Economics & Business Management
ISSN Print: 3079-5214
ISSN Online: 3079-5222
https://doi.org/10.63313/EBM.9159

Carbon Emission Prediction for Sichuan Province Using a SHAP-Explained Machine Learning Framework

ChangZhi Xie
School of Economics and Management, Southwest Petroleum University, Chengdu 610500, China
Email: 953797218@qq.com

How to cite this paper: Xie, C. Z. (2026). Carbon Emission Prediction for Sichuan Province Using a SHAP-Explained Machine Learning Framework. *Economics & Business Management*, 5(1), 49-56. ISSN Print: 3079-5214; ISSN Online: 3079-5222. https://doi.org/10.63313/EBM.9159
Published: 2026-03-05

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Abstract

This study develops a provincial carbon accounting and forecasting framework for Sichuan Province covering 2005–2022 under a consumption-based boundary. Direct fossil fuel emissions and indirect electricity-related emissions are clearly distinguished to ensure accounting consistency and additive closure. On this basis, a Transformer-based machine learning model integrated with SHAP is constructed to predict carbon emissions and identify peak characteristics. The results show high historical fitting accuracy and project that Sichuan's carbon emissions will reach a plateau-type peak around 2031, followed by a gradual decline. SHAP-based interpretation indicates that total energy consumption and the carbon emission factor are the dominant drivers, while economic scale and energy intensity exert secondary but significant effects. Robustness tests confirm the stability of the explanatory structure. The proposed framework integrates accounting consistency, predictive modeling, and interpretability analysis, providing empirical support for regional peak management and low-carbon transition planning.

Keywords

Carbon emission forecasting; Transformer model; SHAP analysis; Sichuan Province

1. Introduction

1.1. Research Background

Recent policy reforms have emphasized the deepening of structural adjustments in the oil and gas sector, with the overarching objective of strengthening national energy security. Institutional reforms have been progressively advanced across upstream, midstream, and downstream segments to address structural inefficiencies, while the strategic petroleum reserve system has been further consolidated. In parallel, the 14th Five-Year Plan for National Work Safety

underscores the need to enhance full-process risk assessment in oil and gas production, promote the adoption of reliable technologies and equipment, and improve emergency response and mitigation mechanisms. Beyond sectoral safety reforms, the commitment to carbon neutrality has provided a clear strategic direction for low-carbon transition across industries. At the national level, a governance framework characterized by "total quantity control, efficiency enhancement, and structural transformation" has gradually taken shape. In this framework, the dual control of total energy consumption and carbon intensity serves as the principal regulatory instrument, with priority industries and regions designated as focal points for coordinated implementation. Compared with earlier single-dimensional energy consumption constraints, the carbon peaking and carbon neutrality agenda places stronger emphasis on lifecycle-wide and system-wide emission mitigation. This approach encompasses low-carbon substitution on the primary energy supply side as well as electrification and efficiency improvements on the end-use side, while requiring measurable and regionally differentiated transition pathways with explicit temporal benchmarks. Against this backdrop, Sichuan Province represents a particularly relevant case due to its distinctive and complex structural characteristics. On the one hand, the province possesses abundant hydropower resources, large installed hydroelectric capacity, and a relatively high share of clean electricity generation. Favorable natural gas endowments further support terminal decarbonization pathways anchored in electricity and gaseous fuels. On the other hand, sustained economic expansion, a large population base, and an extensive industrial chain—encompassing chemicals, metallurgy, building materials, and equipment manufacturing—continue to generate rigid and structurally embedded energy demand.In light of these structural features, this study addresses the following core question: under the overarching constraints of the carbon peaking and carbon neutrality targets, how can a provincial carbon accounting and forecasting framework be constructed that is boundary-consistent, methodologically unified, and empirically verifiable, while explicitly incorporating Sichuan's high clean electricity share, strong seasonal volatility, and diversified industrial structure? Furthermore, how can operational peak identification criteria and validation procedures be established within such a framework? Addressing these issues is essential not only for evaluating Sichuan's pathway toward carbon neutrality but also for providing a replicable analytical paradigm for regions with comparable energy–industrial configurations.

1.2. Literature Review

Guo Chaoxian employed the Logarithmic Mean Divisia Index (LMDI) method to decompose China's carbon emissions [1], constructing a multifactor carbon emission model that incorporated two economic indicators—total economic output and economic structure—as well as three energy-environmental indicators: energy utilization efficiency, energy consumption structure, and carbon emission

coefficients. Based on data from 1995 to 2007, the study demonstrated that these factors constitute the principal drivers of China's carbon emissions. Guo Yungong applied the STIRPAT model to examine the relationship between economic development and carbon emissions in Shanghai [2]. Building upon the original IPAT framework and calibrating the model using empirical data, the study concluded that population growth and accelerated urbanization are the primary determinants of Shanghai's carbon emissions. In her doctoral dissertation, Zhang Lifeng developed a forecasting model for China's energy supply and demand [3]. Through fitting historical data and implementing methodological refinements, she established an integrated energy prediction framework that combined system dynamics models, vector autoregression (VAR), grey system models, and variable-weight combination forecasting models. This framework was used to conduct comprehensive medium- and long-term projections of China's total energy supply, demand, and energy structure.Xu Guangyue, in his doctoral empirical research, employed econometric techniques to investigate the relationships among energy consumption, carbon emissions, and economic growth [4]. The study applied panel unit root and panel cointegration tests, followed by a VAR model to examine dynamic interactions among the three variables. Drawing on the Environmental Kuznets Curve (EKC) theory, the research identified the structural relationship between economic growth, energy consumption, and carbon emissions. In the international literature, Grossman and Krueger proposed the Environmental Kuznets Curve (EKC) hypothesis, which posits an inverted U-shaped relationship between environmental pollution and economic growth. However, analyses focusing solely on fluctuations in economic growth are insufficient to fully explain variations in carbon emissions. To further explore additional determinants of environmental change, Ehrlich introduced the IPAT model [5], which posits that population, affluence, and technology are the direct drivers of environmental impact. Based on the IPAT framework, York and Dietz developed the stochastic STIRPAT model (Stochastic Impacts by Regression on Population, Affluence, and Technology), which has been widely applied in environmental and resource studies [6].Ang argued that the LMDI (Logarithmic Mean Divisia Index) method is a preferred approach for analyzing changes in energy consumption structure [7–8]. Sun applied the LMDI method to analyze carbon emission data from 281 Chinese cities and dynamically simulated China's carbon mitigation pathways [9].

2. Data Sources

The empirical analysis focuses on annual carbon emissions and their associated driving factors within Sichuan Province. The spatial boundary corresponds to the current provincial administrative jurisdiction, without disaggregation to prefectural or county levels. All energy and economic variables are derived from official provincial statistical sources. The sample period spans 2005–2022. A

consumption-based accounting boundary is adopted, whereby carbon dioxide emissions from fossil fuel combustion are attributed to final energy consumption occurring within the provincial territory. Two dependent variables are specified: total energy-related CO_2 emissions (10,000 tons) and carbon intensity per unit of GDP. Explanatory variables are organized into two categories: socio-economic indicators and energy-related indicators. Socio-economic variables include gross domestic product (GDP, 100 million yuan), industrial value added (100 million yuan), permanent resident population (10,000 persons), and urbanization rate (%), capturing the scale of economic activity and structural transformation. Energy-related variables encompass total primary energy consumption, total electricity consumption, and the shares (%) of coal, oil, natural gas, and non-fossil energy in the primary energy mix. In addition, energy intensity is included to reflect efficiency performance. Together, these variables provide a comprehensive representation of the scale, structure, and efficiency dimensions underpinning provincial carbon emissions.

3. Carbon Emission Forecasting Results and Peak Identification

The model predictions exhibit a high degree of consistency with historical observations, particularly during the training and validation period from 2005 to 2022. The predicted curve effectively reproduces the phased evolution of carbon emissions. From 2005 to 2012, emissions experienced rapid growth, rising from 163 million tons to 289 million tons. The model accurately captures this steep upward trajectory. Between 2013 and 2016, the growth rate slowed and emissions entered a plateau phase with moderate fluctuations; the model successfully reflects this transitional pattern. From 2017 to 2022, emissions resumed an upward trend and reached a local peak in 2021 (approximately 308 million tons), followed by a slight decline in 2022. These dynamics are also well reproduced by the model.

In the projection period (after 2023), the model output continues to show short-term growth, reaching a predicted peak in 2031 before gradually entering a downward trajectory, forming a "rise-then-decline" peaking pathway. The predicted peak falls within the preset window of 2030–2035. Notably, although no exogenous constraint on the peaking year was imposed, the model's endogenous learning process yields a trajectory consistent with policy expectations, thereby validating the effectiveness of the shape-prior mechanism.

Compared with the linear extrapolation of the historical trend (green curve), which suggests that emissions would exceed 325 million tons by 2050 without a clear turning point, the Transformer model identifies a pronounced structural inflection in the early 2030s. This divergence indicates that the model's nonlinear integration of multidimensional drivers produces a more realistic and explanatory projection than a simple continuation of past growth rates.

According to the forecast results, Sichuan Province is expected to reach its carbon

emission peak in 2031, with peak emissions of approximately 336 million tons of CO_2. This represents an increase of about 10% compared to the actual 2022 level (approximately 304 million tons). The peaking process exhibits a pronounced plateau characteristic: emissions are projected to reach approximately 334 million tons in 2030, rise slightly to the peak in 2031, and remain nearly unchanged in 2032.This suggests that Sichuan's peaking trajectory is unlikely to follow a sharp turning-point pattern; rather, it may manifest as a multi-year peak interval, characterized as a "plateau-type peaking." This pattern resembles the peaking pathways observed in certain developed regions, where weakening growth momentum precedes a gradual decline phase. Thereafter, emissions are projected to decline gradually: to approximately 322 million tons by 2035, around 275 million tons by 2040, and further to about 182 million tons by 2050. This represents a reduction of roughly 45% from the peak level and a return to nearly 1.1 times the 2005 level (163 million tons of CO_2), indicating a pronounced long-term structural inflection effect.

Overall, the Transformer model not only identifies the timing and magnitude of Sichuan's emission peak but also reveals the multidimensional drivers underlying peak formation. These findings provide important policy implications: continued reductions in energy intensity, accelerated clean energy substitution, and optimization of industrial structure will be critical to ensuring timely achievement of the carbon peaking target and laying a solid foundation for the subsequent carbon neutrality pathway.

4. SHAP Analysis

4.1. Global SHAP Analysis

"Total Energy Consumption (E)" ranks highest among all features, with a SHAP value substantially exceeding those of other variables. This result is consistent with theoretical expectations, as total primary energy consumption directly determines total CO_2 emissions and constitutes the primary driving force behind emission changes. The second most influential variable is the "Carbon Emission Factor per Unit of Energy Consumption (EF)," which reflects the carbon intensity of energy use and captures the regulatory role of energy structure in emissions. A higher EF implies greater carbon emissions per unit of energy consumed, thereby positioning EF as a core determinant in the prediction results. Gross Domestic Product (GDP) ranks third. Although its influence is often transmitted indirectly through energy consumption, GDP—as a proxy for the intensity of economic activity—still demonstrates a significant positive contribution in the learning process. Economic expansion implies rising energy demand, and consequently GDP receives a relatively high global SHAP weight. Energy Intensity per Unit of GDP (EI) ranks fourth. As an indicator of energy efficiency, it reflects the energy required per unit of economic output, and its variation exerts an important moderating effect on emission

dynamics. The “Share of Non-Fossil Energy” ranks fifth, highlighting the significant emission-reduction effect associated with increasing clean energy penetration.

In contrast, the average SHAP values of “Coal Share,” “Oil Share,” and “Natural Gas Share” are comparatively low. This suggests that once the overall carbon intensity is captured by EF, the marginal predictive contribution of disaggregated fuel shares becomes limited. “Secondary Industry Share (IS2),” “Urbanization Rate,” and “Population (P)” also play secondary roles. Although these variables carry socioeconomic explanatory relevance, their influence on emissions is typically indirect, which explains their lower importance in the model attribution.

4.2. Robustness Analysis

To assess the robustness of the SHAP feature importance ranking, subsample partitioning and background distribution sensitivity tests were conducted. First, SHAP values were calculated separately for the historical observation period (2005–2022) and the projection period (2023–2050). In both subsamples, total energy consumption and EF consistently remained the two most influential features. In scenario-based projection samples, the importance of the non-fossil energy share increased, indicating that future emission dynamics will depend more strongly on structural transformation. Bootstrap resampling yielded highly consistent rankings, further supporting the robustness of the SHAP results.Second, alternative baseline distributions were adopted in the SHAP analysis. Mean SHAP values were recalculated using the 2005 level and the 2050 target scenario as reference baselines, respectively. Although the absolute contribution values of individual features changed slightly under different baselines, the overall ranking remained stable and the positions of key variables did not shift materially. This further demonstrates the consistency and credibility of the model’s global interpretability outcomes.

In summary, the SHAP-based global attribution provides stable and theoretically coherent explanations of the model’s input contributions. The core drivers of carbon emission forecasts are concentrated in three dimensions: total energy consumption, energy carbon intensity, and economic scale. By integrating SHAP distribution visualization with quantitative metrics, this section demonstrates both the transparency and logical coherence of the model’s explanatory structure, thereby establishing a solid foundation for subsequent local-level interpretability analysis.

5. Conclusion

This study calculates annual energy-related carbon dioxide emissions in Sichuan Province from 2005 to 2022 based on a provincial consumption-side accounting framework. It clarifies the boundaries and conversion parameters for direct emissions from fossil fuels and indirect emissions from purchased electricity, and establishes a verifiable inventory system along with a self-consistent

intensity–structure identity relationship. At the interpretative level, SHAP is employed to conduct both global and local attribution analyses of the deep learning model, further identifying key interaction mechanisms. The global importance ranking indicates that total primary energy consumption and the carbon emission factor per unit of energy consumption constitute the two principal axes driving emission changes, followed by economic scale and energy intensity. The share of non-fossil energy plays a significant role in promoting medium- and long-term emission decline, whereas variables such as industrial structure, urbanization, and population exert more indirect and auxiliary effects.

Based on these findings, the policy implications can be summarized in three aspects. First, demand-side scale management and supply-side structural transformation must be advanced in coordination. When total energy consumption remains at a high level, increasing the proportion of non-fossil energy and clean electricity can deliver greater absolute emission reduction density. As energy demand growth stabilizes or enters a plateau phase, structural optimization continues to yield substantial marginal benefits but should be complemented by end-use electrification, energy storage deployment, and improvements in system flexibility. Second, sustained reductions in energy intensity per unit of GDP are essential for achieving growth–emissions decoupling. In key sectors, process upgrades, equipment renewal, and digital energy-saving measures should be incorporated into routine governance in a measurable and verifiable manner. Third, natural gas has practical value in suppressing high-carbon fuels during the transition period; however, the long-term strategy must anchor on large-scale substitution by non-fossil energy. Enhancing the availability and affordability of clean energy is crucial to avoid path dependence arising from intra-fossil adjustments. Overall, for Sichuan Province to cross the peak plateau in a timely and stable manner, policy design should rest on efficiency as the foundation, structure as the lever, and total volume as the boundary, thereby strengthening the executability and auditability of the peaking pathway.

References

[1] Guo, C. X. (2010). Decomposition of China's carbon emissions: Based on the LMDI decomposition technique. *China Population, Resources and Environment*, 20(12), 4–9.

[2] Guo, Y. G., Wang, D. D., & Lin, F. C. (2010). A study on the carbon emission footprint of energy utilization in Shanghai. *China Population, Resources and Environment*, 20(02), 103–108.

[3] Zhang, L. F. (2006). *Research on China's energy supply and demand forecasting model and development strategies* (Doctoral dissertation). Capital University of Economics and Business, Beijing.

[4] Xu, G. Y. (2010). *Research on the relationship among China's energy consumption, carbon emissions, and economic growth* (Doctoral dissertation). Huazhong University of Science and Technology, Wuhan.

[5] Li, Y. T., & Li, F. (2009). An analysis of the Environmental Kuznets Curve between economic growth and environmental protection. *Economic Theory and Business

Management*, (2), 35–39.

[6] York, R., E.A. Rosa and T. Dietz, STIRPAT, IPAT and ImPACT: analytic tools for unpacking the driving forces of environmental impacts. Ecological Economics, 2003. 46: p. 351-365.

[7] Ang, B.W., LMDI decomposition approach: A guide for implementation. Energy Policy, 2015. 86: p. 233-238.

[8] Ang, B.W., The LMDI approach to decomposition analysis: a practical guide. Energy Policy, 2005. 33: p. 867-871.

[9] Sun, Q., H. Chen and A. Et, Can Chinese cities reach their carbon peaks on time? Scenario analysis based on machine learning and LMDI decomposition. Applied Energy, 2023. 347: p. 121427-121427.