Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 8, No. 1, 2026

Analysis and Prediction of Factors Influencing Carbon Emissions in Guangdong Province Based on STIRPAT Interpretability Analysis

Yiting Qiu[1,#], Jingmin Lan[1,#], Xiaolu Zou[1,#], Yun Huang[1,#], Ying Tan[1,#], Dang Gan[1,#], Jiaxin Huang[1,#], Jiaqi Zhang[1,#], Zhihan Wang[1,#], Junpeng Yuan[1,*,#], Hang Zeng[2,#], Yiling Zhang[3,#]

[1]School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China

[2]School of Economics and Statistic, Guangzhou University, Guangzhou 510006, China

[3]School of Management, Guangzhou University, Guangzhou 510006, China.

[*]**Corresponding author**: Junpeng Yuan (Email: 13257516286@163.com)

[#]**These authors contributed equally.**

Abstract: This study employs an extended STIRPAT model and SHAP interpretability analysis to identify and quantify the key socioeconomic drivers of carbon emissions in Guangdong Province, China. Ridge regression is applied to mitigate multicollinearity among six variables: population, affluence, industrial structure, openness, technology, and energy mix. Results indicate that population size exerts the strongest positive effect (+2.31 % per 1 % increase), followed by industrial structure (+2.1 %), whereas technology (−0.31 %), openness (−0.07 %), and cleaner energy structure (−0.18 %) significantly curb emissions. A high-performing back-propagation neural network optimized by improved particle swarm optimization (IPSO-BP) is then trained to predict future emissions, with SHAP values revealing the marginal contribution of each driver under different policy scenarios. The integrated framework offers transparent, evidence-based guidance for Guangdong to achieve its carbon-peaking and carbon-neutrality targets and provides a replicable template for other rapidly developing regions.

Keywords: STIRPAT model, Carbon-emission drivers, Ridge regression, IPSO-BP neural network.

1. Introduction

With the acceleration of global industrialisation, greenhouse gas emissions continue to rise, leading to global warming and frequent extreme weather events that pose severe challenges to human survival and development. As the world's largest developing nation, China ranks among the top global emitters of carbon dioxide and faces immense pressure to reduce emissions. Guangdong Province, one of China's most economically developed regions, holds a pivotal position in the nation's economic development while simultaneously grappling with significant carbon emission issues.

Against the backdrop of global climate change mitigation efforts, scientifically analysing the factors influencing Guangdong's carbon emissions and accurately forecasting future trends holds significant importance for formulating rational carbon reduction policies and promoting the province's sustainable economic development. This will not only assist Guangdong in achieving its own carbon peak and carbon neutrality objectives but also provide valuable insights and reference for other regions across the nation[1].

Domestic and international research in carbon emissions measurement has long relied on traditional statistical and accounting-based methods, such as the IPCC Guidelines and the STIRPAT model, which calculate carbon emissions and analyze their trends by combining energy consumption data with emission factors. With technological advancements, methods like machine learning and big data analysis have been introduced to this field, effectively enhancing measurement accuracy. Regarding influencing factors, economic development levels and energy intensity are widely recognized as key drivers of carbon emission growth, while technological innovation and industrial restructuring can significantly suppress emissions. In forecasting research, methodologies have evolved from early time series analysis and regression models to machine learning approaches like Long Short-Term Memory networks and Convolutional Neural Networks. These newer methods better capture the complex nonlinear characteristics and temporal dependencies inherent in carbon emissions.

However, existing research still faces significant limitations, including insufficient accuracy of foundational data, inadequate analysis of regional and sector-specific variations, and issues with prediction models themselves—such as high complexity, substantial computational costs, and weak interpretability. These shortcomings hinder in-depth understanding of regional carbon emission characteristics and undermine the scientific basis for related policy formulation. To address these challenges, this study proposes to employ an extended STIRPAT model integrated with machine learning techniques. By synthesizing multidimensional factors such as economic development, industrial structure, and energy mix, we aim to conduct quantitative analysis and precise forecasting of Guangdong Province's carbon emissions. The research will enhance model interpretability through techniques like SHAP feature analysis, thereby providing robust theoretical and decision-making support for Guangdong to formulate scientifically sound carbon reduction pathways and achieve its carbon peak and carbon neutrality objectives.

2. Analysis and Model Development

2.1. Development of the Extended STIRPAT Model

Prior to establishing the carbon emissions forecasting model, analysing the correlation between various factors and carbon emissions, along with their respective impact levels, is crucial, laying the groundwork for subsequent research. This paper employs the STIRPAT model, a flexible stochastic environmental impact assessment tool designed to investigate the influence of social factors on carbon emissions.

The STIRPAT model, grounded in regression analysis, aims to uncover relationships between environmental impacts (such as carbon emissions) and factors including population, affluence, and technological advancement. It posits that environmental impacts arise from the combined effects of these variables. By quantifying the extent to which these factors influence environmental outcomes, it assists policymakers in better understanding and addressing environmental challenges[2]. Moreover, the STIRPAT model boasts broad applicability, serving not only for carbon emissions assessment but also for environmental impact analysis across multiple domains including energy consumption and water resource utilisation. This comprehensive perspective renders STIRPAT a vital tool for environmental research and policy formulation. Its standard formula is as follows:

$$I = aP^b A^c T^d e \tag{1}$$

Where a represents the constant term; b, c, d denotes the parameter exponent requiring estimation; and e indicates the error term.

This model is based on a multivariable nonlinear framework. Taking the logarithm of both sides of the standard formula yields:

$$\ln I = \ln a + b \ln P + c \ln A + d \ln T + \ln e \tag{2}$$

To construct a more comprehensive model analysing the determinants of carbon emissions, this paper extends the traditional STIRPAT model to account for Guangdong Province's unique geographical and economic characteristics. Considering Guangdong's locational advantages as a coastal province, its frequent international trade activities, and distinctive industrial development patterns, we incorporate the degree of openness to the outside world and industrial structure as key indicators within the model. This extended STIRPAT model more accurately reflects the multifaceted determinants of carbon emissions in Guangdong Province. Its calculation formula is as follows:

$$\ln C = \ln a + b \ln P + c \ln A + d \ln T + e \ln F + f \ln S + g \ln E + \ln e \tag{3}$$

Where: C denotes carbon emissions; P represents population size, expressed as the year-end resident population; A indicates affluence, measured here by per capita GDP; T signifies technological level, represented by energy intensity; F reflects openness to the outside world, measured by the ratio of total goods imports and exports to GDP; S indicates industrial structure, represented by the ratio of secondary industry value-added to GDP; E denotes energy structure, represented by the ratio of coal consumption to total energy consumption[3].

2.2. Development of the Backpropagation Neural Network Model

The BP neural network, formally known as the backpropagation neural network, is an artificial neural network model widely applied in the field of artificial intelligence. It employs supervised learning techniques, primarily for training multi-layer feedforward neural networks. The structure of a BP network typically comprises an input layer, one or more hidden layers, and an output layer. Its core characteristic involves calculating the error between predicted and actual values, then propagating this error backwards from the output layer through each layer to adjust the connection weights. This process aims to minimise overall error, thereby enabling the model to better fit the training data.

2.3. Model Optimisation – Improved Particle Swarm Optimisation (IPSO)

Despite the widespread application of BP neural networks in classification, regression, and pattern recognition tasks, they still face challenges such as susceptibility to local minima and relatively slow training speeds[4]. To address these issues, researchers have proposed a series of improved algorithms and network architectures. This study employs an improved particle swarm optimisation algorithm (IPSO) to enhance the performance of the BP neural network model.

The Particle Swarm Optimisation (PSO) algorithm draws inspiration from the foraging behaviour of birds. By simulating cooperation and information sharing within a swarm, it achieves efficient exploration within the solution space. Each individual within the swarm (referred to as a particle) continuously adjusts its search strategy by learning from its own and other members' optimal historical positions to seek the global optimum. PSO has found extensive application in various fields, including neural network training and function optimisation, owing to its simplicity of implementation and rapid convergence[5].

The enhanced IPSO algorithm further optimises PSO performance through linear weight adjustment and the introduction of mutation operations.

3. Model Solution

3.1. Solution of the STIRPAT Model

Based on the aforementioned analysis, this paper employs the modified STIRPAT model to analyse the factors influencing carbon emissions using the following formula.

$$ln\, C = ln\, a + b\, ln\, P + c\, ln\, A + d + ln\, T + e\, ln\, F + f\, ln\, S + g\, ln\, E + ln\, e \tag{4}$$

First, regression analysis was performed on the collected data using SPSS PRO, with results presented in the table below:

Table 1 Regression Analysis Fitting Results

Model	R^2	F-value	P-value
Value	0.979	1774.158	<0.001

Table 1 presents the regression analysis fitting results. With

an R^2 value of 0.979, an F-value of 1774.158, and a P-value below 0.001, the model demonstrates excellent fitting performance[6].This confirms a strong correlation between the selected six variables and carbon emissions, validating the effectiveness of the variable selection. It indicates that the model exhibits a favourable multiple linear regression fit, demonstrating a strong correlation between the selected six variables and carbon emissions. This confirms the validity of the variable selection.

Secondly, multicollinearity tests were conducted on the six selected variables, with results presented in the table 2 below:

Table 2 Variable Multicollinearity Statistics

Variable	Tolerance	VIF
Total Population	0.013	77.313
GDP per capita	0.012	81.542
Technical level	0.175	56.745
Opening Up	0.112	8.445
Industrial Structure	0.212	3.456
Energy Structure	0.083	9.583

The results in Table 2 indicate that the VIF values for the three variables—total population, per capita GDP, and technological level—exceed 50, suggesting a degree of multicollinearity. Employing ordinary least squares (OLS) regression directly would yield coefficients of questionable accuracy. Therefore, this study employs ridge regression analysis based on OLS to mitigate the mutual influence among variables.

The final coefficient results are presented in the table below:

Table 3 Ling-Yin Regression Results

Variable	Coefficient	Standard Error	Standardised Coefficient	t-statistic
Population size P	2.3154	0.0780	0.2716	14.9714
Affluence Index A	1.2162	0.0111	0.2737	21.1575
Industrial Structure S	0.1074	0.1723	0.1574	9.54783
Foreign Opening-up F	-0.0723	0.1314	-0.1975	-15.6555
Technical Level T	-0.3123	0.1873	-0.1486	-18.4966
Energy Structure E	-0.1792	0.1921	-0.0125	-1.0547
Constant	2.3363	1.2753	0.0000	-3.5784

Table 3 presents the results of the ridge regression analysis. It displays the coefficients, standard errors, standardized coefficients, and t-statistics for each variable, including Population Size P, Affluence Index A, Industrial Structure S, Foreign Opening-up F, Technological Level T, Energy Structure E, and the Constant. The post-training model yields an F-value of 136.5781 and a p-value of 0.000, indicating statistical significance[7]. The adjusted R-squared value (R^2) is 0.9717, signifying that the model exhibits a high degree of goodness of fit. The final linear equation representation of the STIRPAT model is expressed as:

$$ln\,C = 2.3154\,ln\,P + 1.2162\,ln\,A + 0.1074\,ln\,T - 0.0723\,ln\,F - 0.1074\,ln\,S - 0.1792 - 13.6055 \quad (5)$$

4. Conclusions

Based on the equation coefficients, Guangdong Province's total population size exerts the most significant influence on carbon emissions. Total population and carbon emissions exhibit a positive correlation, with a 1% change in population leading to a 2.31% increase in carbon emissions. The impact of per capita GDP is relatively minor compared to total population size; a 1% increase in per capita GDP results in only a 0.22% rise in carbon emissions.

Regarding industrial structure, a 1% increase in the value-added share of the secondary sector leads to a 2.1% rise in carbon emissions. The secondary sector encompasses key carbon-emitting industries such as mining, manufacturing, and construction. To advance the goals of carbon peaking and carbon neutrality, industrial structure requires optimisation and upgrading to reduce emissions.

Negative correlations exist between energy structure, openness to foreign trade, technological advancement, and carbon emissions. Among these, openness to foreign trade—specifically foreign trade volume—exerts a relatively minor influence, with a 1% increase in foreign trade volume reducing carbon emissions by merely 0.07%.

References

[1] Dong Wei, Che Jinxing. Modelling China's Industrial Carbon Emissions Forecast Based on the Kupman Theory [J]. Journal of Nanchang Institute of Engineering, 2025, 44(01): 107-112.

[2] Liu Fengying, Zhong Mingte, Zhu Bifeng. Construction of a Prediction Framework for Carbon Emission Intensity in University Campuses Based on Energy Consumption Drivers [J]. Green Building, 2025, 17(02): 8-14.

[3] Yao Xiaojian, Wu Yuyuan. Analysis of Carbon Emission Influencing Factorsand Scenario Prediction for the Guangdong-Hong Kong-Macao Greater Bay Area [J/OL]. Resources and Industries, 1-17 [2025-04-18]. https://doi.org/10.13776/j.cnki.resourcesindustries.20250217.002.

[4] Guan Minjie, Yuan Yanhong, Ran Muxi, Wang Zi. Influencing Factors and Peak Prediction of Energy Carbon Emissions in Shanxi Province Based on the STIRPAT Model [J]. China Coal, 2021, 47(09): 48-55.

[5] Feng Qisheng, Gao Xinhua. The Role of Excel in Leave-One-Out Cross-Validation Experimental Teaching [J]. Experimental Science and Technology, 2015,13(02): 49-51+65.

[6] Chen Jiaxin, Deng Ao. Analysis of Factors Affecting Carbon Emissions: A Case Study of Guangdong Province [J]. China Environmental Protection Industry, 2025, (07): 59-62+72. DOI: CNKI:SUN:ZHBY.0.2025-07-018.

[7] Kong Xiangbo. Analysis of Operational Efficiency and Influencing Factors of China's Coastal Ports Considering Carbon Emissions [J]. Journal of Inner Mongolia University for Nationalities (Natural Science Edition), 2025, 40(04): 62-69. DOI: 10.14045/j.cnki.15-1220.2025.04.010.