



OPEN Spatiotemporal dynamics of ecological quality and its drivers in Shanxi Province and its planned mining areas

Lulu Chen¹, Huabin Chai¹✉, Weibing Du^{1,2}, Zengzeng Lian¹, Yu Wang¹, Chunyi Li¹, Lailiang Cai¹ & Lei Zhang¹

As a major coal-producing province, understanding the spatiotemporal evolution of ecological quality and its driving factors in Shanxi is essential for promoting environmental protection and sustainable development. This study employs MODIS data to calculate the Remote Sensing Ecological Index (RSEI) for Shanxi Province and its designated mining areas from 2000 to 2023, aiming to investigate the spatial and temporal dynamics of ecological quality. The CatBoost model and Geographically Weighted Regression (GWR) are applied to identify and analyze the underlying driving factors. The results show that ecological quality in both Shanxi Province and its planned mining regions exhibited an overall upward trend between 2000 and 2020, with varying levels of improvement observed across different mining zones. Trend analysis indicates a general enhancement in ecological conditions over the past two decades. RSEI displays significant spatial autocorrelation, characterized by high-value clustering in the southern regions and low-value clustering in the northern and western mining zones and areas with intensive human activity. Key influencing factors include elevation, net primary productivity (NPP), precipitation, and population density. The CatBoost model, supplemented with SHAP (SHapley Additive exPlanations) values, quantifies the relative importance and predictive contribution of each factor to RSEI outcomes. The GWR model further reveals spatial heterogeneity in these relationships, uncovering localized effects, spatial gradient patterns, and clustering phenomena. Additionally, the Hurst index analysis indicates that most areas within Shanxi Province and its designated mining zones are likely to maintain an upward trend in ecological quality in the future. As a comprehensive large-scale and long-term assessment, this study provides valuable theoretical and empirical support for regional planning, ecological monitoring, and the management of mining areas, thereby contributing to sustainable development and ecological conservation efforts.

Keywords Ecological quality change, Driving factors, RSEI, GWR, CatBoost

Shanxi Province is one of the most coal-rich regions in China, holding significant strategic importance for ensuring national energy security and supporting economic growth¹. However, large-scale coal mining activities^{2,3} have caused substantial ecological degradation⁴. Therefore, the protection and restoration of ecosystems in Shanxi and its designated mining areas are essential. These efforts not only contribute to improving local environmental quality but also help mitigate the broader impacts of global climate change⁵. Conducting research and continuous monitoring of ecological quality in these areas⁶ is of great importance for regional ecological conservation and the effective implementation of sustainable development strategies⁷.

Ecological quality is intricately linked to human life, making the rational assessment of regional ecological environment quality particularly significant. The rapid advancement of remote sensing technology⁸ has opened up new possibilities for large-scale, long-term monitoring of dynamic changes in ecological environment quality. By analyzing remote sensing data, real-time monitoring of ecosystem changes can be achieved, allowing for the evaluation of environmental quality and providing support for informed decision-making. For instance, Zhong et al. utilized the Normalized Difference Vegetation Index (NDVI) to assess regional vegetation coverage⁹; Liu et al. identified landscape drivers of urban heat island effects using land surface temperature (LST)¹⁰; and Li et

¹School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China.

²Collaborative Innovation Center of Geo-Information Technology for Smart Central Plains, Zhengzhou 450000, China. ✉email: chaihbpuedu@163.com

al. investigated ecological changes in the Mu Us Sandy Land by combining land cover trends with NDVI data¹¹. Ye et al. quantified ecological quality in the Loess Plateau through soil moisture (SM) measurements¹². While these remote sensing indices provide valuable insights, the assessment based on single indicators is often one-dimensional and insufficient for comprehensive ecological evaluations. As a result, the trend toward utilizing multiple integrated remote sensing-based indicators for ecological quality assessment has become a necessity. Xu et al. proposed the Remote Sensing Ecological Index (RSEI), which integrates greenness, wetness, heat, and dryness¹³. Through statistical methods such as principal component analysis, RSEI can ascertain the weights of each indicator's impact on ecological quality, leading to comprehensive ecological quality assessments. Currently, The RSEI has gained widespread application in ecological assessments across various landscapes, including urban agglomerations¹⁴, islands¹⁰, and mountainous regions¹⁵. This established methodology makes RSEI an appropriate and reliable tool for evaluating ecological quality in Shanxi Province and its designated mining areas. While previous studies have commonly employed the NDVI to characterize vegetation greenness, the Kernel Normalized Difference Vegetation Index (kNDVI) offers distinct advantages. The kernel-based approach¹⁶ of kNDVI effectively mitigates saturation effects, accounts for complex phenological cycles, and better accommodates seasonal variations. Furthermore, kNDVI demonstrates superior performance in resolving mixed pixel issues and exhibits enhanced sensitivity to vegetation dynamics. These technical advantages enable kNDVI to provide more accurate vegetation coverage and health data across diverse ecological conditions, consequently improving the robustness of ecological assessments. Therefore, incorporating kNDVI as the greenness component in RSEI construction for monitoring ecological changes in Shanxi Province represents both a methodologically sound and scientifically rigorous approach.

Ecological quality, as a key indicator of ecosystem health, stability, and resilience¹⁷, is closely linked to human well-being and socio-economic development¹⁸. In the context of globalization, various factors—including topography, climate change, and population density—exert substantial influence on ecological environmental quality¹⁹. Specifically, moderate wind speeds and adequate precipitation contribute to enhanced ecosystem stability and productivity, while elevation and slope affect ecological conditions by altering local climate regimes, soil characteristics²⁰, and light availability. Areas with high population density²¹ often face intensified land development and resource consumption, leading to the degradation of ecosystem services²² and overall ecological quality. Identifying and quantifying the factors that influence ecological quality is essential for implementing targeted measures to prevent environmental pollution and mitigate ecosystem degradation. Such efforts are critical for advancing ecological protection and restoration²³. Therefore, when evaluating the driving forces behind ecological quality, it is imperative to take a comprehensive approach that accounts for the complex interplay among multiple influencing factors, as well as the effects of spatial heterogeneity.

When analyzing the driving factors behind changes in ecological quality, traditional regression methods, correlation analyses, and residual analyses exhibit certain limitations, particularly as these approaches often assume a linear relationship between factors and ecological quality. This assumption may not adequately capture the complex coupling effects among multiple variables. In contrast to conventional analytical techniques, CatBoost, a machine learning algorithm based on gradient boosting decision trees²⁴, effectively addresses classification and regression challenges while learning intricate nonlinear relationships between input and output variables. Furthermore, the application of SHAP (SHapley Additive exPlanations) tools allows for the interpretation of CatBoost model predictions²⁵, quantifying the impact of different features on these predictions and thereby providing deeper insights into large-scale ecological influencing factors. In addition, traditional methods often fail to account for the spatial heterogeneity inherent in geospatial data, which can significantly influence the assessment and prediction of ecological quality²⁶. The Geographically Weighted Regression (GWR) model is specifically designed to address spatial data^{27,28}, allowing it to capture spatial heterogeneity and reveal localized relationships between ecological variables and the RSEI across different regions. Applying the GWR model to analyze the driving forces behind ecological change is essential for understanding the complex spatial dynamics of ecosystems²⁹. While CatBoost excels at modeling nonlinear relationships and handling complex variable interactions, GWR is particularly effective in detecting spatial variation and regional disparities. The combined use of these two methods enables a more comprehensive and accurate analysis of ecological quality changes and their underlying drivers³⁰.

This study employs the RSEI to evaluate the variations in ecological quality across Shanxi Province and its planned mining areas. By calculating the kNDVI, humidity (WET), land surface temperature (LST), and dryness (NDBSI), the RSEI is constructed through PCA to provide a comprehensive assessment of the ecological environment quality in the region. Subsequently, utilizing the CatBoost model and GWR, this research reveals the influence of different factors on the ecological environment, thereby enhancing the reliability of the study's findings. The objectives of this research are to: (1) investigate the dynamic changes in ecological quality (RSEI) in Shanxi Province and its planned mining areas from 2000 to 2020; (2) analyze the trends in RSEI variations within the region; (3) explore the potential factors contributing to changes in ecological quality in Shanxi Province and its mining areas; and (4) reveal the spatial heterogeneity of ecological quality and examine its specific relationships with underlying factors.

Materials and methods

Study area

Shanxi Province (34°34'–40°44'N, 110°14'–114°33'E) is located in northern China, in the eastern part of the Loess Plateau, with a total area of 156,700 square kilometers (Fig. 1). The province encompasses twelve planned mining areas, featuring a wealth of diverse mineral resources and a complex geological environment. The climate is classified as a temperate continental monsoon climate, characterized by significant temperature variations across regions, with temperatures decreasing from south to north and from plains to mountainous areas. The province's geomorphology is predominantly mountainous and hilly, comprising approximately 80% of the total

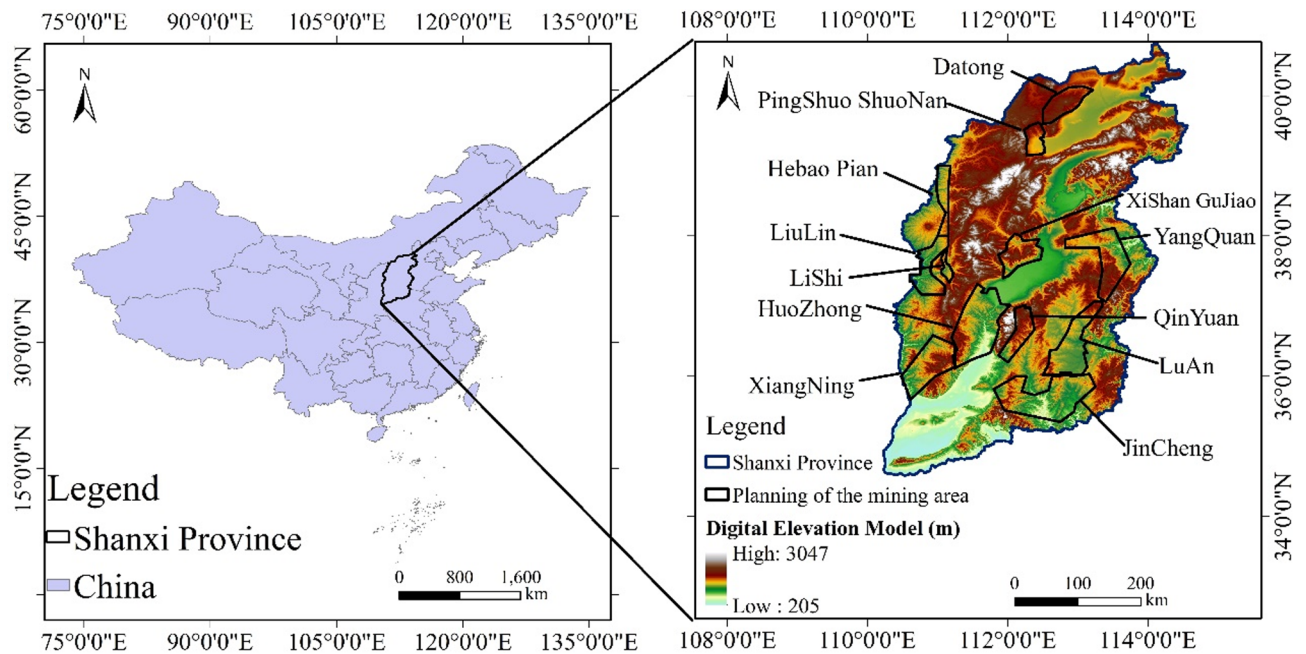


Fig. 1. Location map of the study area. (This image was created using ArcGIS 10.8 software, which is an Esri product and is publicly available at the URL (<https://www.esri.com/>). The base map is based on the standard map (GS2024, No. 0650), and no modifications have been made to the map boundaries.)

area. The study area includes critical regions such as the Taihang and Lüliang mountains, which are rich in resources but face numerous challenges, including resource depletion and ecological restoration.

Technical route

The detailed workflow of this study is illustrated in Fig. 2: (1) RSEI was constructed based on the GEE platform to invert the ecological quality changes in Shanxi Province and its planned mining areas from 2000 to 2020; (2) Theil-Sen analysis and Mann-Kendall trend tests were employed to analyze the RSEI trend; (3) Moran's I was utilized to assess the spatial correlation of RSEI in Shanxi Province; (4) The interactions between RSEI and potential driving factors were analyzed using CatBoost and GWR models.

Datasets and preprocessing

This study employed the MODIS dataset to construct the ecological quality assessment index. The data sources included the MOD09A1 (2000–2023, 500 m resolution, 8-day interval) and MOD11A2 (2000–2023, 500 m resolution, 8-day interval) from the MODIS collection. The JRC annual water classification historical data was utilized for water body masking. Additional datasets were leveraged to extract potential factors that may influence ecological quality in the study area, including Terra Climate (2000–2020, providing climate factor data), NASA's SRTM Digital Elevation data (90 m resolution), population density data provided by East View Cartographic, and MYD17A3HGF (2000–2020, net primary productivity data). All data were acquired from the GEE cloud platform, with all datasets resampled to a resolution of 500 m. Atmospheric correction and relevant preprocessing techniques were applied to enhance data quality. To account for seasonal variation and meteorological conditions, the data time window was set from June 15 to September 15 each year, ensuring consistency in vegetation status and ecological outcomes.

Methods

RSEI indicator construction

RSEI leverages ground information and ecological data obtained through remote sensing technologies, employing mathematical and statistical methods for processing and analysis, thereby facilitating rapid assessment and monitoring of the health status and environmental quality of specific regions³¹. Given that greenness, humidity, heat, and dryness are crucial components of the ecological environment, this study selects these four ecological factors to construct the RSEI index³².

$$RSEI = f(\text{Greenness}, \text{Wetness}, \text{Heat}, \text{Dryness}) \quad (1)$$

where f is the set of functions used to perform PCA.

kNDVI applies a weighted average of NDVI for each pixel and its surrounding pixels to more accurately capture the details and variations in vegetation coverage. Thus, kNDVI serves as a representation of greenness. The calculation formula is as follows:

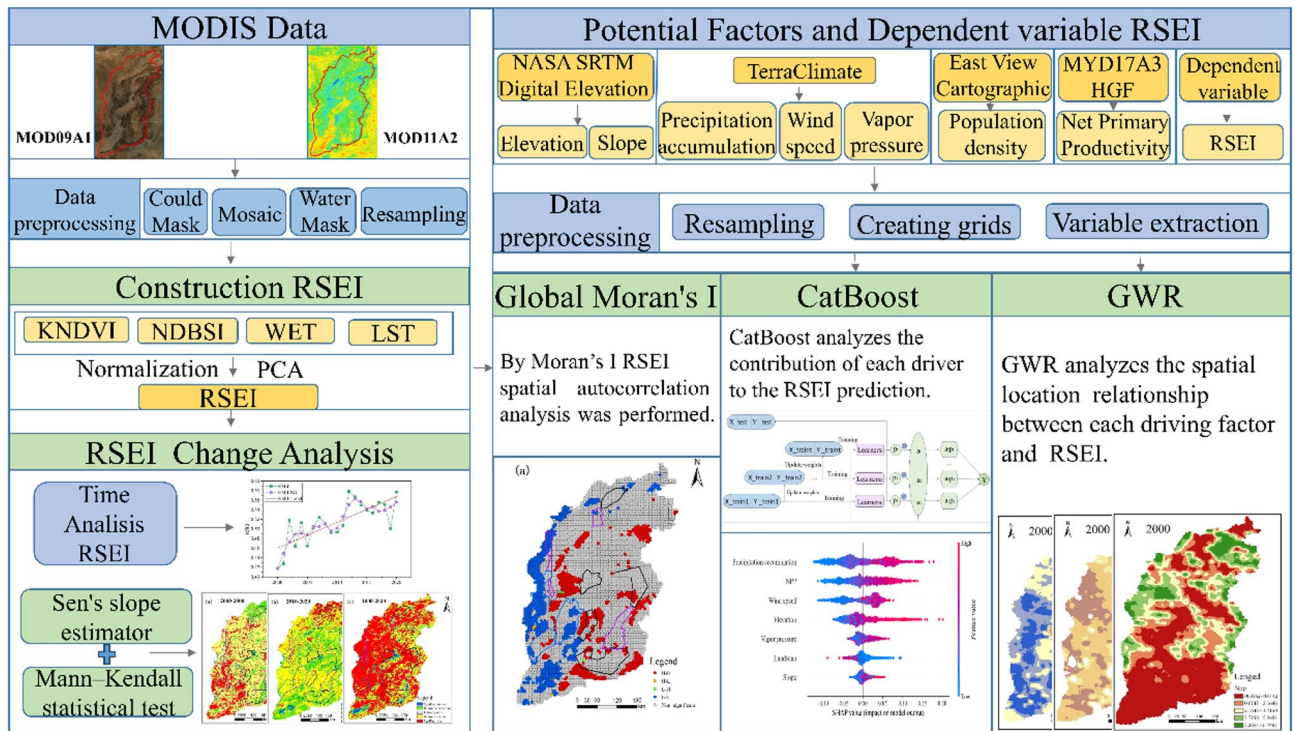


Fig. 2. Technology roadmap. (This maps was created using ArcGIS 10.8 software, which is an Esri product, generated using our dataset, and is publicly available at the URL (<https://www.esri.com/>).)

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (2)$$

$$kNDVI = \tanh \left(\left(\frac{NIR - Red}{2\sigma} \right)^2 \right) = \tanh \left(\left(\frac{NDVI}{2\sigma} \right)^2 \right) \quad (3)$$

where NIR and RED represent the reflectances of the near-infrared ($NIR1$) and red bands of MOD09A1, respectively. The symbol σ denotes a length scale that is proportional to the average reflectance of the near-infrared and red bands, which can be adjusted. When σ is set to $0.5(NIR + RED)$, the calculation formula is as follows:

$$kNDVI = \tanh (NDVI^2) \quad (4)$$

Humidity (WET) is calculated using the formula:

$$WET = 0.1147 \times \rho_1 + 0.2489 \times \rho_2 + 0.2408 \times \rho_3 + 0.3132 \times \rho_4 - 0.3122 \times \rho_5 - 0.6416 \times \rho_6 - 0.5087 \times \rho_7 \quad (5)$$

where ρ_i ($i = 1, 2, \dots, 7$) represents the reflectance of the red, near-infrared 1, blue, green, near-infrared 2, short-wavelength infrared 1, and short-wavelength infrared 2 bands of the MOD09A1 image, respectively.

Dryness consists of the Index of building (IBI) and the index of bare Soil (SI). The formula for calculation is:

$$NDBSI = (IBI + SI) / 2 \quad (6)$$

$$IBI = \frac{2\rho_6/(\rho_6 + \rho_2) - \rho_2/(\rho_2 + \rho_1) + \rho_4/(\rho_4 + \rho_6)}{2\rho_6/(\rho_6 + \rho_2) + \rho_2/(\rho_2 + \rho_1) + \rho_4/(\rho_4 + \rho_6)} \quad (7)$$

$$SI = \frac{(\rho_6 + \rho_1) - (\rho_2 + \rho_3)}{(\rho_6 + \rho_1) + (\rho_2 + \rho_3)} \quad (8)$$

where ρ_i ($i = 1, 2, \dots, 6$) denotes the reflectance of the red, near-infrared 1, blue, green, near-infrared 2, and short-wavelength infrared 1 bands of the MOD09A1 image, respectively.

Heat is expressed in terms of land surface temperature (LST). MOD11A2 LST products providing the necessary data. The raw LST data (LST0) requires unit conversion from Kelvin (K) to degrees Celsius (°C). The formula for this conversion is:

$$LST = 0.02LST_0 - 273.15 \tag{9}$$

In this paper, the RSEI is calculated using the following formula:
 $RSEI = f(kNDVI, Wet, NDBSI, LST)(10)$.
After obtaining the results of the four ecological factors, it is essential to linearly map each indicator's values to the range of [0, 1] for normalization. This process eliminates the impacts caused by differing units and value ranges, allowing for PCA to construct the RSEI.
Furthermore, RSEI is subjected to additional normalization to facilitate comparison and measurement, as shown in the following formula:

$$RSEI = \frac{RSEI_0 - RSEI_{min}}{RSEI_{max} - RSEI_{min}} \tag{11}$$

where $RSEI_0$ is the initial RSEI, and $RSEI_{min}$ and $RSEI_{max}$ represent the minimum and maximum values of $RSEI_0$, respectively. A higher RSEI value, approaching 1, indicates better ecological environmental quality, while a lower RSEI value signifies poorer ecological environmental quality.
To enhance the assessment and comparison of ecological conditions, the RSEI values were categorized into five levels from high to low: excellent (0.8–1), good (0.6–0.8), moderate (0.4–0.6), fair (0.2–0.4), and poor (0–0.2). This classification provides a clearer understanding of ecological quality, with values closer to 1 indicating better ecological environment health.

Sen's slope estimator and Mann–Kendall statistical test
Sen's Slope Estimator can accurately estimate the trend slope in time series data, encompassing both linear and nonlinear trends. Its robustness allows it to effectively handle outliers and data with significant fluctuations. By employing Sen's Slope Estimator, the strength of the trend in each time series can be determined, thereby identifying the rate and magnitude of ecosystem changes. The calculation formula is as follows:

$$\beta = Median \left(\frac{X_j - X_i}{j - i} \right), \forall j > i \tag{12}$$

where β denotes the slope and X_i and X_j denote the data values for year i and year j , respectively.
The MK test is employed to detect the presence and significance of trends in time series data³³. It does not rely on assumptions about the data distribution, making it applicable to various types of ecological data. While Sen's Slope Estimator provides a quantitative assessment of the trends, the Mann–Kendall test confirms the significance and statistical stability of these trends³⁴. The calculation formulas for the MK test are as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n sgn(x_j - x_i) \tag{13}$$

$$sgn(x_j - x_i) = \begin{cases} 1, & x_j - x_i > 0 \\ 0, & x_j - x_i = 0 \\ -1, & x_j - x_i < 0 \end{cases} \tag{14}$$

$$V = n(n - 1)(2n + 5) / 18 \tag{15}$$

$$Z = \begin{cases} \frac{S-1}{\sqrt{V}}, & S > 0 \\ 0, & S = 0 \\ \frac{S+1}{\sqrt{V}}, & S < 0 \end{cases} \tag{16}$$

where n denotes the time series length; sgn is the sign function; S is the statistic; and Z is the trend significance test statistic.
In the MK test, the significance of RSEI changes is determined based on a set significance level α . Typically, α is set at 0.05, which corresponds to a Z value of ± 1.96 . Therefore, when the absolute value of the calculated Z from the MK test exceeds 1.96, the change passes the significance test at a 95% confidence level, indicating that the trend change is significant; conversely, the opposite holds true. The results are categorized into five classes, as shown in the table below (Table 1).

β	Z	RSEI trend
≥ 0.0005	≥ 1.96	Significant rise (SR)
≥ 0.0005	$-1.96 \leq Z < 1.96$	No significant rise (NSR)
$-0.005 \sim -0.005$	$-1.96 \leq Z < 1.96$	Relatively stable (RS)
≤ -0.0005	$-1.96 \leq Z < 1.96$	No significant decrease (NSD)
≤ -0.0005	≤ -1.96	Significant decrease (SD)

Table 1. RSEI trend classification.

Hurst exponent

The Hurst exponent is an indicator used to measure the long-term memory in time series or fractals, aiding in understanding the fundamental dynamics of the series and providing a degree of predictability for changes in ecological quality³⁵. The formula for calculating the Hurst exponent is as follows:

$$H = \log(R/S)/\log(n) \quad (17)$$

where H is the Hurst exponent. In the actual computation, it is typically necessary to perform multiple decompositions and calculate the R/S values at different scales, followed by averaging these R/S values to obtain the final estimate of the Hurst exponent.

The Hurst exponent is commonly employed to analyze the persistence or anti-persistence of time series, with values ranging from 0 to 1. Specifically, when $0 < H < 0.5$, it indicates anti-persistent change; the closer H is to 0, the stronger the anti-persistence. When $0.5 < H < 1$, it suggests persistent change; the closer H is to 1, the stronger the persistence. When H equals 0.5, it denotes that the time series is independent and random, exhibiting uncertainty with no correlation between past and future trends.

Global moran's

Global Moran's I is a commonly used method for spatial autocorrelation analysis, employed to measure the spatial correlation among observations within a spatial dataset and to reveal the spatial relationships between reference units and their neighboring units²⁶. The Moran's I statistic ranges from -1 to $+1$; a value close to $+1$ indicates that the observations of neighboring geographic units tend to be similar, while a value close to -1 suggests that the observations are inclined to be opposite. A Moran's I value near 0 indicates that the data are randomly distributed in space, showing no significant spatial autocorrelation. The formula for calculating Moran's I is as follows:

$$\text{Moran's } I = \frac{m * \sum_{i=1}^m \sum_{j=1}^m w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^m \sum_{j=1}^m w_{ij} (x_i - \bar{x})^2} \quad (18)$$

where m represents the total number of samples, x_i and x_j denote the RSEI values at positions i and j , w_{ij} indicates the spatial weight value between i and j , and \bar{x} is the mean value of the RSEI.

CatBoost

CatBoost is a machine learning algorithm based on gradient-boosted decision trees, demonstrating exceptional performance in handling classification and regression problem³⁶. One notable advantage of CatBoost is its ability to efficiently manage categorical features while achieving high prediction accuracy. Additionally, CatBoost employs regularization techniques to mitigate overfitting, enhancing the robustness of the model when dealing with complex data. The fundamental steps for training a CatBoost model are illustrated in Fig. 3: first, create a CatBoost model and train it using the training dataset; once the model is trained, it can be utilized for predicting new data. Throughout the training and prediction processes, CatBoost simplifies the complexities of data preprocessing and offers an efficient gradient boosting algorithm, assisting users in constructing more accurate models.

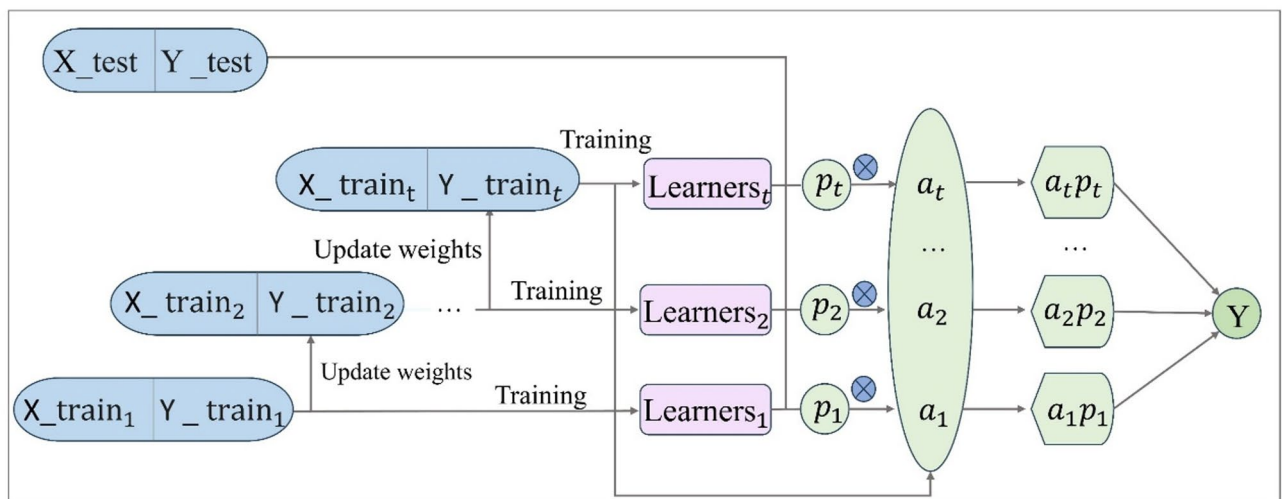


Fig. 3. CatBoost implementation process. In the figure X_{train_t}, Y_{train_t} is the training set, X_{test}, Y_{test} is the test set, p_t is the prediction of the t -th sample, a is the weight of the t -th sample, and Y is the final prediction obtained by weighting the prediction of each learner.

Geographically weighted regression (GWR)

In spatial research, particularly at large scales, conventional multivariate linear regression models may fail to accurately capture the variations and heterogeneity in geographic space. The GWR model, on the other hand, is adept at exploring the relationships between variables in spatial data, effectively reflecting the correlations between dependent and independent variables across geographic space³⁷. Moreover, the GWR model estimates parameters through independent linear regressions within each local area, utilizing local parameter estimates instead of global ones. The kernel type is “bisquare” and the number of neighbors is determined by the optimal bandwidth (gwr_bw). The formula for the GWR model is as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{k,i} + \epsilon_i \tag{19}$$

where y_i is the dependent variable for sample i . (u_i, v_i) are the coordinates of sample i . $\beta_0(u_i, v_i)$ is the intercept term for sample i . $\beta_k(u_i, v_i)$ is the k -th regression parameter for sample i . $x_{k,i}$ is the k -th independent variable for sample i . ϵ_i represents the random error.

Model evaluation

To evaluate the model’s estimation performance, the dataset is divided into 80% training set and 20% testing set. Two evaluation metrics are used: the coefficient of determination (R^2) and the root mean square error (RMSE). Additionally, to assess the model’s generalization ability, we use the 5-fold cross-validation method. Specifically, the data is randomly divided into 5 subsets, and the model is trained and tested on each subset. The final model performance is the average of the results from each fold. The calculation formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{y})^2} \tag{20}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \tag{21}$$

Where x_i is the true value of RSEI, y_i is the estimated value of RSEI, \bar{y} is the mean of the true RSEI values, and n is the number of samples in the test set.

Results

RSEI-based evaluation of ecological environment quality

The average RSEI values and PCA results for Shanxi Province from 2000 to 2020 are presented in Table 2. The contribution of the first principal component (PC1) reached a maximum of 83.13% and a minimum of

Year	RSEI	kNDVI	WET	LST	NDBSI	PC1(%)
2000	0.42	0.60	0.59	0.35	0.57	78.33%
2001	0.43	0.54	0.60	0.44	0.58	83.13%
2002	0.52	0.62	0.48	0.45	0.54	77.65%
2003	0.46	0.63	0.52	0.45	0.58	76.55%
2004	0.51	0.63	0.58	0.40	0.56	71.03%
2005	0.46	0.60	0.56	0.44	0.58	80.48%
2006	0.50	0.62	0.47	0.40	0.57	77.81%
2007	0.52	0.65	0.46	0.43	0.59	77.67%
2008	0.51	0.66	0.48	0.40	0.57	79.07%
2009	0.49	0.61	0.50	0.47	0.60	78.69%
2010	0.49	0.64	0.47	0.41	0.59	80.16%
2011	0.51	0.64	0.45	0.44	0.56	77.17%
2012	0.58	0.68	0.46	0.41	0.54	72.87%
2013	0.57	0.72	0.39	0.36	0.52	65.70%
2014	0.55	0.68	0.42	0.40	0.55	76.28%
2015	0.54	0.67	0.47	0.47	0.57	79.19%
2016	0.53	0.73	0.44	0.38	0.54	70.68%
2017	0.56	0.70	0.46	0.43	0.56	77.60%
2018	0.55	0.74	0.49	0.35	0.54	68.71%
2019	0.50	0.68	0.48	0.38	0.58	76.94%
2020	0.58	0.72	0.43	0.42	0.55	72.35%

Table 2. Mean and PC1 results of RSEI in Shanxi Province (2000–2020).

65.70%, with a multi-year average of 76.14%. This indicates that PC1 concentrates the primary characteristics of each indicator, confirming the scientific validity of using kNDVI, WET, LST, and NDBSI in constructing the RSEI. Furthermore, pixel-by-pixel correlation analysis of RSEI and kNDVI from 2000 to 2020 (Figure S1) revealed a high correlation between the two ($R^2 = 0.90$), which to a certain extent verified the reliability and representativeness of RSEI.

The time-series variation of average RSEI values in Shanxi Province and its planned mining areas from 2000 to 2020 is presented in Fig. 4. During this period, RSEI values in Shanxi ranged from 0.42 to 0.58, with an average annual increase of 1.44%. The lowest value occurred in 2000, while the highest was observed in 2020. Throughout the study period, the average ecological grade consistently remained at a moderate level, indicating a relatively stable and favorable ecological condition in the region. To better illustrate the long-term trend of RSEI averages, the Savitzky-Golay (SG) filter was applied to smooth the time-series curve. The smoothed curve indicates an upward trend from 2000 to 2008 and from 2011 to 2014, while fluctuations or downward trends occurred during other periods. The research found that starting in 2000, ecological restoration projects were initiated in Shanxi Province. The significant rise observed from 2010 to 2014 may be attributed to the comprehensive launch of ecological province construction in 2011 and the in-depth implementation of ecological environment governance and restoration projects, resulting in noticeable improvements in the ecological environment during this period. Overall, from 2000 to 2020, the ecological environment quality in Shanxi Province showed a clear improvement trend, indicating the effectiveness of ecological protection measures.

Figure 5 reveals the spatiotemporal distribution of RSEI values in Shanxi Province and the planned mining areas. From 2000 to 2020, the overall ecological quality in the eastern region of the study area was superior to that of the western region, with areas of better ecological quality concentrated on both sides of the Lüliang and Taihang mountain ranges. Notably, the ecological quality in the Jincheng mining area, Xiangning mining area, and the Xishan gujiao mining area was relatively high, situated in the central and southern parts of Shanxi Province. Conversely, the ecological quality in the Datong mining area, Liulin mining area, Lishi mining area, and Hebao mining area was poorer, predominantly located in the western part of Shanxi Province.

Analyzing the changes in the proportion of areas with different RSEI grades in Shanxi Province, as shown in Fig. 6, reveals that regions with moderate RSEI levels accounted for the highest share, reaching approximately $40\% \pm 3\%$. The proportions of areas classified as fair and poor showed a downward trend, with the poor grade declining from 8% in 2000 to 0% in 2020, and the fair grade decreasing from 41 to 13%. Conversely, the proportions of areas with good and excellent grades exhibited a steady upward trend, with good areas increasing from 14 to 38% and excellent areas rising from 0 to 7%. These results indicate that the ecological quality in Shanxi Province in 2020 is significantly better than in 2000, demonstrating a steady improvement in ecological quality over the past 20 years.

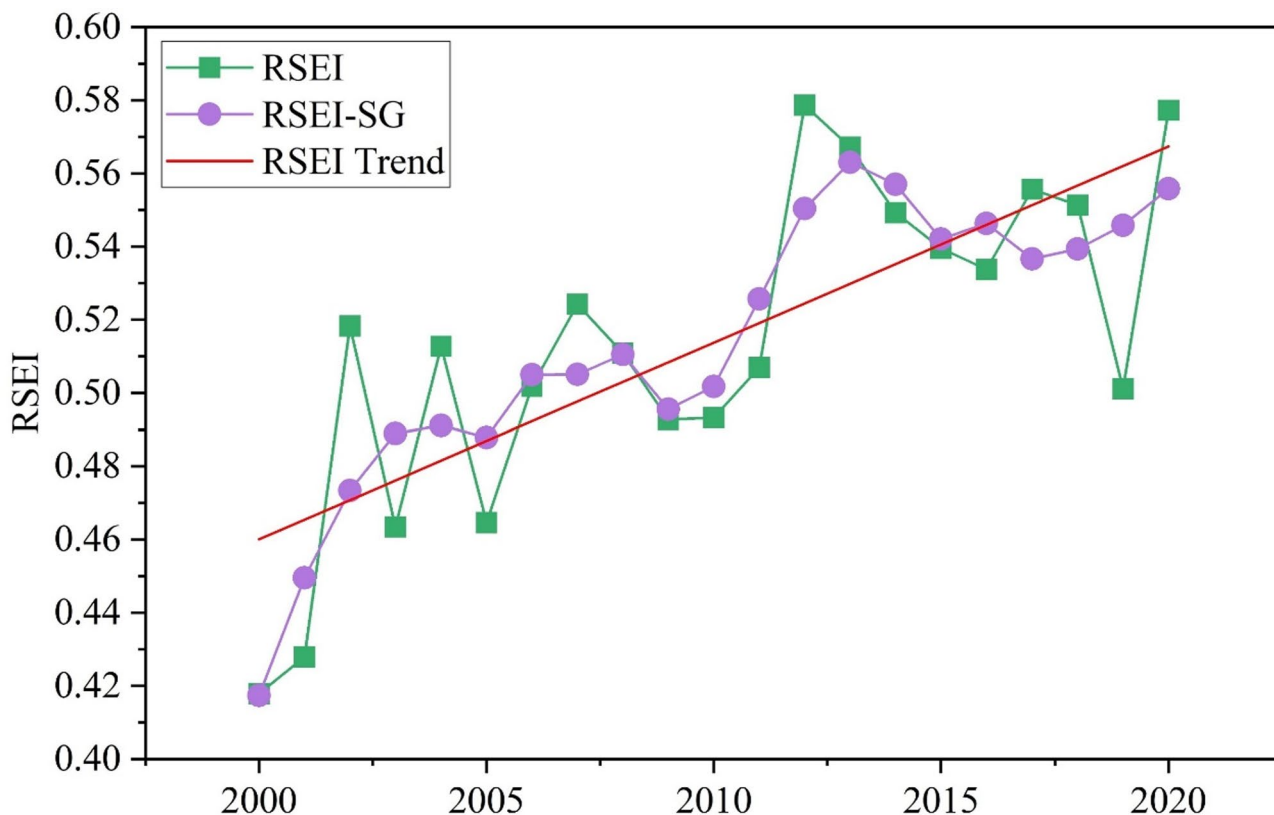


Fig. 4. Time-series changes of the mean RSEI value in Shanxi Province from 2000 to 2020.

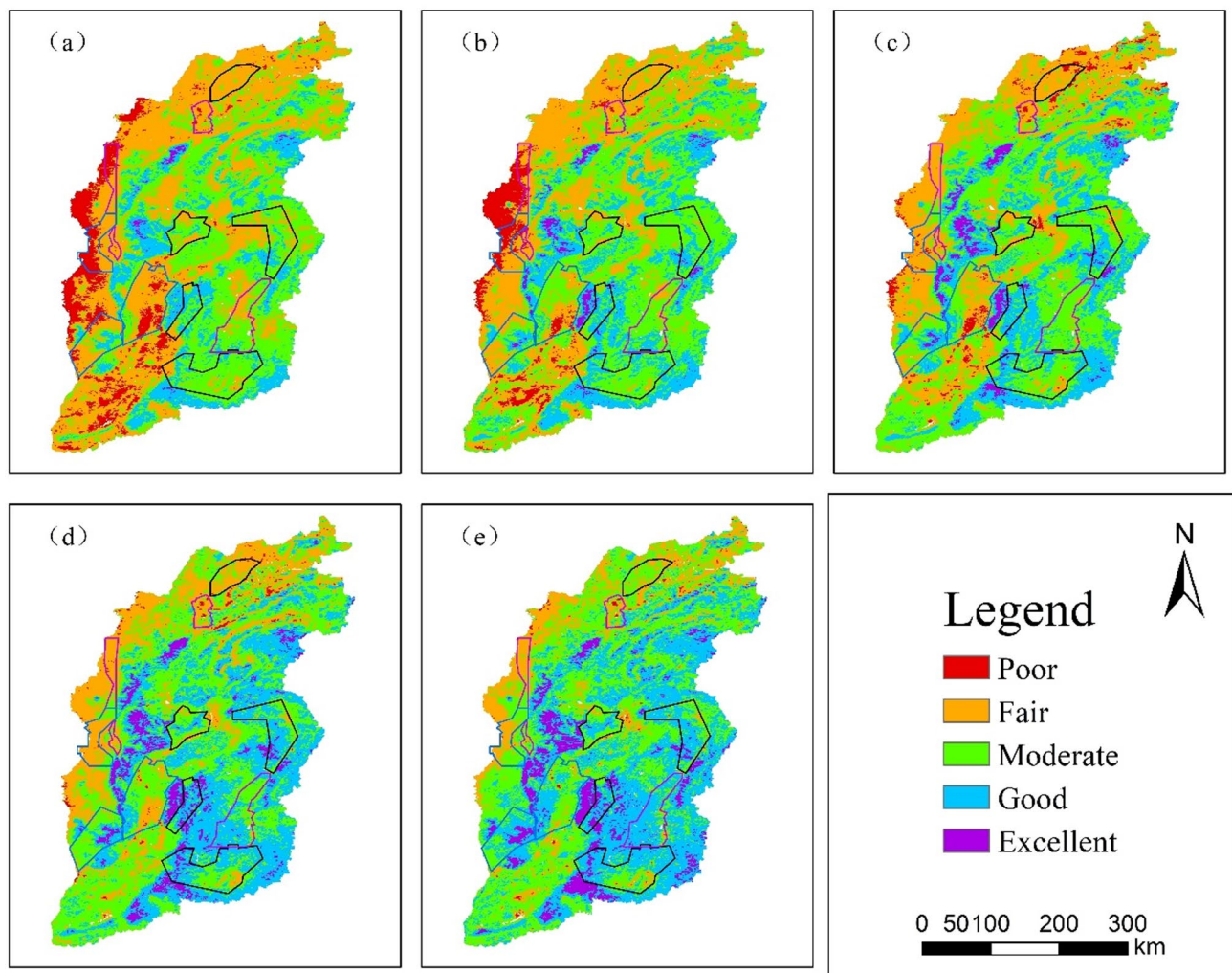


Fig. 5. (a) to (e) represent the spatial and temporal distribution of RSEI in Shanxi Province and planned mining areas in 2000, 2005, 2010, 2015 and 2020, respectively. (This image was created using ArcGIS 10.8 software, which is an Esri product, generated using our dataset, and is publicly available at the URL (<https://www.esri.com/>).)

Analysis of the annual mean RSEI box plots for the twelve planned mining areas (Fig. 7) reveals that the Jincheng mining area demonstrates relatively better ecological quality with a mean RSEI value of 0.32 compared to other mining areas. In contrast, the Liulin mining area exhibits poorer ecological conditions, showing the lowest mean RSEI value of 0.16. The overall ecological assessment indicates that the average RSEI grades across all twelve planned mining areas fall within the “fair” to “poor” categories, highlighting the necessity for implementing ecological protection measures in these regions.

To better understand the trends in ecosystem changes, an assessment and comparison of the ecological conditions across different mining areas were conducted, categorizing the annual growth rates into three Levels based on their values: Level 1 for annual growth rates exceeding 2.5%; Level 2 for rates between 1.5% and 2.5%; and Level 3 for rates ranging from 0 to 1.5%. Among these, the Xiangning, Liulin, and Huozhou mining areas were classified as Level 1, with growth rates of 2.74% and 3.35%. The Pingshuo Shuo’nan, Hebao Piao, Lu’an, and Lishi mining areas fell into Level 2, with growth rates of 1.55%, 2.36%, 1.80%, and 2.29%, respectively. The Datong, Jincheng, Qinyuan, Xishan Gujiao, and Yangquan mining areas were categorized as Level 3, with growth rates of 1.48%, 1.17%, 1.42%, 1.17%, and 1.34%. Overall, the annual average growth rates across the planned mining areas indicate that from 2000 to 2020, the RSEI values in all mining areas displayed an upward trend, reflecting improvements in their ecological conditions.

Spatial and Temporal trends in RSEI

The RSEI values in Shanxi Province exhibit significant spatial variability, with the analysis utilizing Sen’s slope and the MK test to explore the temporal and spatial trends of RSEI, resulting in five distinct groups. As illustrated in Fig. 8, from 2000 to 2010, the overall RSEI in Shanxi Province demonstrated an upward trend, with the area of increase accounting for 79.47% of the total area. Notably, the majority of this increase was no significant rise,

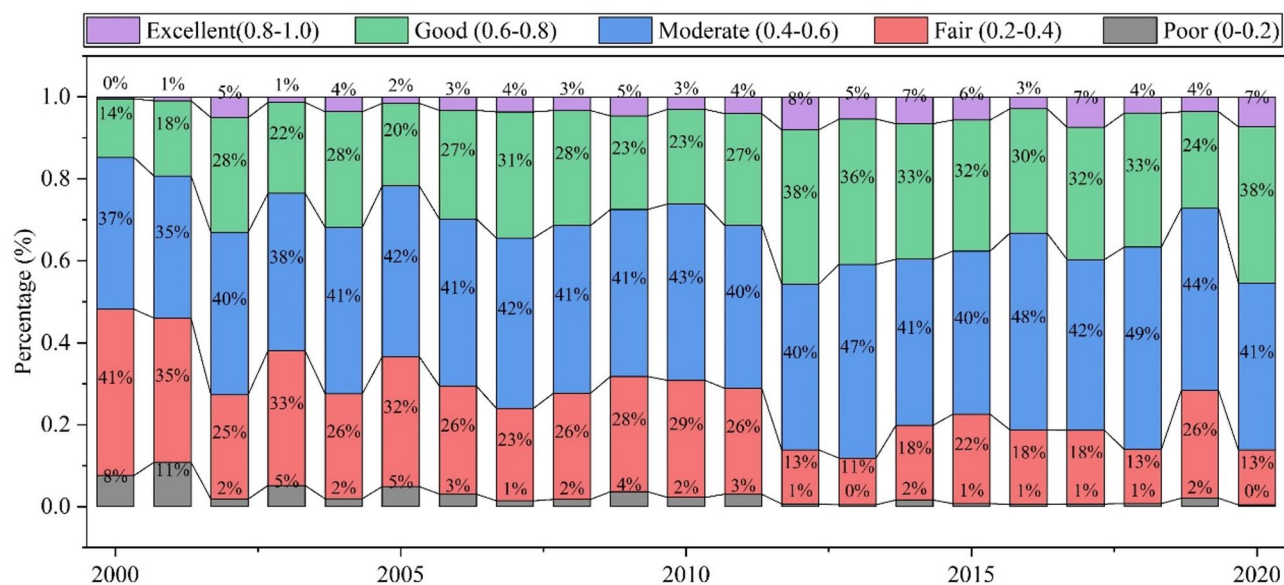


Fig. 6. Percentage change of area occupied by different classes of RSEI in Shanxi Province.

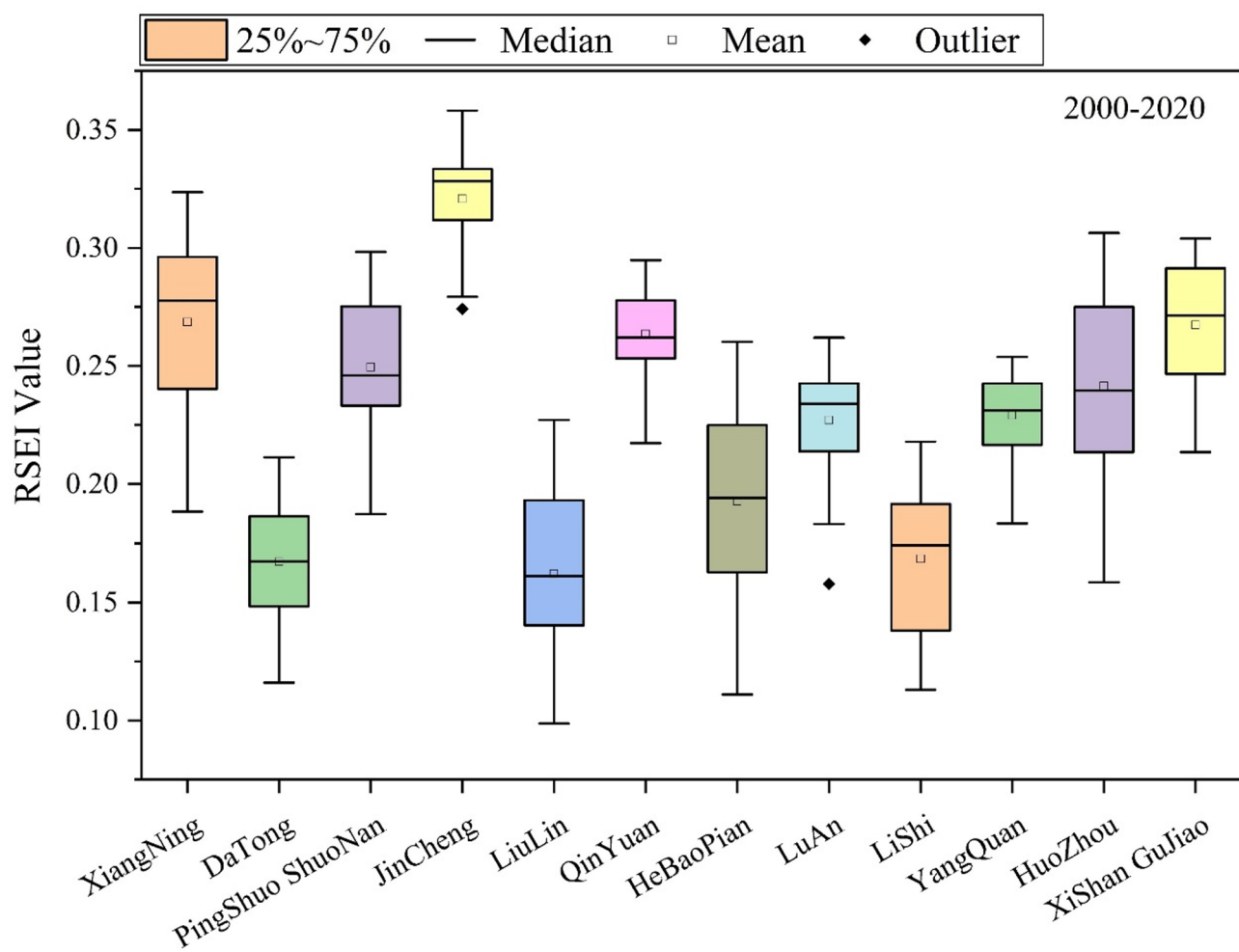


Fig. 7. Annual average change of RSEI in the planned mining area.

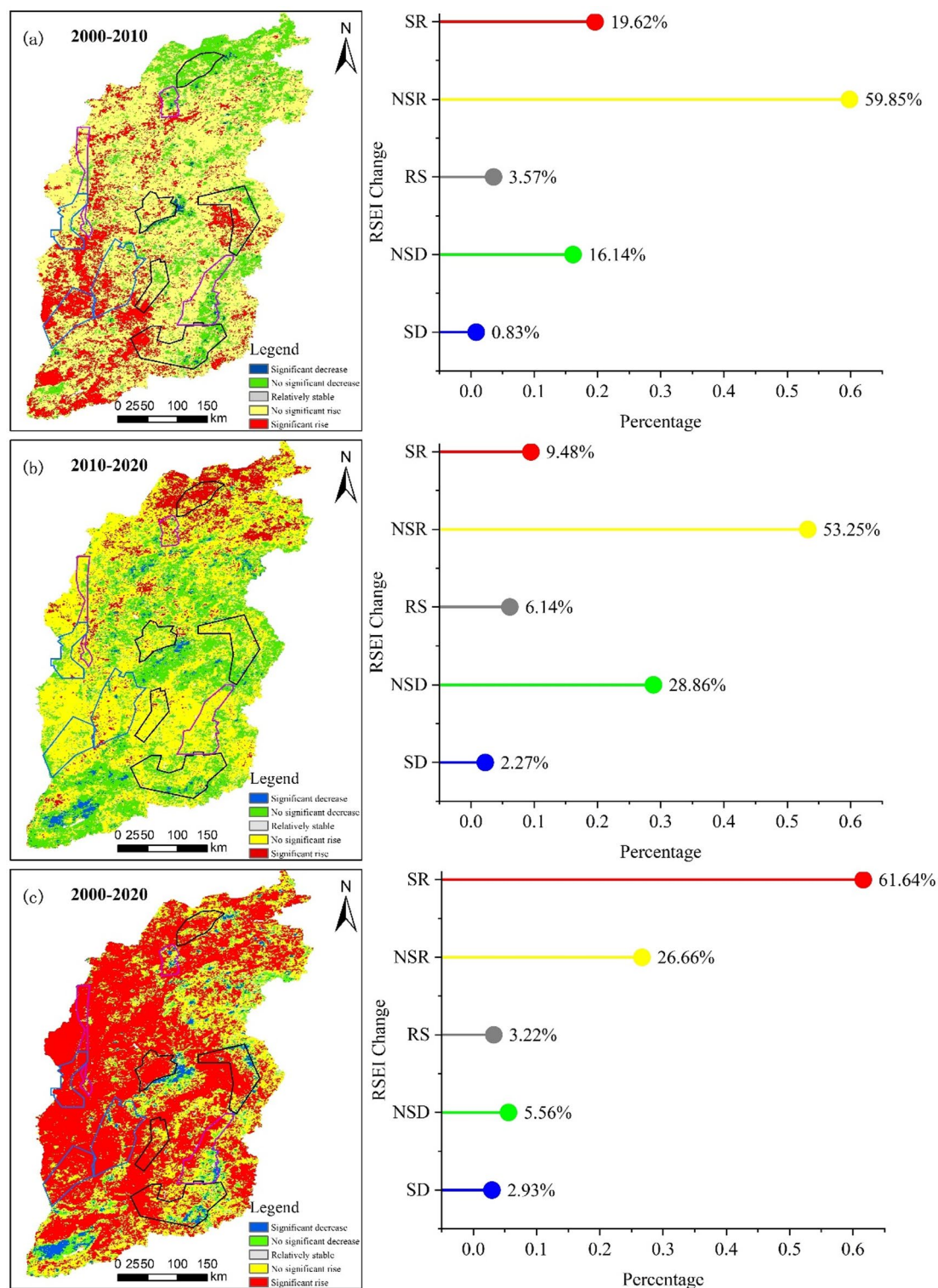


Fig. 8. (a–c) Spatial distribution and area proportions of RSEI trends for 2000–2010, 2010–2020, and 2000–2020, respectively. (This image was created using ArcGIS 10.8 software, which is an Esri product, generated using our dataset, and is publicly available at the URL (<https://www.esri.com/>).)

comprising 59.85%. From 2010 to 2020, the area of increase constituted 62.73%, with a significant rise of 9.48% primarily concentrated in the northern part of Shanxi Province. Over the entire period from 2000 to 2020, the area exhibiting an upward trend accounted for 88.3%, while the area showing a downward trend comprised 8.49%, primarily distributed in the southern and eastern regions. These results indicate a general upward

trajectory in ecological quality across Shanxi Province, reflecting the significant effectiveness of ecological governance measures implemented in the region.

From 2000 to 2020, the area with increased RSEI in the 12 planned mining areas accounted for 87.08%, while the area that relative stable was only 2.89%, and the area with decreased RSEI for 10.03%. The spatiotemporal trends of RSEI in the planned mining areas align closely with the overall trends observed in Shanxi Province. The varying spatial trends of RSEI across different grades of mining areas and their respective proportions are detailed in Table 3. Overall, all grades of mining areas in Shanxi Province demonstrate an upward trend in RSEI, indicating a reduction in ecological vulnerability and an improvement in ecological conditions. Notably, the ecological improvement in Level 1 mining areas is particularly pronounced, significantly surpassing that of Level 2 and Level 3 areas. This suggests that ecological restoration efforts in Level 1 mining areas, such as those in Xiangning, Liulin, and Huozhou, have achieved more substantial results. This trend corresponds with the broader improvement in ecological quality throughout Shanxi Province, reflecting the positive outcomes of environmental protection and restoration initiatives in the region. Since 2007, Shanxi Province has implemented a special plan for the comprehensive management of subsidence areas, gradually restoring the ecological functions of the land in these areas through integrated measures such as surface backfilling, land reclamation, and ecological reconstruction. Additionally, between 2000 and 2020, the province advanced ecological restoration efforts for abandoned open-pit mines in key areas of the Yellow River Basin.

Spatial autocorrelation analysis of the RSEI

Spatial autocorrelation analysis was employed to explore the ecological quality in Shanxi Province by sampling RSEI into a 5 km × 5 km grid. The results, validated at a 1% significance level, revealed significant spatial autocorrelation of RSEI at the 99% confidence level, indicating a clear spatial correlation in ecological quality across various regions of Shanxi Province. The Moran’s I values for the years 2000, 2005, 2010, 2015, and 2020 were 0.78, 0.75, 0.74, 0.74, and 0.70, respectively (Fig. 9), demonstrating a gradual weakening of spatial autocorrelation in ecological quality, but the values remain relatively high, suggesting a strong spatial correlation still exists. The Moran scatter plot delineates spatial clustering or dispersion patterns across four quadrants: HH (high-high), LH (low-high), LL (low-low), and HL (high-low). The results indicate that RSEI’s Moran’s I scatter plot is predominantly concentrated in the HH and LL quadrants, highlighting a substantial positive spatial correlation in ecological quality across Shanxi Province and suggesting that the spatial distribution of RSEI exhibits a degree of aggregation.

To analyze the spatiotemporal variation in the distribution of RSEI within the study area, a Local Indicators of Spatial Association (LISA) cluster map was generated. The LISA clustering results (Fig. 10) demonstrate a clear spatial correlation between ecological quality patterns and mining activities, climatic conditions, and topographic characteristics in Shanxi Province. H-H clusters are predominantly distributed in the central and southeastern regions (e.g., Yuncheng and Jincheng), where higher annual precipitation and the ecological barrier effects of the Taihang and Lüliang Mountains limit mining disturbances. These areas also benefit from fertile river valleys such as the Fenhe and Qinhe, which support vegetation growth.

In contrast, L-L clusters are concentrated in northern coal mining hubs (Datong, Shuozhou) and western mining zones (Lüliang, Liulin, Hebao Pian), overlapping significantly with the open-pit mining areas outlined in the Shanxi Mineral Resources Plan. The temporal analysis from 2000 to 2020 reveals that H-H clusters have expanded and spatially consolidated, while L-L clusters have shifted northward and decreased in area. This trend corresponds with the implementation of ecological restoration projects in the Yellow River Basin, including vegetation planting and mining land rehabilitation. Notably, western mining areas like Liulin and Hebao Pian

Mine grade	RSEI	Trends RSEI	2000–2020	
			Land area (km ²)	Percentage
Level I	Rise	SR	9722.99	84.57
		NSR	1339.99	11.66
	Relative stable	RS	107.73	0.94
		NSD	209.66	1.82
		SD	116.02	1.01
Level II	Rise	SR	3394.31	52.97
		NSR	1822.28	28.44
	Relative stable	RS	248.61	3.88
		NSD	624.83	9.75
		SD	318.22	4.97
Level III	Rise	SR	8664.76	54.63
		NSR	4592.59	28.96
	Relative stable	RS	612.40	3.86
		NSD	1200.77	7.57
		SD	788.91	4.97

Table 3. Area of different Spatial trends in RSEI at various levels of the mine and their percentages.

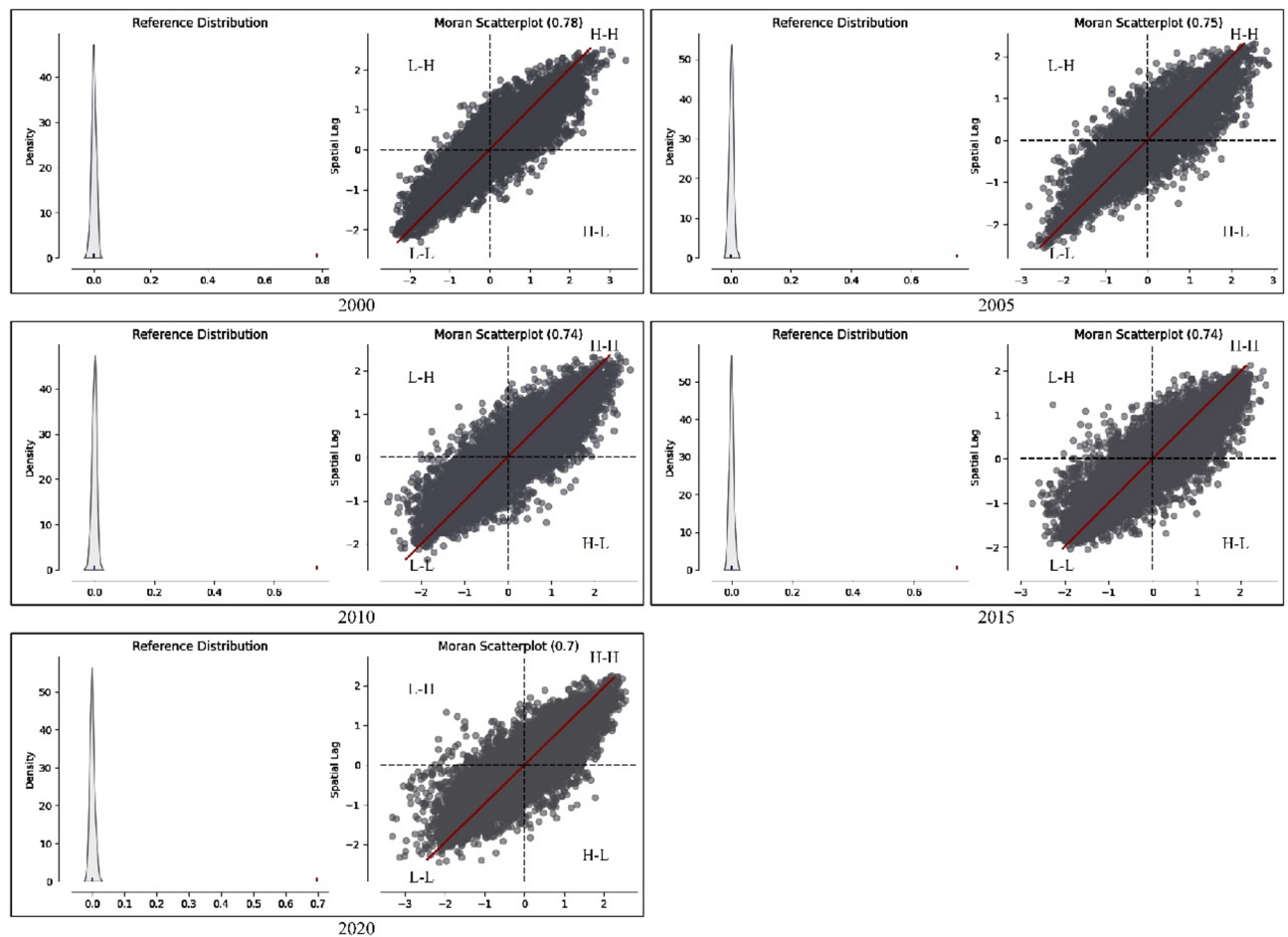


Fig. 9. 2000, 2005, 2010, 2015 and 2020 Moran I scatter plots for the study area.

show reduced L-L clusters, whereas regions such as Xiangning and Jincheng exhibit increased H-H clusters, reflecting the combined effects of mining regulation, climate advantages, and terrain protection.

Potential drivers of change affecting ecological quality

To comprehensively and systematically investigate the driving factors influencing RSEI, various potential drivers were treated as independent variables, with RSEI as the dependent variable. Due to data availability constraints—some datasets being updated only through 2020—five representative years (2000, 2005, 2010, 2015, and 2020) were selected for analysis. The Pearson correlation coefficient method was used to evaluate the linear relationships between RSEI and the selected driving factors for each of these years. The results, validated through significance testing, identify the key factors influencing RSEI, as presented in Table 4. At the same time, taking 2020 as an example, we analyzed the statistical characteristics of each variable (Table S2). The changes in ecological quality within mining areas are also affected by factors such as elevation, wind speed (Ws), Precipitation accumulation (Pa), Vapor pressure (Vp) and NPP. However, in studies of ecological quality changes in mining areas, human activities—particularly coal mining—have a more pronounced impact on the ecological quality of these regions. Therefore, when investigating the potential influencing factors of ecological quality changes, it is essential to consider the overall context of Shanxi Province. A systematic study of Shanxi Province can enhance the understanding of the impacts of coal mining and provide effective strategies for ecological restoration in mining areas, ultimately achieving more efficient results.

Results of driver regression based on catboost and GWR

Table 5 presents the average results of the CatBoost model using 5-fold cross-validation. The model exhibited a high coefficient of determination ($R^2 > 0.670$) and a low root mean square error ($RMSE < 0.083$), highlighting CatBoost's ability to efficiently and accurately predict RSEI evaluation metrics. These results confirm the model's effectiveness in handling the complex relationships between ecological factors and RSEI.

The analysis of the CatBoost model's prediction results using the SHAP tool (Fig. 11) quantitatively assessed the contribution of each feature to the model's outputs. While traditional models can indicate which features are important, they often fail to explain how these features influence the predictions. The key advantage of the SHAP approach lies in its ability to reveal the impact of each feature on a per-sample basis, including whether the effect is positive or negative. The results show that population density exerts a negative influence on RSEI, whereas

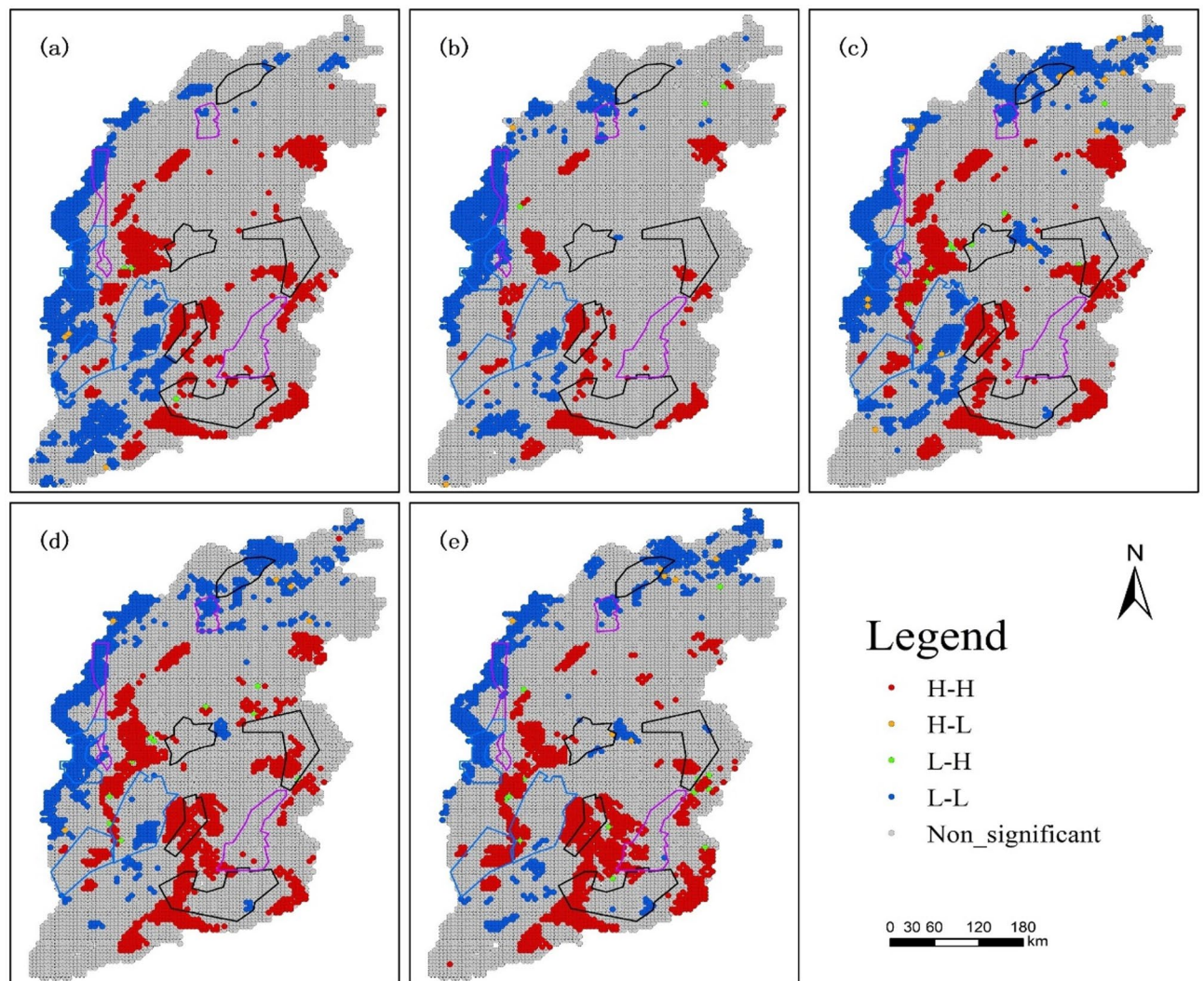


Fig. 10. (a–e) shows the LISA clusters for the study area for the years 2000, 2005, 2010, 2015, and 2020. (This image was created using ArcGIS 10.8 software, which is an Esri product, generated using our dataset, and is publicly available at the URL (<https://www.esri.com/>).)

the other variables have positive effects. Moreover, the analysis highlights the varying degrees to which different features impact RSEI: NPP has a substantial effect, with lower NPP values resulting in smaller model outputs, while higher NPP values lead to larger outputs. Parameters such as Pa, elevation, wind speed, population density, and vapor pressure have moderate impacts on RSEI, whereas slope has a relatively minor effect.

To capture the temporal variation and spatial heterogeneity in the effects of various factors on RSEI, both GWR and OLS models were constructed, and the fitted parameter results for the years 2000, 2005, 2010, 2015, and 2020 were compared and a map of localized R^2 coefficients (Table 6). The results indicated that the GWR model significantly outperformed both the OLS and CatBoost models in terms of RMSE and R^2 metrics. Furthermore, the AICc value for the GWR model was notably lower than that of the OLS model, suggesting that the GWR model can more accurately fit the data when addressing spatially heterogeneous data. By comparing the global R^2 and local R^2 coefficients (Fig. 12), while the global R^2 is usually higher, indicating a better overall model fit, the variation in the local R^2 reveals differences in fit across different regions or sample points. This is crucial for understanding the spatial heterogeneity of the data. This reflects its superior sensitivity and adaptability to spatial variations, further validating its unique advantage in predicting RSEI within the context of spatial correlation.

To examine the spatial and temporal differences in the coefficients between independent and dependent variables in the GWR model, the average coefficients for different years are presented in Table 7. By analyzing these average coefficients, we can reveal the temporal and spatial trends of various environmental factors in relation to RSEI. The positive and negative values of the coefficients indicate the nature of the influence that each driving factor has on RSEI. In all years, elevation, Ws, Vp, and NPP showed a positive correlation with RSEI, while population density exhibited a negative correlation. The absolute values of the coefficients reflect the magnitude of the influence each driving factor has on RSEI, indicating that NPP and population density have a significant impact, whereas slope has a relatively minor effect.

Year	Variables	Mean	Std	Max	Mix	Median
2000	Elevation (m) **	1166.51	361.93	2560.00	290.00	1151.00
	Slope(°)**	14.85	10.99	78.17	0.96	12.69
	Wind speed (m/s) **	210.99	28.15	335.50	151.50	206.50
	Precipitation accumulation (mm) **	85.51	18.17	151.00	53.00	81.00
	Vapor pressure (kPa) **	1411.07	210.88	2058.50	959.25	1397.75
	NPP(kg C/m2)**	3381.19	4264.05	32767.00	301.00	2794.00
	Population density(people/km ²)**	126.17	700.25	28565.00	1.00	28.00
	RSEI	0.42	0.16	0.96	0.04	0.41
2005	Elevation (m) **	1163.17	360.54	2560.00	290.00	1147.00
	Slope(°)**	14.76	10.90	78.17	0.96	12.45
	Wind speed (m/s) **	219.66	27.91	341.00	151.50	216.00
	Precipitation accumulation (mm) **	84.38	22.45	155.50	50.75	78.00
	Vapor pressure (kPa)	1329.07	186.24	1918.00	922.50	1312.00
	NPP(kg C/m2)**	4551.39	4080.77	32767.00	829.00	4082.00
	Population density(people/km ²)**	137.17	622.20	21374.00	1.00	20.00
	RSEI	0.46	0.16	0.94	0.00	0.46
2010	Elevation (m) **	1160.23	359.71	2560.00	290.00	1144.00
	Slope(°)**	14.65	10.82	78.17	0.96	12.35
	Wind speed (m/s) **	214.09	28.16	335.25	151.75	209.50
	Precipitation accumulation (mm) **	93.05	17.15	150.00	59.50	90.75
	Vapor pressure (kPa) **	1392.91	210.46	2086.75	954.25	1382.00
	NPP(kg C/m2)*	3944.63	4139.76	32767.00	978.00	3424.00
	Population density(people/km ²)**	139.56	669.83	23542.00	1.00	18.00
	RSEI	0.49	0.16	0.95	0.05	0.49
2015	Elevation (m) **	1160.65	359.63	2560.00	290.00	1145.00
	Slope(°)**	14.65	10.83	78.17	0.96	12.35
	Wind speed (m/s) **	226.77	26.99	342.25	164.50	223.00
	Precipitation accumulation (mm) **	71.58	11.13	110.75	50.00	69.50
	Vapor pressure (kPa) **	1270.79	186.19	1849.75	868.00	1254.50
	NPP(kg C/m2)**	4351.45	4109.18	32767.00	945.00	3868.00
	Population density(people/km ²)**	148.24	712.48	22910.00	1.00	18.00
	RSEI	0.54	0.16	0.96	0.09	0.54
2020	Elevation (m) **	1155.75	357.55	2560.00	290.00	1139.00
	Slope(°)**	14.49	10.71	78.17	0.96	12.31
	Wind speed (m/s) **	218.49	28.93	340.00	147.50	215.00
	Precipitation accumulation (mm) **	86.27	9.88	120.25	56.50	85.75
	Vapor pressure (kPa) *	1428.38	187.56	2029.75	1028.25	1414.25
	NPP(kg C/m2)*	4348.49	4125.43	32767.00	945.00	3858.00
	Population density(people/km ²)**	147.52	573.86	17263.00	1.00	16.00
	RSEI	0.58	0.15	0.97	0.07	0.58

Table 4. Significance test results of the main drivers of RSEI. **, * represent passing the 1%, 5% significance level, respectively.

Year	R ²	RMSE
2000	0.786 ± 0.010	0.074 ± 0.002
2005	0.770 ± 0.020	0.077 ± 0.004
2010	0.804 ± 0.015	0.073 ± 0.003
2015	0.786 ± 0.011	0.076 ± 0.002
2020	0.706 ± 0.018	0.082 ± 0.001

Table 5. Evaluation results of catboost regression model.

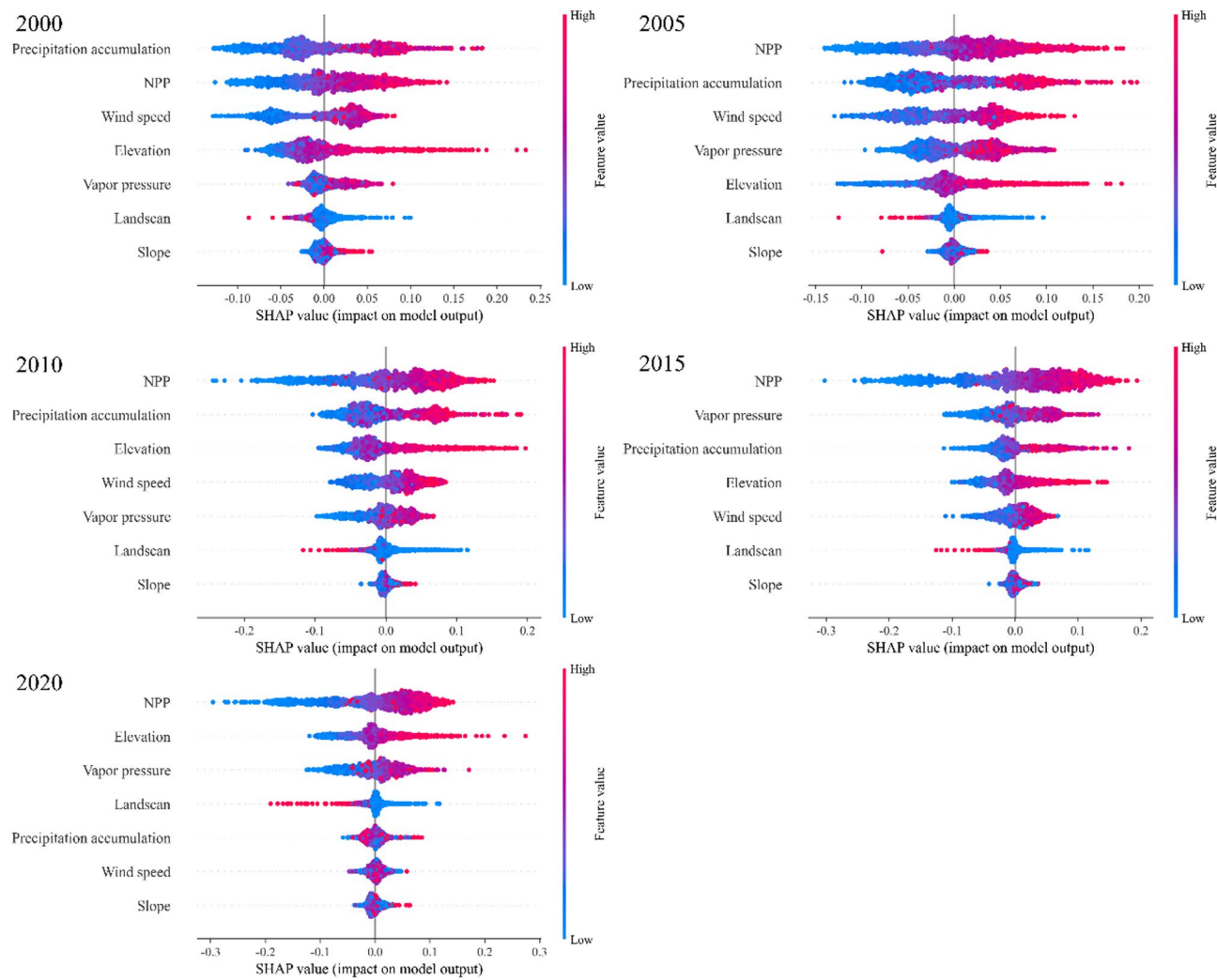


Fig. 11. Scatter plot of SHAP feature density.

Year	R ²		R ² Adjusted		RMSE	AICc		
	GWR	OLS	GWR	OLS		GWR	OLS	GWR
2000	0.887	0.487	0.855	0.487	0.0558	0.1189	-6257.9	-3562.4
2005	0.875	0.386	0.839	0.386	0.062	0.1372	-5425.4	-2435.7
2010	0.879	0.445	0.845	0.445	0.0597	0.1279	-5728	2983.94
2015	0.874	0.416	0.838	0.416	0.061	0.1313	-5556.4	2780.06
2020	0.813	0.36	0.778	0.360	0.0688	0.1274	5341.45	3012.8

Table 6. Model evaluation results of OLS and GWR.

The spatial distribution of the mean regression coefficients for various factors across different years is depicted in Fig. 13. It is evident from the figure that the relationships between elevation, slope, Ws, Pa, Vp, NPP, and population density with RSEI exhibit both temporal and spatial variations. The levels of influence these factors have on RSEI and their corresponding areas of impact differ, with the distribution of regression coefficients not appearing random. Notably, the coefficients for NPP and population density demonstrate a pronounced clustering effect over large areas, indicating that the influence of environmental factors on RSEI is related to spatial location. The GWR model further confirms that the impact of different factors on the ecological environment varies spatially, while within certain ranges, the same influencing factors consistently affect RSEI.

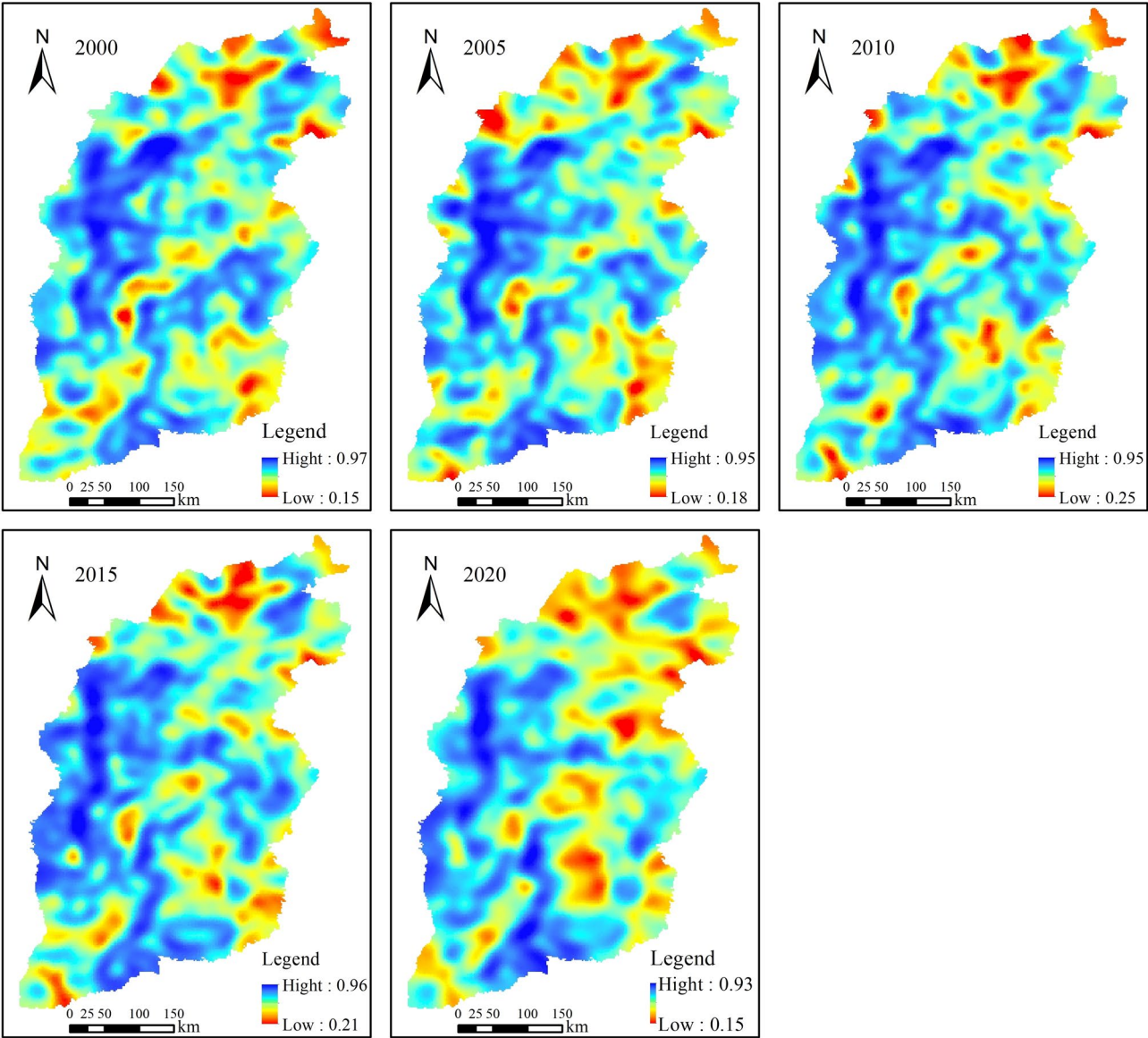


Fig. 12. Local R^2 Coefficient Map. (This image was created using ArcGIS 10.8 software, which is an Esri product, generated using our dataset, and is publicly available at the URL (<https://www.esri.com/>).)

Variables	2000	2005	2010	2015	2020
Elevation	0.372	0.345	0.424	0.314	0.319
Slope	0.003	-0.003	0.003	-0.022	-0.012
Wind speed	0.488	0.332	0.342	0.384	0.320
Precipitation accumulation	0.133	0.074	0.243	0.061	-0.040
Vapor pressure	0.378	0.338	0.428	0.410	0.279
NPP	1.416	1.236	1.690	1.576	1.184
population density	-0.808	-0.414	-0.607	-0.701	-0.714

Table 7. Mean values of regression coefficients of independent variables in the GWR model.

Discussion
Changes in ecological environment quality

Based on long-term RSEI monitoring (2000–2020), the ecological quality of Shanxi Province showed a significant upward trend (38% increase in mean RSEI), with particularly prominent improvements in planned mining areas. This conclusion aligns with studies by Zhao et al. (2019)³⁸, Gong et al.³⁹, and Zhang et al.⁴⁰, all indicating

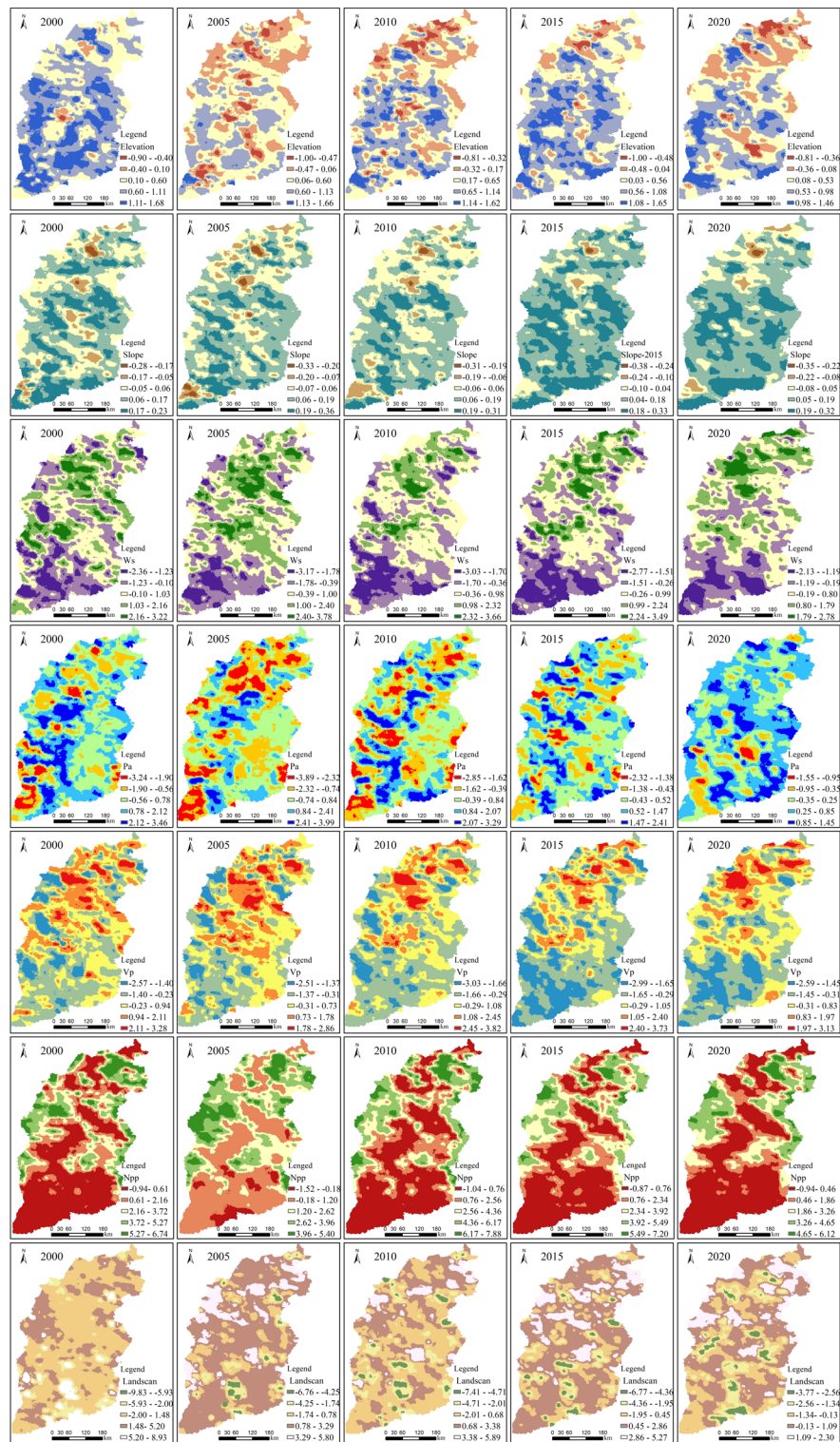


Fig. 13. Spatial distribution of regression coefficients of independent variables of GWR model. (This image was created using ArcGIS 10.8 software, which is an Esri product, generated using our dataset, and is publicly available at the URL (<https://www.esri.com/>).)

that ecological restoration policies—including mine rehabilitation, soil erosion control, and afforestation—have driven quality improvements in Shanxi. For example, Gong et al. reported that post-2000 reclamation projects increased vegetation coverage by approximately 21% in mining areas, consistent with the sustained RSEI improvement observed in this study (87.08% of mining areas showed upward trends)³⁹. However, Xu et al. emphasized a 5–8-year lag in ecological recovery within mining areas, contrasting with the long-term positive

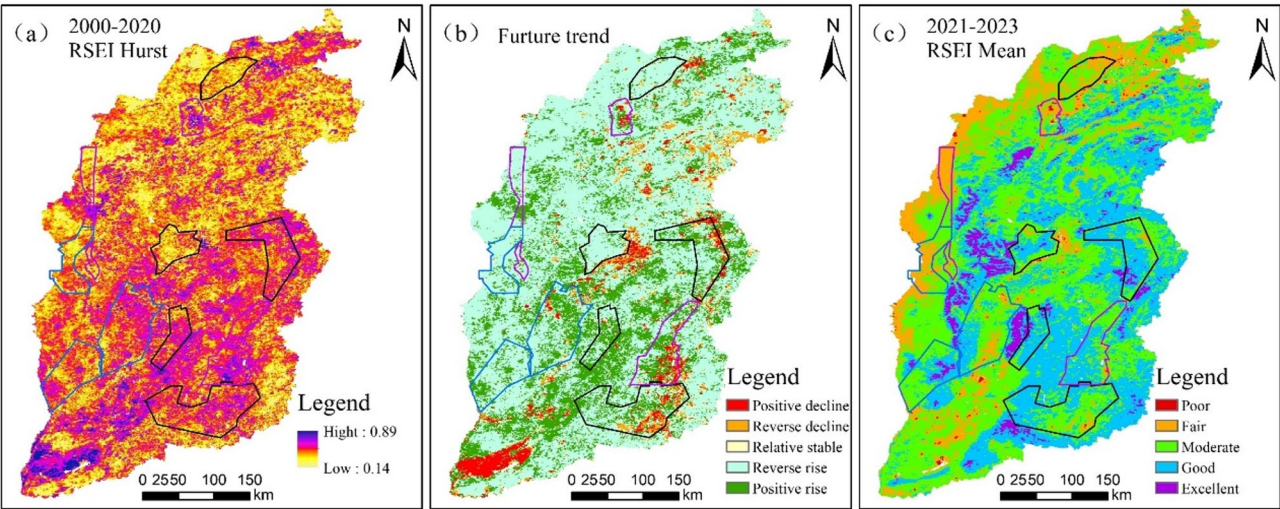


Fig. 14. Future RSEI Changes in Shanxi Province. (This image was created using ArcGIS 10.8 software, which is an Esri product, generated using our dataset, and is publicly available at the URL (<https://www.esri.com/>).)

Future trends	Shanxi Province	Diggings
	Percentage	Percentage
Positive decline	4.38	5.48
Reverse decline	4.11	4.17
Relative stable	3.22	2.87
Reverse rise	63.12	58.38
Positive rise	25.17	29.10

Table 8. Changes in RSEI sustainability grading for Shanxi Province and its planned mining areas.

RSEI	Percentage	Land area(km ²)
Poor(0–0.2)	–0.39	636.75
Fair(0.2–0.4)	–2.35	3842.59
Moderate(0.4–0.6)	–0.63	1030.27
Good(0.6–0.8)	3.99	6508.943
Excellent(0.8–1)	–0.65	1064.64

Table 9. Change in average RSEI, 2021–2023 vs. average RSEI, 2018–2020.

trends identified here⁴¹. This discrepancy may stem from differences in study duration and methodology: Xu et al. relied on short-term observations (<10 years), reflecting initial restoration fluctuations, whereas this study captured cumulative recovery effects through 20-year temporal analysis.

To forecast future RSEI trends, the Hurst exponent was employed, yielding values between 0.14 and 0.89 (mean: 0.46) for Shanxi (Fig. 14a), indicating a likelihood of trend reversals. The overlay analysis of RSEI spatial trends and Hurst index (Fig. 14b) demonstrates that approximately 63.12% of Shanxi Province is likely to experience “reverse rise” (early degradation followed by future improvement) according to statistical results (Table 8), a phenomenon consistent with the “delayed recovery” pattern observed in global mining restoration studies⁴². Despite the overall upward trend in ecological quality, localized degradation remains evident in areas of intensive mining, such as Pingshuo-Shuonan (18.43% showing a declining trend) and Jincheng (19.73%), consistent with findings by Li et al. in northern China³⁰. Validation using RSEI data from 2021 to 2023 showed over 87% consistency with Hurst-based predictions, supporting the reliability of this method and aligning with similar applications by Kang et al.⁴³ on the Loess Plateau. These findings highlight the importance of balancing coal extraction with ecological management to ensure sustained long-term environmental quality.

Figure 14c illustrates the RSEI mean distribution from 2021 to 2023, using the average RSEI values from these three years to validate the accuracy of future trend changes against the averages from 2018 to 2020. This method visually represents the variations between the two time periods and effectively mitigates the influence of short-term fluctuations in long-term trend analysis. The results, as shown in Table 9, reveal that RSEI values for

poor, fair, and moderate grades declined by 0.39%, 2.35%, and 0.63%, respectively, while RSEI for good grades increased by 3.99% and for excellent grades decreased by 0.65%. Overall, the RSEI exhibits an upward trend, consistent with the predictions made by the Hurst index.

Driving factors of ecological quality

As a major coal-producing province, Shanxi's ecological environment is highly sensitive to human activities and vegetation dynamics. The CatBoost model identified NPP as the primary driver of RSEI improvement (32.7% contribution), consistent with Gou et al.'s findings in the Loess Plateau⁴⁴, which highlight vegetation restoration as critical in arid/semi-arid regions. However, this study found that the negative effect of population density on RSEI is more significant than that reported by Xu et al.⁴⁵ in the Qinghai-Tibet-Loess Plateau transition zone (with a difference in SHAP values reaching 1.8 times). This divergence reflects Shanxi's unique resource-dependent context: intensive mining and topographic constraints (e.g., >70% mountainous terrain) likely amplify population pressure on ecosystems⁴⁶.

The combination of the CatBoost model and SHAP tool effectively quantified the influence of various factors on RSEI and intuitively illustrated how each feature affects RSEI, thereby overcoming the limitations of traditional models. Although the CatBoost model demonstrated excellent predictive performance, it did not adequately account for spatial data and geographic relationships, overlooking the spatial autocorrelation present in Shanxi Province. To address the fitting errors caused by spatial autocorrelation, this study introduced the GWR model to analyze the impacts of potential ecological factors on RSEI in Shanxi Province. The GWR model not only confirmed the magnitude and direction of each factor's influence on RSEI but also clearly revealed the spatial heterogeneity of regression coefficients through the spatial distribution maps of GWR coefficients (Fig. 13). The results of the GWR model further confirmed that the effects of ecological factors on RSEI vary across different regions. Moreover, the GWR results showed that the NPP coefficients were relatively low in the southern part of Shanxi Province and the Lüliang Mountain area. This is mainly because the southern region of Shanxi has experienced a faster pace of urbanization and has relatively lower vegetation coverage, making NPP primarily influenced by vegetation cover and topography⁴⁰. This result is consistent with actual conditions and verifies the relationship between NPP and RSEI. The combined findings from the CatBoost and GWR models can assist policymakers in formulating targeted policy measures that take into account the specific needs and conditions of different regions within Shanxi Province, thereby promoting more effective ecological management and restoration efforts.

Limitations and future prospects

This study utilized remote sensing data to evaluate ecological quality trends and drivers in Shanxi, providing a scientific basis for environmental assessment. However, compared to high-precision monitoring studies (e.g.^{47,48}), limitations exist. First, reliance on MODIS imagery—despite its high temporal resolution—limited spatial detail, potentially underestimating micro-scale changes in mining areas. Second, in future studies, we plan to introduce spatial downscaling models and integrate high-resolution data such as Landsat to improve the representation of localized regions. While this investigation considered seven primary influencing factors (elevation, slope, W_s , P_a , V_p , NPP, and population density), ecological quality is fundamentally determined by a more complex interplay of numerous elements. Future studies would benefit from incorporating additional scientifically relevant variables, such as GDP, industrial activity indicators, and energy consumption patterns⁴⁰, to enable a more comprehensive evaluation of their synergistic effects on ecosystem health. As a major coal resource province, incorporating coal mining resource data into ecological evaluation metrics can provide new perspectives. Specifically, the study can reveal the relationship between mining activities and ecological quality by assessing the direct impacts of coal mining on the surrounding ecological environment, such as changes in vegetation coverage and increased soil erosion. Furthermore, by examining the intensity and distribution characteristics of coal mining resources, a more precise analysis of their long-term impacts on regional ecological environments can be achieved, particularly in the context of ecological restoration and management, evaluating the balance between mining and ecological recovery.

Conclusion

This study systematically evaluated the spatiotemporal changes in ecological environment quality and their driving mechanisms in Shanxi Province and its planned mining areas from 2000 to 2020 by constructing a RSEI, integrated with trend analysis, spatial autocorrelation, and machine learning models. The results demonstrate a significant improvement in the ecological quality of Shanxi Province, with the mean RSEI increasing from 0.42 in 2000 to 0.58 in 2020, representing a 38% enhancement—significantly higher than the national average vegetation restoration rate during the same period. Notably, 88.3% of the study area exhibited sustained ecological improvement, while only 8.49% showed degradation, highlighting the province's success in balancing resource exploitation and ecological protection. The planned mining areas also exhibited significant ecological recovery, with 87.08% of the regions showing upward trends, thereby validating the effectiveness of ecological restoration measures implemented in these zones. To further assess the persistence of ecological quality changes, this study integrated Hurst index analysis with validation using recent observational data (2021–2023), confirming a high level of consistency (>87%) with projected improvement trends. Spatial autocorrelation analysis, as indicated by a Global Moran's I exceeding 0.7, verified the presence of strong spatial clustering in ecological quality, thereby supporting subsequent spatially explicit analyses of driving factors. In examining the underlying mechanisms, the CatBoost model identified NPP as the most influential factor contributing to RSEI improvements (contribution rate: 32.7%). In contrast, population density showed a significant negative correlation with RSEI, highlighting the persistent anthropogenic pressure on regional ecosystems. GWR further revealed significant spatial heterogeneity in the ecological effects of different drivers, offering scientific guidance

for region-specific governance strategies. The proposed framework—integrating multi-source data with machine learning-driven interpretation in this study is not only applicable to ecological monitoring and management in resource-dependent regions but can also be extended to sustainability assessments of similar mining areas globally, such as Australia's Bowen Basin and Inner Mongolia's Shendong mining area. These findings provide quantitative support for optimizing the synergy between resource development and ecological conservation, aiding in the formulation of targeted ecological restoration policies.

Data availability

The data is available upon reasonable request to the corresponding authors.

Received: 14 December 2024; Accepted: 8 August 2025

Published online: 23 August 2025

References

- Wang, Y. P. et al. Exploring watershed ecological risk bundles based on ecosystem services: A case study of Shanxi Province China. *Environ. Res.* **245**, 118040 (2024).
- Guo, Y. et al. Study on the influence of mining disturbance on the variation characteristics of vegetation index: A case study of Lingwu Mining Area. *Environ. Dev.* **45**, 100811 (2023).
- Xiao, W. et al. Assessing the ecological impacts of opencast coal mining in Qinghai-Tibet Plateau—a case study in Muli coal field China. *Ecol. Indic.* **153**, 110454 (2023).
- Guan, Y. J. et al. Delimitation of supervision zones based on the soil property characteristics in a reclaimed opencast coal mine dump on the Loess Plateau China. *Sci. Total Environ.* **772**, 145006 (2021).
- Kalnay, E. & Cai, M. Impact of urbanization and land-use change on climate. *Nature* **423**(6939), 528–531 (2003).
- Boori, M. S. et al. Spatiotemporal ecological vulnerability analysis with statistical correlation based on satellite remote sensing in Samara Russia. *J. Environ. Manag.* **285**, 112138 (2021).
- Zhao, F. et al. Assessment of the sustainable development of rural minority settlements based on multidimensional data and geographical detector method: A case study in Dehong China. *Socio-econ. Plan. Sci.* **78**, 101066 (2021).
- Ariken, M. et al. Coupling coordination analysis of urbanization and eco-environment in Yanqi Basin based on multi-source remote sensing data. *Ecol. Indic.* **114**, 106331 (2020).
- Zhong, R. et al. Spatiotemporal variation of enhanced vegetation index in the Amazon Basin and its response to climate change. *Phys. Chem. Earth* **123**, 103024 (2021).
- Liu, C. et al. Spatiotemporal evolution of island ecological quality under different urban densities: A comparative analysis of Xiamen and Kinmen Islands, southeast China. *Ecol. Indic.* **124**, 107438 (2021).
- Li, Y. R. et al. Dynamic analysis of ecological environment combined with land cover and NDVI changes and implications for sustainable urban-rural development: The case of Mu Us Sandy Land China. *J. Clean. Prod.* **142**, 697–715 (2017).
- Ye, L. P. et al. Spatio-temporal dynamics of soil moisture driven by “Grain for Green” program on the Loess Plateau China. *Agric. Ecosyst. Environ.* **269**, 204–214 (2019).
- Xu, H. A remote sensing index for assessment of regional ecological changes. 2013. **33**: p. 889–897.
- Lv, Y. et al. Spatiotemporal evolution and driving factors analysis of the eco-quality in the Lanxi urban agglomeration. *Ecol. Ind.* **156**, 111114 (2023).
- Tang, H. et al. Impact of land cover change on a typical mining region and its ecological environment quality evaluation using remote sensing based ecological index (RSEI). *Sustainability* **14**(19), 12694 (2022).
- Wang, Q. et al. Estimation of vegetation traits with kernel NDVI. *ISPRS J. Photogram. Rem. Sens.* **195**, 408–417 (2023).
- Yang, H. et al. Long-time series ecological environment quality monitoring and cause analysis in the Dianchi Lake Basin China. *Ecol. Indic.* **148**, 110084 (2023).
- Yuan, B. D. et al. Spatiotemporal change detection of ecological quality and the associated affecting factors in Dongting Lake Basin, based on RSEI. *J. Clean. Prod.* **302**, 1026995 (2021).
- Zhang, X., Jia, W. & He, J. Spatial and temporal variation of ecological quality in northeastern China and analysis of influencing factors. *J. Clean. Prod.* **423**, 138650 (2023).
- Wang, F. et al. Succession of bacterial community composition in coastal agricultural soils along a 1000-year reclamation chronosequence in Hangzhou Bay China. *Ecol. Indic.* **121**, 106972 (2021).
- Shukla, A. K. et al. Population growth, land use and land cover transformations, and water quality nexus in the Upper Ganga River basin. *Hydrol. Earth Syst. Sci.* **22**(9), 4745–4770 (2018).
- Tian, P. et al. Impacts of reclamation derived land use changes on ecosystem services in a typical gulf of eastern China: A case study of Hangzhou bay. *Ecol. Indic.* **132**, 108259 (2021).
- Liu, P. et al. Exploring the ecological quality and its drivers based on annual remote sensing ecological index and multisource data in Northeast China. *Ecol. Ind.* **154**, 110589 (2023).
- Hancock, J. T. & Khoshgoftar, T. M. CatBoost for big data: an interdisciplinary review. *J. Big Data* **7**(1), 94 (2020).
- Chang, W. F. et al. An Improved CatBoost-Based Classification Model for Ecological Suitability of Blueberries. *Sensors* **23**(4), 1997 (2023).
- Jing, Y. Q. et al. Assessment of spatial and temporal variation of ecological environment quality in Ebinur Lake Wetland National Nature Reserve, Xinjiang China. *Ecol. Ind.* **110**, 105874 (2020).
- Dubois, E., Larocque, M. & Brunner, P. Impact of land cover changes on Long-Term Regional-Scale groundwater recharge simulation in cold and humid climates. *Hydrol. Processes* **37**(2), e14810 (2023).
- Lu, B. B. et al. Geographically weighted regression with parameter-specific distance metrics. *Int. J. Geogr. Inf. Sci.* **31**(5), 982–998 (2017).
- Liu, J. et al. Analysis of the heterogeneity of urban expansion landscape patterns and driving factors based on a combined Multi-Order Adjacency Index and Geodetector model. *Ecol. Indic.* **136**, 108655 (2022).
- Li, J. et al. Growing-season vegetation coverage patterns and driving factors in the China-Myanmar Economic Corridor based on Google Earth Engine and geographic detector. *Ecol. Indic.* **136**, 108620 (2022).
- Yang, X. Y. et al. Spatiotemporal change and driving factors of the Eco-Environment quality in the Yangtze River Basin from 2001 to 2019. *Ecol. Indic.* **131**, 108214 (2021).
- Xu, H. Q. et al. Detecting ecological changes with a remote sensing based ecological index (RSEI) produced time series and change vector analysis. *Rem. Sens.* **11**(20), 2345 (2019).
- Nyikadzino, B., Chitakira, M. & Muchuru, S. Rainfall and runoff trend analysis in the Limpopo river basin using the Mann Kendall statistic. *Phys. Chem. Earth* **117**, 102870 (2020).
- Yang, Z. et al. Assessment of eco-geo-environment quality using multivariate data: A case study in a coal mining area of Western China. *Ecol. Indic.* **107**, 105651 (2019).

35. Tran, T. V. et al. Characterising spatiotemporal vegetation variations using LANDSAT time-series and Hurst exponent index in the Mekong River Delta. *Land Degrad. Dev.* **32**(13), 3507–3523 (2021).
36. Bentéjac, C., Csörgö, A. & Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **54**(3), 1937–1967 (2021).
37. Wu, D. C. Spatially and temporally varying relationships between ecological footprint and influencing factors in China's provinces Using Geographically Weighted Regression (GWR). *J. Clean. Prod.* **261**, 121089 (2020).
38. Zhao, A. et al. Assessing the effects of drought and “Grain for Green” Program on vegetation dynamics in China's Loess Plateau from 2000 to 2014. *CATENA* **175**, 446–455 (2019).
39. Gong, C., Lyu, F. & Wang, Y. Spatiotemporal change and drivers of ecosystem quality in the Loess Plateau based on RSEI: A case study of Shanxi China. *Ecol. Indic.* **155**, 111060 (2023).
40. Zhang, J. et al. Dynamic monitoring of environmental quality in the loess plateau from 2000 to 2020 using the google earth engine platform and the remote sensing ecological index. *Rem. Sens.* **14**(20), 5094 (2022).
41. Xu, H. et al. A systematic review and comprehensive analysis on ecological restoration of mining areas in the arid region of China: Challenge, capability and reconsideration. *Ecol. Ind.* **154**, 110630 (2023).
42. Li, J. et al. Quantitative evaluation of ecological cumulative effect in mining area using a pixel-based time series model of ecosystem service value. *Ecol. Indic.* **120**, 106873 (2021).
43. Kang, S. Z. et al. Spatiotemporal variation and driving factors of ecological environment quality on the loess plateau in China from 2000 to 2020. *Rem. Sens.* **16**(24), 4778 (2024).
44. Gou, Y. et al. Elucidate the complex drivers of significant greening on the Loess Plateau from 2000 to 2020. *Environ. Dev.* **50**, 100991 (2024).
45. Xu, H. Q. et al. Prediction of ecological effects of potential population and impervious surface increases using a remote sensing based ecological index (RSEI). *Ecol. Ind.* **93**, 730–740 (2018).
46. Liu, L. et al. Analysis of land use change drivers and simulation of different future scenarios: Taking Shanxi Province of China as an example. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph20021626> (2023).
47. Lin, X. C. et al. Time-series simulation of alpine grassland cover using transferable stacking deep learning and multisource remote sensing data in the Google Earth Engine. *Int. J. Appl. Earth Observ. Geoinf.* **131**, 103964 (2024).
48. Zhou, M. L. et al. Revealing the Eco-environmental quality of the yellow river basin: Trends and drivers. *Rem. Sens.* **16**(11), 2018 (2024).

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No. U22A20620), the Surveying and Mapping Science and Technology “Double First-Class” project (Grant No. BZCG202301), and the PI project of the Collaborative Innovation Center of Geo-Information Technology for Smart Central Plains (Grant No. 2023C003).

Author contributions

LLC: Conceptualization, Data curation, Methodology, Writing original draft. HBC: Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Supervision, Writing review&editing. WBD: Data curation, Funding acquisition, Writing–review&editing. ZZL: Data curation, Writing–review&editing. YW: Data curation, Writing–review&editing. CYL: Resources, Data curation. LLC: Resources, Data curation. LZ: Writing–review&editing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-15550-3>.

Correspondence and requests for materials should be addressed to H.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025