Spatiotemporal Evolution and Driving Analysis of Carbon Storage in Hebei Province Based on InVEST-CatBoost-MGWR Models

Zhengyan Zhang, Hui Chao*

Department of Surveying and Mapping Engineering, North China University of Science and Technology, Tangshan, Hebei 063210, China
* Corresponding author

Abstract: Land use and cover change significantly affects regional carbon cycles. From a multi-model coupling perspective, this study focuses on the mechanisms by which land use change influences carbon storage in Hebei Province. The InVEST model was employed to quantify the spatiotemporal dynamics of carbon storage from 2000 to 2025. The CatBoost machine learning algorithm was used to identify key driving factors, whose contributions were interpreted through SHAP values. Additionally, the multiscale geographically weighted regression (MGWR) model was applied to reveal the spatially non-stationary relationships between these factors and carbon storage. The results indicate that carbon storage in Hebei Province experienced a continuous decline from 2000 to 2025, decreasing from 1.887 billion tons to 1.841 billion tons,a total reduction of 46 million tons, or 2.44%. Notably, the decline accelerated markedly in the later period: the reduction was only 0.23% from 2000 to 2015 but reached 2.23% from 2015 to 2025. NDVI was the most important driving factor during 2000–2015, whereas slope emerged as the dominant factor by 2025, reflecting the increasing control of topographic conditions over carbon storage. The influence of each factor exhibited significant spatial heterogeneity. NDVI had the strongest impact in the Taihang Mountain area, while population density showed a significant negative effect in the urban agglomerations of the central Hebei Plain. These findings provide a scientific basis for low-carbon territorial spatial planning and the realization of carbon peak and carbon neutrality goals in Hebei Province.

Keywords: Carbon storage, multi-model coupling, spatial heterogeneity, spatiotemporal dynamics.

1. Introduction

Against the backdrop of intensifying global climate change, achieving the goals of carbon peaking and carbon neutrality has become particularly important. Terrestrial ecosystems reduce atmospheric CO_2 concentrations through carbon sequestration, and land use change is a primary driver of changes in terrestrial ecosystem carbon storage [1]. China has explicitly set the strategic targets of reaching carbon peak by 2030 and achieving carbon neutrality by 2060, requiring all regions to optimize ecological spatial layouts to enhance carbon sink functions. Hebei Province, as a key area in the Beijing-Tianjin-Hebei coordinated development region, has a relatively heavy industrial structure, making its carbon reduction task extremely arduous. Investigating the impacts of land use change on carbon storage can help optimize regional ecological spatial patterns and provide a scientific basis for relevant policy formulation.

In terms of carbon storage estimation, the InVEST model has been widely applied in carbon storage research [2]. Yu Zhuoxuan et al. [3] analyzed the spatiotemporal evolution of carbon storage in the Xiangjiang River Basin based on the PLUS-InVEST model; Lou Jingzhi et al.[4] used the PLUS-InVEST model to simulate and predict carbon storage changes in the karst region of Southwest China; Fan Yi et al.[5] employed the InVEST model to assess ecosystem services in the low-mountain and hilly areas of Fujian Province; and Obateru et al.[6] utilized the InVEST platform to evaluate carbon storage capacity in urban areas of Nigeria. These studies have validated the effectiveness of the InVEST model, yet most of them stop at quantifying carbon storage and provide relatively limited in-depth analysis of driving factors.

Regarding the identification of driving factors, machine learning methods offer new approaches for revealing the spatial heterogeneity of carbon storage. The CatBoost algorithm can effectively handle nonlinear relationships among multidimensional features. Dat et al. [7] used CatBoost regression to estimate soil carbon storage in mangrove ecosystems in northern Vietnam. Hu Zihan et al. [8] found that large-tree biomass was the main driver of changes in tree-layer carbon storage in old-growth forests; Haukenes et al.[9] indicated that vegetation composition, slope, and soil moisture were the most important drivers of organic layer carbon storage. These studies demonstrate that machine learning models can effectively identify key driving factors; however, most research stops at ranking factor importance, lacking quantitative interpretation of factor contributions and failing to reveal the differential effects of drivers across regions.

In terms of spatial heterogeneity analysis, the multiscale geographically weighted regression (MGWR) model provides an effective tool for characterizing the spatial differentiation of the impacts of driving factors. Hao Zixuan et al. [10] applied the InVEST-MGWR model to explore the characteristics and influencing factors of habitat quality changes in the Shanxi section of the Yellow River Basin. While these studies have validated the effectiveness of the MGWR model, most of them use MGWR as a standalone model or couple it with only a single model; they fail to integrate the three methodological components—carbon storage estimation, driving factor identification, and spatial heterogeneity characterization—into a coherent framework, thereby making it difficult to form a comprehensive understanding of the mechanisms influencing carbon storage.

In summary, existing studies predominantly focus on the application of a single model: either using the InVEST model to estimate carbon storage, or employing machine learning methods such as CatBoost to identify driving factors, or using MGWR to analyze spatial heterogeneity. Few studies have coupled these three approaches to establish a complete analytical chain from “spatiotemporal changes in carbon storage” to “identification of core driving factors” to “characterization of spatial heterogeneity.” Moreover, current research lacks in-depth discussion on the questions of “how driving factors operate” and “in which regions their influences are stronger.”

To address these research gaps, this study constructs an InVEST-CatBoost-MGWR coupled analytical framework, taking Hebei Province as the study area, and conducts the following three aspects of research: first, quantifying the spatiotemporal variation characteristics of carbon storage in Hebei Province from 2000 to 2025 based on the InVEST model; second, using the CatBoost algorithm combined with SHAP value analysis to identify core driving factors and quantify their contributions; and third, applying the MGWR model to reveal the spatial heterogeneity of the impacts of each driving factor on carbon storage. Through this coupled analytical framework, this study can provide a scientific basis for low-carbon territorial spatial planning and the achievement of the “dual carbon” goals in Hebei Province, and also offer methodological references for carbon storage research in similar regions.

2. Study Area Overview and Data Sources

2.1.Study area

Hebei Province is located in the northern part of the North China Plain, spanning 113°4'E to 119°53'E in longitude and 36°05'N to 42°37'N in latitude, with a total area of approximately 21.8×10^4 km² [11]. The terrain is higher in the northwest and lower in the southeast, comprising a variety of landform types including plateaus, mountains, hills, and plains, mainly consisting of the Bashang Plateau, the Yanshan Mountains, the Taihang Mountains, and the Hebei Plain (Fig. 1a).

The land use types in Hebei Province are diverse. The northern and western regions are predominantly covered by forestland and grassland, while the central and southeastern plains are mainly occupied by cropland (Fig. 1c). The topography exhibits pronounced relief, with elevations ranging from −56 m to 2,841 m (Fig. 1b), encompassing a complete geographical gradient from coastal lowlands to mountain peaks. In recent years, with the advancement of the Beijing-Tianjin-Hebei coordinated development strategy, the urbanization process in Hebei Province has accelerated, leading to land use changes that have impacted the carbon sink functions of regional ecosystems.

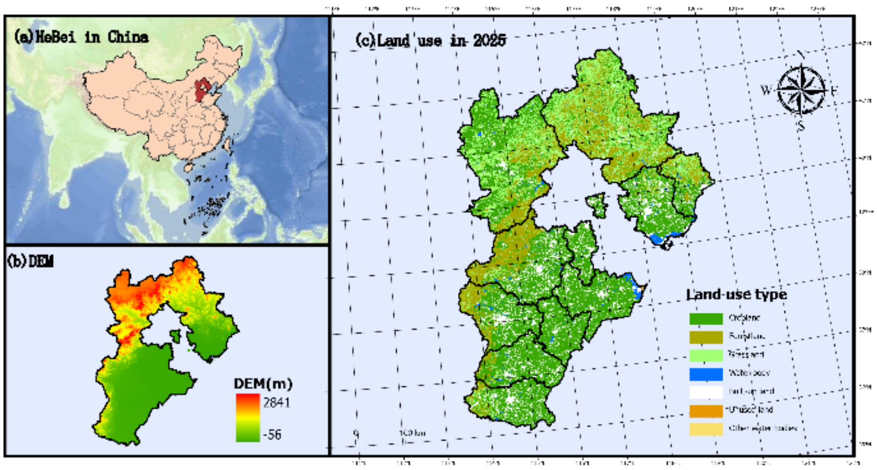

Figure 1. Study area

2.2.Data sources

The datasets employed in this study primarily consist of land use data, carbon density data, natural driving factors, and anthropogenic driving factors (Table 1).

Table 1. Data Sources and Description

Data category	Data name	Resolution	Source
Land use	LUCC data	30m	TPDC, Tianditu
Natural factors	DEM, Slope	30m	RESDC
Natural factors	Temp.,Precip.	1km	RESDC
Natural factors	NDVI	30m	RESDC
Anthropogenic factors	Population density	1km	RESDC
Anthropogenic factors	NTL	1km	RESDC
Administrative division	Hebei boundary	-	Tianditu

(NOTE: TPDC = National Tibetan Plateau Data Center; RESDC = Resource and Environmental Science and Data Center; Tianditu = Tianditu (National Geographic Information Public Service Platform of China); NTL = Nighttime light index; Temp. = Temperature; Precip. = Precipitation.)

3. Methodology

3.1.Invest model carbon storage module

The InVEST model is an ecosystem service assessment tool jointly developed by The Nature Conservancy, Stanford University, and the World Wildlife Fund [12]. The model evaluates changes in ecosystem services under different land use scenarios and presents the assessment results in a spatially explicit manner, addressing the difficulty of visualizing assessment results in previous studies. The model includes multiple modules, such as carbon storage, water yield, and habitat quality. Among them, the carbon storage module estimates ecosystem carbon storage based on land use type and carbon density data [13]. The carbon storage module comprises four carbon pools: aboveground biomass carbon (including trunks, branches, leaves, etc.), belowground biomass carbon (roots), soil carbon (soil organic matter), and dead organic matter carbon (litter). The calculation formula [14] is as follows:

$$C_i=C_{i\_above}+C_{i\_below}+C_{i\_soil}+C_{i\_dead} \tag{1}$$

$$C_{total}=\sum_{i=1}^{n} C_i*S_i \tag{2}$$

where C_i is the total carbon storage (t/ha); C_{i_above}、C_{i_below}、C_{i_soil}、C_{i_dead} are the aboveground biomass carbon density, belowground biomass carbon density, soil carbon density, and dead organic matter carbon density of land use type i, respectively (t/ha); C_{total} denotes the total carbon storage of the study area (t/ha); n is the number of land use types and S_i is the area of land use type i.

Table 2. Carbon density of different land use types in Hebei Province[15,18]

LUT	AGC	BGC	SC	DOMC	TCD
Cropland	8.3	0.8	92.9	0	102.0
Forestland	22.615	5.93	126.75	3.005	158.3
Grassland	1.175	4.835	58.62	0.355	65.0
Water body	2.29	0	17.16	0	19.45
Built-up land	7.61	4.51	42.17	0	54.29
Unused land	9.1	14.2	22.63	0	45.93

(NOTE: LUT = Land use type; AGC = Aboveground carbon; BGC = Belowground carbon; SC = Soil carbon; DOMC = Dead organic matter carbon; TCD = Total carbon density. Other water bodies are merged into Water body due to identical carbon density values.)

3.2.Catboost machine learning model

CatBoost is an ensemble learning algorithm based on gradient-boosted decision trees[19]. Through iterative construction of multiple weak learners (decision trees), the final model is represented as the weighted sum of all weak learners. It can effectively handle categorical features, reduce overfitting bias, and has significant advantages in capturing nonlinear relationships among variables [20]. Compared with traditional algorithms such as Random Forest and XGBoost, CatBoost employs a symmetric tree structure, enabling more efficient processing of high-dimensional features and exhibiting better robustness to missing values[21]. The iterative update expression of the model can be summarized as follows [22]:

$$F^n(x)=F^{n-1}(x)+\alpha h^n \tag{3}$$

Where $F^n(x)$ is the model after the n-th iteration,α is the step size updated at each round,and h^n is the objective function in each iteration. In this study, the CatBoost model was used to analyze the contribution of each driving factor listed in Table 1 to the spatial heterogeneity of carbon storage [23]. The model parameters were set as follows [23]: the number of iterations was 1,000, the learning rate was 0.03, and the tree depth was 6. The dataset was divided into a training set (70%) and a test set (30%) for model training and validation.

3.3.Shap value interpretation method

SHAP is a model interpretation method based on cooperative game theory. This framework considers the prediction result of a model for each sample as the outcome of the joint contribution of all features, and assigns a contribution value (SHAP value) to each feature. For a given sample, the sum of the SHAP values of all features, plus the average model prediction, exactly equals the prediction for that sample [24]. The SHAP values can be used to visually interpret the contributions of driving factors to the model output, and this approach has been applied in driving-factor analysis of spatiotemporal changes in carbon storage [25].

Let the model being explained [26] be $f(x)$,with N input variables denoted as $x=(x_1,x_2,\dots,x_i)$,The explanation model $g(x')$ is given as follows:

$$g(x')=\varphi_0+\sum_{i=1}^{M} \varphi_i x_i'=f(x) \tag{4}$$

Where M is the total number of features;φ_0 is the average predicted value of the model;φ_i is the contribution value (SHAP value) of the i-th feature;and x' is the simplified input, which has a mapping relationship with the input x.

3.4.Multiscale geographically weighted regression model

The MGWR model is an improvement upon the traditional GWR model, with its key feature being that it allows each explanatory variable to have its own independent spatial scale of influence. The traditional GWR model uses a single bandwidth for all variables, making it difficult to reflect the differences in spatial impacts of various factors at different scales. In contrast, the MGWR model selects optimal bandwidths for each variable individually, enabling a more accurate characterization of the multiscale spatial heterogeneity of the impacts of driving factors on carbon storage, and demonstrating better adaptability in handling complex geographical data [27]. The formula is as follows:

$$Y_i=\beta_0(u_i,v_i)+\sum_{j=1}^{p}\beta_i(u_i,v_i,b_j)X_{ij}+\varepsilon_i \quad (5)$$

Where Y_i is the carbon storage; (u_i,v_i) are the coordinates of the i-th sampling point; and $\beta_i(u_i,v_i,b_j)$ is the regression coefficient of the j-th variable under bandwidth b_j.

4. Results and Analysis

4.1. Spatiotemporal characteristics of carbon storage in hebei province

From 2000 to 2025, carbon storage in Hebei Province exhibited a continuous declining trend, decreasing from 1.887 billion tons to 1.841 billion tons, with a cumulative reduction of 0.046 billion tons (46 million tons), equivalent to a decrease of 2.44% (Table 3). From the perspective of different stages, carbon storage decreased by 0.004 billion tons (0.23%) from 2000 to 2015, and by 0.042 billion tons (2.23%) from 2015 to 2025, indicating a marked acceleration in the decline rate. Notably, from 2020 to 2025, carbon storage decreased by 0.015 billion tons (0.81%), further reflecting the intensifying loss of carbon storage in recent years.

Table 3. Total carbon storage in Hebei Province from 2000 to 2025

Year	TC (ton)	TC (10^4 ton)	TC (10^8 ton)
2000	1,887,058,560.36	188,705.86	18.87
2005	1,885,084,903.42	188,508.49	18.85
2010	1,884,787,661.47	188,478.77	18.85
2015	1,882,636,980.50	188,263.70	18.83
2020	1,855,764,417.47	185,576.44	18.56
2025	1,841,420,757.64	184,142.08	18.41

(Note: TC = Total carbon storage.)

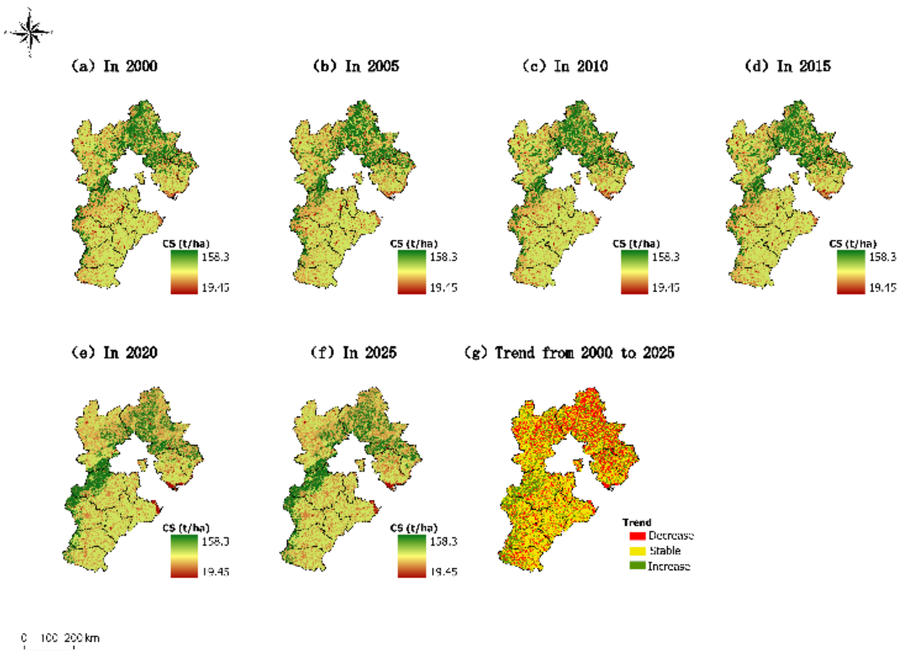

Figure 2. Spatial distribution of carbon storage in Hebei Province from 2000 to 2025.

(a) Carbon storage in 2000

(b) Carbon storage in 2005

(c) Carbon storage in 2010

(d) Carbon storage in 2015

(e) Carbon storage in 2020

(f) Carbon storage in 2025

(g) Trend of carbon storage change from 2000 to 2025

In terms of spatial distribution, carbon storage in Hebei Province exhibited a distinct pattern of "high in the northwest and low in the southeast" (Fig. 2a–f). High-value areas (>100 t/ha) were mainly concentrated in the Taihang Mountains and the Yanshan Mountains, where forestland and grassland predominated, vegetation coverage was high, and carbon density was large. Low-value areas (<50 t/ha) were primarily distributed in the urban agglomerations of the central Hebei Plain and coastal areas, where cropland and built-up land were the dominant land use types, resulting in relatively low carbon storage.

In terms of temporal trends (Fig. 2g), carbon storage in Hebei Province showed an overall declining trend from 2000 to 2025, with a marked acceleration after 2015. The most significant loss of carbon storage occurred in the urban agglomerations of the central Hebei Plain.

4.2.Identification of driving factors for spatial heterogeneity of carbon storage

4.2.1. Catboost model performance

From 2000 to 2015, the R^2value remained stable between 0.53 and 0.54, with RMSE ranging from 17.25 to 17.37, indicating good fitting performance. However, from 2020 to 2025, the R^2value declined to approximately 0.39, while the RMSE increased to over 25, suggesting that carbon storage changes have become more complex in recent years, and the explanatory capacity of the existing driving factors has been challenged in the context of rapid changes. In the future, more driving factors (e.g., land use policies, extreme climate events, etc.) should be incorporated to enhance model adaptability.

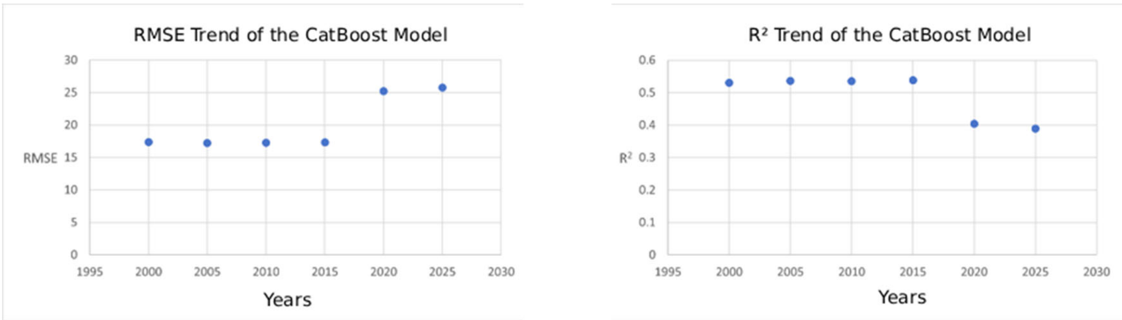

Figure 3. CatBoost Model Fitting Performance

4.2.2. Feature importance analysis

The feature importance ranking derived from the CatBoost model (Fig. 4) shows that from 2000 to 2015, NDVI was the most important driving factor affecting carbon storage in Hebei Province, with its importance score increasing from 24.75% to 25.92%. From 2020 to 2025, the importance of NDVI declined (21.46% in 2020 and 19.44% in 2025), while the importance of slope increased significantly, reaching 20.35% in 2020 (ranking second) and 22.50% in 2025 (ranking first), thus becoming the new dominant factor. Temperature and population density constituted the second tier of influencing factors, followed by DEM and precipitation, while the influence of nighttime light was relatively minor.

Figure 4. Feature importance bar chart (2000–2025).

Table 4. Feature importance of driving factors (%)

Driving factors	2000	2005	2010	2015	2020	2025
NDVI	24.75	25.17	25.45	25.92	21.46	19.44
Temperature	15.43	15.53	16.09	16.52	15.74	14.22
Population	14.26	14.65	13.77	14.04	11.94	11.92
DEM	12.65	12.56	12.78	12.64	13.43	12.53
Precipitation	12.41	12.36	12.31	12.06	8.12	9.34
Slope	11.85	11.17	11.09	10.39	20.35	22.50
NightLight	8.65	8.56	8.50	8.42	8.97	10.06

The feature importance analysis indicates that natural factors predominantly control the spatial distribution of carbon storage in Hebei Province, with vegetation condition being the most critical controlling factor. The importance of NDVI showed a continuous increase from 2000 to 2015, but declined from 2020 to 2025, while the importance of temperature remained relatively stable. The importance of slope jumped to 20.35% in 2020 and further increased to 22.50% in 2025, making it the most important factor, reflecting the significantly enhanced controlling role of topographic conditions in recent years. The importance of population density declined, suggesting that the direct impact of urbanization on carbon storage has weakened. However, carbon storage continued to decline at an accelerated rate, implying that indirect effects, such as construction land expansion and land use change, have become increasingly prominent.

4.2.3. Shap value analysis

SHAP values further quantified the marginal contributions of each driving factor to carbon storage (Fig. 5, Table 5). The SHAP value of NDVI increased from 7.18 in 2000 to 7.47 in 2015, further rose to 8.30 in 2020, and slightly declined to 8.01 in 2025, indicating that the positive driving effect of vegetation condition on carbon storage has generally strengthened over time. Notably, the SHAP value of DEM increased substantially in 2020 (7.59) and 2025 (7.13), making it the second largest contributing factor after NDVI. Similarly, the SHAP value of slope increased from 2.52 in 2000 to 5.01 in 2025, which is consistent with the feature importance results.

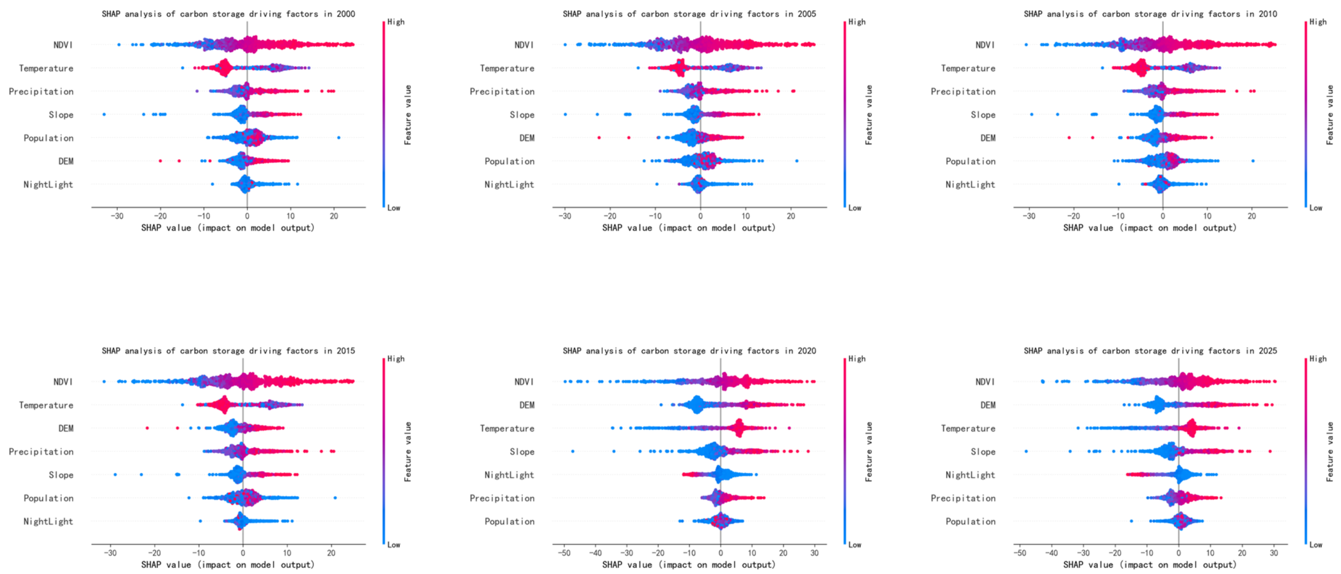

Figure 5. SHAP value distribution of driving factors by year

Table 5. SHAP values by year

Driving factors	2000	2005	2010	2015	2020	2025
NDVI	7.18	7.42	7.40	7.47	8.30	8.01
Temperature	5.70	5.31	5.34	5.28	7.26	6.06
Population	2.32	2.36	2.31	2.32	1.74	1.62
DEM	2.14	2.43	2.43	2.63	7.59	7.13
Precipitation	2.57	2.60	2.55	2.54	2.01	2.58
Slope	2.52	2.58	2.50	2.34	4.84	5.01
NightLight	1.24	1.18	1.18	1.19	2.71	3.13

The SHAP dependence plot (Fig. 5) reveals the nonlinear influence of each driving factor on carbon storage. Taking NDVI as an example, when NDVI values are low (blue points), the SHAP values are concentrated on the negative side, exerting a negative effect on carbon storage; when NDVI values are high (red points), the SHAP values are concentrated on the positive side, exerting a positive effect, indicating a clear positive correlation. Regarding the distribution of SHAP values for population density, the results show that in areas with low population density (blue points), the SHAP values are distributed across both positive and negative sides, suggesting that the direction of influence of population density on carbon storage exhibits spatial differentiation—positive in some regions and negative in others. This may be related to regional development patterns, vegetation cover, and other factors.

4.3.Spatial heterogeneity analysis of driving factors

4.3.1. Diagnostics of the mgwr model

The MGWR model results for 2025 (Table 6) show that the model goodness-of-fit $R^2$is 0.9993, and the adjusted R^2 is 0.9992, indicating that the model can explain 99.9% of the spatial variation in carbon storage in Hebei Province, and the influence of each driving factor on carbon storage is highly regular in space. Compared with the CatBoost model (R^2=0.3889), the MGWR model's explanatory power improved by approximately 61 percentage points after accounting for spatial heterogeneity, fully demonstrating the complexity of the spatial relationships of carbon storage and the necessity of considering spatial heterogeneity.

Table 1 MGWR model diagnostics

Indicator	Value
R^2	0.9993
Adjusted R^2	0.9992

4.3.2. Statistics of driving factor coefficients

In 2025, the mean coefficient of NDVI was 0.9478, ranging from 0.7570 to 1.1580, all positive values, indicating that NDVI generally exerts a positive influence on carbon storage with significant spatial variations in influence intensity. The mean coefficient of temperature was −0.1193, ranging from −0.8491 to 1.1761, showing a predominantly negative influence pattern. The mean coefficient of population density was −0.0189, ranging from −0.5840 to 0.8546, suggesting that its influence direction exhibits pronounced spatial differentiation. Notably, the slope coefficients ranged from −0.1034 to 0.9215, with a mean of 0.0892, indicating that positive effects dominated.

Table 7. MGWR coefficient statistics

Driving factors	Mean	Min	Max	SD
NDVI	0.9478	0.7570	1.1580	0.0408
Temperature	-0.1193	-0.8491	1.1761	0.1297
Population	-0.0189	-0.5840	0.8546	0.0736
Precipitation	0.1145	-0.3380	0.2891	0.0464
Slope	0.0892	-0.1034	0.9215	0.0697
DEM	0.0187	-0.2846	0.2141	0.0374
NightLight	-0.0064	-0.7232	1.0303	0.0669

(NOTE: Min = Minimum; Max= Maximum; SD= Standard deviation.)

As shown in Table 8, the optimal number of neighboring elements for NDVI in 2025 was 156, with 20,000 significant points, accounting for 100% of the total, indicating that NDVI had a significant influence across the entire province. Precipitation (87.36%) and DEM (73.23%) also had relatively high proportions of significant points, whereas the proportions for population density (8.86%) and nighttime light (17.26%) were relatively low, suggesting that the influences of anthropogenic factors were only significant in localized areas.

Table 8. MGWR model bandwidth and significance summary

Driving factor	Optimal bandwidth	Significant points	Significant proportion (%)
NDVI	156	20000	100.00
Temperature	32	4742	23.71
Population	40	1772	8.86
Precipitation	127	17472	87.36
Slope	62	8905	44.52
DEM	133	14647	73.23
NightLight	56	3451	17.26

(NOTE: Bandwidth (optimal number of neighbors))

4.3.3. Spatial heterogeneity of the influence of driving factors

The spatial distribution map of coefficients generated by the MGWR model (Fig. 6) visually displays the spatial differences in the impacts of various driving factors on carbon storage.

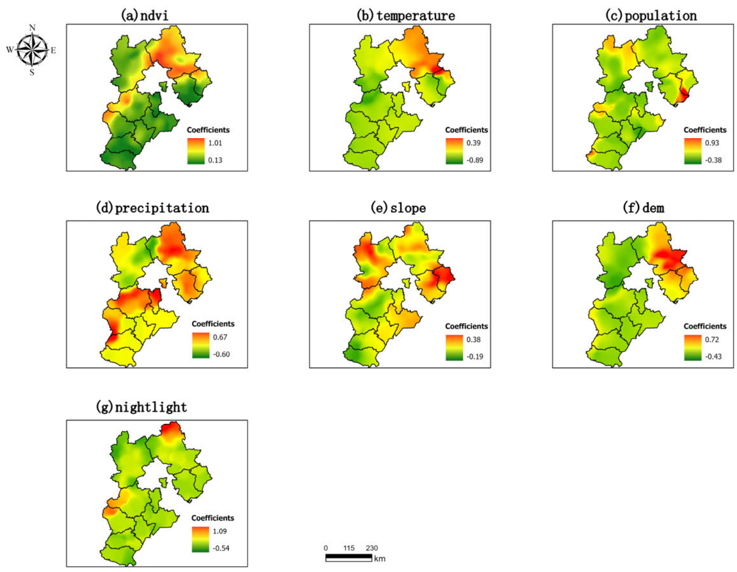

Figure 6. Spatial distribution of MGWR coefficients in 2025

(a) NDVI coefficient

(b) Precipitation coefficient

(c) Population density coefficient

(d) Temperature coefficient

(e) Slope coefficient

(f) DEM coefficient

(g) Nighttime light coefficient

The spatial distribution of NDVI coefficients (Fig. 6a) shows that the positive influence of NDVI on carbon storage is widespread across the province, but with pronounced spatial heterogeneity. High-value areas (>1.0) were mainly concentrated in the Taihang Mountains, where vegetation coverage is high and ecological conditions are favorable, resulting in a stronger positive correlation between NDVI and carbon storage. Low-value areas (0.75–0.85) were distributed in the northern Hebei Plateau and the urban agglomerations of the central Hebei Plain, possibly related to vegetation type or urbanization.

The spatial distribution of precipitation coefficients (Fig. 6b) exhibited a predominantly positive pattern, with coefficients ranging from –0.338 to 0.289. Positive influence areas were widely distributed across the central Hebei Plain and most of the Taihang Mountains, indicating that precipitation acts as a promoting factor for carbon storage. Negative influence areas were only sporadically distributed in localized parts of the northern Hebei Plateau, possibly related to poor drainage or soil water oversaturation.

The spatial distribution of population density coefficients (Fig. 6c) exhibited a pronounced negative pattern. High negative areas (< –0.3) were concentrated in the urban centers of Shijiazhuang, Tangshan, Baoding, and other cities, indicating that urban population agglomeration exerts a strong inhibitory effect on carbon storage. Positive influences were observed in sporadic areas of the northern Hebei mountainous region and the Taihang Mountains (with coefficients reaching up to 0.85), possibly due to better vegetation protection around population agglomeration areas in these regions.

The spatial distribution of temperature coefficients (Fig. 6d) showed a predominantly negative pattern. Negative influence areas were widely distributed across most of Hebei Province, indicating that rising temperatures generally inhibit carbon storage, possibly because high temperatures exacerbate water stress and suppress vegetation growth. Positive influence areas were only sporadically distributed in localized parts of the northern Hebei Plateau and the Taihang Mountains, where thermal conditions are relatively insufficient and temperature increases are favorable for vegetation growth.

The spatial distribution of slope coefficients showed that the positive influence of slope on carbon storage was mainly concentrated in the Taihang Mountains and the Yanshan Mountains, where the great topographic relief limits human activities and favors vegetation preservation. In plain areas, slope coefficients were close to zero, indicating no significant influence.

The spatial distribution of DEM coefficients (Fig. 6f) exhibited pronounced spatial differentiation. Most areas showed positive influences, primarily distributed in the central Hebei Plain, suggesting that slight increases in elevation favor carbon storage accumulation, possibly due to improved drainage conditions. Negative influence areas were concentrated in the northern Hebei mountainous region, where higher elevations corresponded to lower carbon storage, possibly related to low temperatures and restricted vegetation growth at high altitudes.

The spatial distribution of nighttime light coefficients (Fig. 6g) also exhibited pronounced spatial differentiation. Negative influence areas were concentrated in localized parts of the Taihang Mountains and the Yanshan Mountains, where, despite the presence of some lighting, surrounding areas maintain high vegetation cover, showing a statistical negative correlation, possibly related to vegetation disturbance from tourism development or village agglomeration. Positive influence areas were widely distributed in the central and southern Hebei urban agglomerations and the eastern Hebei urban agglomerations, where urbanization levels are high, and nighttime light showed a positive correlation with carbon storage, possibly related to urban greening improvements and ecological spatial planning.

4.3.4. Analysis of regional differentiation characteristics

Based on the spatial distribution of all factor coefficients, the driving mechanisms of carbon storage in Hebei Province can be divided into three main regions:

(1) Taihang Mountain Ecological Zone: In this region, NDVI coefficients are the highest (>1.0), slope coefficients are positive, and temperature coefficients are negative, indicating that carbon storage is primarily controlled by vegetation and topography. The significant ecological conservation effects make this region the most important carbon sink functional area in Hebei Province.

(2) Northern Hebei Plateau Zone: In this region, NDVI coefficients are relatively low (0.75–0.85), precipitation coefficients are partially negative, and temperature and DEM coefficients are predominantly negative. This indicates that carbon storage is constrained by high altitude, low temperature, and negative precipitation effects, resulting in weakened positive effects of vegetation and relatively limited carbon sink function.

(3) Central Hebei Plain–Urban Agglomeration Zone: In this region, NDVI coefficients are moderate, while population density coefficients and slope coefficients are negative, and precipitation coefficients are positive. Population agglomeration and construction land expansion are the main driving factors of carbon storage decline, suggesting that urban ecological spatial planning should be strengthened.

5. Discussion

5.1.Comparison with previous studies

This study found that NDVI is the most important factor affecting carbon storage in Hebei Province, which is consistent with the findings of Liu Yan [28], Chu Xinglin [29], Zhang Yusen [30], and Zhang Qidi [31] in their respective study areas—vegetation cover is the core controlling factor influencing carbon storage. Compared with the study by Zhou Mengjun [32] on tropical mangrove forests, this study reveals the uniqueness of the driving mechanisms of carbon storage in temperate regions: the influence of NDVI is more prominent (25% importance), followed by temperature (16%), while the influence of precipitation is relatively moderate (12%). This may be related to the fact that Hebei Province is located in a semi-humid to semi-arid transition zone, where thermal conditions are more limiting than moisture conditions.

Compared with similar studies on carbon storage estimation based on remote sensing data, Liu Fen et al. [33] used multiple stepwise regression to model aboveground carbon storage in shrublands of Inner Mongolia, achieving R^2 values of 0.61, 0.74, and 0.86, respectively; Shen Gaoyun et al. [34] applied a sequential Gaussian cosimulation method to spatially estimate forest carbon storage in Xianju County, Zhejiang Province, with an R^2 of 0.6203. In this study, the CatBoost model achieved R^2 values stable between 0.53 and 0.54 from 2000 to 2015, but declined to approximately 0.39 from 2020 to 2025. This may be related to the intensifying changes in carbon storage and the strengthening nonlinear relationships among driving factors in recent years, and also indicates that the current driving factor system has limited explanatory capacity for recent carbon storage changes, suggesting that future studies should further enrich the set of driving factors. Notably, the MGWR model in this study achieved an R^2 of 0.9993, representing an improvement of approximately 61 percentage points over the CatBoost model.

This is consistent with the findings of Han Xin et al. and Li Keke et al. Han X [35] et al. found that the MGWR model exhibited higher R^2 explanatory power than traditional regression models in a study on urban street restorativeness. Li Keke [36] et al. also demonstrated that MGWR can greatly improve the explanatory power of socioeconomic variables for erosion rates in a study on soil erosion in China. All these studies confirm that model explanatory power is significantly enhanced after accounting for spatial heterogeneity.

5.2.Advantages and limitations of the model approach

This study couples the InVEST, CatBoost, and MGWR models, sequentially addressing three core questions: "How does carbon storage change?","Which factors are driving the changes?", and "In which regions are these driving factors more influential?" The InVEST model provides reliable baseline data on carbon storage; CatBoost combined with SHAP values identifies the core driving factors; and the MGWR model, through multiscale bandwidth settings, reveals the spatial differences in factor influences, uncovering a composite driving mechanism characterized by the dominance of natural factors and the localized influence of anthropogenic factors.

However, this study still has certain limitations. First, the carbon density parameters are derived from literature rather than field measurements, which may introduce certain errors. Second, the social data (population) are spatially gridded products that may deviate from actual distributions. Third, the study period covers 2000 to 2025; future research could combine the FLUS model to conduct multi-scenario projections. Fourth, the proportions of significant points for some factors in the MGWR model (population density, nighttime light) are relatively low, which may be related to data accuracy or the localized influence characteristics of these factors themselves; subsequent studies could consider higher-resolution socioeconomic data. Fifth, this study selected seven driving factors, which may not fully encompass all key factors affecting carbon storage (such as land use structure, policy factors, etc.). This could be one of the reasons why the R^2 of the CatBoost model in this study is slightly lower than that in some comparable studies, and future research should further enrich the set of driving factors.

5.3.Implications for carbon management in hebei province

The findings of this study have several important implications for carbon management in Hebei Province. First, NDVI is the core driving factor, indicating that vegetation protection is the fundamental approach to enhancing carbon storage. Priority should be given to ecological restoration and vegetation recovery in the Taihang Mountains, while vegetation protection in the northern Hebei Plateau should also be strengthened. Second, slope became the most important factor in 2025, suggesting that the spatial differentiation between mountainous and plain areas in terms of carbon storage has become increasingly pronounced—carbon storage in the plains is declining rapidly while that in the mountains remains relatively stable. This implies that the expansion of construction land in plain areas should be controlled during urbanization to protect the carbon sink function in mountainous areas. Third, the negative influence of population density is mainly concentrated in urban agglomeration areas, highlighting the need to strengthen urban ecological planning in the context of rapid urbanization. Fourth, the spatial heterogeneity of the influences of various factors indicates that carbon sequestration management should be tailored to local conditions. Specifically, vegetation protection should be prioritized in the Taihang Mountains, ecological restoration in the northern Hebei Plateau, and ecological spatial planning in urban agglomerations.

6. Conclusion

Based on multi-source data and the InVEST-CatBoost-MGWR coupled model, this study systematically analyzed the spatiotemporal variation characteristics and driving mechanisms of carbon storage in Hebei Province from 2000 to 2025, and drew the following main conclusions:

(1) From 2000 to 2025, carbon storage in Hebei Province exhibited a continuous declining trend, decreasing from 1.887 billion tons to 1.841 billion tons, with a cumulative reduction of 0.046 billion tons, equivalent to a decrease of 2.44%. The decline was only 0.23% from 2000 to 2015, but reached 2.23% from 2015 to 2025, indicating a marked acceleration in the decline rate. The spatial distribution exhibited a pattern of "high in the northwest and low in the southeast," with high-value areas concentrated in the Taihang Mountains and the Yanshan Mountains, and low-value areas distributed in the urban agglomerations of the central Hebei Plain.

(2) From 2000 to 2015, NDVI was the most important driving factor affecting carbon storage (importance: 24.8%–25.9%; SHAP value: 7.18–7.47), followed by temperature (15.4%–16.5%) and population density (14.0%–14.7%). From 2020 to 2025, NDVI remained among the top factors, but the importance of slope increased significantly, and slope became the most important factor in 2025 (importance: 22.5%; SHAP value: 5.01), reflecting the enhanced controlling role of topographic conditions on carbon storage in recent years.

(3) The influences of driving factors exhibited significant spatial heterogeneity. The 2025 MGWR model results showed that NDVI had a significant influence across the entire province (100% significant points), followed by precipitation (87.36%) and DEM (73.23%) with relatively extensive influences, while population density (8.86%) and nighttime light (17.26%) were only significant in localized areas. NDVI exerted the strongest influence in the Taihang Mountains, while population density exhibited a significant negative influence in the urban agglomerations of the central Hebei Plain.

Acknowledgment

This work was supported by the College Students' Innovation and Entrepreneurship Training Program of North China University of Science and Technology (Grant No. [X2025092]). We thank Hui Chao for his guidance and support throughout this study.

References

[1] Zhang, L., Yang, J. W., Cao, C. J., et al. (2026). Land use change and its carbon sequestration capacity in the Yellow River Basin (Henan section). Yellow River, 48(3), 107–112+118. (In Chinese)

[2] Jin, M. Z. (2020). Carbon storage function and value assessment of alpine grassland ecosystem in Maduo County,

Sanjiangyuan region [Doctoral dissertation]. Lanzhou University, China. (In Chinese)

[3] Yu, Z. X., & Liu, Y. (2025). Spatiotemporal evolution and multi-scenario prediction of carbon storage in the Xiangjiang River Basin based on PLUS and InVEST models. Journal of Agricultural Resources and Environment, (2025), 1–15. (In Chinese)

[4] Lou, J. Z. (2025). Carbon storage change and prediction in the karst region of Southwest China based on the PLUS-InVEST model. Southern Agricultural Machinery, 56(7), 67–70. (In Chinese)

[5] Fan, Y., Dou, P. F., Tian, Y. F., et al. (2025). Ecological protection and restoration zoning at the county level in the low-mountain and hilly areas of Fujian based on the supply-demand relationship of ecosystem services. Chinese Journal of Ecology, (2025), 1–13. (In Chinese)

[6] Obateru, O. R., Okhimamhe, A. A., Fashae, A. O., et al. (2025). Assessing the status of ecosystem regulating services in the urbanising Rainforest and Guinea savanna ecological regions of Nigeria using InVEST models. Urban Climate, 61, 102410.

[7] Dat, T. P., Naoto, Y., Trang, T. T. N., et al. (2021). Improvement of mangrove soil carbon stocks estimation in North Vietnam using Sentinel-2 data and machine learning approach. GIScience & Remote Sensing, 58(1), 68–87.

[8] Hu, Z. H., Wang, L. F., Shang, R. G., et al. (2025). Changes and driving factors of tree layer carbon storage in old-growth forests in Taiyanghe Provincial Nature Reserve, Yunnan. Chinese Journal of Applied Ecology, (2025), 1–11. (In Chinese)

[9] Haukenes, L. V., Asplund, J., Nybakken, L., et al. (2025). Disentangling drivers of organic layer and charcoal carbon stocks in boreal pine and spruce forests with different fire histories. Forest Ecosystems, 14, 100334.

[10] Hao, Z. X., Ma, J. K., Wang, A., et al. (2025). Changes and influencing factors of habitat quality in the Shanxi section of the Yellow River Basin based on the InVEST-MGWR model. Chinese Journal of Applied Ecology, (2025), 1–12. (In Chinese)

[11] Zhang, J. Y., Yang, X. B., & Meng, X. S. (2025). Spatiotemporal distribution characteristics of forest carbon storage in the Beijing-Tianjin-Hebei region. Journal of Northeast Forestry University, 53(7), 75–83. (In Chinese)

[12] Huang, C. H., Yang, J., & Zhang, W. J. (2013). Research progress of ecosystem service function assessment models. Chinese Journal of Ecology, 32(12), 3360–3367. (In Chinese)

[13] Zhang, P. (2022). Multi-scale spatiotemporal distribution of land use and ecosystem services based on InVEST and PLUS models [Doctoral dissertation]. Hebei University of Engineering, China. (In Chinese)

[14] Luo, L., Wu, L. K., Wang, Q. N., et al. (2025). Spatiotemporal evolution and prediction of mangrove wetland carbon storage in Hainan Island based on the PLUS-InVEST model. Science of Soil and Water Conservation, (2025), 1–14. (In Chinese)

[15] Yue, S. J. (2024). Characteristics of ecosystem service function changes and future multi-scenario simulation in the Beijing-Tianjin-Hebei region based on PLUS and InVEST models [Master's thesis]. Henan Agricultural University, China. (In Chinese)

[16] Li, L. Y., Chen, L., Han, M. J., et al. (2024). Spatiotemporal evolution and multi-scenario simulation of ecosystem carbon storage in Hebei Province. Chinese Journal of Applied Ecology, 35(12), 3247–3256. (In Chinese)

[17] Li, J. P., Xia, S. X., Yu, X. B., et al. (2020). Carbon storage of terrestrial ecosystems in Hebei Province based on the InVEST model. Journal of Ecology and Rural Environment, 36(7), 854–861. (In Chinese)

[18] Niu, X. G., Zhu, X. L., Liu, M. Y., et al. (2024). Spatiotemporal evolution characteristics and driving factors of ecosystem carbon storage and habitat quality in Hebei Province. Bulletin of Soil and Water Conservation, 44(6), 353–365. (In Chinese)

[19] Liu, F. Y. (2020). Application of root cause localization for anomalous faults based on CatBoost model [Doctoral dissertation]. Lanzhou University, China. (In Chinese)

[20] Shi, J. (2023). Application of BO-CatBoost in earthquake prediction in the Western region [Master's thesis]. Guangxi Normal University, China. (In Chinese)

[21] Du, Y. X., Hu, Z. J., Chen, W. N., et al. (2021). Transient stability assessment method of power system based on improved CatBoost. Electric Power Automation Equipment, 41(12), 115–122. (In Chinese)

[22] Guo, W., Huang, Z., Zhou, M. Z., et al. (2026). Prediction of inrush current of no-load transformer based on CatBoost algorithm. Electric Drive, 56(2), 34–42. (In Chinese)

[23] Lai, J., Qi, S., Chen, J., et al. (2025). Exploring the spatiotemporal variation of carbon storage on Hainan Island and its driving factors: Insights from InVEST, FLUS models, and machine learning. Ecological Indicators, 172, 113236.

[24] Yi, F. L., Chen, D. R., Yang, H., et al. (2024). Classification of mild cognitive impairment using XGBoost-SHAP interpretable machine learning framework. Chinese Journal of Health Statistics, 41(3), 423–429. (In Chinese)

[25] Ma, Z. L., Han, M., Kong, X. L., et al. (2026). Interpretable driving analysis of carbon storage in terrestrial ecosystems of the Yellow River Delta based on SHAP-XGBoost model. Environmental Science, (2026), 1–28. (In Chinese)

[26] Huang, T., Peng, M. M., Wan, C. F., et al. (2026). Intelligent prediction of fatigue life of corroded steel wires in bridge cables based on SSA-LSTM-SHAP. Engineering Mechanics, (2026), 1–7. (In Chinese)

[27] Li, W. L., Li, F. R., Zhen, Z., et al. (2025). Estimation of forest carbon storage in Liaoning Province based on multiscale geographically weighted regression model. Journal of Central South University of Forestry and Technology, 45(9), 130–140. (In Chinese)

[28] Liu, Y., Tian, L. H., & Shi, D. R. (2025). Spatiotemporal variation characteristics of carbon storage and its driving factors in the Qinghai Lake Basin. Qinghai Agricultural and Forestry Science and Technology, (3), 1–7+19. (In Chinese)

[29] Chu, X. L. (2025). Spatiotemporal evolution characteristics and multi-scenario prediction of carbon storage in the Tumen River Basin under land use change [Doctoral dissertation]. University of Jinan, China. (In Chinese)

[30] Zhang, Y. S. (2024). Ecosystem service changes, trade-offs/synergies, and driving factors in Eastern Tibet [Doctoral dissertation]. Liaoning Technical University, China. (In Chinese)

[31] Zhang, Q. D. (2024). Spatiotemporal evolution and prediction analysis of ecosystem carbon storage in the Danjiangkou Reservoir area [Doctoral dissertation]. Henan University, China. (In Chinese)

[32] Zhou, M. J. (2025). Changes in distribution and carbon storage of mangroves in Hainan Island over the past 30 years [Doctoral dissertation]. Hainan Tropical Ocean University, China. (In Chinese)

[33] Liu, F., Wei, J. S., Zhou, M., et al. (2016). Remote sensing models of carbon storage in shrublands of Armeniaca sibirica and Caragana microphylla based on Landsat 8 OLI data. Forest Resources Management, (1), 112–117. (In Chinese)

[34] Shen, G. Y., & Zhang, M. Z. (2015). Estimation of spatial distribution of forest carbon density at multi-scale regions

based on sequential Gaussian cosimulation. Journal of Southwest Forestry University, 35(2), 55–62. (In Chinese)

[35] Xin, H., Lei, W., Jie, H., et al. (2023). Restorative perception of urban streets: Interpretation using deep learning and MGWR models. Frontiers in Public Health, 11, 1141630.

[36] Li, K. K., Zhou, Z. H., & Wang, Z. (2024). Socioeconomic spatial driving factors of soil erosion in China: An analysis based on the multiscale geographically weighted regression model. Journal of Huazhong Agricultural University, 43(6), 29–38. (In Chinese)