buildings

MDPI

Article

Predicting Urban Carbon Peak Through an Extended STIRPAT Model: A Scenario Analysis of Bozhou, China

Shu-Guang Zhu [1,2,3], Chang-Qing Wang [1,2], Yong-Chang Shen [4], Zheng Guo [3], Yun Wan [1,2,*], Ying Qin [5], Li-Pin Xue [1,2], Yu-Qi Feng [1,2] and Bai Sun [1,2]

1 Engineering Research Center of Building Energy Efficiency Control and Evaluation, Ministry of Education, Anhui Jianzhu University, Hefei 230601, China; zhushuguang@ahjzu.edu.cn (S.-G.Z.)
2 School of Environment and Energy Engineering, Anhui Jianzhu University, Hefei 230022, China
3 Energy Saving Research Institute, Anhui Jianzhu University, Hefei 230601, China
4 School of Economics and Management, Anhui Jianzhu University, Hefei 230022, China
5 School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei 230601, China
* Correspondence: cloud_wan@163.com

Abstract

Cities play an important role in achieving carbon emission peaking and neutrality targets, yet identifying feasible mitigation pathways remains challenging for prefecture-level cities with limited historical data. This study develops a context-specific extended STIRPAT framework to examine the relationships between urban carbon emissions and nine socioeconomic and sector-related indicators in Bozhou, China. Based on annual data from 2010 to 2022, grey relational analysis was used to support indicator selection, and ridge regression was applied to address severe multicollinearity. Ten scenarios, comprising one baseline and nine alternative development pathways, were constructed to assess carbon emission trajectories during 2023–2035. Population exhibited the largest standardized ridge regression coefficient among the selected explanatory variables. Scenario analysis indicates that carbon emissions under the baseline scenario (S1) reach a projected maximum of 16.78 Mt CO_2 in 2030. The energy consumption adjustment scenario (S6) and the coordinated development scenario (S8) reach their projected maxima in 2025, at 15.60 and 15.69 Mt CO_2, respectively. Under the integrated scenarios S9 and S10, emissions decline continuously throughout 2023–2035 without reaching a new peak during the projection period. Compared with S1, projected emissions in 2030 are 10.65%, 9.06%, 24.74%, and 37.99% lower under S6, S8, S9, and S10, respectively. A supplementary Monte Carlo analysis further illustrates the sensitivity of these projections to uncertainty in the assumed explanatory-variable growth rates. The findings should be interpreted as conditional scenario outcomes rather than precise forecasts or causal estimates of policy effectiveness.

Keywords: Bozhou city; STIRPAT model; scenario analysis; carbon peak; carbon reduction potential

Check for updates

Academic Editor: Theodosiou Theodoros

Received: 20 September 2026
Revised: 2 November 2026
Accepted: 4 November 2026
Published: 8 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Climate change, driven largely by anthropogenic greenhouse gas emissions, poses substantial challenges to environmental sustainability and socioeconomic development [1–3]. Under the Paris Agreement, countries have established nationally determined contributions (NDCs) to guide their emission reduction efforts [4]. China has committed to peaking its carbon dioxide emissions before 2030 and achieving carbon neutrality before 2060 [5,6]. Meeting these targets requires effective mitigation strategies at multiple spatial and administrative scales. Cities are particularly important in this regard, accounting for a substantial

proportion of global energy-related CO_2 emissions and serving as major centers of economic activity, infrastructure development, and energy consumption [7–11]. Understanding their emission trajectories and potential mitigation pathways is therefore essential for supporting urban low-carbon transitions.

A range of quantitative approaches has been applied to investigate carbon emission drivers and assess future emission trajectories. For example, Qiao et al. [12] employed a discrete grey model to project emissions across APEC economies, while Peng et al. [13] examined the relationship between economic development and environmental pressure using the Environmental Kuznets Curve framework in Fujian Province, China. Among the available approaches, the Stochastic Impacts by Regression on Population, Affluence, and Technology (STIRPAT) model provides a flexible framework for examining associations between environmental impacts and socioeconomic factors. Building on the IPAT identity, STIRPAT was developed by Dietz and Rosa [14] and subsequently extended by York et al. [15]. It has been widely applied in carbon emission research, including studies of emission peaking in Jiangsu Province [16] and industrial-sector emissions [17]. Subsequent research has extended the framework by incorporating additional explanatory variables, regression techniques, and alternative development scenarios to examine regional and urban carbon emission pathways [18–23].

Despite these advances, applying such approaches to individual prefecture-level cities remains challenging, particularly where historical observations are limited and relevant indicators differ in availability and definition. The selection of explanatory variables must balance theoretical relevance, sectoral representation, and data continuity, while strong correlations among socioeconomic and infrastructure-related indicators can complicate coefficient estimation and interpretation. Furthermore, scenario-based projections depend on assumptions about future development trajectories and should be interpreted as conditional assessments rather than deterministic forecasts. These considerations highlight the need for transparent, locally adapted analytical frameworks that clearly distinguish empirical associations, scenario assumptions, and their associated limitations.

The accounting boundary of urban carbon emissions also warrants consideration. Urban emission inventories may adopt territorial or consumption-based perspectives, which capture different aspects of urban activities and their environmental impacts. Wiedmann et al. [24], for example, examined three-scope greenhouse gas inventories for 79 C40 cities, illustrating the importance of considering emissions associated with activities beyond direct territorial sources. The present study uses reported city-level CO_2 emissions and does not independently construct a consumption-based inventory. Its findings are therefore interpreted within the accounting boundary of the emission dataset rather than as a comprehensive assessment of emissions embodied in urban consumption.

Against this background, Bozhou, a prefecture-level city in Anhui Province, China, was selected to examine urban carbon emission trajectories under alternative development assumptions. Using annual observations from 2010 to 2022, this study develops an extended STIRPAT model incorporating nine socioeconomic and sector-related indicators. Grey relational analysis is employed to support indicator selection, while ridge regression is used to address multicollinearity among the explanatory variables. Ten scenarios are then constructed to assess potential emission trajectories and peak years over the period 2023–2035. The study aims to provide a transparent, context-specific assessment of urban emission pathways by integrating indicator screening, regression analysis, and scenario comparison within a single analytical framework. Its contribution lies primarily in the locally adapted application of established methods, rather than the development of a new forecasting methodology.

2. Study Area and Data Sources

2.1. Study Area

Bozhou is a prefecture-level city located in northwestern Anhui Province, China (115°53′–116°49′ E, 32°51′–35°05′ N; Figure 1). The city has a warm-temperate, semi-humid monsoon climate, characterized by hot, rainy summers and relatively cold, dry winters. Agriculture, including the cultivation of traditional Chinese medicinal herbs, forms an important part of the local economy.

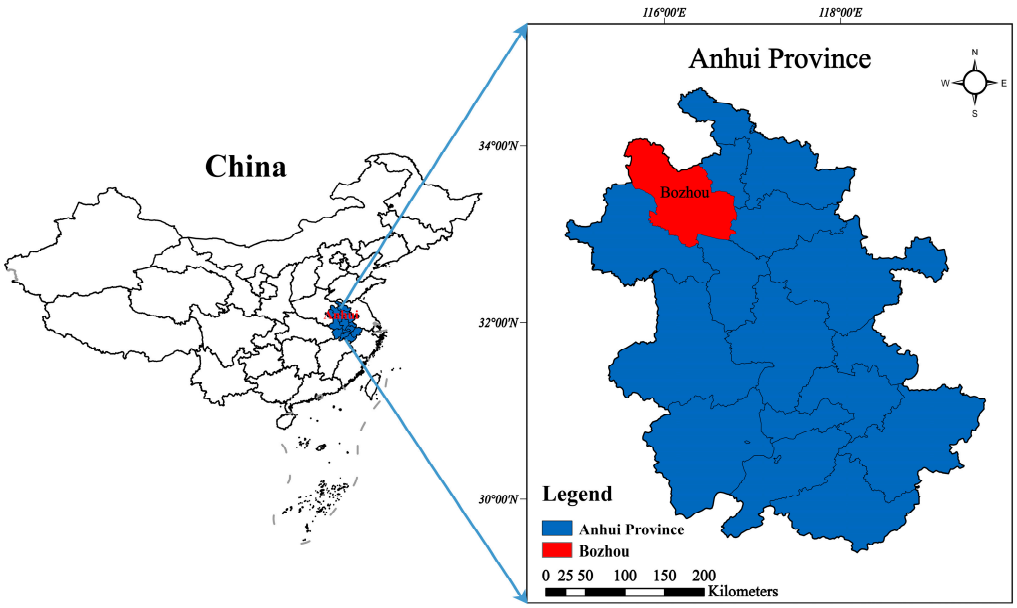


Figure 1. Location of the Study area.

Bozhou was selected as the study area because of its ongoing socioeconomic development and the need to assess potential carbon emission pathways at the prefecture-city level. In 2023, its gross domestic product (GDP) reached RMB 221.58 billion, with a GDP per capita of RMB 44,941 [25]. According to the historical emission dataset used in this study, the city's CO_2 emissions increased from 14.1926 million tonnes in 2021 to 14.8961 million tonnes in 2022 (Figure 2). These figures illustrate the challenge of reconciling continued economic development with carbon emission mitigation. The study therefore examines Bozhou's potential emission trajectories under alternative socioeconomic and sector-related development scenarios, using 2010–2022 as the historical observation period.

2.2. Data Sources

The annual carbon dioxide emissions data for Bozhou during 2010–2022 were compiled from two sources. Data for 2010–2019 were obtained from the China Cities Carbon Emissions Dataset (CEADs) [26], while data for 2020–2022 were obtained from a research institution through the research team. The annual observations are provided in the Supplementary Materials. The available documentation does not allow the underlying inventory methodology and accounting boundary to be independently verified. Data for the explanatory variables were compiled primarily from the Bozhou Statistical Yearbook (2011–2023 editions) [25], supplemented by the Anhui Statistical Yearbook and official statistical bulletins on Bozhou's economic and social development. The definitions, measurement units, and data sources of the variables included in the final model are presented in Table 1. The dependent variable represents the reported annual city-level CO_2 emissions series rather than a comprehensive consumption-based carbon footprint. The available source documentation does not

establish whether emissions associated with imported electricity or industrial processes are included. Land-use carbon removals were not independently quantified in this study, and forest coverage was treated as an explanatory indicator rather than a measured carbon offset. Accordingly, the precise inventory boundary and its implications for the interpretation of the results remain subject to the documentation of the original emissions data.

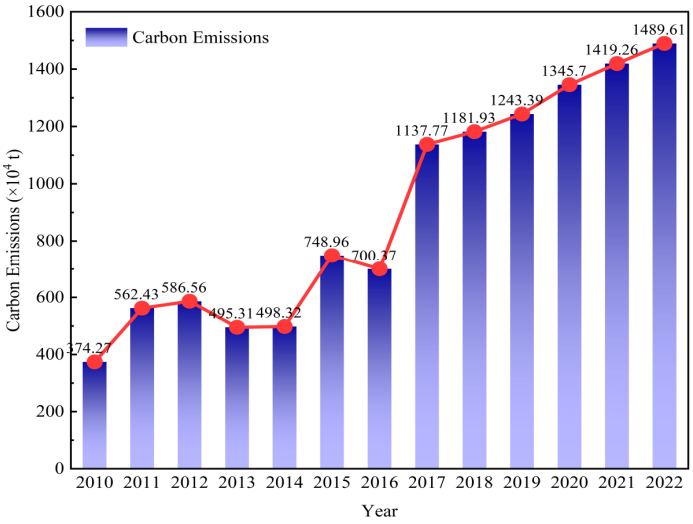


Figure 2. Carbon emissions in the study area from 2010 to 2022.

Table 1. Description of variables and data sources.

Variable	Symbol	Description	Unit	Data Source
Carbon emissions	C	Carbon dioxide emissions	10^4 tonnes	CEADs (2010–2019); research institution data (2020–2022)
Population	P	Resident population	10^4 persons	Bozhou Statistical Yearbook
Affluence	A	GDP per capita	10^4 RMB/Person	Bozhou Statistical Yearbook
Urbanization rate	U	Urbanization rate	%	Bozhou Statistical Yearbook
Transportation route length	L	Transportation route length	10^4 km	Bozhou Statistical Yearbook
Annual floor area under construction	B	Annual floor area under construction	10^4 m^2	Bozhou Statistical Yearbook
Annual wastewater treatment volume	W	Annual volume of wastewater treated	10^4 m^3	Bozhou Statistical Yearbook
Total energy consumption	E	Total energy consumption	10^4 tons of standard coal	Bozhou Statistical Yearbook
Industrial proportion	I_s	Industrial proportion	%	Bozhou Statistical Yearbook
Forest coverage rate	G	Forest coverage rate	%	Bozhou Statistical Yearbook

To identify suitable explanatory variables, an initial set of 29 candidate indicators was compiled based on the STIRPAT framework, previous studies, and the availability of

consistent city-level statistics. Population, affluence, and urbanization were considered core socioeconomic dimensions, while additional candidate indicators represented six urban development dimensions: construction, transportation, industry, municipal services, energy consumption, and carbon sinks. Candidate indicators were assessed in terms of their conceptual relevance to urban carbon emissions and the availability of comparable annual observations throughout the study period.

Grey relational analysis (GRA) was used to assess the similarity between the temporal patterns of candidate indicators and carbon emissions. The annual carbon emission series was treated as the reference sequence, and the candidate indicator series were treated as comparison sequences. Following normalization, grey relational coefficients and the corresponding grey relational grades were calculated. The distinguishing coefficient (ρ) was set to 0.5 for the primary analysis, with additional calculations at ρ = 0.3 and 0.7 to examine the sensitivity of the rankings. The complete indicator screening results are reported in Appendix A. Because several socioeconomic indicators and carbon emissions exhibited common temporal trends during 2010–2022, their grey relational grades may partly reflect shared trends rather than independent explanatory relationships. Consequently, the GRA rankings were not interpreted as evidence of causal importance, and the final indicator selection was also informed by the theoretical structure of the extended STIRPAT framework and sectoral relevance.

The final model included nine explanatory variables: population (P), GDP per capita (A), urbanization rate (U), annual floor area under construction (B), transportation route length (L), annual volume of wastewater treated (W), total energy consumption (E), industrial proportion (Is), and forest coverage rate (G). The six sector-related indicators were selected to represent different dimensions of urban development, alongside the three core socioeconomic variables. In particular, B represents the scale of building construction activity rather than the existing building stock or operational energy consumption of buildings. The selected indicators are treated as explanatory proxies; their grey relational grades do not establish causal relationships with carbon emissions.

3. Model Construction and Validation

3.1. Theoretical Foundation and Model Extension

The STIRPAT model extends the IPAT identity by introducing a stochastic term and estimable parameters, thereby allowing the relationships between environmental impacts and socioeconomic factors to be examined empirically [27,28]. Its general multiplicative form is expressed as follows:

$$C = aP^bA^cT^de \quad (1)$$

where C represents environmental pressure and carbon dioxide emissions, P represents the population, A represents affluence, and T represents technological level. Coefficients b, c, and d are to be estimated, a is the constant term, and e is the error term.

Although the conventional STIRPAT model incorporates population, affluence, and technology, the additional indicators used in this study represent a broader range of urban development characteristics. Specifically, they capture urbanization, building construction activity, transportation infrastructure, wastewater treatment, energy consumption, industrial structure, and ecological conditions. These indicators should not be interpreted exclusively as measures of technological progress. Accordingly, the extended model is specified as a multidimensional empirical adaptation of the STIRPAT framework rather than a decomposition of its technology component into independent sectoral effects:

$$C = aP^bA^cU^dB^eL^fW^gE^hG^iI_s{}^j\zeta \quad (2)$$

where U denotes the urbanization rate; B represents the annual floor area under construction; L denotes transportation route length; W represents the annual volume of wastewater treated; E denotes total energy consumption; I_s represents the industrial proportion; and G denotes the forest coverage rate. Population (P) is measured using the resident population, while affluence (A) is represented by GDP per capita. The definitions, measurement units, and data sources of all variables are provided in Table 1. The coefficients associated with the explanatory variables are estimated empirically, and ε represents the multiplicative error term.

Taking natural logarithms of both sides of Equation (2) yields an additive specification:

$$\ln C = \ln a + b \ln P + c \ln A + d \ln U + e \ln B + f \ln L + g \ln W + h \ln E + i \ln I_s + j \ln G + \ln \zeta \quad (3)$$

For estimation, the model can be written in linear regression form:

$$\ln C = \alpha + \beta_1 \ln P + \beta_2 \ln A + \beta_3 \ln U + \beta_4 \ln B + \beta_5 \ln L + \beta_6 \ln W + \beta_7 \ln E + \beta_8 \ln I_s + \beta_9 \ln G + u \quad (4)$$

where $\alpha = \ln(a)$ is the intercept, β_1–β_9 are the regression coefficients, and $u = \ln(\varepsilon)$ is the additive error term. Under the log–log specification, each coefficient represents the conditional elasticity of carbon emissions with respect to the corresponding explanatory variable, holding the other variables constant. Given the limited number of annual observations and the strong correlations among the explanatory variables, these coefficients are interpreted as exploratory statistical associations rather than causal effects.

3.2. Multicollinearity Diagnosis and Ridge Regression

Before estimating the extended STIRPAT model, multicollinearity among the nine explanatory variables was assessed using variance inflation factors (VIFs) obtained from ordinary least squares (OLS) regression. As shown in Table 2, all nine predictors had VIF values exceeding the conventional threshold of 10, indicating severe multicollinearity. Such high levels of collinearity can substantially increase the variance of OLS coefficient estimates and make individual parameter estimates sensitive to small changes in the data. Accordingly, the OLS results were used primarily to diagnose multicollinearity rather than to draw substantive conclusions about the effects of individual predictors.

Table 2. Ordinary least squares regression results and multicollinearity diagnostics.

Model	Non-Standardized Coefficients		Standard Coefficient	t	Significance	Collinearity Statistics	
	B	Standard Error	Beta			Tolerance	VIF
constant	−123.790	54.524		−2.270	0.044		
P	19.841	8.653	0.589	2.293	0.043	0.004	228.232
A	0.627	0.519	0.606	1.207	0.253	0.001	871.206
U	−1.844	1.631	−0.655	−1.131	0.282	0.001	1160.944
B	0.066	0.080	0.101	0.827	0.426	0.019	51.322
L	1.251	1.002	0.434	1.248	0.238	0.002	419.176
E	0.838	0.529	0.218	1.584	0.141	0.015	65.570
G	−0.520	1.009	−0.067	−0.515	0.617	0.017	59.193
Is	3.040	1.521	0.180	1.999	0.071	0.035	28.209
W	−0.095	0.321	−0.064	−0.296	0.773	0.006	160.104

Note: OLS denotes ordinary least squares; B denotes the unstandardized regression coefficient; Beta denotes the standardized regression coefficient; SE denotes standard error; VIF denotes variance inflation factor; and tolerance is the reciprocal of VIF. The OLS results are presented primarily to diagnose multicollinearity rather than to support inference about individual predictors. B in the coefficient column denotes the unstandardized regression coefficient, whereas B in the model denotes annual floor area under construction.

Given the severe multicollinearity identified in the OLS analysis, ridge regression was adopted to estimate the extended STIRPAT model. Ridge regression introduces an L2

penalty into the estimation procedure, thereby shrinking coefficient estimates and reducing their sensitivity to correlations among explanatory variables [29]. This improvement in numerical stability, however, comes at the cost of introducing estimation bias. Moreover, the short annual time series and potential common trends among the variables may affect the reliability of the estimated relationships. Ridge regression does not, by itself, address non-stationarity, serial dependence, or spurious regression. The resulting coefficients should therefore be interpreted as exploratory conditional associations rather than evidence of causal relationships.

The ridge regularization parameter (k) was selected using the VIF-based procedure implemented in SPSSPRO. For the full-sample model, this procedure yielded k = 0.106, which was adopted to reduce the severity of multicollinearity. Figure 3 presents the ridge trace, illustrating changes in the standardized regression coefficients as the regularization parameter increases. The coefficient trajectories became less sensitive to changes in k as the penalty increased. The selected value was therefore used for coefficient estimation and subsequent scenario analysis. Importantly, k = 0.106 was selected according to a collinearity-control criterion rather than by minimizing out-of-sample prediction error.

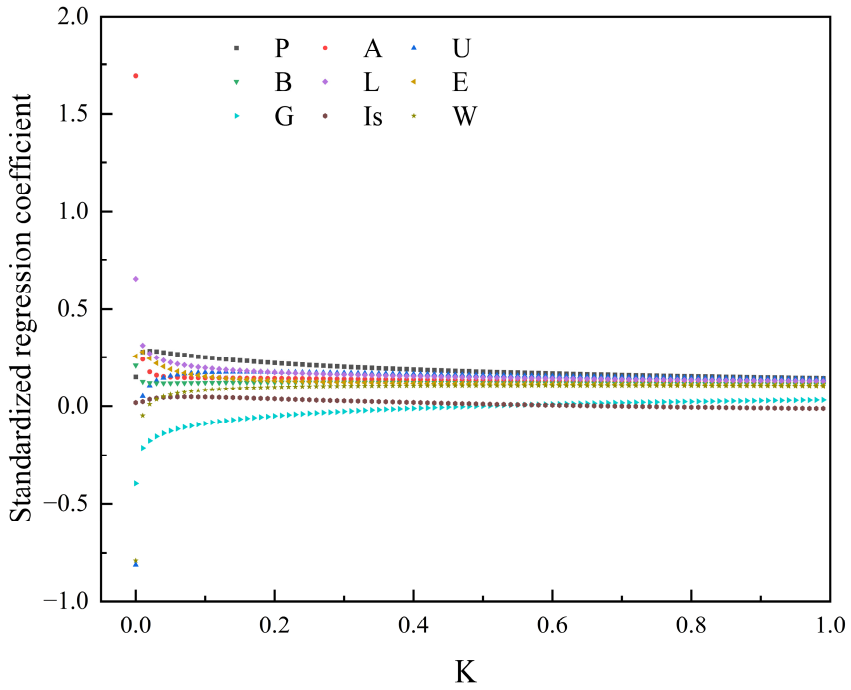


Figure 3. Ridge trace of standardized regression coefficients.

The ridge regression (k = 0.106) results are summarized in Table 3. The model yielded an adjusted R^2 of 0.951. It is reported here as a descriptive measure, whereas model reliability primarily relies on the ridge trace plot (Figure 3) and coefficient stability across k values (Table 4). The standardized regression coefficients (Beta) were used to compare the relative magnitudes and directions of associations among explanatory variables measured on different scales. Population (P) exhibited the largest absolute standardized coefficient (Beta = 0.248), followed by transportation route length (L, Beta = 0.194) and urbanization rate (U, Beta = 0.173). Forest coverage (G) showed a negative standardized coefficient (Beta = −0.084). These standardized coefficients describe relative associations within the fitted model and should not be interpreted as elasticities or independent causal effects. It should be noted that, due to the biased estimation nature of ridge regression, the standard errors and t-values reported in traditional OLS outputs are not directly interpretable for sig-

nificance testing. Therefore, our interpretation focused on the direction, relative magnitude, and stability of the coefficients across the ridge parameter path, as reflected in Figure 3.

Table 3. Ridge regression coefficient estimates at k = 0.106.

K = 0.106	Non-Standardized Coefficients		Standardization Coefficient	R^2	Adjust R^2
	B	Standard Error	Beta		
constant	−56.538	14.3	-	0.988	0.951
P	9.303	2.383	0.248		
A	0.156	0.047	0.145		
U	0.591	0.168	0.173		
B	0.084	0.06	0.122		
L	0.571	0.137	0.194		
E	0.566	0.281	0.151		
G	−1.337	1.174	−0.084		
Is	0.574	0.771	0.046		
W	0.138	0.123	0.085		

Note: B and Beta represent the unstandardized and standardized ridge regression coefficients, respectively. The ridge parameter was set to k = 0.106 based on the VIF criterion implemented in SPSSPRO. The coefficients are reported for descriptive interpretation and scenario projection. Conventional *t*-tests and *p*-values are not reported because ridge regression involves penalized estimation.

Table 4. Sensitivity of standardized ridge regression coefficients to the ridge parameter.

Predictor	k = 0.05	k = 0.10	k = 0.106	k = 0.15
Population (P)	0.251	0.249	0.248	0.244
GDP per capita (A)	0.148	0.146	0.145	0.142
Urbanization (U)	0.176	0.174	0.173	0.170
Annual floor area under construction (B)	0.124	0.123	0.122	0.119
Transport route (L)	0.196	0.195	0.194	0.191
Wastewater treatment (W)	0.087	0.086	0.085	0.083
Energy consumption (E)	0.153	0.152	0.151	0.148
Secondary industry (Is)	0.048	0.047	0.046	0.044
Forest coverage (G)	−0.082	−0.083	−0.084	−0.086

Note: All reported values are standardized ridge regression coefficients (Beta). The reference ridge parameter is k = 0.106, while k = 0.05, 0.10, and 0.15 are used for sensitivity comparison. The table describes coefficient variation across the selected ridge parameters and does not constitute a formal assessment of statistical uncertainty. The alternative ridge parameters (k = 0.05, 0.10, 0.106, and 0.15) were used to examine the local sensitivity of standardized regression coefficients. These values do not represent the complete SPSSPRO parameter-search grid. The analysis does not directly assess sensitivity to changes in the estimation period or predictor set.

To assess the sensitivity of the regression results to the choice of estimation method, partial least squares regression (PLSR) was employed as an alternative to ridge regression. Both methods were applied to the same 13 annual observations from 2010 to 2022 and the same nine explanatory variables. PLSR reduces multicollinearity by extracting latent components that account for the covariance between explanatory variables and the dependent variable. Leave-one-out cross-validation (LOOCV) was used to examine predictive performance, with root mean squared error (RMSE) and mean absolute error (MAE) serving as evaluation metrics. The number of PLS components was examined over a range of candidate values. Given the limited sample size, this comparison was intended as a supplementary sensitivity assessment rather than a definitive validation of model reliability.

Prior to ridge regression estimation, the log-transformed explanatory variables were standardized to have a mean of zero and a standard deviation of one. The standardized coefficients were used to compare the relative magnitudes of associations across predictors, while the coefficients expressed on the original log-transformed scale were reported for

interpretation within the log–log model. This distinction is important because standardized coefficients describe associations in standard deviation units, whereas unstandardized coefficients correspond to conditional elasticities in the specified model.

The ridge regression coefficient estimates obtained at k = 0.106 are presented in Table 3. Among the nine explanatory variables, population (P) had the largest positive standardized coefficient ($\beta = 0.248$), followed by transportation route length (L; $\beta = 0.194$) and urbanization rate (U; $\beta = 0.173$). Forest coverage rate (G) was negatively associated with carbon emissions ($\beta = -0.084$). The unstandardized coefficient for population was 9.303, an unusually large estimate that warrants particular caution. Given the limited sample size, severe multicollinearity, and potential common temporal trends, this coefficient should not be interpreted as a reliable causal elasticity or as evidence that population is the dominant underlying determinant of emissions. The standardized coefficients are used here to describe relative associations within the fitted model, not to establish causal importance.

The ridge regression model yielded an in-sample coefficient of determination (R^2) of 0.988, indicating a close fit to the historical observations. Here, R^2 refers to the conventional proportion of variation explained by the fitted values and is calculated without including the regularization penalty in the residual sum of squares. The software also reported an adjusted R^2 of 0.951; however, its interpretation requires caution because the effective degrees of freedom in penalized regression differ from those in ordinary least squares. Given the 13 annual observations and nine explanatory variables, these statistics should be regarded as descriptive measures of historical fit rather than evidence of reliable out-of-sample predictive performance.

To assess the sensitivity of the coefficient estimates to the choice of regularization parameter, the model was re-estimated at k = 0.05, 0.10, and 0.15, in addition to the reference value of k = 0.106. As shown in Table 4, the standardized coefficients varied modestly across the examined values, and the ranking of the three largest positive coefficients—population, transportation route length, and urbanization rate—remained unchanged. These results suggest that the relative coefficient ranking was not highly sensitive to the choice of k within the selected range. Nevertheless, this analysis does not establish robustness to alternative predictor sets, estimation methods, or temporal dependence.

Given the trending nature of the annual observations from 2010 to 2022, additional exploratory diagnostics were considered to assess potential non-stationarity and serial correlation. The augmented Dickey–Fuller (ADF) test was used to examine the stationarity of the logarithmically transformed variables, while the Durbin–Watson statistic and a lag-one Ljung–Box test were used to examine residual serial correlation. Owing to the small sample size, these diagnostics were interpreted cautiously and were not treated as definitive tests of long-run equilibrium relationships.

Based on the ridge regression estimates reported in Table 3, the fitted log–log model is expressed as follows:

$$\begin{aligned} \ln C = {} & -56.538 + 9.303 \ln P + 0.156 \ln A + 0.591 \ln U + 0.084 \ln B + 0.571 \ln L \\ & + 0.138 \ln W + 0.566 \ln E + 0.574 \ln I_s - 1.337 \ln G \end{aligned} \quad (5)$$

In Equation (5), the unstandardized coefficients describe conditional elasticities within the fitted log–log specification. However, their magnitudes should be interpreted cautiously because ridge regularization introduces estimation bias and the available time series provides limited information for estimating nine correlated predictors. Accordingly, the fitted equation is used to generate conditional emission trajectories under the specified scenarios, rather than to establish causal relationships between individual predictors and carbon emissions.

3.3. Model Fitting and Robustness Assessment

The historical fitting performance of the extended STIRPAT model was evaluated by comparing the observed carbon emissions with the fitted values obtained from ridge regression using the complete 2010–2022 dataset. As illustrated in Figure 4, the fitted values generally followed the observed emission trajectory, suggesting that the model captured the overall variation in carbon emissions during the estimation period. However, this comparison represents an assessment of in-sample fit rather than independent predictive validation.

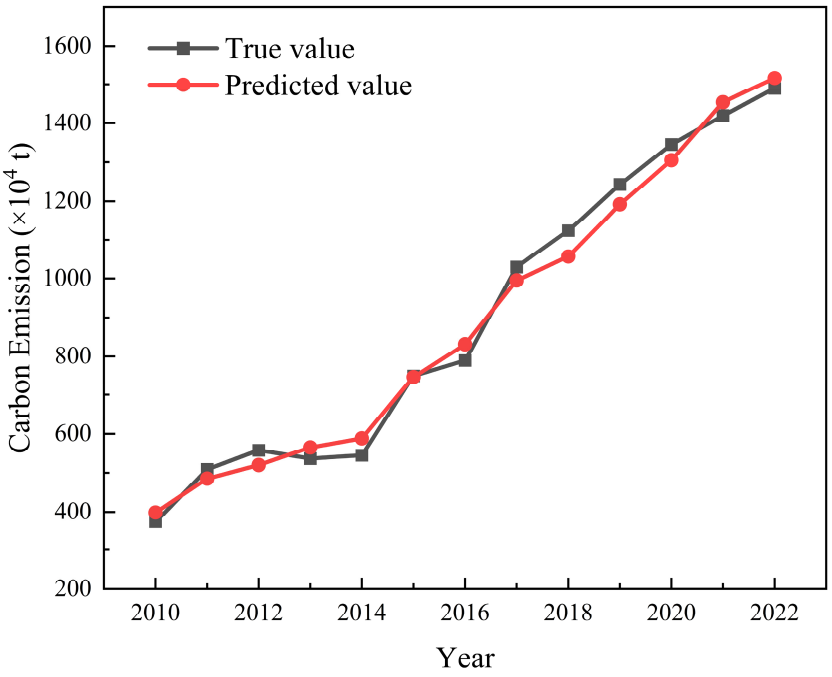


Figure 4. Observed and in-sample fitted CO_2 emissions in Bozhou City, 2010–2022.

To examine the fitting accuracy in greater detail, the observed and fitted emissions for the final three years of the study period (2020–2022) were compared, as presented in Table 5. The absolute percentage errors were 3.01%, 2.45%, and 1.78%, respectively, resulting in a mean absolute percentage error (MAPE) of 2.41%. These relatively small errors indicate a close fit to the selected historical observations. Nevertheless, because these observations were included in model estimation, the reported MAPE should not be interpreted as a measure of out-of-sample predictive accuracy.

Table 5. In-sample fitting results for 2020–2022.

Year	Actual (10^4 t)	Predicted (10^4 t)	APE (%)
2020	1345.70	1305.25	3.01
2021	1419.26	1454.02	2.45
2022	1489.61	1516.14	1.78
MAPE			2.41

Note: Fitted values were obtained from the ridge regression model estimated using the full 2010–2022 dataset. APE denotes absolute percentage error, and MAPE denotes mean absolute percentage error. The reported MAPE describes in-sample fitting accuracy for the three selected years and does not represent out-of-sample predictive performance.

As a supplementary assessment of model robustness, partial least squares regression (PLSR) was considered as an alternative to ridge regression. PLSR constructs latent components from the explanatory variables and is suitable for regression settings involving highly correlated predictors. The comparison was based on the same 13 annual observations and

nine explanatory variables, with leave-one-out cross-validation (LOOCV) used to evaluate prediction errors across model specifications. As reported in Table 6, the three-component PLSR specification achieved the lowest cross-validation errors among the examined PLSR models, with an RMSE of 0.0764 and an MAE of 0.0684, compared with 0.0868 and 0.0777, respectively, for the ridge regression specification. Although the five-component PLSR model exhibited a higher in-sample R^2 of 0.994, its cross-validation errors were greater than those of the three-component specification. These results suggest that increasing model complexity does not necessarily improve cross-validated performance. Ridge regression was retained as the primary framework for scenario analysis to maintain consistency with the extended STIRPAT specification and to preserve the direct interpretation of its log–log coefficients as elasticities. However, the lower LOOCV errors of the three-component PLSR model indicate that ridge regression cannot be considered superior in predictive accuracy on the basis of the available evidence. Sensitivity to predictor selection was also examined by comparing the original nine-variable extended STIRPAT specification with a parsimonious five-variable model, as reported in Appendix C. This comparison complements the ridge-parameter sensitivity analysis in Table 4 by assessing the influence of model specification on the estimated results.

Table 6. Comparison of Ridge Regression and Partial Least Squares Regression Using Leave-One-Out Cross-Validation.

Model	Specification	LOOCV RMSE	LOOCV MAE
Ridge	$\alpha = 0.106$	0.0868	0.0777
PLSR	2 components	0.1114	0.0905
PLSR	3 components	0.0764	0.0684
PLSR	4 components	0.0897	0.0806
PLSR	5 components	0.0951	0.0784

Despite the relatively close historical fit and the supplementary regression comparison, the reliability of long-term projections remains subject to important limitations. The dataset comprises only 13 annual observations, while the extended STIRPAT model includes nine explanatory variables. Strong multicollinearity and potential common temporal trends may further affect coefficient stability and model generalizability. In addition, LOOCV does not preserve the chronological ordering of the observations and may therefore provide an optimistic assessment of forecasting performance when temporal dependence is present. Consequently, the model is used primarily to compare conditional emission pathways under alternative development assumptions, rather than to generate precise or independently validated long-term forecasts.

To examine the sensitivity of the regression results to variable selection, a reduced five-variable specification was constructed and compared with the original nine-variable model. The reduced model retained population (P) and GDP per capita (A) as the fundamental demographic and economic components of the STIRPAT framework, together with urbanization rate (U), transportation route length (L), and total energy consumption (E), representing major dimensions of urban development, transportation infrastructure, and energy use. The remaining four variables—annual floor area under construction (B), annual wastewater treatment volume (W), industrial proportion (Is), and forest coverage rate (G)—were excluded from the reduced specification.

Both specifications were estimated using the same 13 annual observations from 2010 to 2022. The log–log functional form and ridge regression approach were retained, while the regularization parameter was determined separately for each specification. The comparison focused on in-sample model fit, changes in standardized coefficients, and the relative ranking of the retained predictors. This analysis was intended to assess the sensitiv-

ity of the estimated associations to predictor selection rather than to identify a uniquely optimal model.

As shown in Table 7, the reduced specification achieved an in-sample R^2 of 0.982, compared with 0.988 for the full model, indicating that the reduction in the number of predictors was accompanied by only a modest decrease in historical fitting performance. Population retained the largest positive standardized coefficient in both specifications, with its coefficient changing from 0.248 to 0.255. The standardized coefficient of transportation route length decreased from 0.194 to 0.183, while that of urbanization increased from 0.173 to 0.223. Consequently, urbanization ranked second in the reduced specification, whereas transportation route length ranked second in the full specification.

Table 7. Sensitivity of ridge regression results to variable selection.

Variable/Statistic	Full Model (9 Predictors)	Reduced Model (5 Predictors)
Population (P)	0.248	0.255
GDP per capita (A)	0.145	0.172
Urbanization rate (U)	0.173	0.223
Transportation route length (L)	0.194	0.183
Total energy consumption (E)	0.151	0.175
Number of predictors	9	5
Ridge parameter (k)	0.106	0.110
R^2	0.988	0.982
Adjusted R^2 (software-reported)	0.951	0.969

Note: The coefficient values are standardized ridge regression coefficients (Beta). The reported adjusted R^2 values are software outputs and should be interpreted cautiously for penalized regression. Both specifications were estimated using 13 annual observations.

The standardized coefficients of GDP per capita and total energy consumption also increased moderately after variable reduction. These differences indicate that, although the ranking of population remained unchanged, the relative magnitudes and rankings of other predictors were sensitive to the selected explanatory variables. Moreover, the unstandardized population coefficient remained relatively large in the reduced specification (9.518), compared with 9.303 in the full model. This result does not resolve the concerns regarding the interpretation of population elasticity, particularly in the presence of a short time series and strongly correlated predictors.

Overall, the reduced-variable comparison provides supplementary evidence that some descriptive features of the fitted model were maintained after removing four predictors. However, it does not establish that the estimated coefficients are causally interpretable or that the full model is robust to all alternative specifications. Furthermore, the similar in-sample fitting statistics do not demonstrate equivalent out-of-sample predictive performance.

To assess the influence of temporal trends, the nine-variable ridge regression model was compared with a simple linear time-trend model, specified as $\ln(C_t) = \alpha + \gamma t + \varepsilon_t$, where $t = \text{Year} - 2010$. Both models were trained using observations from 2010 to 2019 and evaluated on the subsequent three years (2020–2022). For the ridge regression model, standardization and regularization-parameter selection were performed using the training data only.

4. Scenario Analysis and Projection

4.1. Parameter Setting Justification

The future trajectories of the nine explanatory variables were specified using three types of information: historical observations for Bozhou, relevant regional and municipal policy

objectives, and explicit assumptions concerning alternative development pathways [30,31]. Historical trends were used to establish empirical reference points, while policy objectives were considered where their definitions and measurement units were compatible with the corresponding model variables [32–36]. For each explanatory variable, a medium (M) pathway was established as the reference assumption, with low (L) and high (H) pathways representing alternative rates of change. The resulting parameters were not intended to reproduce official forecasts; rather, they were designed to provide a consistent basis for comparing conditional carbon emission trajectories. Table 8 summarizes the principal considerations underlying the parameter settings, and the numerical adjustment rules are provided in Appendix B.

Table 8. Data sources and justification for scenario parameters.

Factor	Basis for the Medium Pathway	Low-Pathway Assumption	High-Pathway Assumption
P	Historical population trends, including the decline observed during 2020–2022	Faster population decline	Slower decline or population stabilization
A	Historical per capita economic growth and regional economic development objectives	Slower per capita economic growth	Faster per capita economic growth
U	Historical urbanization trends and the 2022 urbanization rate of 44.1%	Slower urbanization	Faster urbanization
B	Historical changes in annual construction activity	Lower growth in construction activity	Higher growth in construction activity
L	Historical changes in transportation infrastructure	Slower infrastructure expansion	Faster infrastructure expansion
W	Historical wastewater treatment volumes and assumptions regarding future service demand	Lower growth or decline in treatment volume	Higher growth in treatment volume
E	Historical energy demand and relevant energy-related policy objectives	Lower growth in total energy consumption	Higher growth in total energy consumption
I_s	Historical changes in the industrial structure	Lower industrial proportion	Higher industrial proportion
G	Historical forest coverage and relevant ecological development objectives	Slower increase in forest coverage	Faster increase in forest coverage

Note: The low, medium, and high pathways represent alternative assumptions concerning the future trajectories of the explanatory variables. They do not necessarily correspond to low, medium, and high carbon emissions. Historical observations provide empirical reference points, while policy objectives are used only where their statistical definitions are compatible with the model variables. The specific annual change rates are presented in Table 9.

The medium-pathway growth rates were determined by considering historical trends alongside plausible changes in future socioeconomic and environmental conditions. The historical compound annual growth rates and 2022 baseline values are reported in Table A4. These historical rates were used as reference information rather than directly extrapolated

into the projection period. The differences between historical growth rates and future assumptions reflect anticipated changes in population dynamics, economic development, construction activity, infrastructure expansion, energy demand, and ecological conditions. For each variable, the low and high pathways were specified through variable-specific adjustments to the medium-pathway rates across three periods: 2023–2025, 2026–2030, and 2031–2035. The detailed numerical settings and adjustment rules are documented in Table 9 and Appendix B. These parameters represent predefined scenario assumptions rather than statistically estimated forecasts or confidence intervals. Compared with the historical CAGR values, the medium pathways assume slower growth in GDP per capita, urbanization, construction activity, and energy consumption, reflecting alternative assumptions about future development rather than a continuation of past trends. The negative population growth assumptions reflect the more recent population decline, while the negative wastewater treatment volume growth assumptions represent a particularly strong departure from historical expansion and should be regarded as exploratory rather than empirically established forecasts. These differences highlight the conditional nature of the scenario settings and the need to assess their feasibility when interpreting projected emissions.

Table 9. Assumed annual change rates of explanatory variables under low, medium, and high pathways, 2023–2035.

Factor	Unit	2023–2025			2026–2030			2031–2035		
		L	M	H	L	M	H	L	M	H
P	%	−0.33	−0.12	0.09	−0.38	−0.17	0.04	−0.43	−0.22	−0.01
A	%	5.74	6.96	8.18	5.54	6.76	7.98	5.34	6.56	7.78
U	%	0.5	0.7	0.9	0.4	0.65	0.85	0.4	0.6	0.8
L	%	1.84	2.7	3.56	1.89	2.75	3.61	1.94	2.8	3.66
B	%	6.6	9.8	12.7	4.6	7.8	10.7	2.6	5.8	8.7
W	%	−3.56	−0.96	1.64	−3.66	−1.06	1.54	−3.76	−1.16	1.44
E	%	−1.7	0.76	3.21	−2.7	−0.24	2.21	−3.7	−1.24	1.21
Is	%	−0.1	0.1	0.3	−0.15	0.05	0.25	−0.2	0	0.2
G	%	0.10	0.20	0.30	0.15	0.25	0.35	0.20	0.30	0.40

Note: L, M, and H denote the low, medium, and high pathways, respectively. All values represent assumed annual relative rates of change, expressed in percent. For variables originally measured as percentages, including urbanization rate (U), industrial proportion (Is), and forest coverage rate (G), the scenario rates refer to relative changes in the corresponding indicator values rather than percentage-point changes. Annual values are calculated as $X(t) = X(t-1) \times [1 + r(t)/100]$, where $r(t)$ is the assumed annual change rate for the corresponding period. These rates are predefined scenario assumptions rather than confidence intervals or official forecasts.

4.2. Scenario Design and Rationale

Scenario analysis was employed to examine how alternative trajectories of socioeconomic and sector-related variables may affect future carbon emissions in Bozhou. Based on the parameter-setting principles described in Section 4.1, three development pathways were defined for each explanatory variable: low (L), medium (M), and high (H). These pathways represent alternative annual rates of change in the corresponding variables rather than predefined levels of carbon emissions or policy effectiveness. To reflect potential changes in development conditions over time, the projection period was divided into three stages: 2023–2025, 2026–2030, and 2031–2035. The assumed annual change rates for the nine explanatory variables are presented in Table 9. The scenario assumptions for B and W represent changes in annual construction activity and wastewater treatment volume, respectively. The resulting emission projections should not be interpreted as direct estimates of building operational energy savings or improvements in wastewater treatment efficiency.

The explanatory variables were differentiated according to their roles in the scenario framework. Population (P) and GDP per capita (A) were treated as socioeconomic development indicators, while urbanization rate (U) and industrial proportion (Is) represented

changes in urban and economic structure. Annual floor area under construction (B), transportation route length (L), annual wastewater treatment volume (W), and total energy consumption (E) were used to characterize construction activity, infrastructure development, municipal service demand, and energy use, respectively. Forest coverage rate (G) was included as an indicator of ecological conditions rather than a direct measure of carbon sequestration. Although some of these variables may be influenced by public policies, they are not equivalent to specific policy instruments. Consequently, differences between scenario projections reflect the combined effects of the assumed variable trajectories and the estimated model relationships, rather than independently identified causal effects of individual policies.

A supplementary Monte Carlo simulation was conducted to assess the sensitivity of the deterministic scenario projections to uncertainty in the assumed annual growth rates of the explanatory variables. The analysis used the deterministic trajectories generated by the extended STIRPAT model and the scenario assumptions presented in Tables 9 and 10 as reference pathways. For each scenario, 5000 simulations were performed.

Table 10. Definitions of the ten policy scenarios.

Scenario	Category	P	A	U	B	L	W	E	Is	G
S1	Baseline (BAU)	M	M	M	M	M	M	M	M	M
S2	Single-factor adjustment (construction)	M	M	M	L	M	M	M	M	M
S3	Single-factor adjustment (transport)	M	M	M	M	L	M	M	M	M
S4	Single-factor adjustment (municipal)	M	M	M	M	M	L	M	M	M
S5	Single-factor adjustment (industry)	M	M	M	M	M	M	M	L	M
S6	Single-factor adjustment (energy)	M	M	M	M	M	M	L	M	M
S7	Single-factor adjustment (forest coverage)	M	M	M	M	M	M	M	M	H
S8	Integrated (coordinated)	H	H	H	L	L	L	L	L	H
S9	Integrated (balanced)	M	M	M	L	L	L	L	L	H
S10	Integrated (enhanced low-carbon)	L	L	L	L	L	L	L	L	H

Note: L, M, and H denote the low, medium, and high pathways, respectively, for each explanatory variable. P = population; A = GDP per capita; U = urbanization rate; B = annual floor area under construction; L = transportation route length; W = annual wastewater treatment volume; E = total energy consumption; Is = industrial proportion; and G = forest coverage rate. The pathway designations refer to variable-specific assumptions and do not directly indicate low, medium, or high carbon emissions.

For each explanatory variable and projection stage (2023–2025, 2026–2030, and 2031–2035), the scenario-specific annual growth rate was used as the mode of a triangular distribution. The lower and upper bounds were specified by subtracting and adding half the difference between the corresponding high- and low-pathway growth rates, respectively. Growth-rate perturbations were sampled independently across variables and projection stages, while each sampled stage-specific rate remained constant within its corresponding stage.

The sampled deviations from the central growth-rate assumptions were propagated through the fitted log-linear STIRPAT model, holding the estimated regression coefficients fixed. The simulation was anchored to the deterministic scenario trajectories, so that zero perturbation reproduced the original deterministic projections. The 5th, 50th, and 95th percentiles were calculated for annual emissions and the maximum annual emissions over 2022–2035. A fixed random seed of 20261021 was used to ensure reproducibility.

The resulting uncertainty intervals are conditional on the specified growth-rate distributions, independence assumptions, and fitted regression coefficients. They do not represent comprehensive predictive confidence intervals and do not incorporate uncertainty in coefficient estimation or model specification.

4.3. Policy Scenario Combinations

Ten scenarios (S1–S10) were constructed by combining the low (L), medium (M), and high (H) development pathways defined in Section 4.2. The scenarios were organized into three categories: a baseline scenario, six single-factor adjustment scenarios, and three integrated scenarios. Each scenario represents a predefined combination of changes in socioeconomic, infrastructure-related, energy, and environmental indicators. These combinations are intended to examine differences in projected carbon emission trajectories rather than to quantify the causal effects of specific policy interventions. The detailed pathway assignments are presented in Table 10.

Baseline scenario (S1): All nine explanatory variables follow their medium pathways (M). This scenario provides a common reference against which the alternative combinations are compared. It represents the assumed medium-development trajectory rather than an official forecast or a future without any policy measures.

Single-factor adjustment scenarios (S2–S7): Each scenario changes the pathway of one explanatory variable while holding the remaining eight variables at their medium settings:

S2 (Construction Activity Adjustment): Annual floor area under construction (B) follows the low pathway (L), while all other variables remain at M.

S3 (Transportation Infrastructure Adjustment): Transportation route length (L) follows the low pathway (L), while all other variables remain at M.

S4 (Wastewater Treatment Volume Adjustment): Annual wastewater treatment volume (W) follows the low pathway (L), while all other variables remain at M. This assumption represents a lower trajectory of wastewater treatment volume and does not imply improved treatment efficiency, service coverage, or pollutant removal.

S5 (Industrial Structure Adjustment): Industrial proportion (Is) follows the low pathway (L), while all other variables remain at M.

S6 (Energy Consumption Adjustment): Total energy consumption (E) follows the low pathway (L), while all other variables remain at M.

S7 (Forest Coverage Enhancement): Forest coverage rate (G) follows the high pathway (H), while all other variables remain at M. This scenario examines changes associated with the forest coverage indicator rather than directly estimating carbon sequestration.

Integrated scenarios (S8–S10): Three scenarios combine changes in multiple explanatory variables to examine alternative development pathways:

S8 (Coordinated Development): Population (P), GDP per capita (A), and urbanization rate (U) follow the high pathways (H). Annual floor area under construction (B), transportation route length (L), wastewater treatment volume (W), total energy consumption (E), and industrial proportion (Is) follow the low pathways (L), while forest coverage rate (G) follows the high pathway (H). This scenario combines relatively rapid socioeconomic development with lower trajectories for selected activity-related indicators and increased forest coverage.

S9 (Balanced Development): Population (P), GDP per capita (A), and urbanization rate (U) follow the medium pathways (M). The five activity-related and structural indicators (B, L, W, E, and Is) follow the low pathways (L), while forest coverage rate (G) follows the high pathway (H). This scenario examines a combination of moderate socioeconomic development and lower growth in selected activity-related indicators.

S10 (Enhanced Low-Emission Pathway): Population (P), GDP per capita (A), urbanization rate (U), annual floor area under construction (B), transportation route length (L), wastewater treatment volume (W), total energy consumption (E), and industrial proportion (Is) all follow the low pathways (L), while forest coverage rate (G) follows the high pathway (H). This scenario represents the most restrictive combination of the selected socioeconomic and activity-related growth assumptions.

The integrated scenarios were designed to explore the implications of alternative combinations of development assumptions, rather than to reproduce specific policy targets. Their general rationale is consistent with the broader objectives of carbon emission peaking, energy conservation, and ecological improvement. Nevertheless, the annual change rates assigned to the variables are analytical assumptions and should not be interpreted as officially prescribed targets. Furthermore, because the model variables are statistical proxies rather than direct policy instruments, differences among the projected emission trajectories cannot be attributed exclusively to policy effectiveness.

4.4. Projection Results and Policy Implications

Based on the extended STIRPAT model and the scenario assumptions described in Sections 4.1–4.3, carbon emissions in Bozhou were projected for 2023–2035 under ten alternative development scenarios. The deterministic results indicate substantial differences in projected emission levels and the timing of the maximum emissions across scenarios. Figure 5 presents the trajectories of S1, S6, S7, S8, S9, and S10 to illustrate the contrasting development pathways.

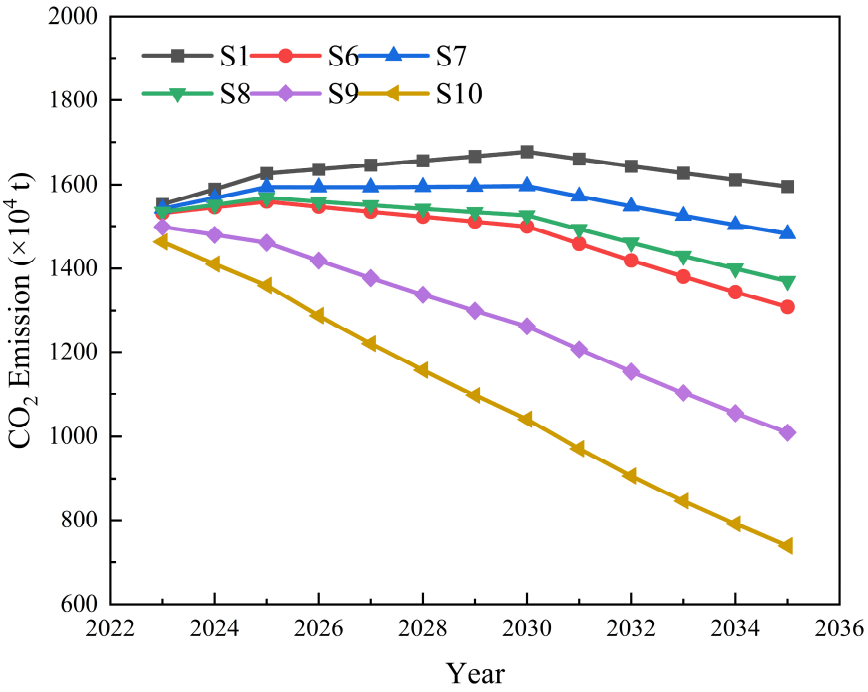


Figure 5. Deterministic CO_2 emission trajectories under six selected development scenarios in Bozhou, 2023–2035.

Under the baseline scenario (S1), emissions increase from 14.90 Mt CO_2 in 2022 to 16.78 Mt CO_2 in 2030 before declining to 15.96 Mt CO_2 in 2035. Thus, the baseline scenario reaches its projected maximum in 2030 rather than continuing to increase throughout the projection period. Among the single-factor adjustment scenarios, S6 produces a comparatively pronounced reduction. Emissions under S6 reach a projected maximum of 15.60 Mt CO_2 in 2025 and subsequently decline to 13.09 Mt CO_2 in 2035. These results indicate that the assumed trajectory of total energy consumption has an important influence on the modeled emission pathway, although the differences should not be interpreted as direct causal estimates of energy conservation policy effectiveness. The construction activity adjustment scenario (S2) further illustrates how a lower assumed growth rate of annual floor area under construction (B) is associated with changes in projected municipal CO_2

emissions, while the other explanatory variables remain at their medium pathways. This comparison reflects the statistical relationship between construction activity and citywide emissions, rather than measured building-sector emissions or the direct effects of building energy-efficiency measures.

The integrated scenarios exhibit more pronounced differences in projected emission trajectories. Under S8, emissions reach their projected maximum of 15.69 Mt CO_2 in 2025 and decline to 13.70 Mt CO_2 by 2035. In contrast, S9 and S10 show declining emissions during 2023–2035. However, S9 emissions in 2023 (14.9853 Mt CO_2) slightly exceed the historical 2022 level (14.8961 Mt CO_2), whereas S10 remains below that level. Under S9, projected emissions decrease to 12.63 Mt CO_2 in 2030 and 10.09 Mt CO_2 in 2035. S10 produces the lowest projected emissions among the ten scenarios, reaching 10.41 Mt CO_2 in 2030 and 7.39 Mt CO_2 in 2035.

Relative to S1, projected emissions in 2030 are 10.65%, 9.06%, 24.74%, and 37.99% lower under S6, S8, S9, and S10, respectively. These differences reflect the combined assumptions concerning socioeconomic development, construction activity, transportation infrastructure, wastewater treatment volume, energy consumption, industrial structure, and forest coverage. In particular, the reductions under S10 reflect both slower socioeconomic growth and lower trajectories for selected activity-related variables, rather than the isolated effects of emission mitigation measures.

To facilitate comparison across the ten development scenarios, Table 11 summarizes the deterministic CO_2 emission projections for 2030 and 2035, the corresponding percentage differences relative to the baseline scenario, and the projected maximum emissions and their timing. The table also distinguishes between the maximum within the projection period (2023–2035) and the overall maximum when the observed 2022 emission level is included.

Table 11. Summary of deterministic CO_2 emission projections under ten development scenarios.

Scenario	2030 Emissions	Change vs. S1 (%)	2035 Emissions	Change vs. S1 (%)	Peak Year (2023–2035)	Peak Emissions	2022 Remains Overall Maximum
S1	1678.09	0.00	1596.07	0.00	2030	1678.09	No
S2	1644.68	−1.99	1544.24	−3.25	2030	1644.68	No
S3	1614.87	−3.77	1499.54	−6.05	2030	1614.87	No
S4	1629.49	−2.90	1521.60	−4.67	2030	1629.49	No
S5	1621.44	−3.38	1507.79	−5.53	2030	1621.44	No
S6	1499.34	−10.65	1308.54	−18.01	2025	1559.88	No
S7	1597.19	−4.82	1481.78	−7.16	2030	1597.19	No
S8	1526.12	−9.06	1370.20	−14.15	2025	1569.29	No
S9	1262.85	−24.74	1009.20	−36.77	2023	1498.53	No
S10	1040.62	−37.99	739.43	−53.67	2023	1462.98	Yes

Note: Emissions are expressed in 10^4 t CO_2. Change vs. S1 represents the percentage difference relative to the baseline scenario (S1) in the corresponding year. The peak year and peak emissions refer to the maximum annual emissions during the projection period (2023–2035). The final column indicates whether the observed 2022 emissions (1489.61×10^4 t CO_2) remain the overall maximum during 2022–2035. All values are deterministic projections.

Although S10 achieves the lowest modeled emission levels, its interpretation requires caution. In addition to lower growth in energy consumption and selected activity-related indicators, S10 assumes lower trajectories for population, GDP per capita, and urbanization. Consequently, the difference between S10 and the baseline reflects a combination of socioeconomic and sector-related assumptions rather than the isolated effect of mitigation measures. Similarly, the lower projected emissions under S9 cannot be attributed exclusively to improvements in energy efficiency or ecological conditions. The integrated scenarios therefore provide conditional comparisons of alternative development pathways, rather

than direct estimates of policy effectiveness. However, S10's simultaneous low-growth assumptions may involve trade-offs with economic development, infrastructure provision, and public service needs, while their implementation would require coordination across multiple sectors. Therefore, despite its lowest projected emissions, S10 should not be considered an automatically preferable policy option without further assessment of socioeconomic impacts and implementation feasibility. Their practical relevance depends on the feasibility, consistency, and socioeconomic implications of the underlying assumptions.

To examine the sensitivity of the deterministic projections to uncertainty in the assumed annual change rates, a supplementary Monte Carlo analysis was conducted using 5000 iterations for each scenario. The analysis allowed the growth rates of the explanatory variables to vary within the predefined scenario-specific ranges while keeping the fitted regression coefficients fixed. Table 12 summarizes the median maximum annual emissions over 2022–2035 and the median emissions projected for 2035, together with their corresponding 5th–95th percentile intervals. The historical reference year was included when identifying maximum annual emissions so that scenarios exhibiting continuous declines from 2023 onward could be distinguished from those reaching new peaks during the projection period. The Monte Carlo analysis was designed to evaluate uncertainty in the assumed future growth rates rather than the full predictive uncertainty of the regression model. The estimated ridge regression coefficients were held fixed, and growth-rate perturbations were sampled independently across explanatory variables and projection stages. Given the potential correlations among socioeconomic and sector-related indicators, these assumptions may not fully represent coherent future development pathways. Accordingly, the reported 5th–95th percentile intervals are conditional scenario-based uncertainty ranges, rather than comprehensive predictive confidence intervals.

Table 12. Monte Carlo uncertainty analysis of projected carbon emissions under ten development scenarios.

Scenario	Median Maximum Emissions	Maximum P5–P95	Median 2035 Emissions	2035 P5–P95
S1	1691.71	1570.59–1858.51	1596.28	1404.18–1817.59
S2	1660.64	1554.45–1820.24	1542.08	1361.72–1757.97
S3	1638.53	1536.23–1789.72	1500.03	1318.99–1708.35
S4	1647.07	1547.22–1800.64	1518.61	1341.00–1725.68
S5	1642.8	1543.67–1797.96	1508.93	1317.61–1711.40
S6	1570.81	1506.92–1665.83	1311.29	1149.28–1491.97
S7	1623.56	1529.57–1765.45	1480.81	1296.43–1685.06
S8	1581.6	1510.01–1689.94	1371.92	1199.72–1562.72
S9	1498.88	1489.61–1537.82	1009.86	886.60–1145.92
S10	1489.61	1489.61–1489.61	739.68	648.68–839.30

Note: Results are based on 5000 Monte Carlo simulations for each scenario. Maximum annual emissions are calculated over 2022–2035, including the observed 2022 emissions of 1489.61×10^4 t CO_2. P5 and P95 denote the 5th and 95th percentiles of the simulated outcomes, respectively. For S10, 95.80% of simulated trajectories attain their maximum in 2022. Consequently, both the 5th and 95th percentiles of the maximum-emission distribution equal the observed 2022 value. This does not imply that all simulated trajectories have identical maximum emissions or that future emission projections are free from uncertainty. The reported intervals are conditional on the fitted regression coefficients and the assumed growth-rate distributions.

For S10, the median maximum annual emissions and the corresponding 5th–95th percentile interval are both 1489.61×10^4 t CO_2. This result arises because the calculation includes the observed emissions in 2022, which constitute the maximum over 2022–2035 in 95.80% of simulated trajectories. Therefore, the identical percentile values reflect the concentration of simulated maximum emissions at the historical reference level, rather than the absence of uncertainty in future projections. Indeed, the projected emissions for

2035 under S10 exhibit a median of 739.68 × 10^4 t CO_2 and a 5th–95th percentile interval of 648.68–839.30 × 10^4 t CO_2, demonstrating that substantial uncertainty remains in future emission levels.

As shown in Table 13, the timing of maximum annual emissions varies across the ten development scenarios. Under the baseline scenario (S1), 86.34% of simulations attain their maximum emissions during 2023–2030, while 13.66% attain their maximum during 2031–2035. Under S6 and S8, 99.78% and 99.34% of simulations, respectively, reach their maximum within 2023–2030. In contrast, 28.44% of S9 simulations and 95.80% of S10 simulations have their maximum emissions in 2022, indicating that subsequent emissions remain below the historical reference level in these simulations. These frequencies describe the location of the maximum within the specified time horizon rather than the probability of a confirmed and sustained emission peak. The results are conditional on the fitted regression coefficients and assumed scenario-specific growth-rate distributions.

Table 13. Distribution of maximum-emission years under ten development scenarios based on Monte Carlo simulations.

Scenario	Maximum in 2022 (%)	Maximum in 2023–2030 (%)	Maximum in 2031–2035 (%)
S1	0.00	86.34	13.66
S2	0.00	91.64	8.36
S3	0.00	94.88	5.12
S4	0.00	94.02	5.98
S5	0.00	94.54	5.46
S6	0.18	99.78	0.04
S7	0.00	95.64	4.36
S8	0.02	99.34	0.64
S9	28.44	71.56	0.00
S10	95.80	4.20	0.00

Note: Results are based on 5000 Monte Carlo simulations for each scenario. The maximum-emission year is identified over 2022–2035, including the historical reference year. The three categories are mutually exclusive, and their percentages sum to 100% for each scenario. The year of maximum emissions within this period does not necessarily represent a sustained turning point in the emission trajectory. All percentages are conditional on the fitted regression coefficients and assumed growth-rate distributions.

Under S1, the median maximum annual emissions are 16.92 Mt CO_2, with a 5th–95th percentile interval of 15.71–18.59 Mt CO_2. The single-factor adjustment scenarios exhibit partially overlapping uncertainty intervals, suggesting that their projected differences should not be interpreted as precisely distinguishable outcomes under the assumed uncertainty distributions.

The integrated scenarios S9 and S10 produce substantially lower projected emissions in 2035. Under S9, the median projected emissions in 2035 are 10.10 Mt CO_2, with a 5th–95th percentile interval of 8.87–11.46 Mt CO_2. Under S10, the corresponding median is 7.40 Mt CO_2, with an interval of 6.49–8.39 Mt CO_2. For S10, the observed 2022 emission level remains the maximum over 2022–2035 in 95.80% of the Monte Carlo simulations.

These results illustrate the sensitivity of future emissions to uncertainty in the assumed development pathways. However, the intervals are conditional on the prescribed growth-rate distributions and fixed regression coefficients. They should not be interpreted as comprehensive predictive confidence intervals, particularly given the short historical sample and potential instability of the estimated regression relationships.

To examine the sensitivity of scenario projections to the regression method, the three-component PLSR model was applied to the same scenario inputs used for ridge regression. The comparison focused on the projected peak years and emission reductions in 2030 relative to the baseline scenario (S1). As summarized in Table A9 in Appendix C, the results

indicate that the estimated peak timing and reduction magnitudes are sensitive to model selection. In particular, differences under S6 and S8 suggest that some scenario-specific conclusions are not robust across the two estimation methods. Therefore, the projected emission trajectories should be interpreted as conditional, model-dependent outcomes rather than independently validated forecasts.

Overall, the results indicate that alternative development assumptions may lead to substantially different carbon emission trajectories in Bozhou. Lower assumed energy consumption growth is associated with reduced projected emissions in the single-factor comparison, while integrated scenarios combining changes in multiple socioeconomic, infrastructure-related, and environmental indicators produce larger differences from the baseline. Nevertheless, the projected advantages of individual scenarios must be considered alongside the socioeconomic assumptions required to achieve them. The findings therefore provide a quantitative basis for exploring potential emission pathways and identifying influential scenario assumptions, rather than establishing definitive forecasts of carbon peaking or causal estimates of specific policy interventions.

5. Conclusions and Discussion

5.1. Discussion

This study employed an extended STIRPAT framework to investigate the statistical associations between socioeconomic and sector-related indicators and carbon emissions in Bozhou and to explore potential emission trajectories under alternative development assumptions. The findings provide a city-level perspective on the application of the STIRPAT framework to carbon emission scenario analysis. However, the estimated relationships and projected outcomes require cautious interpretation because of the limited historical observations, correlations among explanatory variables, and uncertainty associated with future development assumptions.

First, the population exhibited a relatively large positive coefficient in the estimated model. Previous studies have also identified population-related factors as relevant to carbon emission dynamics [37,38]. Nevertheless, the estimated population coefficient of 9.303 should not be interpreted as reliable evidence of an exceptionally large causal population elasticity. The model was estimated using only 13 annual observations and nine explanatory variables, and the coefficient may be sensitive to multicollinearity, common temporal trends, and model specification. In particular, population changes may be correlated with urbanization, economic activity, and infrastructure development, making it difficult to isolate an independent population effect. Moreover, because GDP per capita is calculated as total GDP divided by population, the simultaneous inclusion of population and GDP per capita may introduce overlapping demographic and economic information, further complicating the interpretation of their individual coefficients. Therefore, the result is more appropriately interpreted as an indication of the statistical sensitivity of the fitted model to population variation than as evidence that population is the dominant determinant of carbon emissions. The findings support the need to consider demographic changes alongside economic and structural factors, rather than relying on population control as a direct emission reduction strategy [39].

Second, the scenario analysis illustrates the potential importance of considering multiple development dimensions simultaneously. Compared with the baseline scenario (S1), the single-factor adjustment scenarios generally produced smaller changes in projected emissions, while several integrated scenarios generated more pronounced differences. This broad pattern is consistent with the motivation for examining multiple sectors in urban carbon emission studies [40,41]. Under the reconstructed deterministic projections, S6 reaches its maximum emission level in 2025. In contrast, S9 and S10 exhibit declining

trajectories throughout 2023–2035, although S9 slightly exceeds the 2022 historical reference level in 2023. The larger emission reductions projected under S9 and S10 demonstrate the combined influence of the assumed socioeconomic, energy-related, infrastructure, and environmental trajectories. However, these results do not establish that integrated policy interventions necessarily produce synergistic emission reductions, because the model does not independently identify policy effects or explicitly estimate interactions among policy instruments. Moreover, S10 assumes relatively low growth in population, GDP per capita, and urbanization, which may involve socioeconomic trade-offs. Its lower projected emissions should therefore not be interpreted as sufficient evidence that S10 represents the most feasible or desirable development strategy.

Third, the projected emission trajectories for Bozhou differ from those reported in previous studies of Jiangsu and Fujian. Under S9 and S10, the reconstructed projections indicate declining emissions throughout 2023–2035, whereas the cited studies reported emission peaks during 2025–2028 for Jiangsu and before 2030 for Fujian [37,42]. Such differences may be related to regional variations in demographic trends, economic structures, energy demand, and infrastructure development. Nevertheless, these potential explanations were not independently tested in the present study. Furthermore, differences in historical periods, model specifications, emission accounting boundaries, and scenario assumptions limit direct comparisons of projected peak timing across regions. Consequently, the comparison should be regarded as contextual evidence of regional heterogeneity rather than proof that Bozhou has a stronger inherent capacity for early carbon peaking.

Fourth, forest coverage rate (G) exhibited a negative estimated coefficient of −1.337, suggesting an inverse statistical association with territorial carbon emissions within the fitted model. However, forest coverage is an indirect indicator of ecological conditions rather than a direct measure of carbon sequestration. Because land-use carbon removals were not deducted from the dependent variable, the estimated coefficient cannot be interpreted as the quantity of carbon absorbed by forests or as a direct offset to fossil-fuel-related emissions. Forest coverage may also be correlated with land-use patterns, urban expansion, and other socioeconomic characteristics. Accordingly, the scenario results involving increased forest coverage should be understood as conditional model projections, while the actual contribution of forest carbon sinks would require a separate assessment based on forest area, biomass, carbon density, and relevant land-use changes [43].

Finally, the model comparison and uncertainty analysis highlight the need to distinguish statistical model performance from the reliability of long-term projections. The comparison between ridge regression and partial least squares regression provides an additional assessment of sensitivity to the estimation method, but neither approach can fully resolve the limitations associated with the short historical series. Similarly, the Monte Carlo analysis demonstrates how uncertainty in the assumed growth rates propagates into projected carbon emissions, while holding the estimated regression coefficients constant. The resulting 5th–95th percentile intervals therefore do not account for uncertainty in coefficient estimation, model specification, or future structural changes. Taken together, these considerations indicate that the projected emission trajectories should be interpreted as conditional scenario outcomes rather than precise forecasts. The principal contribution of the analysis lies in illustrating how alternative combinations of urban development assumptions may be associated with different emission pathways and in identifying issues that warrant further empirical investigation.

5.2. Conclusion and Suggestions

This study developed an extended STIRPAT framework to examine the statistical relationships between socioeconomic and sector-related indicators and carbon emissions in

Bozhou, China, using annual data from 2010 to 2022. Nine explanatory variables were incorporated to represent demographic, economic, urbanization, construction, transportation, municipal service, energy, industrial, and ecological dimensions. The ridge regression results revealed associations between carbon emissions and several explanatory variables, particularly population, transportation route length, urbanization rate, and total energy consumption. However, given the limited historical sample and the correlations among the explanatory variables, the estimated coefficients should be interpreted as exploratory statistical associations rather than stable causal effects or independently identified emission elasticities. In particular, the negative coefficient of forest coverage (G) represents a conditional statistical association and should not be interpreted as a quantitative estimate of carbon sequestration or forest carbon removals.

The scenario analysis indicates that alternative assumptions concerning socioeconomic development and sector-related activities can produce substantially different carbon emission trajectories in Bozhou. Under the baseline scenario (S1), projected emissions reach a maximum of 16.78 Mt CO_2 in 2030 and subsequently decline to 15.96 Mt CO_2 in 2035. The single-factor energy consumption adjustment scenario (S6) reaches a projected maximum of 15.60 Mt CO_2 in 2025, while the coordinated development scenario (S8) reaches a maximum of 15.69 Mt CO_2 in the same year. The integrated scenarios S9 and S10 exhibit continuously declining emissions throughout 2023–2035, without reaching a new maximum during the projection period. For the built environment, the S2 scenario highlights the potential relevance of construction activity trajectories to municipal carbon emissions, but does not quantify emissions from building construction or operation separately.

Compared with S1, projected emissions in 2030 are 10.65%, 9.06%, 24.74%, and 37.99% lower under S6, S8, S9, and S10, respectively. These results suggest that alternative combinations of socioeconomic growth, energy consumption, construction activity, infrastructure development, industrial structure, and ecological conditions may be associated with markedly different emission pathways. However, the projected reductions are conditional on the scenario assumptions and cannot be attributed exclusively to individual mitigation policies. Furthermore, the lowest projected emissions under S10 do not necessarily indicate the most preferable development pathway, as its socioeconomic trade-offs and implementation feasibility require further evaluation.

The supplementary Monte Carlo analysis indicates that uncertainty in future explanatory-variable growth rates can lead to substantial variation in projected emissions. These findings reinforce the need for cautious interpretation of deterministic scenario outcomes. Given the limited historical sample, possible coefficient instability, and simplified assumptions concerning future development, the results should be regarded as exploratory scenario projections rather than precise long-term forecasts. Future research should incorporate longer observation periods, more detailed sectoral data, chronological validation, and broader assessments of model and parameter uncertainty.

5.3. Study Limitations and Future Directions

Several limitations of this study should be acknowledged. First, the extended STIRPAT model was estimated using only 13 annual observations and nine explanatory variables, resulting in a high risk of overfitting and coefficient instability. Although ridge regression can alleviate multicollinearity, it cannot compensate for the limited sample size or fully address potential common temporal trends and non-stationarity. Furthermore, grey relational analysis identifies similarities in data patterns rather than causal relationships, and selecting indicators using the same short time series may introduce additional uncertainty. Consequently, the estimated coefficients should be interpreted as exploratory statistical associations rather than reliable causal effects.

Second, several explanatory variables serve as indirect proxies for sectoral activities. Annual floor area under construction does not represent the existing building stock, transportation route length does not directly measure transportation emissions, and wastewater treatment volume does not indicate treatment efficiency. Similarly, forest coverage rate cannot directly quantify carbon sequestration. The assessment of model performance is also constrained by the available data. Historical fitting errors do not establish out-of-sample predictive accuracy, while leave-one-out cross-validation does not fully account for the temporal structure of the observations. Moreover, the Monte Carlo analysis considers uncertainty in assumed variable growth rates but excludes uncertainty in regression coefficients and model specification.

Finally, the scenario projections depend on predefined assumptions about future socioeconomic and sectoral development and the continued applicability of historical model relationships. They do not explicitly incorporate implementation costs, technological constraints, or socioeconomic trade-offs. Therefore, the projected emission reductions should be interpreted as conditional outcomes rather than precise forecasts or direct estimates of policy effectiveness. Future research should incorporate longer time series, more detailed sectoral data, chronological model validation, and comprehensive uncertainty assessments. Integrating technical and economic feasibility evaluations would further improve the practical relevance of the scenario analysis.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/buildings16214458/s1: Bozhou_Comment18_Supplementary_Data.xlsx (historical data, model parameters, scenario assumptions, and simulation results); Bozhou_Comment18_MC_Reproduction.py (Python script for reproducing the Monte Carlo simulation procedure); Bozhou_Comment18_README.txt (description of the supplementary files and reproduction instructions); Bozhou_MC_replicated_annual.csv (annual Monte Carlo replication results); and Bozhou_MC_replicated_summary.csv (summary of the Monte Carlo replication results).

Author Contributions: S.-G.Z.: Conceptualization, Methodology, Formal Analysis, Resources, Writing—Original Draft, Writing—Review & Editing, Supervision, Project Administration, Funding Acquisition; C.-Q.W.: Conceptualization, Methodology, Formal Analysis, Data Curation, Writing—Original Draft, Writing—Review & Editing, Visualization; Y.-C.S.: Conceptualization, Methodology, Formal Analysis, Supervision; Z.G.: Investigation, Formal Analysis, Project Administration, Supervision; Y.W.: Conceptualization, Methodology, Formal Analysis, Writing—Original Draft, Writing—Review & Editing, Supervision, Funding Acquisition; Y.Q.: Conceptualization, Methodology, Formal Analysis, Supervision; L.-P.X.: Conceptualization, Methodology, Formal Analysis, Supervision; Y.-Q.F.: Conceptualization, Methodology, Formal Analysis; B.S.: Conceptualization, Methodology, Formal Analysis, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program (No. 2023YFC3807700), Scientific Research Foundation for Anhui Provincial Key Research and Development Project (Grant No. 2023t07020011), Natural Science Research Project of the Anhui Educational Committee (Grant No. 2022AH050255), Top Disciplinary Talents of the Anhui Educational Committee (Grant No. gxbjZD2022030), Anhui Province Housing and Urban Rural Construction Science and Technology Project (Grant No. 2023-YF046), Scientific Research Program of Anhui Provincial Department of Education (Grant No. 2022AH010018), National Natural Science Foundation of China (Grant No. 31800387), Scientific Research Start-up Foundation for Introduction of Talent, Anhui Jianzhu University, China (Grant No. 2020QDZ32), Technology Public Relations Project (Grant No. HYB20220267) and Technology Public Relations Project (Grant No. HYB20220266).

Data Availability Statement: The historical dataset for 2010–2022, model specifications, fitted coefficients, and annual deterministic projections for 2023–2035 are provided in the Supplementary Materials. The Monte Carlo simulation settings and reproduction code are also included.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Grey Relational Analysis for Variable Screening

Appendix A.1. Candidate Indicators

A total of 29 candidate indicators were considered for grey relational analysis (GRA), covering seven dimensions: basic socioeconomic factors, construction, transportation, municipal services, industry, energy, and ecological conditions. Population and GDP per capita were included as the fundamental demographic and affluence components of the STIRPAT framework, while the remaining indicators represented urban development and sector-related characteristics. Table A1 presents the candidate indicators by dimension. The candidate list was aligned with the indicators reported in the GRA results to ensure consistency between the screening framework and the final variable selection.

Table A1. Candidate indicators by dimension.

Dimension	Candidate Indicators	Number
Basic socioeconomic factors	Population; GDP per capita; urbanization rate; fixed asset investment	4
Construction	Annual floor area under construction; building energy intensity; labor productivity; technology equipment rate; power equipment rate; building energy consumption; gross output value of construction	7
Transportation	Transportation route length; transport energy consumption; transport energy intensity; private vehicle ownership; public bus fleet; passenger volume	6
Municipal services	Annual wastewater treatment volume; municipal solid waste treatment rate; tap water coverage rate	3
Industry	Secondary industry share; industrial solid waste utilization rate; industrial energy intensity; industrial wastewater compliance rate	4
Energy	Total energy consumption; energy intensity	2
Ecological conditions	Forest coverage rate; green coverage area; green space per capita	3
Total		29

Appendix A.2. Grey Relational Analysis Method

Grey relational analysis (GRA) was employed to evaluate the similarity between the temporal patterns of candidate indicators and carbon emissions. Carbon emissions were treated as the reference sequence, while each candidate indicator was treated as a comparison sequence. The original observations were normalized using min–max normalization to reduce the influence of differences in measurement units. Grey relational coefficients were then calculated using the absolute differences between the normalized reference and comparison sequences and a distinguishing coefficient (ρ). The grey relational grade of each indicator was obtained by averaging its annual grey relational coefficients.

The primary analysis adopted $\rho = 0.5$, while sensitivity analyses were performed using $\rho = 0.3$ and $\rho = 0.7$. Higher grey relational grades indicate greater similarity in the observed

temporal patterns but do not establish causal relationships or independent explanatory importance. Because the analysis was based on a short time series, the resulting rankings were used as supplementary information for indicator selection rather than as the sole criterion for model specification.

Appendix A.3. GRA Results

Table A2 presents the grey relational grades and rankings for all 29 candidate indicators.

Table A2. Grey relational grades and rankings.

Rank	Candidate Indicator	Grey Relational Grade (0.3)	Grey Relational Grade (0.5)	Grey Relational Grade (0.7)	Sector
1	GDP per capita	0.881	0.925	0.945	Basic
2	Annual floor area under construction	0.853	0.906	0.931	Construction
3	Annual wastewater treatment volume	0.841	0.898	0.925	Municipal
4	Transportation route length	0.836	0.895	0.923	Transportation
5	Transport energy consumption	0.821	0.884	0.914	Transportation
6	Municipal solid waste treatment rate	0.808	0.875	0.907	Municipal
7	Urbanization rate	0.803	0.872	0.905	Basic
8	Labor productivity	0.785	0.859	0.895	Construction
9	Tap water coverage rate	0.773	0.850	0.888	Municipal
10	Total energy consumption	0.763	0.843	0.883	Energy
11	Private vehicle ownership	0.741	0.827	0.870	Transportation
12	Forest coverage rate	0.736	0.823	0.867	Carbon sink
13	Green coverage area	0.705	0.799	0.848	Carbon sink
14	Secondary industry share	0.690	0.788	0.839	Industry
15	Public bus fleet	0.682	0.781	0.833	Transportation
16	Green space per capita	0.655	0.760	0.816	Carbon sink
17	Population	0.645	0.752	0.809	Basic
18	Industrial solid waste utilization rate	0.621	0.732	0.793	Industry
19	Building energy intensity	0.614	0.726	0.788	Construction
20	Gross output value of construction	0.607	0.720	0.783	Construction
21	Fixed asset investment	0.596	0.711	0.775	Basic
22	Energy intensity	0.568	0.687	0.754	Energy
23	Power equipment rate	0.535	0.657	0.728	Construction
24	Building energy consumption	0.520	0.644	0.717	Construction
25	Transport energy intensity	0.497	0.622	0.697	Transportation
26	Industrial energy intensity	0.482	0.608	0.685	Industry
27	Industrial wastewater compliance rate	0.472	0.598	0.676	Industry
28	Passenger volume	0.458	0.585	0.664	Transportation
29	Technology equipment rate	0.445	0.572	0.652	Construction

Note: The indicators are ranked according to their reported grey relational grades at $\rho = 0.5$. Values for $\rho = 0.3$ and $\rho = 0.7$ are presented for sensitivity assessment. Higher grey relational grades indicate greater similarity between the normalized indicator series and the carbon emission reference series; they do not establish causal relationships. The numerical grades are retained from the original GRA results.

Appendix A.4. Variable Selection

The final explanatory-variable set was established by combining the theoretical structure of the extended STIRPAT framework with the GRA results and the availability of continuous annual data. Population (P) and GDP per capita (A) were retained as the fundamental demographic and affluence components, while urbanization rate (U) was included to represent urban development. Six additional variables were selected to charac-

terize construction activity (B), transportation infrastructure (L), municipal wastewater treatment activity (W), energy consumption (E), industrial structure (Is), and ecological conditions (G).

The GRA rankings were used to inform the selection of representative indicators, but the final choices were not based exclusively on the highest grey relational grades. Theoretical relevance, consistency of statistical definitions, and availability of complete annual observations were also considered. Consequently, the selected variables should be regarded as measurable proxies for the corresponding development dimensions rather than direct measures of all sectoral emission processes. Table A3 summarizes the final nine explanatory variables.

Table A3. Final selected variables by sector.

Sector	Selected Variable	GRA Grade	Data Source
Basic	Population (P)	0.752	Bozhou Statistical Yearbook
Basic	GDP per capita (A)	0.925	Bozhou Statistical Yearbook
Basic	Urbanization rate (U)	0.872	Bozhou Statistical Yearbook
Construction	Annual floor area under construction (B)	0.906	Bozhou Statistical Yearbook
Transportation	Transportation route length (L)	0.895	Bozhou Statistical Yearbook
Municipal	Annual wastewater treatment volume (W)	0.898	Bozhou Statistical Yearbook
Energy	Total energy consumption (E)	0.843	Bozhou Statistical Yearbook
Industry	Secondary industry share (Is)	0.788	Bozhou Statistical Yearbook
Carbon sink	Forest coverage rate (G)	0.823	Bozhou Statistical Yearbook

Note: GDP per capita (A) is derived from GDP and population. Wastewater treatment volume (W) represents the annual quantity of wastewater treated, measured in 10^4 m^3. It is an indicator of wastewater treatment activity rather than a direct measure of treatment efficiency, treatment coverage, or the sewage treatment rate. Therefore, W is interpreted as an operational scale indicator, and its estimated coefficient should not be regarded as the causal effect of improvements in wastewater treatment performance.

As a supplementary theory-informed benchmark, the original nine-variable specification was compared with a parsimonious five-variable STIRPAT model retaining population, GDP per capita, urbanization rate, transportation route length, and total energy consumption. The reduced specification emphasizes the core demographic, economic, urbanization, infrastructure, and energy-demand dimensions while excluding additional sectoral proxies. Its regression estimates and scenario-projection comparisons are presented in Appendix C. This comparison evaluates the sensitivity of the results to a simpler, conceptually motivated predictor set rather than establishing the superiority of either specification.

Appendix B

Numerical Rules for Scenario Parameter Settings

Table A4. Historical baseline values, compound annual growth rates, and assumed medium-pathway growth rates.

Variable	2022 V2022.	Historical CAGR 2010–2022 (%)	M 2023–2025 (%)	M 2026–2030 (%)	M 2031–2035 (%)
P	496.40	0.19	−0.12	−0.17	−0.22
A	4.2258	11.35	6.96	6.76	6.56
U	44.10	3.53	0.70	0.65	0.60
B	1593.20	19.63	9.80	7.80	5.80
L	1.6864	3.38	2.70	2.75	2.80
W	8801.00	7.03	−0.96	−1.06	−1.16
E	534.15	3.65	0.76	−0.24	−1.24
Is	27.00	0.09	0.10	0.05	0.00
G	19.40	0.49	0.20	0.25	0.30

Note: Historical CAGR values were calculated from the 2010 and 2022 observations using CAGR = (X2022/X2010)^(1/12) − 1. The 2022 values provide the starting points for the scenario projections. The medium-pathway rates are predefined assumptions rather than direct extrapolations of historical CAGR values. All growth rates represent relative annual changes, including those for variables expressed as percentages.

The historical compound annual growth rates were calculated using the observations for 2010–2022 according to

$$g_i^{hist} = \left[\left(\frac{X_{i,2022}}{X_{i,2010}} \right)^{\frac{1}{12}} - 1 \right] \times 100\%$$

The historical rates were used as reference information rather than direct forecasts. For each explanatory variable and forecast period, the low and high scenario rates were calculated by applying specified downward and upward deviations to the medium rate:

$$L_{i,t} = M_{i,t} - d_{i,t}^{L}$$

$$H_{i,t} = M_{i,t} + d_{i,t}^{H}$$

The medium-pathway rates and their corresponding numerical adjustments are summarized in Table A5. These settings reproduce the annual change rates reported in Table 9 and represent predefined scenario assumptions rather than statistically estimated confidence intervals. All adjustments are expressed in percentage points of the assumed annual change rates.

Table A5. Medium scenario rates and numerical deviations used to construct low and high pathways.

Variable	Medium Rates (Three Periods)	Downward Deviation	Upward Deviation
P	−0.12/−0.17/−0.22	0.21	0.21
A	6.96/6.76/6.56	1.22	1.22
U	0.70/0.65/0.60	0.20/0.25/0.20	0.20
B	9.80/7.80/5.80	3.20	2.90
L	2.70/2.75/2.80	0.86	0.86
W	−0.96/−1.06/−1.16	2.60	2.60
E	0.76/−0.24/−1.24	2.46	2.45
Is	0.10/0.05/0.00	0.20	0.20
G	0.20/0.25/0.30	0.10	0.10

Appendix C. Reduced-Variable Ridge Regression Results

Table A6. Ridge regression coefficient estimates for the reduced five-variable STIRPAT model.

Predictor	B	SE	Beta
Intercept	−59.334	12.159	—
Population (P)	9.518	2.035	0.255
GDP per capita (A)	0.185	0.050	0.172
Urbanization rate (U)	0.762	0.167	0.223
Transportation route length (L)	0.538	0.145	0.183
Total energy consumption (E)	0.655	0.222	0.175

Note: Ridge parameter k = 0.106; R^2 = 0.982; software-reported adjusted R^2 = 0.969. SE denotes the standard error reported by SPSSPRO. Conventional significance tests are omitted because their validity for penalized estimates has not been established here.

To assess sensitivity to common temporal trends, a supplementary ridge regression was estimated using first differences of the logarithmically transformed annual observations. The analysis included 12 annual changes from 2011 to 2022 and the same nine explanatory variables as the original model. The explanatory variables were standardized, and an unpenalized intercept and a regularization parameter of $\alpha = 0.106$ were used for the exploratory model. This parameter is not assumed to be directly equivalent to the SPSSPRO ridge parameter.

The first-difference model yielded an in-sample R^2 of 0.940. As shown in Table A7, seven of the nine explanatory variables retained the same coefficient signs as those in the original model. However, the coefficients for urbanization and wastewater treatment volume changed from positive to negative, indicating sensitivity to the treatment of temporal trends.

Table A7. Comparison of regression coefficients between the original and first-difference ridge models.

Predictor	Original Model (B)	First-Difference Model
Population (P)	9.303	10.734
GDP per capita (A)	0.156	1.255
Urbanization rate (U)	0.591	−3.188
Construction area (B)	0.084	0.068
Transportation route length (L)	0.571	1.753
Total energy consumption (E)	0.566	0.501
Forest coverage (G)	−1.337	−3.351
Industrial structure (Is)	0.574	2.123
Wastewater treatment volume (W)	0.138	−0.962

Note: The original model coefficients are reported on the logarithmic level scale, whereas the first-difference coefficients are estimated using annual logarithmic changes. The comparison focuses on coefficient directions rather than direct equivalence of coefficient magnitudes. The first-difference results are exploratory because of the limited sample size.

Table A8. Sensitivity of Selected Scenario Projections to Model Specification.

Scenario	Year	Nine-Variable Model (Mt CO_2)	Five-Variable Model (Mt CO_2) *
S1	2030	16.78	17.24
S1	2035	15.96	17.42
S6	2030	14.9934	15.13
S6	2035	13.09	14.07
S9	2030	12.63	14.59
S9	2035	10.09	13.26
S10	2030	10.41	12.05
S10	2035	7.39	9.73

The comparison indicates that projected emission magnitudes are sensitive to the selection of explanatory variables, although the lower projected emissions under S10 relative to S1 are preserved in both specifications.

The comparison between ridge regression and PLSR indicates that projected peak timing and classification may depend on model specification. Differences in the estimated peak years under selected scenarios demonstrate that the peak projections are not fully robust to alternative coefficient estimates. In contrast, the declining trajectories under S9 and S10 are more consistent across the two specifications. These findings provide a complementary model-specification sensitivity assessment, although they do not quantify the full distribution of coefficient uncertainty or account for correlated future predictor pathways.

Table A9. Sensitivity of projected peak timing and 2030 emission reductions to regression method.

Scenario	Ridge Peak Year	PLSR Peak Timing	Ridge Reduction (%)	PLSR Reduction (%)
S1	2030	2030	Reference	Reference
S6	2025	2030	10.65	7.45
S8	2025	No peak by 2035	9.06	−1.16
S9	2023	No new peak; declining	24.74	17.95
S10	2023	No new peak; declining	37.99	33.60

Note: Peak year refers to the year with the highest projected annual emissions during 2023–2035, excluding the observed 2022 reference year. Emission reductions in 2030 are calculated relative to S1 under the corresponding regression model. Negative reduction values indicate higher projected emissions than the respective baseline scenario. The PLSR results are based on the alternative three-component PLSR specification.

Appendix D. Chronological Validation Results

Table A10. Comparison of Out-of-Sample Predictive Performance Between the Ridge Regression and Time-Trend Models.

Model	RMSE (10^4 t)	MAE (10^4 t)	MAPE (%)
Nine-variable ridge regression	229.21	196.48	13.54
Linear time-trend model	203.62	174.07	11.99

Note: Both models were trained using 2010–2019 observations and evaluated on 2020–2022. The ridge regression model was re-estimated using the training period only. RMSE, MAE, and MAPE were calculated on the original carbon-emission scale.

References

1. Ahima, R.S. Global warming threatens human thermoregulation and survival. *J. Clin. Investig.* **2020**, *130*, 559–561. [CrossRef]
2. O'Neill, B.C. Envisioning a future with climate change. *Nat. Clim. Change* **2023**, *13*, 874–876. [CrossRef]
3. Davidson, J.P.L.; Kemp, L. Climate catastrophe: The value of envisioning the worst-case scenarios of climate change. *WIREs Clim. Change* **2023**, *15*, e871. [CrossRef]
4. Alcaraz, O.; Sureda, B.; Turon, A.; Ramírez, C.; Gebellí, M. Equitable mitigation to achieve the 1.5 °C goal in the Mediterranean Basin. *Clim. Change* **2021**, *165*, 62. [CrossRef]
5. Cheng, J.; Tong, D.; Zhang, Q.; Liu, Y.; Lei, Y.; Yan, G.; Yan, L.; Yu, S.; Cui, R.Y.; Clarke, L.; et al. Pathways of China's PM2.5 air quality 2015–2060 in the context of carbon neutrality. *Natl. Sci. Rev.* **2021**, *8*, nwab078. [CrossRef]
6. Chen, J.; Cui, H.; Xu, Y.; Ge, Q. Long-term temperature and sea-level rise stabilization before and beyond 2100: Estimating the additional climate mitigation contribution from China's recent 2060 carbon neutrality pledge. *Environ. Res. Lett.* **2021**, *16*, 074032. [CrossRef]
7. Hyvönen, J.; Koivunen, T.; Syri, S. Possible bottlenecks in clean energy transitions: Overview and modelled effects—Case Finland. *J. Clean. Prod.* **2023**, *410*, 137317. [CrossRef]
8. Wiedmann, T.; Wood, R.; Minx, J.C.; Lenzen, M.; Guan, D.; Harris, R. A carbon footprint time series of the UK—Results from a multi-region input–output model. *Econ. Syst. Res.* **2010**, *22*, 19–42. [CrossRef]

9. Liang, X.; Li, X. Carbon emission causal discovery and multi-step forecasting for global cities. *Cities* **2024**, *148*, 104881. [CrossRef]
10. Haris, N.I.N.; Al Edrus, S.S.O.; Raof, N.A.; Wondi, M.H.; Khan, W.R.; Sien, L.S.; Ilyas, R.A.; Norrrahim, M.N.F.; Sawatdeenarunat, C. Toward low-carbon cities: A review of circular economy integration in urban waste management and its impact on carbon emissions. *Wiley Interdiscip. Rev. Energy Environ.* **2024**, *13*, e535. [CrossRef]
11. Shan, Y.; Huang, Q.; Guan, D.; Hubacek, K. China CO_2 emission accounts 2016–2017. *Sci. Data* **2020**, *7*, 54. [CrossRef]
12. Qiao, Z.; Meng, X.; Wu, L. Forecasting carbon dioxide emissions in APEC member countries by a new cumulative grey model. *Ecol. Indic.* **2021**, *125*, 107593. [CrossRef]
13. Peng, Q.; Zhan, L. Research on the Evolution Relationship between Agricultural Carbon Emissions and Economic Growth in Fujian Province Built on the EKC Model. *J. Phys. Conf. Ser.* **2021**, *1852*, 042052. [CrossRef]
14. Dietz, T.; Rosa, E.A. Effects of population and affluence on CO_2 emissions. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 175–179. [CrossRef]
15. York, R.; Rosa, E.A.; Dietz, T. STIRPAT, IPAT and ImPACT: Analytic tools for unpacking the driving forces of environmental impacts. *Ecol. Econ.* **2003**, *46*, 351–365. [CrossRef]
16. Wang, Y.; Dong, L. Research on carbon peak prediction of various prefecture-level cities in Jiangsu Province based on factors influencing carbon emissions. *Sustainability* **2024**, *16*, 7105. [CrossRef]
17. Song, X.; Xu, Y. Modeling carbon emission trends in Zhengzhou's industrial sector: A STIRPAT-based study. *Environ. Res. Commun.* **2025**, *7*, 025022. [CrossRef]
18. Wang, Y.; Fang, X.; Yin, S.; Chen, W. Low-carbon development quality of cities in China: Evaluation and obstacle analysis. *Sustain. Cities Soc.* **2021**, *64*, 102553. [CrossRef]
19. Mao, S.; Li, H.; Cai, X.; Liu, F.; Zhong, H.; Song, Y. Evaluation of the development level of low-carbon tourism in Chinese cities. *Front. Environ. Sci.* **2022**, *10*, 865946. [CrossRef]
20. Wang, H.; Li, B.; Khan, M.Q. Prediction of Shanghai Electric Power Carbon Emissions Based on Improved STIRPAT Model. *Sustainability* **2022**, *14*, 13068. [CrossRef]
21. Fang, H.; Li, H. Analysis of Influencing Factors and Prediction of the Peak Value of Industrial Carbon Emission in the Sichuan-Chongqing Region. *Sustainability* **2024**, *16*, 4532. [CrossRef]
22. Chen, Y.; Wu, Y.; Chen, N.; Kang, C.; Du, J.; Luo, C. Calculation of Energy Consumption and Carbon Emissions in the Construction Stage of Large Public Buildings and an Analysis of Influencing Factors Based on an Improved STIRPAT Model. *Buildings* **2022**, *12*, 2211. [CrossRef]
23. Li, X.; Wu, J.; Lin, C. Decarbonizing provincial construction industry under the "dual carbon" goals: Assessing reduction capacities and charting optimal pathways. *Build. Environ.* **2025**, *272*, 112639. [CrossRef]
24. Wiedmann, T.; Chen, G.; Owen, A.; Lenzen, M.; Doust, M.; Barrett, J.; Steele, K. Three-scope carbon emission inventories of global cities. *J. Ind. Ecol.* **2020**, *25*, 735–750. [CrossRef]
25. Bozhou Municipal Bureau of Statistics. Bozhou Statistical Yearbook 2024. 2024. Available online: https://tj.bozhou.gov.cn/News/show/671492.html (accessed on 3 November 2026).
26. Carbon Emission Accounts and Datasets (CEADs). China City Carbon Emissions Dataset. Available online: https://www.ceads.net/ (accessed on 3 November 2026).
27. Vélez-Henao, J.-A.; Font Vivanco, D.; Hernández-Riveros, J.-A. Technological change and the rebound effect in the STIRPAT model: A critical view. *Energy Policy* **2019**, *129*, 1372–1381. [CrossRef]
28. Gani, A. Fossil fuel energy and environmental performance in an extended STIRPAT model. *J. Clean. Prod.* **2021**, *297*, 126526. [CrossRef]
29. Choi, S.H.; Jung, H.-Y.; Kim, H. Ridge Fuzzy Regression Model. *Int. J. Fuzzy Syst.* **2019**, *21*, 2077–2090. [CrossRef]
30. Ma, Y. A Study on the Path of Carbon Emission Peaking in Northeast China in 2030. Ph.D. Thesis, Jilin University, Changchun, China, 2018.
31. Liu, D.; Xiao, B. Can China achieve its carbon emission peaking? A scenario analysis based on STIRPAT and system dynamics model. *Ecol. Indic.* **2018**, *93*, 647–657. [CrossRef]
32. Zhou, J.; Jin, B.; Du, S.; Zhang, P. Scenario Analysis of Carbon Emissions of Beijing-Tianjin-Hebei. *Energies* **2018**, *11*, 1489. [CrossRef]
33. Wang, L.; Xue, W. Analysis of carbon emission drivers and multi-scenario projection of carbon peaks in the Yellow River Basin. *Sci. Rep.* **2023**, *13*, 13684. [CrossRef]
34. Bozhou Municipal People's Government. National Carbon Peak Pilot (Bozhou) Implementation Plan. 2024. Available online: https://www.ndrc.gov.cn/fggz/hjyzy/tdftzh/202408/t20240822_1392541.html (accessed on 3 November 2026).
35. Bozhou Municipal Bureau of Ecology and Environment. 2023 Bozhou City Work Plan for the Synergistic Efficiency of Pollution Reduction and Carbon Reduction [Government Report]. 2023. Available online: https://sthjj.bozhou.gov.cn/XxgkContent/show/2449602.html (accessed on 3 November 2026).

36. Bozhou Municipal Development and Reform Commission. The 14th Five-Year Plan for National Economic and Social Development and the Long-Range Objectives Through the Year 2035 of Bozhou City Government Report. 2021. Available online: https://fgw.bozhou.gov.cn/News/show/483899.html (accessed on 3 November 2026).
37. Xie, Y.; Qin, W. Carbon peak prediction and scenario analysis in Jiangsu Province: Based on the STIRPAT model. *J. Nanjing For. Univ.* **2023**, *23*, 54–62.
38. Zhu, M.M. Prediction of Peak Industrial Carbon Emissions and Research on Carbon Reduction Strategies in Jiangsu Province. Master's Thesis, Anhui University of Technology, Ma'anshan, China, 2022.
39. Pan, K.; Liu, B.; Luo, J.; Wang, Q.; Li, J.; Tang, L.; Xia, X.; Wei, Y. Carbon peak prediction for differentiated cities from a low-carbon perspective: Key factors, scenario analysis, and low-carbon pathways. *Ecol. Indic.* **2024**, *167*, 112629. [CrossRef]
40. Zhang, S.; Wang, M.; Zhu, H.; Jiang, H.; Liu, J. Impact factors and peaking simulation of carbon emissions in the building sector in Shandong Province. *J. Build. Eng.* **2024**, *87*, 109141. [CrossRef]
41. Tang, Y.; Jiang, H. Analysis of the decoupling effect and driving factors of carbon emissions from the transportation sector in Guangdong Province. *Sci. Rep.* **2023**, *13*, 18634. [CrossRef]
42. Zhao, L.; Pan, W.; Lin, H. Can Fujian achieve carbon peak and pollutant reduction targets before 2030? Case study of 3E system in southeastern China based on system dynamics. *Sustainability* **2022**, *14*, 11364. [CrossRef]
43. Zhao, D.; Cai, J.; Xu, Y.; Liu, Y.; Yao, M. Carbon sinks in urban public green spaces under carbon neutrality: A bibliometric analysis and systematic literature review. *Urban For. Urban Green.* **2023**, *86*, 128037. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.