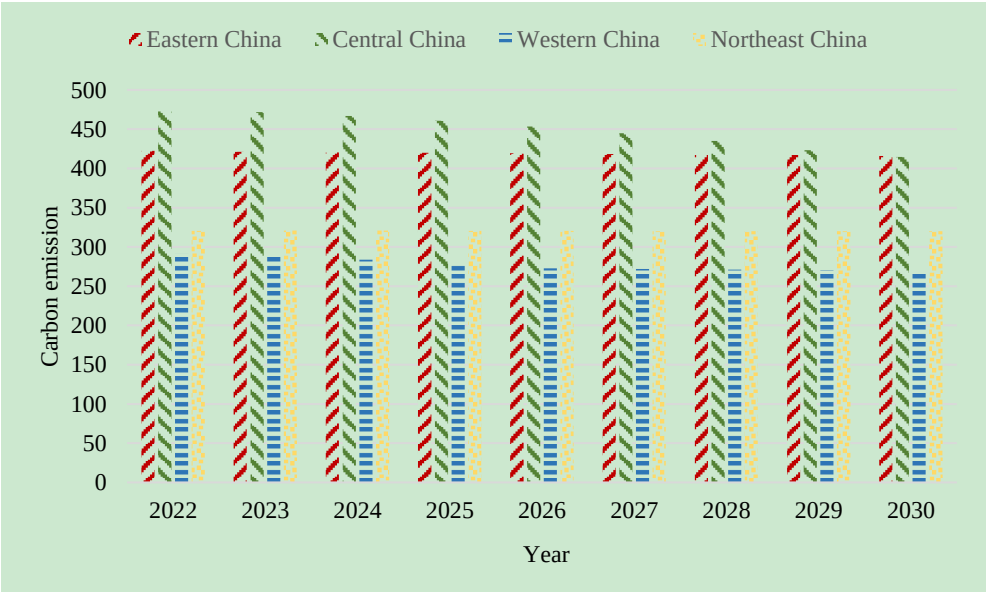


Fig.10 Carbon emission under baseline scenario for four economic regions

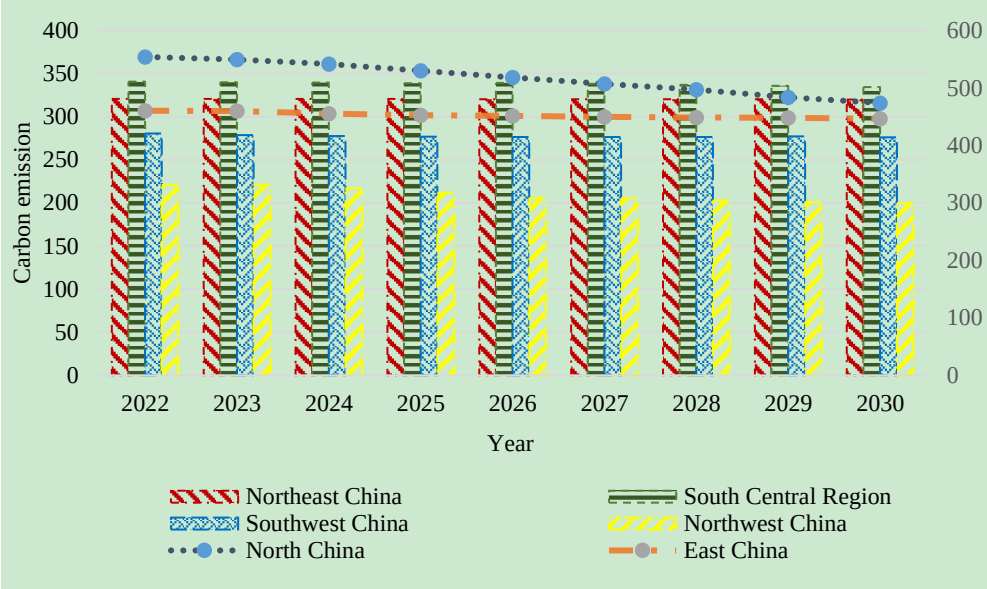


Fig.11 Carbon emissions under baseline scenario for six geographical regions

Taking into account the disparities between different regions, a combination chart is utilized to illustrate the carbon emission forecast data of the six major geographical regions, as depicted in Fig. 11. In this chart, line plots are employed to showcase the data for the Northern and Eastern regions, given their substantially higher carbon emission levels compared to other regions. This choice aims to accentuate their elevated emission levels and respective trends. A declining trend in carbon emissions across various regions of China over the next fourteen years can be observed. The Eastern region shows an overall decrease in carbon emissions, albeit with a relatively small reduction rate of approximately 2.07%. This trend may reflect the challenges faced by the region in terms of economic development pressure and industrial restructuring. In contrast, the Central region exhibits the most significant reduction efforts, with carbon emissions decreasing from 472.41 to 420.52, representing a reduction rate of about 11.00%. This outcome may be attributed to stricter environmental policies and technological innovations driving emissions reductions. However, the Western and Northeastern regions demonstrate relatively smaller reductions in carbon emissions, at approximately 4.45% and 0.07%, respectively. This could be related to the economic structure and resource utilization methods in these regions.

According to the data presented in Fig. 12, the carbon emissions in the northern region exhibit a decreasing trend over the years, decreasing from 553.98 to 484.97, representing a reduction of approximately 12.48%. Specifically, the data between 2023 and 2033 demonstrate relative stability, suggesting that the region may have reached its peak carbon emissions during this period and maintained a stable reduction trend thereafter. In contrast, the carbon emissions in the northeast region decreased slightly from 320.63 to 320.45, with a reduction of about 0.06%. Despite the small reduction, the overall trend still shows a decline. The carbon emissions in the eastern region exhibit slight fluctuations but demonstrate an overall decreasing trend, with a reduction rate of approximately 4.17%. This trend may reflect the effectiveness of emission reduction policies and technological innovations in the region. Both the central and southwestern regions show a downward trend in carbon emissions, although with considerable fluctuations. However, the overall reduction trend remains positive. It is worth noting that the northwestern region experiences fluctuations in carbon emissions, with an overall increasing trend, indicating a growth rate of approximately 5.47%. This underscores the challenges faced by the region in carbon emission reduction efforts.

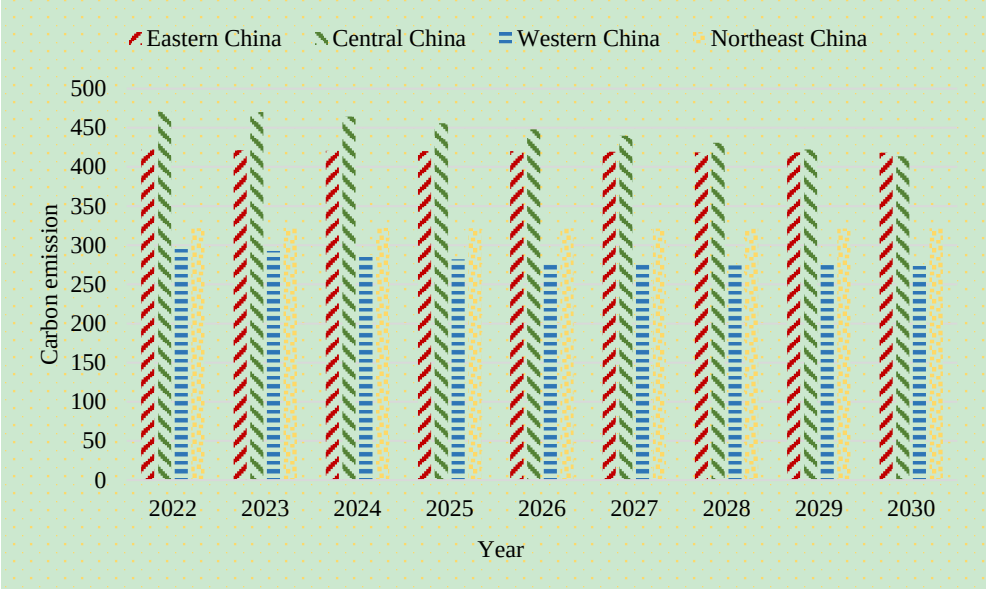


Fig. 12 Carbon emissions under low carbon scenario for four economic regions

Under the low-carbon scenario, predictions were conducted after parameter selection, and the results are displayed in Fig. 12. In the low-carbon scenario, the carbon emissions in Eastern China exhibit an overall declining trend, decreasing from 422.45 in 2022 to 413.70 in 2035, with a reduction rate of approximately 1.89%. This indicates that despite a decrease, the rate of decline in carbon emissions in the Eastern region is relatively slow under the low-carbon scenario. Similarly, Central China shows a downward trend in carbon emissions over the years, resembling the trend in the baseline scenario. This suggests that Central China can still maintain a significant reduction in carbon emissions under the low-carbon scenario. Furthermore, both Western China and Northeast China experience an overall decrease in carbon emissions under the low-carbon scenario, with reductions of approximately 5.55% and 0.55%, respectively. Although the reduction rates have improved compared to the baseline scenario, the reduction rates in these regions remain relatively modest.

Under the low-carbon scenario, carbon emissions in different regions demonstrate a declining trend, albeit with varying reduction rates. Central China stands out as the region with the most significant reduction effect, while the Eastern region shows relatively weaker emission reduction effects. This highlights the necessity for strengthening emission reduction policies and technological innovations across regions to better achieve carbon emission reduction goals under a low-carbon development pathway.

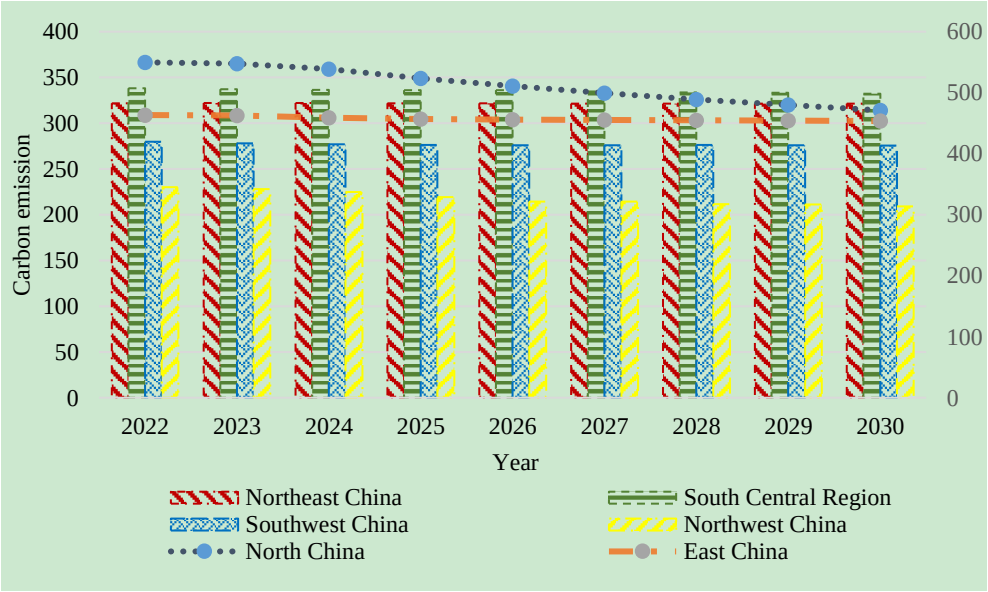


Fig.13 Carbon emissions under low carbon scenario for six geographical regions

From Fig. 13, the fluctuation of carbon emissions in each geographica region is small, showing a slow upward trend. In the context of the low-carbon scenario, North China has notably reduced its carbon emissions by 13.07%, indicating a proactive response and significant efforts toward carbon reduction in this region. Northeast China's carbon emissions remain relatively stable with minimal fluctuations. East China and the Central South Region have witnessed reductions in carbon emissions by 2.90% and 1.05%, respectively, indicating some progress in their emission reduction efforts. Southwest China shows a slight increase in carbon emissions, likely influenced by regional economic structures and policy factors. Conversely, Northwest China has experienced a substantial reduction in carbon emissions by 9.94%, reflecting proactive measures and effective policies implemented to address climate change in the region.

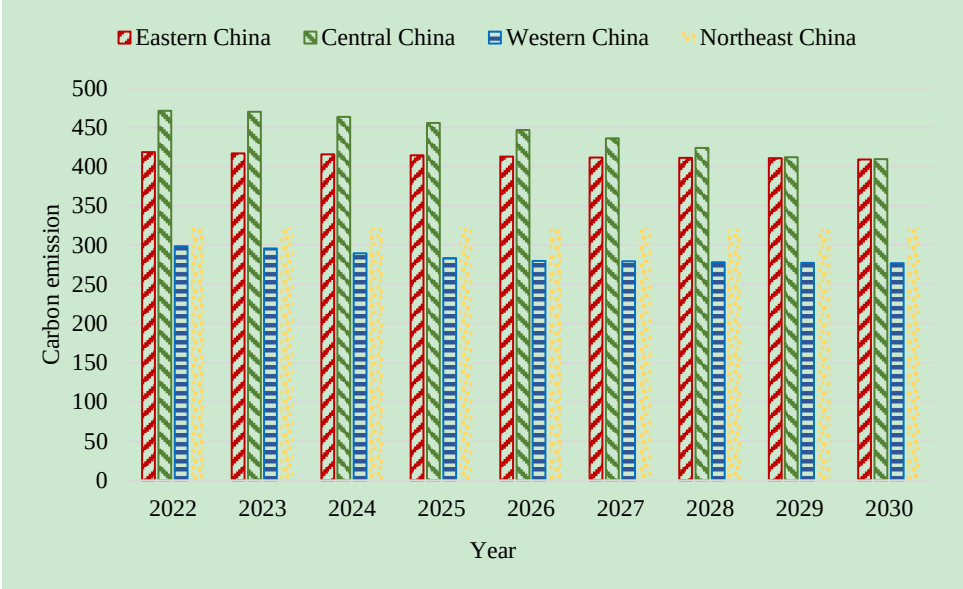


Fig.14 Carbon emission under technology development scenario for four economic regions

Under the technology development scenario, the carbon emission forecasts for major economic regions exhibit complex trends. The detailed results are presented in Fig.14. Eastern China is projected to continue reducing its carbon emissions in the coming years, though the pace of reduction is relatively slower. However, with advancements in technology and innovation, it is expected that the region's emission reduction efforts will gradually strengthen. Compared to the low-carbon scenario, the reduction rate in Eastern China remains limited, with only a 2.35% decrease. In contrast, the central region shows a relatively higher reduction rate of 11.06% under this scenario, possibly benefiting from increased technological innovation and industrial transformation. Similarly, the western region demonstrates a reduction rate of 4.30%. Though lower than that in the low-carbon scenario, it still shows a steady downward trend, likely due to the application of technology and innovation in energy utilization and production processes. Conversely, the northeastern region exhibits minimal changes in carbon emissions, with a reduction rate of only 0.07%, indicating the challenges and pressures the region faces in technological development.

However, more proactive carbon mitigation measures are required to achieve the carbon peak. Overall, the various regions are making positive strides toward the carbon peak target, with the Central region showing a higher likelihood of reaching it earlier. This trend underscores the potential of technological innovation and development in reducing carbon emissions, but ongoing efforts are crucial to ensure the attainment of this goal in each region.

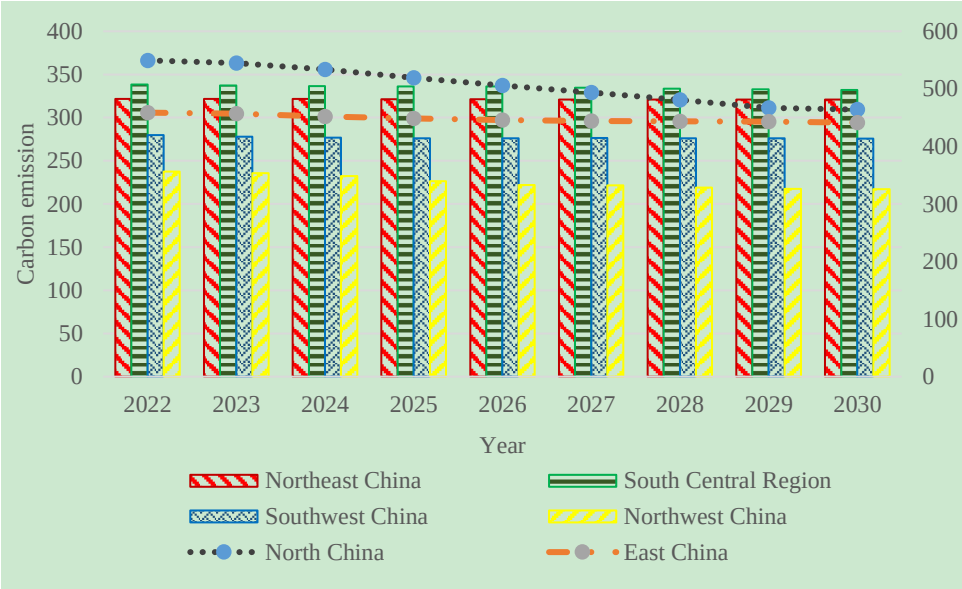


Fig.15 Carbon emission under technology development scenario for six geographical regions

The overall emission reduction trends for the six geographica regions exhibit diverse trends, as illustrated in Fig. 15. North China demonstrates a consistent decline in carbon emissions over the coming years, indicating significant reduction efforts, with an overall decrease of 13.08% from 2022 to 2035. Conversely, Northeast China shows relatively stable carbon emissions, reflecting potential sluggishness in technological advancement in this region. While East China showcases a downward trend in carbon emissions year by year, the pace of reduction is comparatively slow, with a total decrease of 4.35%. The carbon emissions in the Central and Southern regions exhibit substantial declines annually, with overall reductions of 7.10% and 13.08%, respectively, during the same period. Southwest China presents a marginal growth in carbon emissions, likely influenced by regional economic structures and policies, albeit with minimal increments. The Northwest region witnesses fluctuations in carbon emissions over the next few years, yet with an overall decrease of 6.24%. These trends underscore the partial success of carbon emission reduction efforts across various regions under the scenario of technological development, emphasizing the continued need for intensified technological innovation and emissions reduction policies to further drive down carbon emissions.

5.3.2 Comparison of carbon peaking prediction under the three scenarios

This section aims to undertake a complete comparison and analysis of the carbon peaking scenarios in all 30 provinces of China, considering diverse situations. The methodology involves inputting provincial indicator data into the TSF-EBRB model for calculations, thereby determining the carbon peaking timing for each province under different scenarios. The results of the peak carbon forecasts for all 30 provinces in China will be summarized in Table 12. Additionally, the peak attainment ratios for each province under various scenarios will be depicted in Fig. 15, while Fig. 16 will visually display the carbon emissions for each province under different scenarios from 2022 to 2030.

Table12 Peak time projections under different scenarios.

Peak Time	Baseline Scenario	Low Carbon Scenario	Technology Development Scenario
2022-2030	Beijing, Shanxi, Fujian, Zhejiang, Shandong, Shaanxi, Inner Mongolia, Jiangsu, Jiangxi, Tianjin, Hainan, Hebei, Heilongjiang, Anhui, Henan,	Beijing, Shanghai, Shaanxi, Fujian, Zhejiang, Shandong, Shanxi, Inner Mongolia, Jiangsu, Jiangxi, Tianjin, Hainan, Hebei, Heilongjiang, Anhui,	Beijing, Shanghai, Shanxi, Fujian, Zhejiang, Shandong, Shaanxi, Inner Mongolia, Jiangsu, Jiangxi, Tianjin, Hainan, Hebei, Heilongjiang, Anhui, Henan, Hunan,

	Hunan, Guangxi, Guizhou, Xinjiang,	Henan, Hunan, Guangxi, Guizhou, Xinjiang,	Guangxi, Guizhou, Xinjiang, Heilongjiang, Jilin, Chongqing, Gansu
2030 onward	Liaoning, Heilongjiang, Jilin, Hubei, Guangdong, Chongqing, Sichuan, Yunnan, Qinghai, Ningxia, Gansu, Shanghai,	Liaoning, Heilongjiang, Jilin, Hubei, Guangdong, Chongqing, Sichuan, Yunnan, Qinghai, Ningxia, Gansu	Liaoning, Hubei, Guangdong, Sichuan, Yunnan, Qinghai, Ningxia

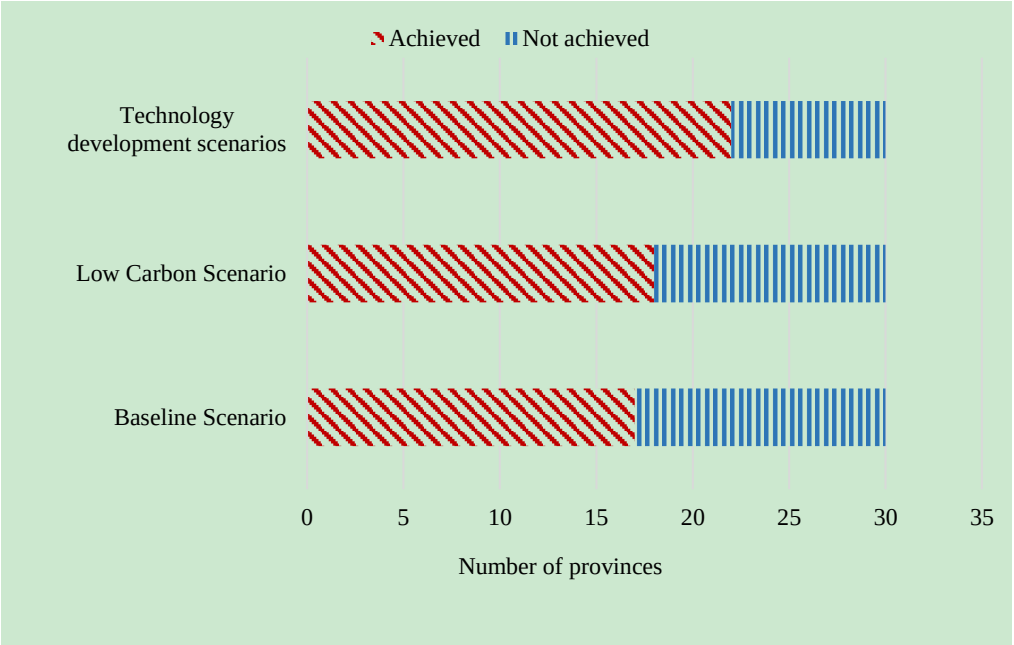


Fig.16 Peaking by province under different scenarios for the period 2022-2030.

The data presented in Table 12 highlights that the more economically developed regions along the eastern coast of China are expected to achieve peak carbon emissions at an earlier stage, while some less developed provinces may not attain this milestone before 2030. The majority of provinces are projected to reach peak carbon emissions during the 14th Five-Year Plan period. From 2022 to 2030, it is anticipated that average annual carbon emissions will gradually decline after reaching their peak, exhibiting an "inverted U-shape" pattern.

From Fig. 16, the number of provinces meeting the peak carbon emissions target is visually represented, revealing substantial variations in progress among individual provinces in achieving peak carbon emissions across different scenarios. In the baseline scenario, some provinces have already reached their peak carbon emissions under current policies and economic conditions, while others have not, signifying the continuous growth of carbon emissions in certain regions. However, in the low-carbon scenario, more provinces have achieved peak carbon emissions, demonstrating the positive impact of low-carbon policies and measures on carbon reduction. Of particular note is the technology development scenario, in which the highest number of provinces have reached peak carbon emissions, underscoring the significance of technological innovation and development in carbon reduction. This comprehensive analysis underscores the interplay of policy, technology, and economic factors in the context of carbon peaking, providing valuable insights for future carbon reduction decisions and a more effective response to the challenges of climate change.

The above data emphasize the importance of policies and technologies and their impact on the Peak Carbon Goal in different regions. Certain provinces may find it relatively easier to achieve peak carbon emissions due to factors such as abundant renewable energy resources or a more advanced technological infrastructure. Conversely, other provinces may require more substantial policy support and increased technological investments to ensure that those provinces are on track to achieve peak carbon emissions by 2030. This highlights the need for tailored strategies and interventions that account for the

unique characteristics and challenges faced by different regions in China's pursuit of carbon emissions reduction targets.

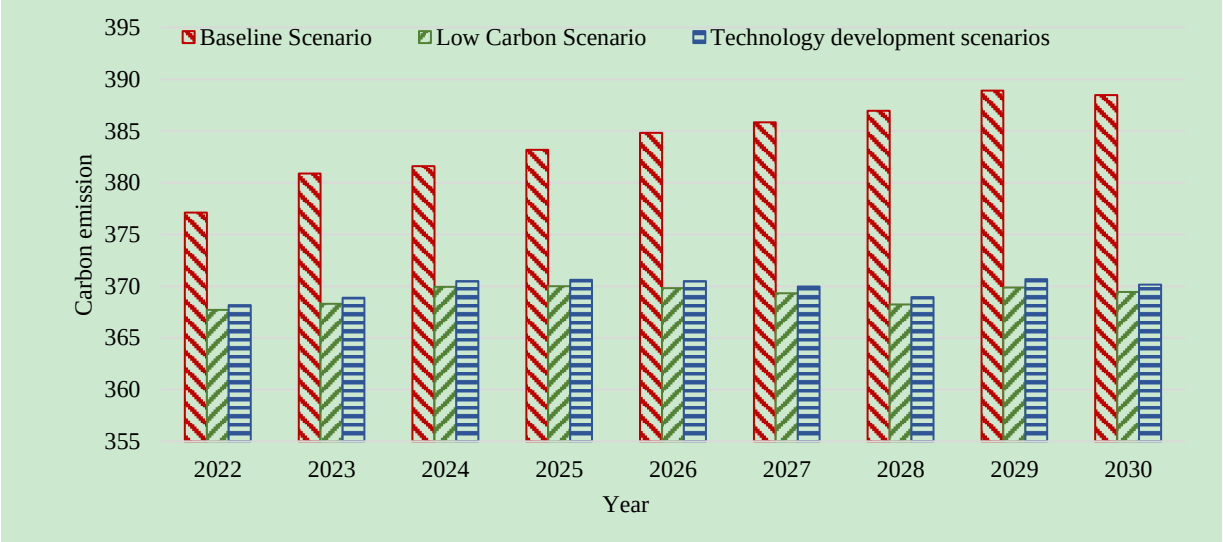


Fig.17 Comparison of carbon emissions under different scenarios between 2022-2030.

The side-by-side comparison presented in Fig. 17 reveals significant disparities in carbon emissions under the three different scenarios, along with variations in the timing of peak emissions. Notably, the baseline scenario exhibits higher carbon emissions compared to the other two scenarios, underscoring the efficacy of current energy conservation and emission reduction policies. In contrast, the low-carbon scenario shows an earlier peak in carbon emissions, indicating a notable reduction in emissions due to the anticipated impact of emission reduction measures.

The scenario of research and technological development shows a somewhat reduced level of carbon emissions compared to the low-carbon scenario, but also reaching the peak earlier in time. This suggests that investments in science and technology can contribute to reducing carbon emissions to a certain extent. Indeed, the variation in emission reduction effects across regions underscores the need for customized strategies aligned with each region's unique development context. This approach can ensure sustainable economic development while effectively curbing the growth of carbon emissions, ultimately working toward the goal of achieving carbon peaking.

6. Conclusions

In this study, an improved EBRB model, known as the TSF-EBRB model, was introduced for time series data analysis. This model was applied to predict the timing of carbon peaks across various Chinese provinces. The empirical analysis encompassed data spanning from 2004 to 2021, covering 30 Chinese provinces. The research started by conducting a thorough examination of the data, which was then followed by the assessment of the efficacy of the TSF-EBRB model. Predictions were generated for the carbon peak in each province up to 2030, considering three distinct scenarios: the baseline scenario, the low-carbon scenario, and the scientific and technical progress scenario. The primary outcomes of the research are succinctly outlined as follows:

(1) The introduction of the TSF-EBRB model enables us to more accurately consider the temporal correlation when predicting the carbon emission peaks of various provinces in China, thereby enhancing the accuracy and reliability of the forecasts. By integrating time series analysis with parameter optimization methods, this model effectively captures the temporal dynamics inherent in carbon emission data. Comparative analysis of prediction errors across 30 provinces in China demonstrates that the TSF-EBRB model outperforms traditional models such as EBRB, GM (1,1), ARIMA, and SVM in terms of prediction accuracy. The robustness of this approach underscores its value in guiding decision-making and strategic planning.

(2) The forecasting results indicate significant regional disparities in the timing of carbon emissions peaks across Chinese provinces. While many provinces are expected to reach their peaks around 2030, some regions may face challenges in achieving this target. Therefore, policymakers need to tailor emission reduction strategies to the specific needs and circumstances of each region, ensuring the effectiveness and fairness of intervention measures.

(3) This study introduces the TSF-EBRB model, which makes significant theoretical contributions to the application of the EBRB model in time series data prediction. By effectively leveraging the temporal correlation in time series data and introducing parameter optimization methods, the TSF-EBRB model improves the accuracy and stability of predicting the peak carbon emissions in various provinces of China. Moreover, the model's strong interpretability and comprehensibility enable clear explanation and understanding of the prediction results. Therefore, the TSF-EBRB model provides new theoretical insights and methods for research in time series data prediction, making important theoretical contributions to achieving accurate, robust, and interpretable time series forecasting.

Based on the research results of this paper, governments and enterprises can derive the following economic benefits from this research:

(1) Based on the high accuracy and reliability of the TSF-EBRB model, the government should consider the correlation and dynamic of the data when considering the carbon emission prediction. In order to ensure the timeliness and accuracy of prediction models, it is recommended that the government establish a dedicated carbon emission prediction body or department, whose responsibilities include not only monitoring and updating the prediction models, it also includes the regular collection, analysis and assessment of up-to-date carbon emission data and the adjustment of model parameters to reflect the long-term dynamics of these data. In addition, the government should develop and implement relevant data maintenance and update policies to ensure the quality and integrity of carbon emission data, therefore, it can provide reliable forecast data and basis for making scientific and reasonable carbon emission management policy.

(2) According to the TSF-EBRB model, the government should formulate different policies to reduce carbon emissions according to the peak time of carbon emissions in different provinces. For provinces that are expected to peak around 2030, the government can take more stringent measures to speed up the process of reducing emissions; for provinces facing challenges, the government can provide more support and policy preferences, help them meet their emissions targets.

(3) Enterprises can leverage the TSF-EBRB model's accurate predictions to adjust business strategies, reducing carbon emission risks and improving competitiveness. By integrating temporal correlations into their planning processes, enterprises can optimize production and resource allocation, lowering energy consumption and mitigating emissions. This not only reduces costs but also enhances overall operational efficiency and market competitiveness.

Although the TSF-EBRB model has been demonstrated to be effective in carbon emission prediction through experimental analysis, it is important to note that the model does not account for carbon emission efficiency or environmental regulatory policies. In future research, these factors should be integrated into the existing model to create a more comprehensive and effective carbon emission prediction framework.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Nos. 72301071 and 72001043), and the Natural Science Foundation of Fujian Province of China (Nos. 2022J01178 and 2020J05122).

References

Bi, W., Gao, F., Zhang, A., & Bao, S. (2022). A framework for extended belief rule base reduction and training with the greedy strategy and parameter learning. *Multimedia Tools and Applications*, *81*(8), 11127-11143. https://doi.org/10.1007/s11042-022-12232-4

Bukhari, A. H., Raja, M. A. Z., Sulaiman, M., Islam, S., Shoaib, M., & Kumam, P. (2020). Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting. *IEEE Access*, *8*, 71326-71338. https://doi.org/10.1109/ACCESS.2020.2985763

Calzada, A., Liu, J., Wang, H., & Kashyap, A. (2015). A New Dynamic Rule Activation Method for Extended Belief Rule-Based Systems. *IEEE Transactions on Knowledge and Data Engineering*, *27*(4), 880-894. https://doi.org/10.1109/TKDE.2014.2356460

Cao, Y., Zhou, Z., Tang, S., Ning, P., & Chen, M. (2023). On the Robustness of Belief-Rule-Based Expert Systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *53*(10), 6043-6055. https://doi.org/10.1109/TSMC.2023.3279286

Chang, L. L., Zhou, Z. J., Chen, Y. W., Liao, T. J., Hu, Y., & Yang, L. H. (2018). Belief Rule Base Structure and Parameter Joint Optimization Under Disjunctive Assumption for Nonlinear Complex System Modeling. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *48*(9), 1542-1554. https://doi.org/10.1109/TSMC.2017.2678607

Chen, M.-Y., & Linkens, D. A. (2004). Rule-base self-generation and simplification for data-driven fuzzy models. *Fuzzy Sets and Systems*, *142*(2), 243-265. https://doi.org/10.1016/S0165-0114(03)00160-X

Chen, Y.-W., Yang, J.-B., Xu, D.-L., Zhou, Z.-J., & Tang, D.-W. (2011). Inference analysis and adaptive training for belief rule based systems. *Expert Systems with Applications*, *38*(10), 12845-12860. https://doi.org/10.1016/j.eswa.2011.04.077

Cheng, S., Fan, Q., & Dagestani, A. A. (2023). Opening the black box between strategic vision on digitalization and SMEs digital transformation: the mediating role of resource orchestration. *Kybernetes*. https://doi.org/10.1108/K-01-2023-0073

Dagestani, A. A., & Qing, L. (2022). The Impact of Environmental Information Disclosure on Chinese Firms' Environmental and Economic Performance in the 21st Century: A Systematic Review. *IEEE Engineering Management Review*, *50*(4), 203-214. https://doi.org/10.1109/EMR.2022.3210465

Dagestani, A. A., Shang, Y., Schneider, N., Cifuentes-Faura, J., & Zhao, X. (2023). Porter in China: A quasi-experimental view of market-based environmental regulation effects on firm performance. *Energy Economics*, *126*, 106966. https://doi.org/10.1016/j.eneco.2023.106966

Dagestani, A. A., Tariq, S., Khan, M., Kamal, M., & Rehman, M. u. (2024). Assessment of nighttime air quality over an urban location in Indo-Gangetic plain using remote sensing observations. *Atmospheric Pollution Research*, *15*(2), 101982. https://doi.org/10.1016/j.apr.2023.101982

Dietz, T., & Rosa, E. A. (1997). Effects of population and affluence on CO2 emissions. *94*(1), 175-179. https://doi.org/doi:10.1073/pnas.94.1.175

Du, Y., Chen, K., Chen, S., & Yin, K. (2022). Prediction of Carbon Emissions Trading Price in Fujian Province: Based on BP Neural Network Model [Original Research]. *10*. https://doi.org/10.3389/fenrg.2022.939602

Fu, C., Hou, B., Xue, M., Chang, L., & Liu, W. (2023). Extended Belief Rule-Based System With Accurate Rule Weights and Efficient Rule Activation for Diagnosis of Thyroid Nodules. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *53*(1), 251-263. https://doi.org/10.1109/TSMC.2022.3180174

Fu, Y.-G., Zhuang, J.-H., Chen, Y.-P., Guo, L.-K., & Wang, Y.-M. (2020). A framework for optimizing extended belief rule base systems with improved Ball trees. *Knowledge-Based Systems*, *210*, 106484. https://doi.org/10.1016/j.knosys.2020.106484

González-Torres, M., Pérez-Lombard, L., Coronel, J. F., & Maestre, I. R. (2021). Revisiting Kaya Identity to define an Emissions Indicators Pyramid. *Journal of Cleaner Production*, *317*, 128328. https://doi.org/10.1016/j.jclepro.2021.128328

Hao, J., Gao, F., Fang, X., Nong, X., Zhang, Y., & Hong, F. (2022). Multi-factor decomposition and multi-scenario prediction decoupling analysis of China's carbon emission under dual carbon goal. *Science of The Total Environment*, *841*, 156788. https://doi.org/https://doi.org/10.1016/j.scitotenv.2022.156788

Huang, J., Li, X., Wang, Y., & Lei, H. (2021). The effect of energy patents on China's carbon emissions: Evidence from the STIRPAT model. *Technological Forecasting and Social Change*, *173*, 121110. https://doi.org/10.1016/j.techfore.2021.121110

Jian-Bo, Y., Jun, L., Jin, W., How-Sing, S., & Hong-Wei, W. (2006). Belief rule-base inference methodology using the evidential reasoning Approach-RIMER. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *36*(2), 266-285. https://doi.org/10.1109/TSMCA.2005.851270

Jiang, H., Yin, J., Wei, D., Luo, X., Ding, Y., & Xia, R. (2024). Industrial carbon emission efficiency prediction and carbon emission reduction strategies based on multi-objective particle swarm optimization-backpropagation: A perspective from regional clustering. *Science of The Total Environment*, *906*, 167692. https://doi.org/10.1016/j.scitotenv.2023.167692

Jiang, X., & Sun, H.-W. (2023). Learning performance of uncentered kernel-based principal component analysis. *International Journal of Wavelets, Multiresolution and Information Processing, 21*(03), 2250059. https://doi.org/10.1142/s021969132250059x

Jiao, M., Wang, D., & Qiu, J. (2020). A GRU-RNN based momentum optimized algorithm for SOC estimation. *Journal of Power Sources*, *459*, 228051. https://doi.org/10.1016/j.jpowsour.2020.228051

Koopmans, M. (2021). On the distinction between fractal and seasonal dependencies in time series data. *Fractals*, *29*. https://doi.org/10.1142/S0218348X21501802

Li, Z.-Z., Li, R. Y. M., Malik, M. Y., Murshed, M., Khan, Z., & Umar, M. (2021). Determinants of Carbon Emission in China: How Good is Green Investment? *Sustainable Production and Consumption*, *27*, 392-401. https://doi.org/10.1016/j.spc.2020.11.008

Liu, J., Yang, Q., Ou, S., & Liu, J. (2022). Factor decomposition and the decoupling effect of carbon emissions in China's manufacturing high-emission subsectors. *Energy*, *248*, 123568. https://doi.org/10.1016/j.energy.2022.123568

Liu, X., Wang, X., & Meng, X. (2023). Carbon Emission Scenario Prediction and Peak Path Selection in China. *Energies*, *16*, 2276. https://doi.org/10.3390/en16052276

Luo, H., Li, Y., Gao, X., Meng, X., Yang, X., & Yan, J. (2023). Carbon emission prediction model of prefecture-level administrative region: A land-use-based case study of Xi'an city, China. *Applied Energy*, *348*, 121488. https://doi.org/10.1016/j.apenergy.2023.121488

Ma, H., Sun, W., Wang, S., & Kang, L. (2019). Structural contribution and scenario simulation of highway passenger transit carbon emissions in the Beijing-Tianjin-Hebei metropolitan region, China. *Resources, Conservation and Recycling*, *140*, 209-215. https://doi.org/10.1016/j.resconrec.2018.09.028

Nie, W., & Duan, H. (2023). A novel multivariable grey differential dynamic prediction model with new structures and its application to carbon emissions. *Engineering Applications of Artificial Intelligence*, *122*, 106174. https://doi.org/10.1016/j.engappai.2023.106174

Niu, L., Zhagn, L., Xi, F., & Wang, J. (2023). Influencing factors and scenario forecasting of carbon emissions in Liaoning Province, China. *The journal of applied ecology*, *34*(2), 499-509. https://doi.org/10.13287/j.1001-9332.202302.001

Pan, X., Guo, S., Xu, H., Tian, M., Pan, X., & Chu, J. (2022). China's carbon intensity factor decomposition and carbon emission decoupling analysis. *Energy*, *239*, 122175. https://doi.org/https://doi.org/10.1016/j.energy.2021.122175

Qing, L., Alnafrah, I., & Dagestani, A. A. (2023). Does green technology innovation benefit corporate financial performance? Investigating the moderating effect of media coverage. *Corporate Social Responsibility and Environmental Management*, 1-19. https://doi.org/10.1002/csr.2659

Qing, L., Li, P., Mehmood, U., & Dagestani, A. A. (2024a). Uncovering the potential impacts of financial inclusion and human development on ecological sustainability in the presence of natural resources and government stability: Evidence from G-20 nations. *Resources Policy*, *88*, 104446. https://doi.org/10.1016/j.resourpol.2023.104446

Ren, F., & Long, D. (2021). Carbon emission forecasting and scenario analysis in Guangdong Province based on optimized Fast Learning Network. *Journal of Cleaner Production*, *317*, 128408. https://doi.org/10.1016/j.jclepro.2021.128408

Romanuke, V. (2022). Arima Model Optimal Selection for Time Series Forecasting. *Maritime Technical Journal*, *224*, 28-40. https://doi.org/10.2478/sjpna-2022-0003

Shen, B., Yang, X., Xu, Y., Ge, W., Liu, G., Su, X., &Ran, Q. (2023). Can carbon emission trading pilot policy drive industrial structure low-carbon restructuring: new evidence from China. *Environmental Science and Pollution Research*, *30*(14), 41553-41569. https://doi.org/10.1007/s11356-023-25169-4

Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*, 132306. https://doi.org/10.1016/j.physd.2019.132306

Sippel, S., Meinshausen, N., Fischer, E. M., Székely, E., & Knutti, R. (2020). Climate change now detectable from any single day of weather at global scale. *Nature Climate Change*, *10*(1), 35-41. https://doi.org/10.1038/s41558-019-0666-7

Sun, W., & Huang, C. (2022). Predictions of carbon emission intensity based on factor analysis and an improved extreme learning machine from the perspective of carbon emission efficiency. *Journal of Cleaner Production*, *338*, 130414. https://doi.org/10.1016/j.jclepro.2022.130414

Wang, Q.-Z., Wang, N.-N., Tseng, M.-L., Huang, Y.-M., & Li, N.-L. (2020). Waste tire recycling assessment: Road application potential and carbon emissions reduction analysis of crumb rubber modified asphalt in China. *Journal of Cleaner Production*, *249*, 119411. https://doi.org/10.1016/j.jclepro.2019.119411

Wen, J., Yang, J., Jiang, B., Song, H., & Wang, H. (2021). Big Data Driven Marine Environment Information Forecasting: A Time Series Prediction Network. *IEEE Transactions on Fuzzy Systems*, *29*(1), 4–18. https://doi.org/10.1109/tfuzz.2020.3012393

Wu, R., Wang, J., Wang, S., & Feng, K. (2021). The drivers of declining CO2 emissions trends in developed nations using an extended STIRPAT model: A historical and prospective analysis. *Renewable and Sustainable Energy Reviews*, *149*, 111328. https://doi.org/10.1016/j.rser.2021.111328

Xu, D.-L., Liu, J., Yang, J.-B., Liu, G.-P., Wang, J., Jenkinson, I., & Ren, J. (2007). Inference and learning methodology of belief-rule-based expert system for pipeline leak detection. *Expert Systems with Applications*, *32*(1), 103-113. https://doi.org/10.1016/j.eswa.2005.11.015

Yang, H., & O'Connell, J. F. (2020). Short-term carbon emissions forecast for aviation industry in Shanghai. *Journal of Cleaner Production*, *275*, 122734. https://doi.org/10.1016/j.jclepro.2020.122734

Yang, L.-H., Liu, J., Wang, Y.-M., & Martínez, L. (2019). New activation weight calculation and parameter optimization for extended belief rule-based system based on sensitivity analysis. *Knowledge and Information Systems*, *60*(2), 837-878. https://doi.org/10.1007/s10115-018-1211-0

Yang, L.-H., Wang, Y.-M., & Fu, Y.-G. (2018). A consistency analysis-based rule activation method for extended belief-rule-based systems. *Information Sciences*, *445-446*, 50-65. https://doi.org/10.1016/j.ins.2018.02.059

Yang, L.-H., Wang, Y.-M., Su, Q., Fu, Y.-G., & Chin, K.-S. (2016). Multi-attribute search framework for optimizing extended belief rule-based systems. *Information Sciences*, *370-371*, 159-183. https://doi.org/10.1016/j.ins.2016.07.067

Yang, L.-H., Ye, F.-F., Hu, H., Lu, H., Wang, Y.-M., & Chang, W.-J. (2024). A data-driven rule-base approach for carbon emission trend forecast with environmental regulation and efficiency improvement. *Sustainable Production and Consumption*, *45*, 316-332. https://doi.org/10.1016/j.spc.2023.12.030

Ye, F.-F., Wang, S., Nicholl, P., Yang, L.-H., & Wang, Y.-M. (2020). Extended belief rule-based model for environmental investment prediction with indicator ensemble selection. *International Journal of Approximate Reasoning*, *126*, 290-307. https://doi.org/10.1016/j.ijar.2020.08.013

Ye, F.-F., Yang, L.-H., Lu, H.-T., & Wang, Y.-M. (2022). A novel data-driven decision model based on extended belief rule base to predict China's carbon emissions. *Journal of Environmental Management*, *318*, 115547. https://doi.org/10.1016/j.jenvman.2022.115547

Zhang, Q., Zhao, B., He, W., Zhu, H., & Zhou, G. (2024). A behavior prediction method for complex system based on belief rule base with structural adaptive. *Applied Soft Computing*, *151*, 111118. https://doi.org/10.1016/j.asoc.2023.111118

Zhang, S., Huo, Z., & Zhai, C. (2022). Building Carbon Emission Scenario Prediction Using STIRPAT and GA-BP Neural Network Model. *14*(15), 9369. https://www.mdpi.com/2071-1050/14/15/9369

Zhou, Z.-J., Hu, C.-H., Yang, J.-B., Xu, D.-L., & Zhou, D.-H. (2009). Online updating belief rule based system for pipeline leak detection under expert intervention. *Expert Systems with Applications*, *36*(4), 7700-7709. https://doi.org/10.1016/j.eswa.2008.09.032