

Article

Integrated Driving Mechanisms of the Thermal Environment, Air Pollution, and Carbon Sequestration Capacity in Henan Province, China

Shaowei Zhang ¹, Chen Li ¹, Shennian Zhang ², Ling Song ¹, Chenming Zhang ^{1,*}  and Pu Jia ¹ 

¹ School of Human Settlements, North China University of Water Resources and Electric Power, Zhengzhou 450046, China; zhangshaowei@ncwu.edu.cn (S.Z.); x20241141054@stu.ncwu.edu.cn (C.L.); songling@ncwu.edu.cn (L.S.); z20241141079@stu.ncwu.edu.cn (P.J.)

² School of Architecture, Xi'an University of Architecture and Technology, Xi'an 710055, China; zsx132465798@xauat.edu.cn

* Correspondence: zhangchenming@ncwu.edu.cn

Abstract

Rapid urbanization and climate change have intensified the interconnected challenges of surface heating, air pollution, and declining ecosystem functions, with important implications for regional sustainability. Taking Henan Province, China, as the study area, this study selected 2013, 2018, and 2023 as representative years and used land surface temperature (LST), fine particulate matter (PM_{2.5}), ozone (O₃), and net primary productivity (NPP) to characterize the thermal environment, air pollution, and carbon sequestration capacity. Pearson correlation analysis, multiple linear regression, and XGBoost-SHAP were integrated to examine bivariate associations, independent linear associations, factor importance, nonlinear responses, and potential threshold characteristics associated with natural, ecological, and anthropogenic factors. The results showed marked spatial differences in the four environmental variables. The multiple linear regression models explained 57.4–69.0% of the variation in LST, 23.8–72.0% in O₃, 81.0–84.8% in PM_{2.5}, and 57.4–62.5% in NPP. Natural factors generally showed relatively large and temporally stable standardized coefficients. Precipitation and potential evapotranspiration were positively associated with LST, whereas elevation and precipitation were negatively associated with PM_{2.5} and O₃. NDVI showed an environmentally favorable pattern, being negatively associated with LST, PM_{2.5}, and O₃ but positively associated with NPP. Anthropogenic variables generally exhibited smaller and less temporally stable coefficients. The XGBoost models demonstrated good predictive performance, particularly for PM_{2.5}, with R^2 values of 0.945, 0.920, and 0.905 in 2013, 2018, and 2023, respectively. SHAP analysis identified DEM, PRE, PET, and NDVI as the main contributors to model predictions and revealed nonlinear responses and potential threshold characteristics. These findings indicate that coordinated management of vegetation cover, hydrothermal conditions, and urban development can support heat mitigation, air pollution control, ecosystem productivity, and more sustainable, climate-resilient, and low-carbon development in rapidly urbanizing regions.



Academic Editor: Mohammad Aslam Khan Khalil

Received: 10 June 2026

Revised: 26 June 2026

Accepted: 27 June 2026

Published: 2 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: regional sustainability; land surface temperature; net primary productivity; XGBoost-SHAP; atmospheric environmental quality

1. Introduction

Rapid urbanization and climate change have placed increasing pressure on regional ecosystems, where surface heat, air pollution, and carbon sequestration capacity are often

altered at the same time. Rising land surface temperature (LST) reflects intensified surface heating and changes in surface energy exchange, while PM_{2.5} and O₃ represent particulate pollution and photochemical pollution, respectively. Net primary productivity (NPP), as an indicator of vegetation productivity, also reflects the contribution of vegetation to regional carbon sequestration. These environmental responses rarely change in isolation. In this setting, land development can modify surface thermal conditions, while population concentration and energy consumption may increase pollution pressure and further affect ecosystem functions. Natural background conditions, including topography, precipitation, potential evapotranspiration, and vegetation cover, also regulate heat accumulation, pollutant dispersion, and vegetation growth. These interrelated environmental challenges have important implications for regional sustainability because they simultaneously affect climate adaptation, public health, ecosystem functioning, and low-carbon development. An integrated assessment of the thermal environment, air pollution, and carbon sequestration capacity is therefore essential for identifying shared driving factors and supporting sustainable urban and regional management [1–4].

Research on urban heat, air pollution, and carbon sinks has expanded in recent years, with increasing attention being paid to the relationships of land-use change, urban expansion, vegetation cover, and climatic conditions with their spatial and temporal variations [5–9]. However, several limitations remain [10–13]. Existing studies often examine heat, air pollution, or carbon sequestration separately, leaving their shared driving factors and interrelated spatial patterns insufficiently understood [14–18]. Pearson correlation analysis can reveal basic bivariate relationships but cannot isolate the independent association of each driving factor after the other variables are controlled for [19,20]. Multiple linear regression addresses this limitation by quantifying the direction, magnitude, and statistical significance of the association between each explanatory variable and the environmental response [21–25]. However, linear models may still overlook nonlinear responses and potential threshold effects, which can be further examined using interpretable machine-learning methods [26–30].

Henan Province is characterized by a large population, a strong agricultural base, and rapid economic development in central China. It also occupies an important position in the Central Plains Urban Agglomeration and in the ecological protection and high-quality development strategy of the Yellow River Basin. The province shows marked terrain contrasts, with mountains mainly distributed in the west and plains extending across the east. This contrasting natural and anthropogenic setting makes Henan Province a suitable case for examining the relationships among surface heat, air pollution, and carbon sequestration capacity. Using 2013, 2018, and 2023 as representative years, this study develops an analytical framework that integrates the thermal environment, air pollution, and carbon sequestration capacity. Remote sensing, meteorological, topographic, and socioeconomic spatial data were combined to support the analysis.

This study selected 2013, 2018, and 2023 as representative cross-sectional years rather than as a continuous annual time series. These three years provide approximately five-year intervals and cover the early, middle, and recent stages of regional urbanization and environmental governance during 2013–2023. The year 2013 was used as the initial observation year, 2018 represented an intermediate stage, and 2023 represented the most recent stage available for all major datasets used in this study. Therefore, the analysis focuses on spatial differences and inter-period changes among three representative time points, rather than on continuous annual or seasonal variability. To avoid overinterpretation, the temporal conclusions in this study are interpreted as changes among representative years.

This study has three aims. First, it characterizes the spatiotemporal patterns of LST, PM_{2.5}, O₃, and NPP and examines their bivariate relationships with the selected driving factors in Henan Province. Second, it applies multiple linear regression to quantify the

independent linear associations of natural, ecological, and anthropogenic factors with the four environmental response variables and to compare these associations across 2013, 2018, and 2023. Third, it uses XGBoost-SHAP to evaluate factor importance and identify nonlinear responses and potential threshold effects. The results can provide scientific support for the coordinated management of heat, air pollution, and carbon sequestration in central China under continuing urbanization.

2. Study Area, Data, and Methods

2.1. Overview of the Study Area

Henan Province is located in central-eastern China. It lies in the middle and lower reaches of the Yellow River. The province extends from about $31^{\circ}23'$ to $36^{\circ}22'$ N and from $110^{\circ}21'$ to $116^{\circ}39'$ E. It covers about 167,000 km². Henan borders Anhui and Shandong in the east, Hebei and Shanxi in the north, Shaanxi in the west, and Hubei in the south (Figure 1).

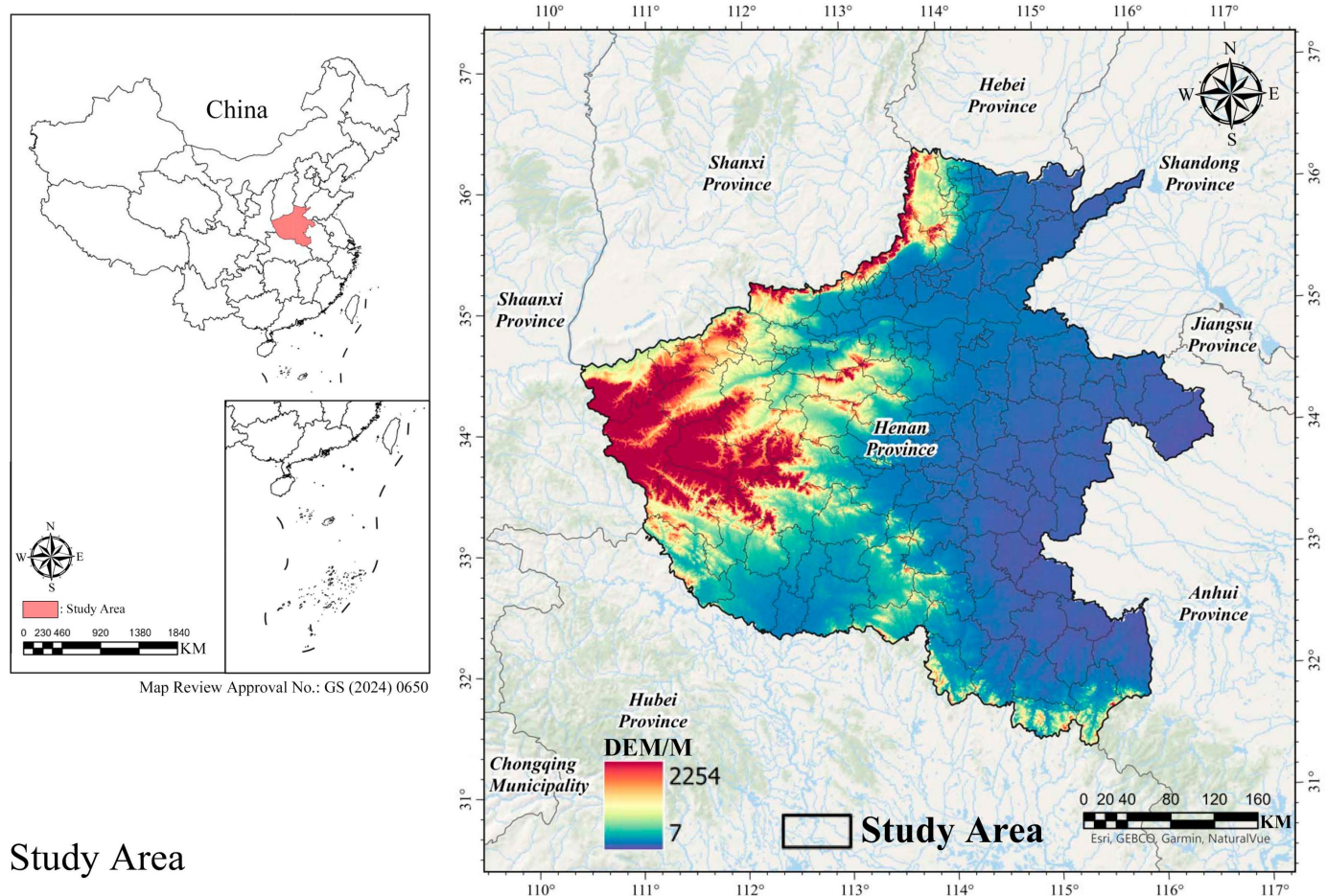


Figure 1. Location of the study area.

Henan plays an important role in the Central Plains Urban Agglomeration. It is also important for national food security and for ecological protection in the Yellow River Basin [31]. The terrain is generally higher in the west and lower in the east. Mountains and hills are mainly found in the western, northern, and southern parts of the province. The central and eastern parts are mainly covered by the Huang-Huai-Hai Plain. This terrain pattern produces marked natural heterogeneity across the province.

Henan Province lies in the transition zone between the northern subtropical and warm temperate zones and has a typical continental monsoon climate. It is characterized by

distinct seasons, the concurrent occurrence of rainfall and heat, and considerable interannual climatic variability. Topography, climate, and land use jointly shape the province's ecological conditions. The western and southern mountainous areas generally have higher vegetation cover, better ecological backgrounds, and stronger vegetation productivity, allowing them to play a more important role in environmental regulation. By contrast, the central and eastern plains are flat, densely populated, and more strongly influenced by urban construction and agricultural activities, where human disturbance is relatively concentrated. These natural and human contrasts make Henan a suitable case for examining the spatial and temporal variations in land surface temperature, air pollution, and carbon sequestration capacity, as well as their underlying driving factors.

In recent years, Henan Province has experienced continued urbanization and industrial development, accompanied by marked changes in land use, intensified surface heating, the local co-occurrence of PM_{2.5} and O₃ pollution, and increasing pressure on ecological space protection [32,33]. Human activities, including population concentration, built-up land expansion, road network growth, building volume, and nighttime light intensity, may affect the spatial patterns of LST, PM_{2.5}, O₃, and NPP. Natural and ecological conditions also play important roles. In this setting, precipitation, potential evapotranspiration, topography, and vegetation cover can influence thermal conditions, pollutant dispersion, and ecological productivity. Henan Province was therefore selected as the study area to analyze the spatial and temporal patterns of LST, PM_{2.5}, O₃, and NPP, and to examine how these patterns are shaped by natural conditions and human activities. The results may provide a reference for ecological management and sustainable development in central China.

2.2. Data Sources and Variable Description

Using 2013, 2018, and 2023 as representative years, this study examined the spatial and temporal variations in LST, PM_{2.5}, O₃, and NPP. It builds a spatial dataset from several data sources. The dataset is used to examine the links among the thermal environment, air pollution, and carbon sequestration capacity in Henan Province. The dependent variables are land surface temperature (LST), fine particulate matter concentration (PM_{2.5}), ozone concentration (O₃), and net primary productivity (NPP). LST is used to describe surface heat conditions. PM_{2.5} is used to describe particulate pollution. O₃ is used to describe ozone pollution. NPP is used to describe vegetation productivity and carbon sequestration capacity. Among these variables, LST shows the level of surface heat and thermal pressure. PM_{2.5} and O₃ show two different types of air pollution. NPP shows vegetation growth. It also serves as an important indicator of regional carbon sequestration capacity.

This study selected explanatory variables from three aspects: natural conditions, ecological regulation, and human activities. These variables were used to explain the spatial differences in the thermal environment, air pollution, and carbon sequestration capacity. Natural background factors include the digital elevation model (DEM), precipitation (PRE), and potential evapotranspiration (PET), which represent topographic conditions, water supply, and regional hydrothermal background, respectively. Ecological regulation factors include the normalized difference vegetation index (NDVI) and water-area extent (WATER), which reflect vegetation cover and the regulatory capacity of blue–green spaces. Human activity factors include nighttime light intensity (NTL), population density (POP), building volume (BV), built-up land area (BLA), and road density (RD), which represent urbanization level, construction intensity, population agglomeration, and traffic activity pressure.

The data sources are listed in Table 1. LST, PM_{2.5}, PRE, and NTL data were obtained from the National Tibetan Plateau Data Center (TPDC). The PM_{2.5} data were obtained from the ChinaHighPM_{2.5} dataset, which is part of the ChinaHighAirPollutants (CHAP) dataset [28]. ChinaHighPM_{2.5} is a long-term, seamless, high-resolution, and high-quality

ground-level PM2.5 dataset for China. The dataset was generated using artificial intelligence and big-data fusion methods by integrating ground-based observations, satellite remote-sensing products, atmospheric reanalysis data, emission inventories, and other auxiliary predictors. The product provides daily, monthly, and yearly PM2.5 concentrations at a spatial resolution of 1 km. In this study, the yearly 1 km product was used for 2013, 2018, and 2023; therefore, PM2.5 refers to the annual mean ground-level PM2.5 concentration. According to the product description, the dataset achieved reliable cross-validation performance, with an R^2 of approximately 0.90 and an RMSE of $10.76 \mu\text{g m}^{-3}$, supporting its use for regional-scale spatial analysis. Nevertheless, PM2.5 values used in this study are gridded estimates rather than direct station observations, and the results should therefore be interpreted as spatial statistical associations rather than direct atmospheric chemical-process simulations.

Table 1. Driving indicator system and data sources.

Classification	Variable	Spatial Resolution	Unit	Source
Explained variable	Land Surface Temperature (LST)	1 km	$^{\circ}\text{C}$	National Tibetan Plateau Data Center/TPDC, https://data.tpdc.ac.cn
	Annual ground-level PM2.5 concentration (PM2.5)	1 km	$\mu\text{g m}^{-3}$	ChinaHighPM2.5/CHAP, National Tibetan Plateau Data Center, https://data.tpdc.ac.cn
	Annual ground-level MDA8 O3 concentration (O3)	1 km	$\mu\text{g m}^{-3}$	ChinaHighO3/CHAP, Zenodo, https://zenodo.org
	Net Primary Productivity (NPP)	500 m	$\text{kg C m}^{-2} \text{yr}^{-1}$	MODIS MOD17A3HGF V6.1 annual NPP product, https://lpdaac.usgs.gov
Explanatory variable	Water area (WATER)	30 m	%	Zenodo: https://zenodo.org
	Built-up land area (BLA)	30 m	%	
	Precipitation (PRE)	1 km	mm yr^{-1}	National Tibetan Plateau Data Center: https://data.tpdc.ac.cn
	Potential evapotranspiration (PET)	1 km	mm yr^{-1}	GeoData: https://www.geodata.cn
	Normalized Difference Vegetation Index (NDVI)	250 m	Dimensionless	Geographic Remote Sensing Ecological Network Platform, http://www.gisrs.cn
	Nighttime Light (NTL)	1 km	$\text{nW cm}^{-2} \text{sr}^{-1}$	National Tibetan Plateau Data Center/TPDC, https://data.tpdc.ac.cn
	Digital elevation model (DEM)	30 m	m	Geospatial Data Cloud: https://www.gscloud.cn
	Building Volume (BV)	1 km	m^3 per grid cell	JRC Global Human Settlement Layer, https://ghsl.jrc.ec.europa.eu
	Road Density (RD)	1 km	km km^{-2}	OpenStreetMap/Geofabrik, https://download.geofabrik.de
	Population Density (POP)	1 km	Persons km^{-2}	Population dataset/DOI: https://doi.org/10.48690/1532445

The O3 data were obtained from the ChinaHighO3 dataset, another component of the ChinaHighAirPollutants (CHAP) dataset [34]. ChinaHighO3 provides long-term, seamless, high-resolution ground-level ozone data for China and was generated from multiple big-data sources, including ground-based measurements, satellite remote-sensing products, atmospheric reanalysis, and model simulations, using artificial intelligence methods that consider the spatiotemporal heterogeneity of air pollution. The dataset provides daily, monthly, and yearly 1 km ground-level maximum daily 8 h average ozone concentration products. In this study, the yearly 1 km product, CHAP_O3_Y1K_V3, was used for 2013, 2018, and 2023. Thus, O3 represents annual ground-level MDA8 O3 concentration, rather than simple daily mean ozone or daily maximum ozone. The ChinaHighO3 product reports a daily cross-validation R^2 of 0.89, an RMSE of $15.77 \mu\text{g m}^{-3}$, and an MAE of $10.48 \mu\text{g m}^{-3}$, indicating good reliability for regional-scale 1 km spatial analysis. Similar to PM2.5, the O3 dataset is a model-derived gridded product, and the results are interpreted as statistical relationships rather than direct photochemical process simulations.

NPP data were derived from the MODIS MOD17A3HGF V6.1 annual NPP product [7,9]. The source of O₃ data is shown in Table 1. Water and built-up land data were obtained from Zenodo land-cover data. PET data were obtained from the GeoData platform, NDVI data from the Geographic Remote Sensing Ecological Network Platform, DEM data from the Geospatial Data Cloud, BV data from the JRC Global Human Settlement Layer, RD was extracted from OpenStreetMap/Geofabrik road network data, and POP data were obtained from a gridded population density dataset. The original spatial resolutions of these datasets vary, including 30 m, 250 m, 500 m, and 1 km.

To ensure spatial comparability among variables, all raster datasets were preprocessed using a unified spatial framework. First, all datasets were projected to the WGS_1984_UTM_Zone_49N coordinate system and clipped using the administrative boundary of Henan Province. Second, all variables were converted to a 1 km × 1 km grid. For continuous raster variables, including LST, PM_{2.5}, O₃, NPP, PRE, PET, NDVI, NTL, DEM, BV, and POP, bilinear interpolation was used during resampling. For categorical land-cover data, nearest-neighbor resampling was used when class labels needed to be preserved. WATER and BLA were derived by reclassifying the original land-cover data into water and built-up land classes, respectively, and then calculating the area proportion of each class within each 1 km grid cell. Road density was calculated as the total length of the road network within each 1 km grid cell divided by the grid-cell area and was expressed in km km^{−2}. Railways were not included in the road-density variable because RD was intended to represent road-network intensity and road-related human activity. Finally, missing and abnormal values were checked, invalid pixels were removed, and all dependent and explanatory variables were standardized using Z-score normalization.

2.3. Methods

The technical framework of this study is shown in Figure 2 and consists of four main stages: data collection and preprocessing, Pearson correlation analysis, multiple linear regression analysis, and XGBoost-SHAP interpretation. Multisource spatial data for 2013, 2018, and 2023 were integrated to construct an analytical dataset for the thermal environment, air pollution, and carbon sequestration capacity in Henan Province. LST, PM_{2.5}, O₃, and NPP were used as dependent variables, while WATER, BLA, PRE, PET, NDVI, NTL, DEM, BV, RD, and POP were used as explanatory variables. All raster datasets were projected, clipped, resampled to a spatial resolution of 1 km × 1 km, and standardized before analysis.

Pearson correlation analysis was first used to identify the direction and strength of the bivariate linear associations between the environmental response variables and the explanatory variables. Multiple linear regression models were then constructed separately for LST, PM_{2.5}, O₃, and NPP in each study year to quantify the independent linear association of each explanatory variable after controlling for the other variables [22–25]. Finally, XGBoost-SHAP was applied to evaluate factor importance and identify influence directions, nonlinear responses, and potential threshold effects.

These three statistical approaches provide complementary evidence. Pearson correlation analysis describes bivariate associations, multiple linear regression quantifies independent linear associations under multivariable control, and XGBoost-SHAP captures factor importance and nonlinear response patterns. The results are interpreted as statistical associations rather than strict causal effects [35,36].

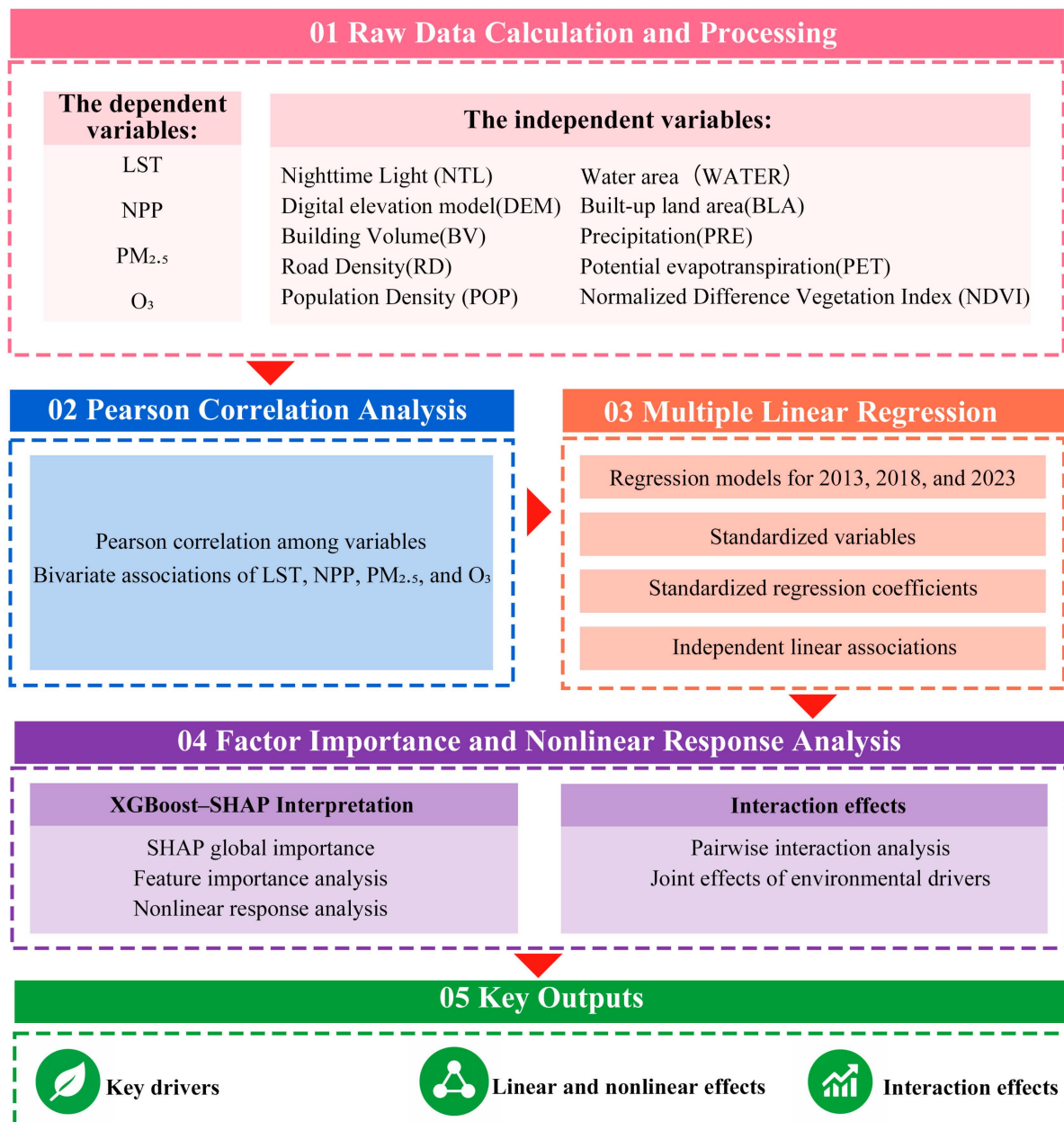


Figure 2. The framework of this study.

2.3.1. Pearson Correlation Analysis

This study used Pearson correlation analysis to evaluate the pairwise linear relationships between LST, PM_{2.5}, O₃, NPP, and the ten driving factors [12,13,24]. The Pearson correlation coefficient, r , is calculated as follows:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

where r ranges from -1 to 1 . A value of 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no linear correlation. X_i and Y_i denote the i -th observed values of variables X and Y , respectively; $\bar{X}$ and $\bar{Y}$ denote their corresponding mean values. The significance level was set at $p < 0.05$.

2.3.2. Multiple Linear Regression

Multiple linear regression was used to quantify the independent linear associations between the driving factors and LST, PM2.5, O3, and NPP [22–25]. Separate models were constructed for each response variable in 2013, 2018, and 2023, resulting in 12 regression models. The explanatory variables included BV, DEM, NDVI, NTL, POP, PET, WATER, BLA, RD, and PRE. Before model estimation, all dependent and independent variables were standardized using Z-score normalization, allowing the estimated coefficients to be interpreted as standardized regression coefficients and compared across predictors.

$$Y_i = \beta_0 + \sum_{j=1}^{10} \beta_j X_{ij} + \varepsilon_i \quad (2)$$

where Y_i represents the standardized value of LST, PM2.5, O3, or NPP; X_{ij} represents the standardized driving factors; β_0 is the intercept; β_j is the standardized regression coefficient; and ε_i is the residual term. The coefficient of determination (R^2), adjusted R^2 , F-statistic, and t-statistics were used to evaluate model performance and coefficient significance. Variance inflation factors (VIFs) were calculated to assess multicollinearity [23,36]. A VIF below 5 was considered indicative of the absence of serious multicollinearity. Statistical significance was evaluated at $p < 0.05$.

2.3.3. XGBoost-SHAP Explainable Machine Learning

Although Pearson correlation analysis and multiple linear regression can identify bivariate and multivariable linear associations, respectively, they cannot fully characterize nonlinear responses across different ranges of the explanatory variables. Therefore, XGBoost-SHAP explainable machine learning was incorporated to improve predictive performance and clarify the relative importance, influence direction, nonlinear responses, and potential threshold effects of the key driving factors [27–30,37].

XGBoost is an ensemble learning algorithm developed from gradient boosting. It builds a strong predictive model by iteratively combining multiple weak learners and correcting the errors generated in previous rounds. Compared with traditional regression models, XGBoost is better suited to capturing nonlinear relationships and complex interactions among variables, while maintaining strong predictive accuracy and generalization ability. Therefore, it has been widely used in ecological-environmental and geospatial modeling studies [38–40]. Its objective function can be expressed as follows:

$$L(\Phi) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^m \Omega(f_k) \quad (3)$$

where $\hat{y}_i$ denotes the predicted value of the i -th sample X_i ; $l(y_i, \hat{y}_i)$ denotes the loss function for an individual sample; and $\sum_{k=1}^m \Omega(f_k)$ is the regularization term, which is used to constrain model complexity and reduce overfitting.

Although XGBoost has strong predictive capability, its model structure is relatively complex and tends to exhibit the typical “black-box” characteristics of machine learning models. To improve the interpretability of the model results, this study combined XGBoost with SHAP (SHapley Additive exPlanations). Based on the concept of Shapley values from cooperative game theory, SHAP quantifies the marginal contribution of each driving factor to model predictions, thereby identifying the relative importance and influence direction

of different variables on LST, PM2.5, O3, and NPP. The Shapley value of feature P can be expressed as follows:

$$\phi_P = \sum_{S \subseteq \{x_1, \dots, x_p\} \setminus \{x_p\}} \frac{|S|!(P-|S|-1)!}{P!} (f(S \cup \{x_p\}) - f(S)) \quad (4)$$

where ϕ_P denotes the Shapley value of feature P ; S denotes any subset of model features; X_p denotes the value vector of feature P ; p denotes the number of input features; and $f(S)$ denotes the prediction result based on the feature subset S . The overall model output $f(X)$ can be obtained as a linear combination of the Shapley values of all features, which is expressed as follows:

$$f(x) = \phi_0 + \sum_{p=1}^p \phi_P Z'_i \quad (5)$$

where ϕ_0 denotes the expected prediction value of the model; Z'_i denotes a binary feature variable, where $Z'_i = 1$ indicates that the feature is present and $Z'_i = 0$ indicates that the feature is absent. A larger absolute Shapley value indicates a greater contribution of the corresponding variable to the model prediction. A positive Shapley value indicates that the variable promotes the prediction result, whereas a negative Shapley value indicates that it suppresses the prediction result.

In this study, separate XGBoost regression models were constructed for LST, PM2.5, O3, and NPP using ten natural, ecological, and anthropogenic explanatory variables. These variables were selected a priori based on their environmental relevance, data availability, and previous studies, and all variables were included in the initial models. The samples were randomly divided into training and testing sets at a ratio of 7:3 using a fixed random seed of 42. The models contained 600 boosting trees, with a learning rate of 0.03, a maximum tree depth of 4, and a minimum child weight of 2. The subsample and feature-sampling ratios were set to 0.85 and 0.90, respectively. The L2 regularization parameter was set to 1.0, and the squared-error regression objective was adopted. The hyperparameters were selected with reference to commonly used settings reported in previous studies, with the aim of balancing predictive performance, model complexity, and computational efficiency [28,29,37,38]. No additional automatic feature-selection procedure was performed. Model performance was evaluated using R^2 , RMSE, and MAE. The built-in XGBoost importance scores and SHAP values were subsequently used to quantify the relative contribution of each explanatory variable.

3. Results

3.1. Spatiotemporal Differentiation Characteristics of LST, PM2.5, O3, and NPP in Henan Province

As shown in Figure 3, LST, PM2.5, O3, and NPP in Henan Province exhibited clear spatiotemporal differentiation characteristics from 2013 to 2023. Overall, high LST values were mainly distributed in the central and eastern plains and in local areas with intensive construction activities, while low-value areas were mainly located in the western mountains and areas with better vegetation cover. In 2013, high LST values showed a patchy distribution and were relatively evident in parts of the central and southern regions. In 2018, the high-LST areas spread toward the central-eastern plains. This pattern shows that surface heat pressure increased in these areas. In 2023, high LST values were still mainly found in parts of the central-eastern and southern regions. These high-value areas became more fragmented. The low-temperature areas in the western mountains changed little. These changes show that topography has a clear effect on the surface thermal pattern in Henan Province. They also show that high-temperature areas are often linked with areas

of stronger human activity. Urban growth and built-up land expansion may be important reasons for the increase in local surface heat.

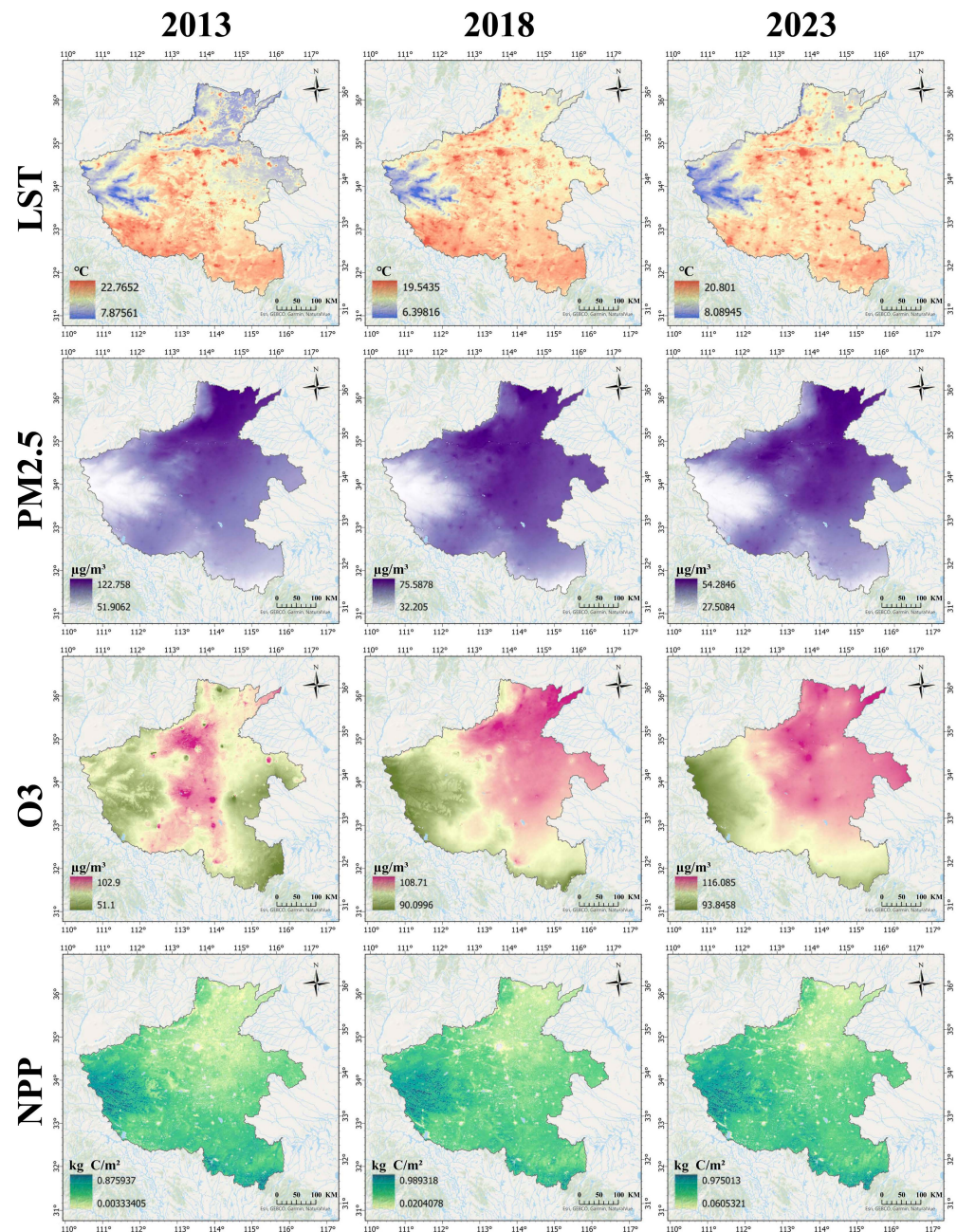


Figure 3. Spatial distributions of LST, PM2.5, O3, and NPP in 2013, 2018, and 2023.

PM2.5 showed a clear decrease from 2013 to 2023. Its high-value areas also became smaller over time. In 2013, PM2.5 levels were generally high across Henan Province. High values were widely found in the northern, eastern, and central plain areas. In 2018, the high-PM2.5 areas became much smaller. Low-value areas in the west became more obvious. In 2023, PM2.5 levels continued to decrease. High values were mainly found in parts of northern and northeastern Henan. This pattern shows that particulate pollution was reduced during the study period. Unlike PM2.5, O3 showed a certain increasing and expanding trend. In 2013, high O3 values were mainly distributed as local patches around central urban areas and in parts of the plains. In 2018, the high-value areas of O3 expanded markedly and extended toward the central-northern and eastern regions. In 2023,

O₃ remained at a relatively high level, with high-value areas becoming more continuous in the central, northern, and eastern plains, indicating an enhanced spatial influence of ozone pollution. The differentiated changes between PM_{2.5} and O₃ suggest that the air pollution structure in Henan Province may have gradually shifted from particulate-dominated pollution to compound pollution characterized by the coexistence of PM_{2.5} and O₃, with an increasing influence of O₃.

The spatial distribution of NPP differed markedly from that of LST, PM_{2.5}, and O₃. In 2013, high NPP values were mainly distributed in the western and southern mountainous areas, while low-value areas were mostly located in the central-eastern plains with intensive construction activities and agricultural development. In 2018, high NPP areas remained concentrated in the western and southern mountains, and ecological productivity increased in some areas. In 2023, the overall spatial pattern of NPP remained stable, with high-value areas continuing to be located in the western mountains and southern regions with stronger ecological backgrounds, while values in the central-eastern plains and built-up land concentration areas remained relatively low. Overall, high NPP areas in Henan Province showed strong spatial stability and were mainly controlled by topography, vegetation cover, and ecological background conditions. Overall, the four environmental variables show clear spatial links in Henan Province. High LST and O₃ values mainly appear in the central-eastern plains and in areas with stronger human activities. PM_{2.5} decreased during the study period, and its high-value areas became smaller. High NPP values remained mainly in the western and southern mountains. This pattern shows that heat, air pollution, and carbon sequestration capacity are spatially connected. These spatial patterns provide a basis for the subsequent Pearson correlation analysis, multiple linear regression analysis, and XGBoost-SHAP interpretation.

3.2. Pearson Correlations Between Driving Factors and Environmental Variables

This study used Pearson correlation analysis to examine the basic linear relationships between the driving factors and the environmental variables in Henan Province. The environmental variables included LST, PM_{2.5}, O₃, and NPP. The analysis was conducted for 2013, 2018, and 2023. The results were shown with correlation bubble matrix plots. Overall, the relationships between the environmental variables and the driving factors differed across variables and years. NDVI, DEM, PRE, and PET exhibited relatively prominent correlations with LST, PM_{2.5}, O₃, and NPP, whereas WATER showed generally weak correlations with the environmental variables.

In 2013, LST was strongly positively correlated with PET and PRE, with correlation coefficients of 0.521 and 0.502, respectively, indicating that thermal conditions and precipitation background had important effects on the spatial differentiation of land surface temperature (Figure 4). LST was negatively correlated with DEM ($r = -0.403$), suggesting that areas with higher elevation generally had lower surface temperatures. PM_{2.5} was significantly negatively correlated with DEM, PRE, and NDVI, with the strongest correlation observed for DEM ($r = -0.677$), followed by PRE ($r = -0.536$) and NDVI ($r = -0.496$). This indicates that topography, precipitation, and vegetation cover had certain mitigating effects on particulate pollution. O₃ was negatively correlated with NDVI and PRE, with correlation coefficients of -0.361 and -0.318 , respectively, but positively correlated with BLA ($r = 0.291$), suggesting that built-up land expansion may be associated with enhanced ozone pollution. NPP was positively correlated with DEM and NDVI, with correlation coefficients of 0.495 and 0.447, respectively, but negatively correlated with BLA ($r = -0.440$), indicating that vegetation cover and topographic conditions promoted ecological productivity, whereas built-up land expansion may weaken regional ecological functions.

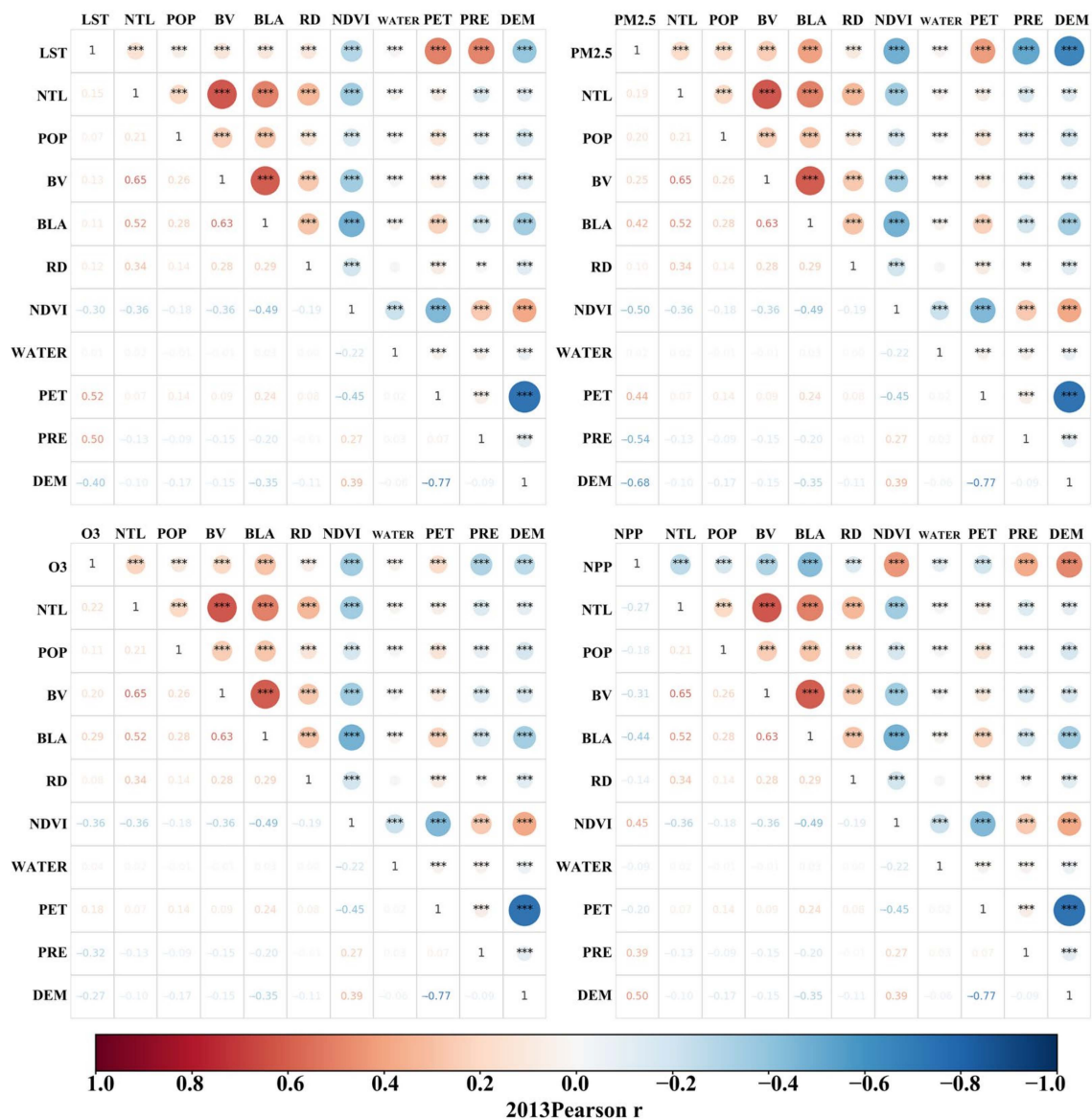


Figure 4. Pearson correlation coefficients for 2013. ** and *** indicate significance at $p < 0.01$ and $p < 0.001$, respectively.

In 2018, the relationships among the variables became stronger (Figure 5). LST had a strong negative relationship with DEM ($r = -0.646$). It had a positive relationship with PET ($r = 0.587$). It also had a negative relationship with NDVI ($r = -0.511$). These results show that topography, potential evapotranspiration, and vegetation cover all affected the surface thermal environment. PM2.5 was significantly positively correlated with PET ($r = 0.708$), and significantly negatively correlated with DEM and NDVI, with correlation coefficients of -0.706 and -0.652 , respectively. This suggests that PM2.5 pollution was influenced not only by meteorological background conditions, but also by the joint regulation of topography and vegetation cover. O3 was negatively correlated with DEM and NDVI, with correlation coefficients of -0.653 and -0.521 , respectively, and positively correlated with PET ($r = 0.528$), indicating that ozone pollution was closely related to thermal conditions, while vegetation cover and elevation exerted inhibitory effects. NPP showed the strongest positive correlation with NDVI ($r = 0.647$), followed by DEM ($r = 0.554$), and was negatively correlated with BLA ($r = -0.464$), indicating that ecological productivity was mainly supported by vegetation cover and natural geographic conditions, while being inhibited by built-up land expansion.

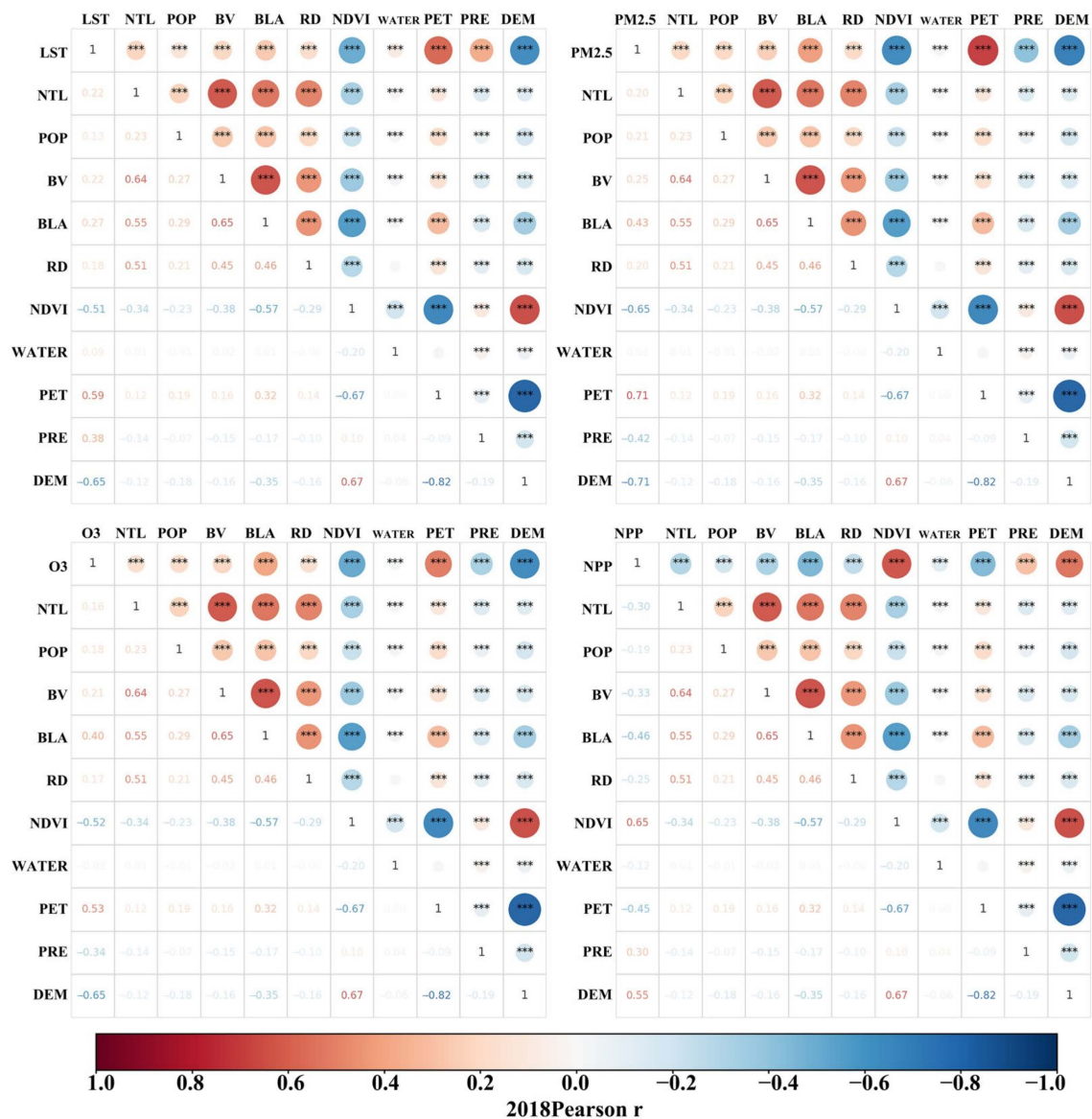


Figure 5. Pearson correlation coefficients for 2018. *** indicates significance at $p < 0.001$.

In 2023, the ecological regulatory role of NDVI became more prominent (Figure 6). LST was significantly negatively correlated with DEM ($r = -0.690$), positively correlated with PET ($r = 0.564$), and negatively correlated with NDVI ($r = -0.472$), indicating that topography and vegetation remained important factors affecting spatial differences in the thermal environment. PM2.5 was positively correlated with PET ($r = 0.721$), and significantly negatively correlated with NDVI and DEM, with correlation coefficients of -0.712 and -0.646 , respectively, suggesting that the negative regulatory effect of vegetation cover on PM2.5 was further strengthened in 2023. O3 showed strong negative correlations with DEM and NDVI, with correlation coefficients of -0.620 and -0.616 , respectively, and a positive correlation with PET ($r = 0.602$), indicating that ozone pollution was promoted by thermal conditions but inhibited by topography and vegetation cover. NPP was positively correlated with NDVI and DEM, with correlation coefficients of 0.633 and 0.498 , respectively, and negatively correlated with BLA ($r = -0.465$), further indicating that built-up land expansion had a persistent negative impact on ecological productivity.

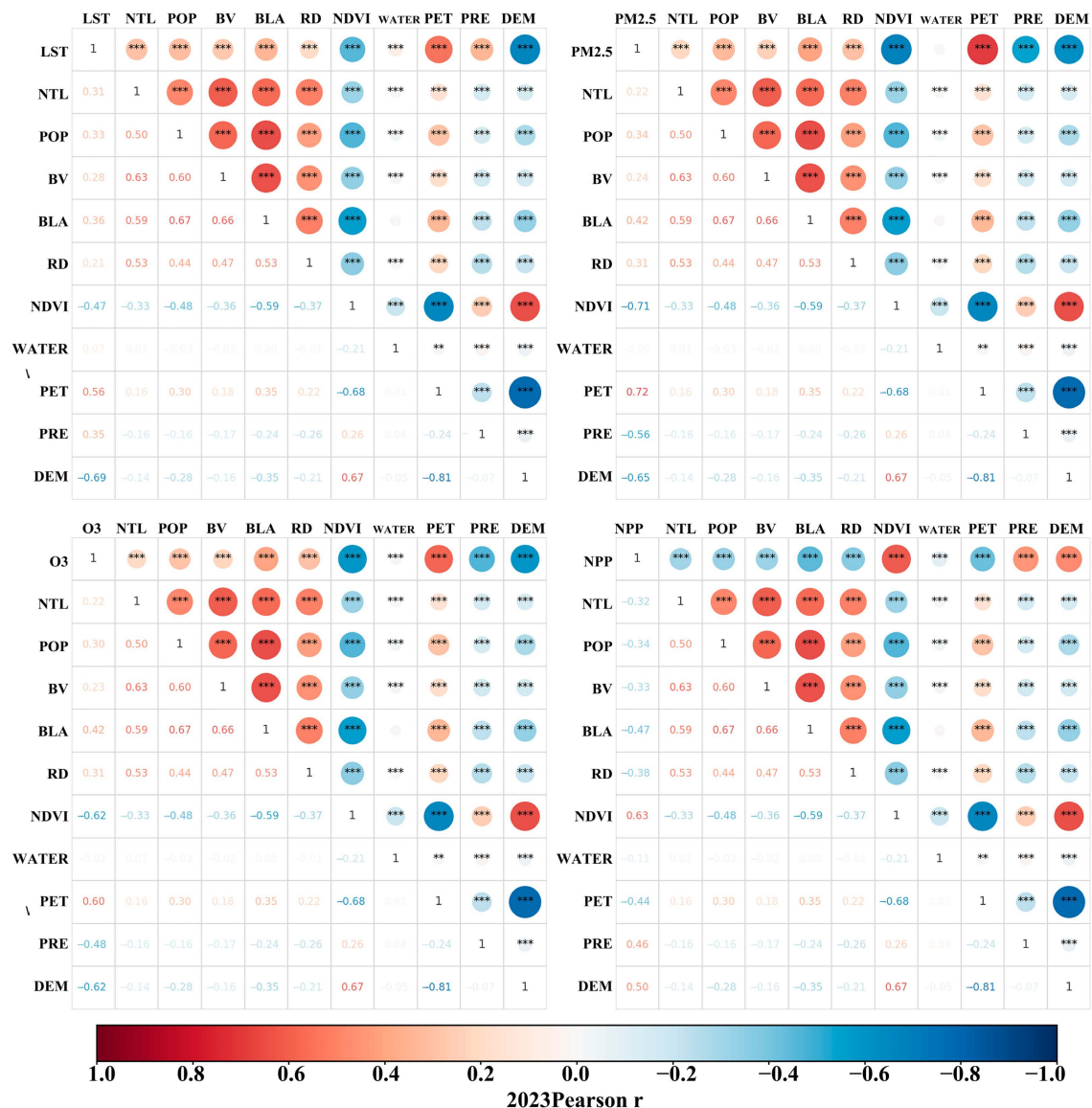


Figure 6. Pearson correlation coefficients for 2023. ** and *** indicate significance at $p < 0.01$ and $p < 0.001$, respectively.

Across the three years, NDVI showed the most stable role among the ecological factors. NDVI had negative relationships with LST, PM2.5, and O3. It had a positive relationship with NPP. This result shows that vegetation cover can help reduce surface heat and air pollution. It can also support ecological productivity. DEM, PRE, and PET were also important natural factors in Henan Province. These factors affected the spatial patterns of the environmental variables. DEM was usually negatively related to LST, PM2.5, and O3. It was positively related to NPP. PET was usually positively related to LST, PM2.5, and O3. This means that stronger heat conditions and higher evapotranspiration demand may increase heat pressure and pollution risk. In contrast, WATER showed generally weak correlations with the four environmental variables, indicating that the linear explanatory power of water area alone is limited at the provincial scale. Pearson correlation analysis shows that the spatial differentiation of LST, PM2.5, O3, and NPP in Henan Province is not controlled by a single urbanization factor, but is jointly influenced by natural geographic background, climatic conditions, and ecological cover. These results provide a basis for the

subsequent multiple linear regression analysis and XGBoost-SHAP interpretation of factor importance and nonlinear responses.

3.3. Linear Associations Identified by Multiple Linear Regression

The multiple linear regression models were statistically significant and showed varying explanatory power across environmental response variables and years. All VIF values were below 5, ranging from 1.092 to 4.610, indicating that serious multicollinearity was not present. Figure 7 presents the standardized regression coefficients, while the complete model statistics are provided in Appendix A.

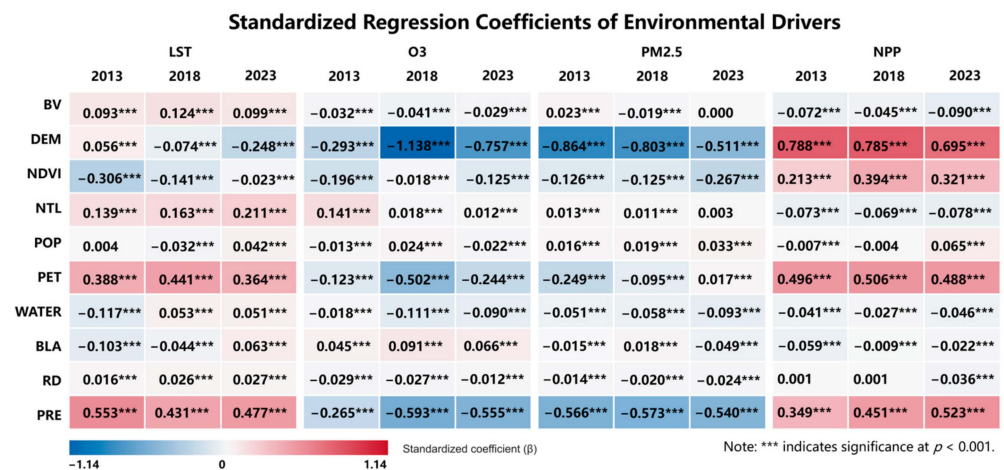


Figure 7. Standardized regression coefficients of the driving factors for LST, O3, PM2.5, and NPP in 2013, 2018, and 2023.

The explanatory power of the LST models increased from 0.574 in 2013 to 0.690 in 2023. PRE and PET showed consistently positive associations with LST, with standardized coefficients of 0.431–0.553 and 0.364–0.441, respectively. NTL was also positively associated with LST, and its coefficient increased from 0.139 in 2013 to 0.211 in 2023. In contrast, NDVI was negatively associated with LST in all three years, although its coefficient weakened from -0.306 to -0.023 . The coefficient of DEM changed from positive in 2013 to negative in 2018 and 2023, indicating temporal variation in the relationship between elevation and surface temperature.

The PM2.5 models exhibited the highest explanatory power, with R^2 values of 0.848, 0.828, and 0.810 in 2013, 2018, and 2023, respectively. DEM and PRE were consistently the strongest negative predictors, with coefficients ranging from -0.864 to -0.511 and from -0.573 to -0.540 , respectively. NDVI and WATER were also negatively associated with PM2.5 in all three years. The negative coefficient of NDVI became stronger in 2023, reaching -0.267 , suggesting an increasing statistical association between vegetation cover and lower particulate pollution.

The explanatory power of the O3 model was relatively low in 2013 ($R^2 = 0.238$) but increased substantially in 2018 ($R^2 = 0.720$) and remained high in 2023 ($R^2 = 0.686$). DEM, PRE, PET, NDVI, and WATER generally showed negative associations with O3. DEM had the largest negative coefficient, particularly in 2018 ($\beta = -1.138$). The direction of several anthropogenic variables, including NTL and POP, varied across years, indicating that their linear relationships with O3 were less temporally stable.

The NPP models explained 57.4–62.5% of the observed variation. DEM, PET, PRE, and NDVI were consistently positively associated with NPP. DEM showed the largest positive coefficients (0.695–0.788), followed by PET (0.488–0.506) and PRE (0.349–0.523). NDVI also exhibited stable positive coefficients ranging from 0.213 to 0.394. By contrast, BV,

NTL, WATER, and BLA were generally negatively associated with NPP, suggesting that intensified development and certain land-cover conditions were associated with reduced ecological productivity.

Overall, the regression results showed that natural background factors generally had relatively large and temporally stable standardized coefficients across the four environmental responses. NDVI exhibited environmentally favorable associations, being negatively associated with LST, PM_{2.5}, and O₃ but positively associated with NPP. In comparison, the coefficients of most anthropogenic variables were smaller and less stable across years.

3.4. Factor Importance Ranking Based on XGBoost-SHAP

3.4.1. Predictive Performance of the XGBoost Models

Figure 8 and Table 2 show the predictive performance of the XGBoost models for LST, PM_{2.5}, O₃, and NPP in 2013, 2018, and 2023. Overall, the models showed good predictive performance for all four dependent variables, indicating that the selected driving factors effectively explained the spatial differences in the thermal environment, air pollution, and carbon sequestration capacity in Henan Province.

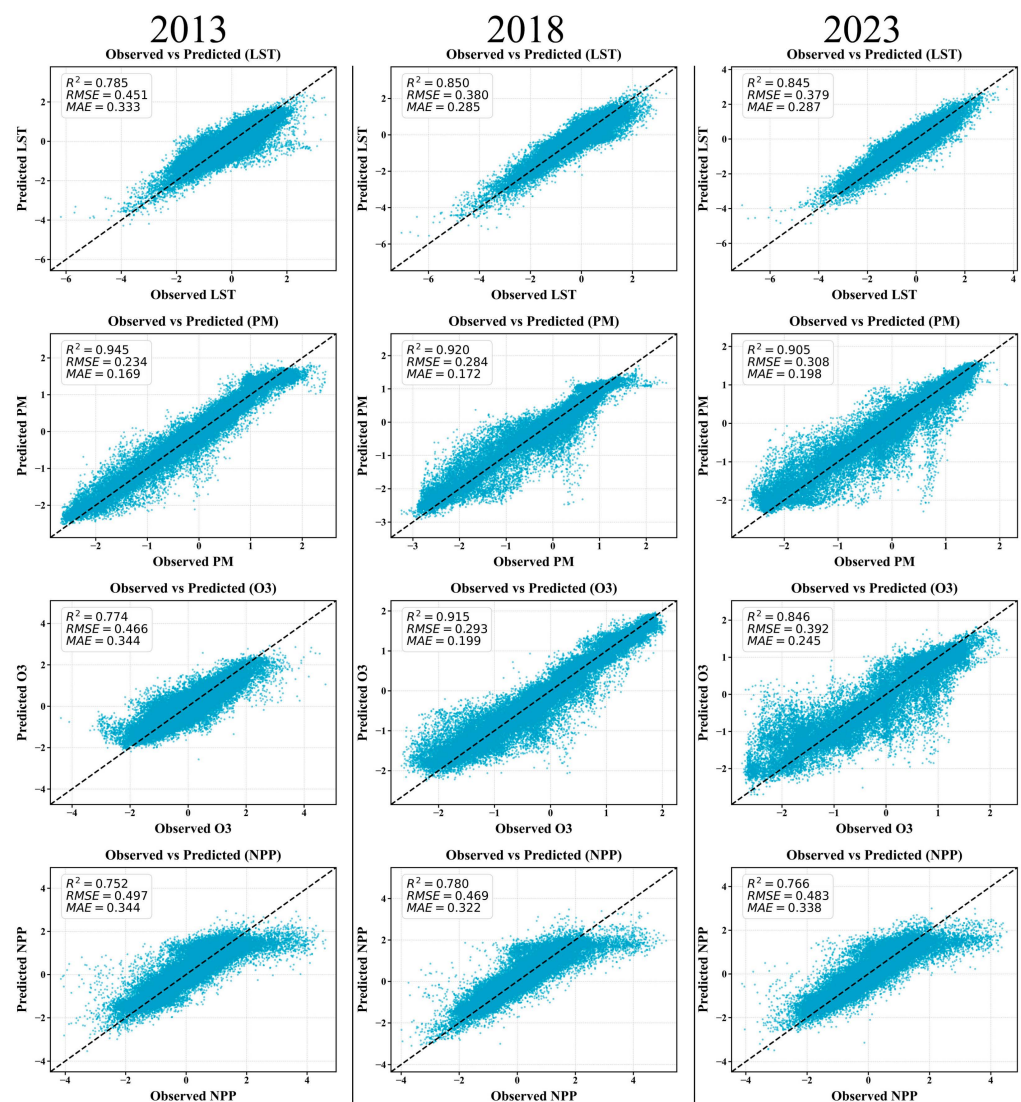


Figure 8. Observed and Predicted Values Fitted by the XGBoost Models in 2013, 2018, and 2023. Blue dots represent sample observations, and the dashed line represents the 1:1 reference line.

Table 2. Predictive accuracy and dominant SHAP factors of the XGBoost models.

Year	Response Variable	R ²	RMSE	MAE	Dominant SHAP Factor	Contribution Rate/Mean SHAP Value
2013	LST	0.785	0.451	0.333	PRE	39.19%
	PM2.5	0.945	0.234	0.169	DEM	44.33%
	O3	0.774	0.466	0.344	DEM	46.07%
	NPP	0.752	0.497	0.344	DEM	39.00%
2018	LST	0.850	0.380	0.285	DEM	34.46%
	PM2.5	0.920	0.284	0.172	DEM	35.94%
	O3	0.915	0.293	0.199	DEM	54.05%
	NPP	0.780	0.469	0.322	PRE	31.55%
2023	LST	0.845	0.379	0.287	DEM	33.45%
	PM2.5	0.905	0.308	0.198	PRE	31.45%
	O3	0.846	0.392	0.245	DEM	36.28%
	NPP	0.766	0.483	0.338	PRE	32.38%

According to the R² results, the PM2.5 models achieved the best predictive performance, with R² values of 0.945, 0.920, and 0.905 in 2013, 2018, and 2023, respectively.

This indicates that XGBoost accurately captured the nonlinear relationships between PM2.5 and natural background, ecological conditions, and human activity factors. The O3 model showed the highest predictive performance in 2018, with an R² value of 0.915, while the values for 2013 and 2023 were 0.774 and 0.846, respectively, indicating interannual differences in the explanatory ability for ozone pollution. The R² values for LST were 0.785, 0.850, and 0.845, showing relatively stable model performance. By contrast, the R² values for NPP ranged from 0.752 to 0.780. Although these values were lower than those for PM2.5 and LST, they still indicate good explanatory ability, suggesting that carbon sequestration capacity is jointly affected by multiple factors and has a more complex spatial variation mechanism.

In terms of error metrics, the RMSE and MAE values of all models were generally low, and most scatter points of observed and predicted values were distributed near the 1:1 reference line, further indicating good predictive stability and reliability. Therefore, the XGBoost models can be used for subsequent SHAP interpretation analysis to reveal the relative contributions and nonlinear influence mechanisms of key driving factors.

3.4.2. Factor Importance Ranking Based on SHAP

Figures 9–11 show the SHAP interpretation results for LST, PM2.5, O3, and NPP in 2013, 2018, and 2023. Overall, DEM, PRE, PET, and NDVI were the core driving factors affecting the “thermal environment-air pollution-carbon sequestration capacity” system in Henan Province, and they showed high stability across different years and dependent variables.

DEM and PRE had relatively high contribution rates in most models, indicating that topography and precipitation background are fundamental natural factors shaping the spatial differences in the thermal environment, air pollution, and carbon sequestration capacity in Henan Province.

Specifically, the dominant factor for LST was PRE in 2013, with a contribution rate of 39.19%, while DEM became the dominant factor in 2018 and 2023, with contribution rates of 34.46% and 33.45%, respectively. This suggests that the main controlling factors of the thermal environment gradually shifted from moisture conditions to the combined effects of topographic background and surface processes. For PM2.5, the dominant factor was DEM in 2013 and 2018, with contribution rates of 44.33% and 35.94%, respectively, but shifted to PRE in 2023, with a contribution rate of 31.45%.

This indicates that particulate pollution is jointly regulated by topographic dispersion conditions and hydrothermal background. For O₃, DEM was the primary factor in all three years, with contribution rates of 46.07%, 54.05%, and 36.28%, respectively, suggesting that the spatial differences in ozone pollution are closely related to topography and associated meteorological conditions. For NPP, the dominant factor was DEM in 2013, with a contribution rate of 39.00%, but shifted to PRE in 2018 and 2023, with contribution rates of 31.55% and 32.38%, respectively, indicating that the influence of precipitation conditions on ecological productivity and carbon sequestration capacity gradually increased.

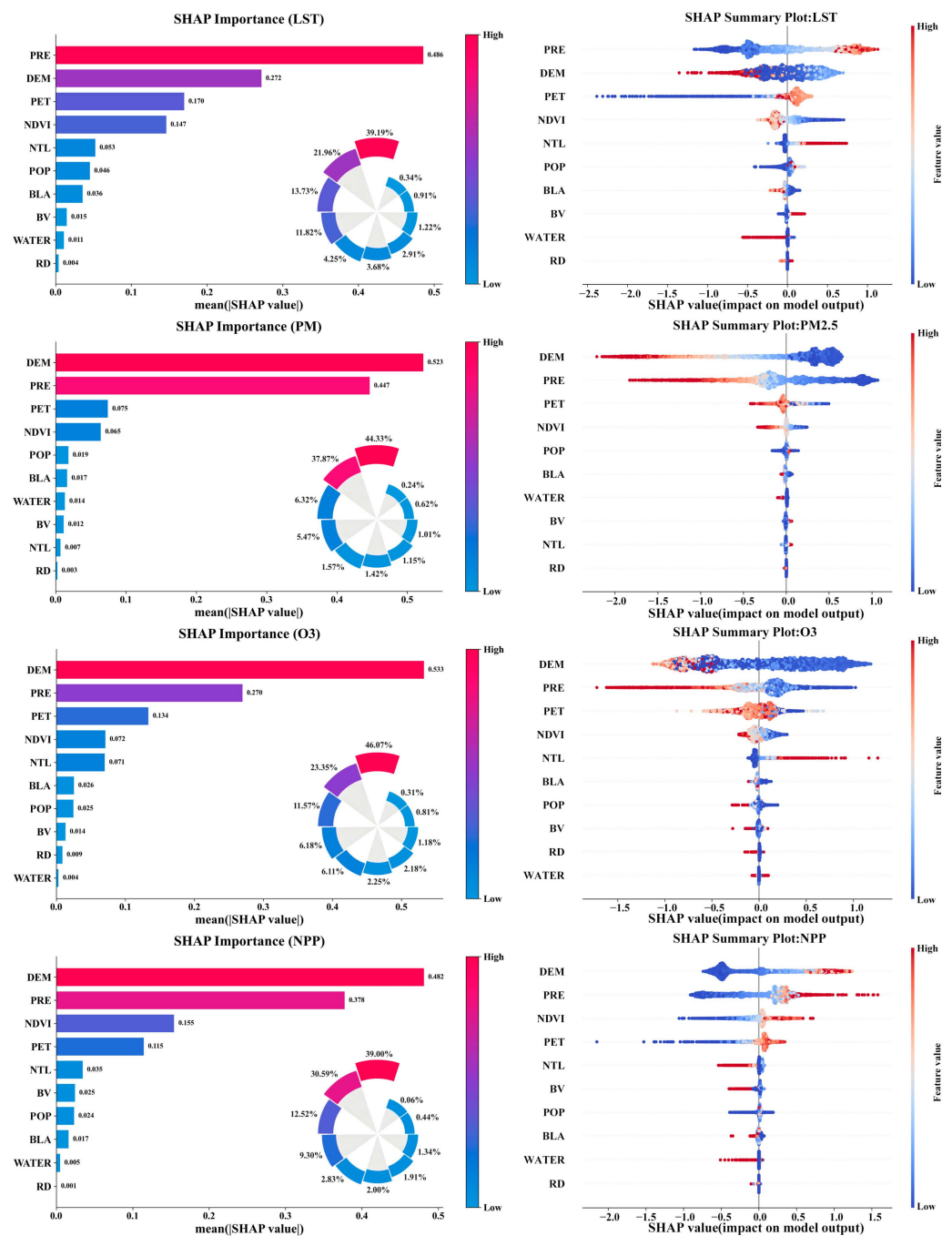


Figure 9. SHAP interpretation results of key driving factors for LST, PM_{2.5}, O₃, and NPP in 2013.

The SHAP beeswarm plots show the influence direction of different factors. NDVI usually made negative contributions to LST, PM_{2.5}, and O₃. It made positive contributions to NPP. This result shows that vegetation cover can help reduce surface heat and air pollution. It can

also improve carbon sequestration capacity. PET usually had positive effects on LST, PM2.5, and O3. This means that stronger heat conditions and higher evapotranspiration demand may increase heat pressure and pollution risk. Human activity factors, such as NTL, BLA, and POP, had lower contributions than natural factors in most models. However, they still helped explain some local differences. This result shows that urbanization factors also affect the links among the thermal environment, air pollution, and carbon sequestration capacity.

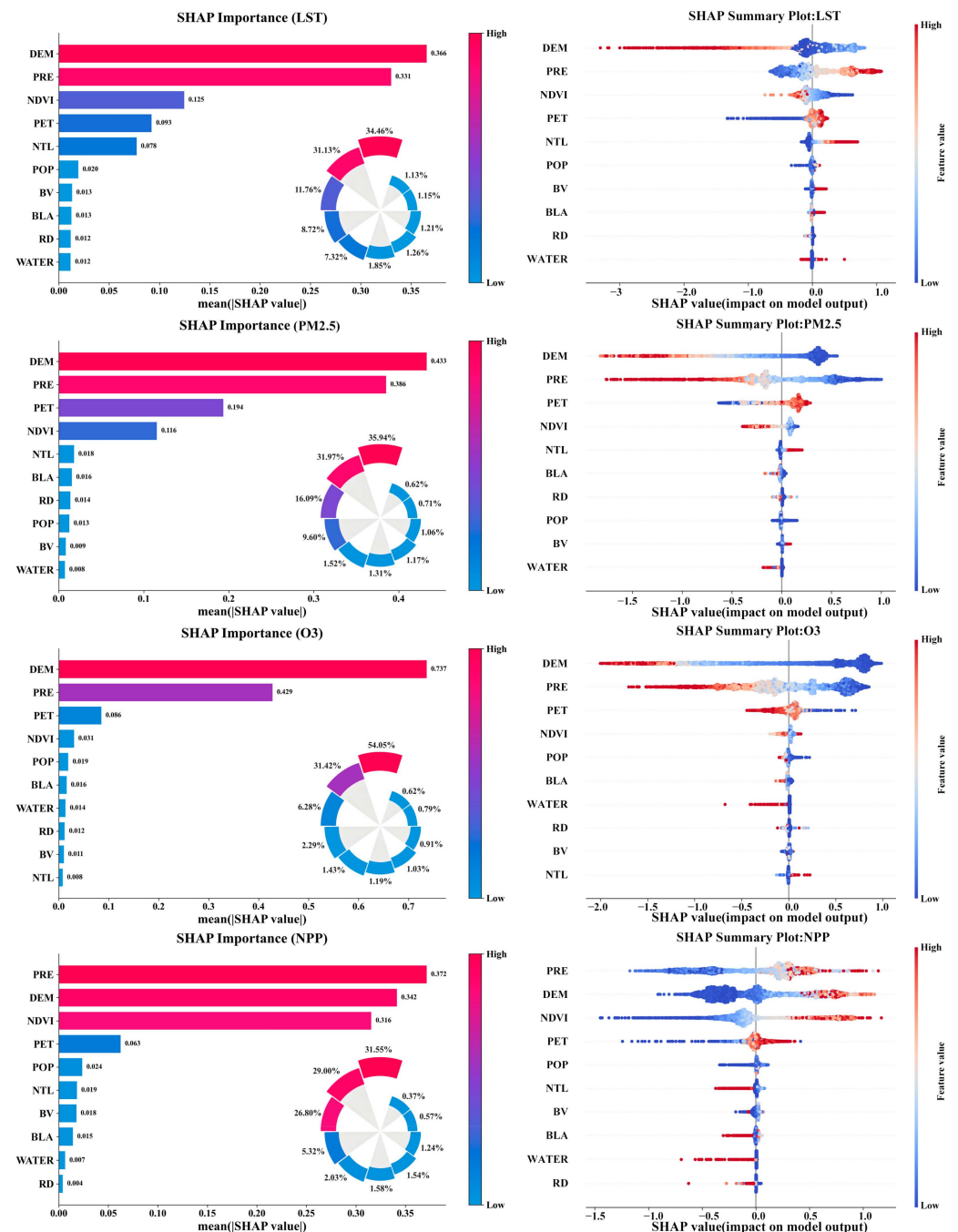


Figure 10. SHAP interpretation results of key driving factors for LST, PM2.5, O3, and NPP in 2018.

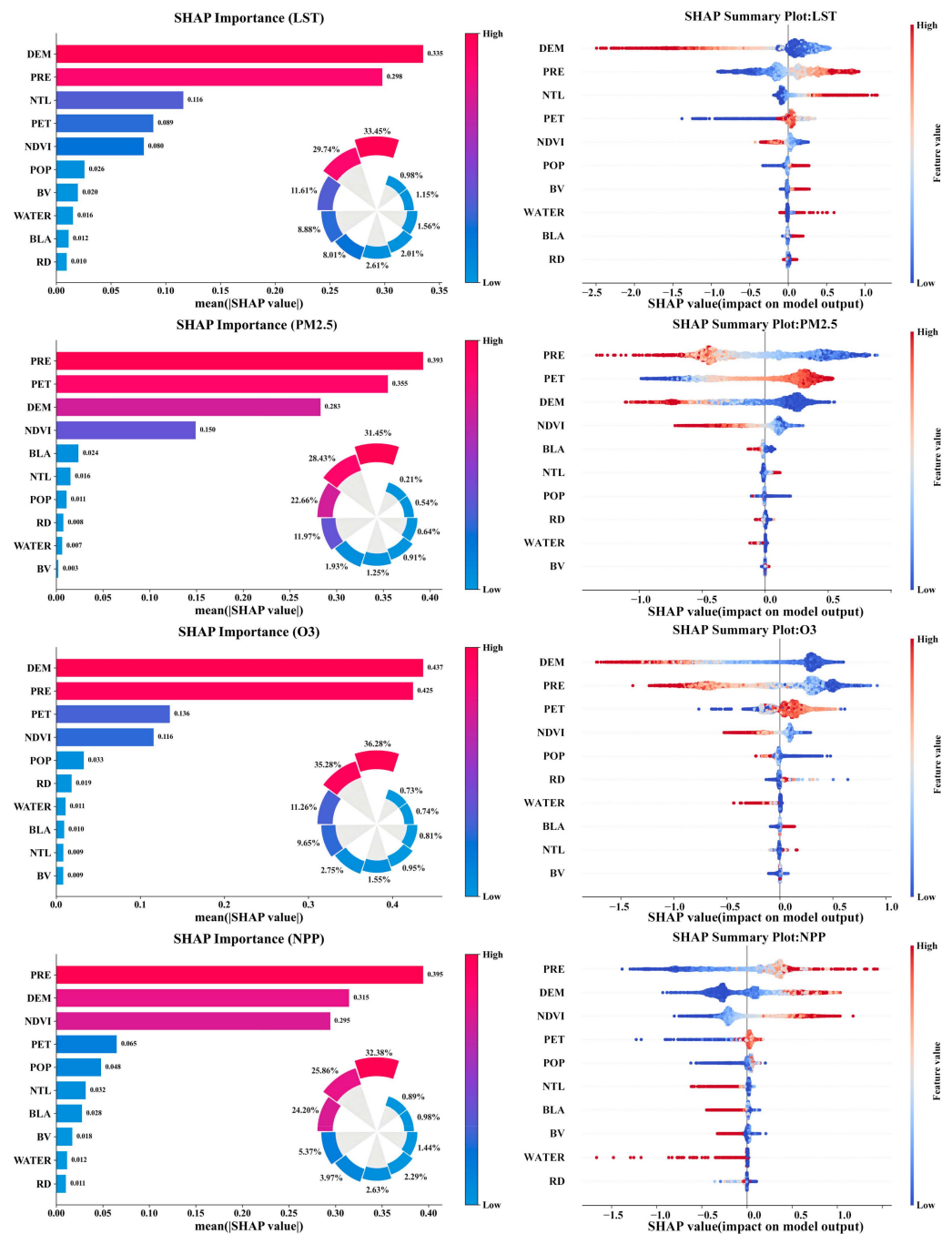


Figure 11. SHAP interpretation results of key driving factors for LST, PM2.5, O3, and NPP in 2023.

3.5. Nonlinear Influence Mechanisms of Key Driving Factors

Figures 12–14 show the SHAP dependence results for 2013, 2018, and 2023. These figures show how key driving factors affect LST, PM2.5, O3, and NPP at different value ranges. Pearson correlation analysis only shows linear relationships. SHAP dependence plots were used to further examine nonlinear responses, showing how each factor's contribution changed with increases or decreases in its value. These plots can also help identify threshold effects, nonlinear increases, and weaker marginal effects.

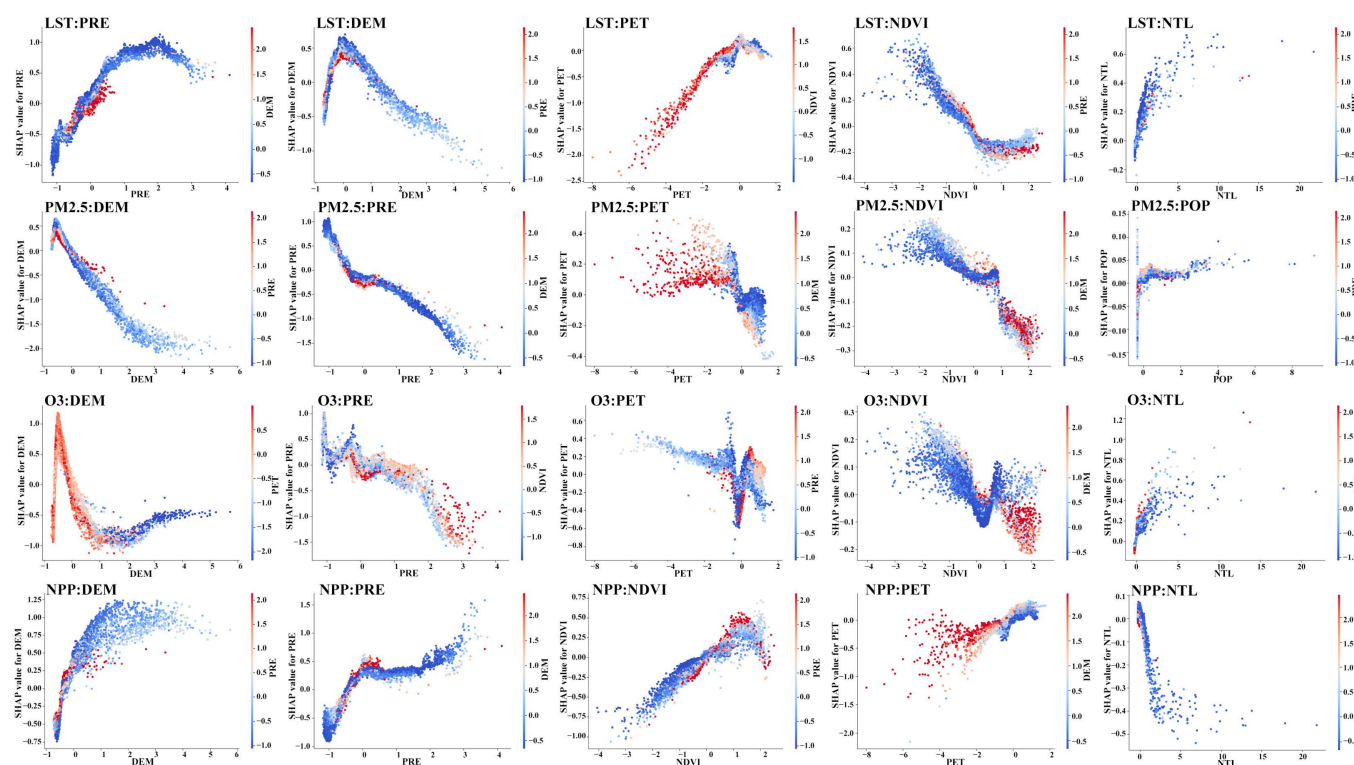


Figure 12. Nonlinear response plots of key driving factors in 2013.

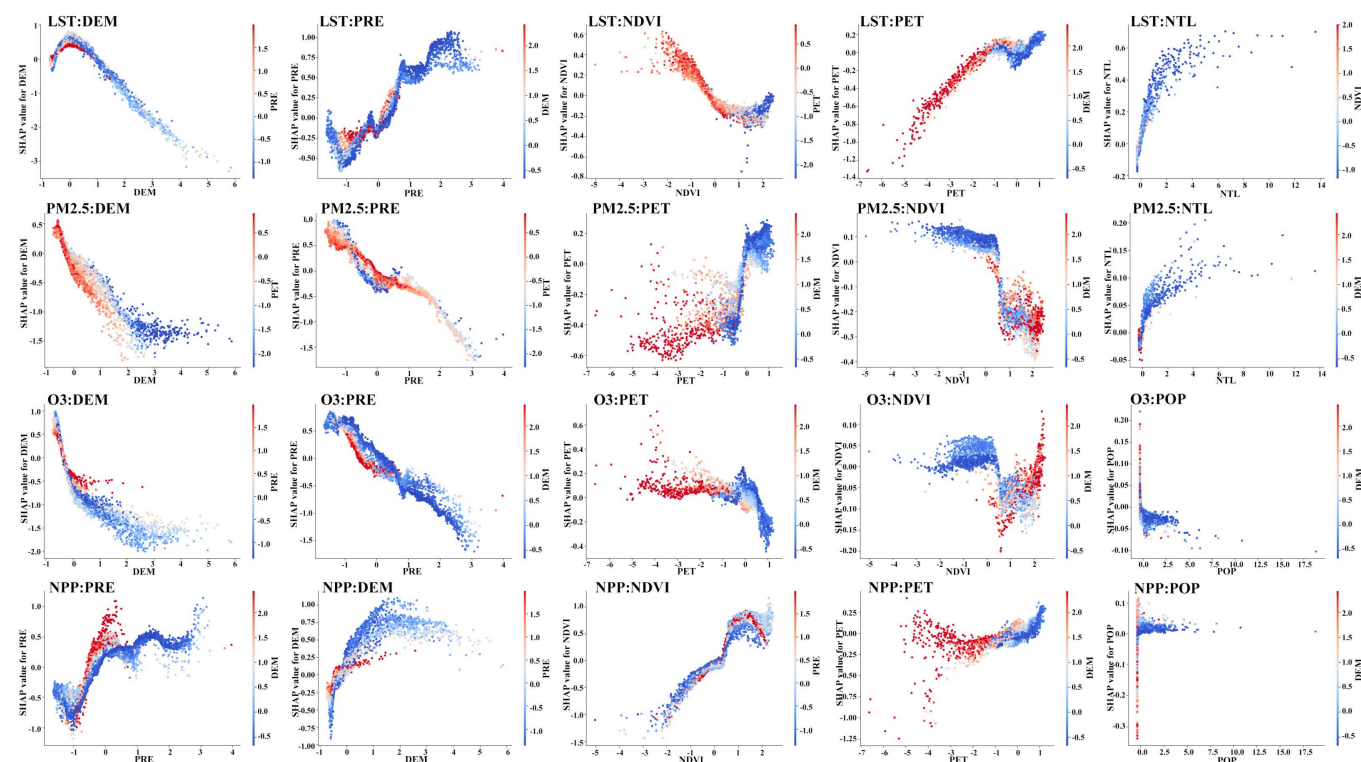


Figure 13. Nonlinear response plots of key driving factors in 2018.

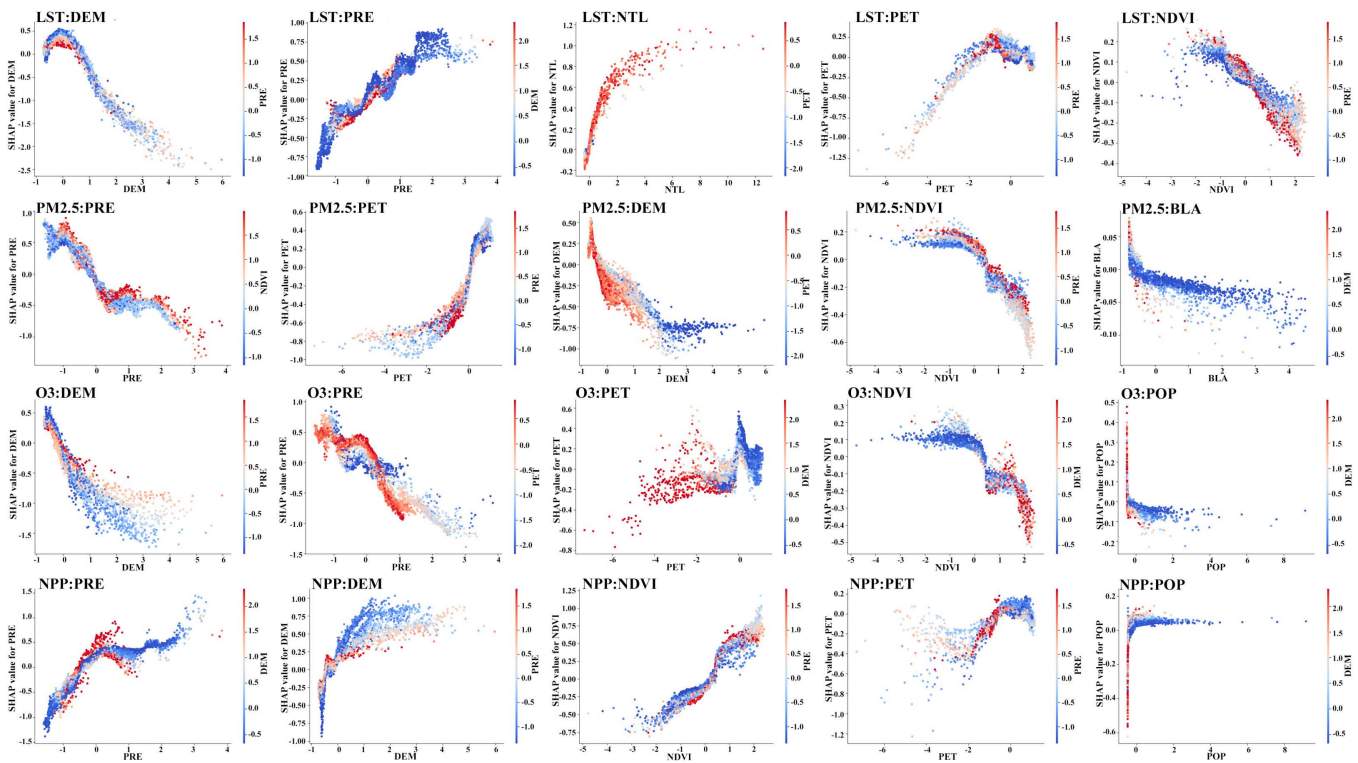


Figure 14. Nonlinear response plots of key driving factors in 2023.

The 2013 SHAP results show nonlinear changes between the environmental variables and the driving factors. Some factors also showed threshold effects and interactions. Overall, NDVI mainly had negative effects on LST and PM2.5. This means that higher vegetation cover can help reduce surface temperature and particulate pollution. NDVI also had a positive effect on NPP. PRE usually reduced PM2.5 and O3, but it increased NPP. DEM affected LST, PM2.5, O3, and NPP in nonlinear ways. Its effects changed at some points along the elevation gradient. PET mostly had positive effects on LST and NPP. Its effects on PM2.5 and O3 were less stable. NTL usually increased LST and O3, but it reduced NPP. This result shows that stronger human activity may raise surface temperature and ozone risk, while reducing vegetation productivity. These results show that both natural factors and human activities affected ecological and environmental changes in 2013. The same factor also had different effects on different target variables.

The 2018 SHAP results show that environmental variables responded to natural factors and human activities in nonlinear ways. Some factors also showed threshold effects and interactions. Overall, DEM mainly had negative effects on LST, PM2.5, and O3. As elevation increased, the SHAP values for these variables decreased. DEM had a positive effect on NPP. This effect increased at first and then became stable. PRE mainly reduced PM2.5 and O3, but it increased NPP. This result shows that more precipitation can help reduce air pollution and support vegetation productivity.

NDVI mainly contributed negatively to LST and PM2.5, suggesting that higher vegetation cover helps reduce land surface temperature and particulate matter concentration, while it had a significant positive effect on NPP. PET generally had positive effects on LST and NPP, whereas its responses to PM2.5 and O3 showed segmented changes and fluctuations. NTL mostly contributed positively to LST, PM2.5, and O3, indicating that intensified human activities may aggravate thermal environmental pressure and pollution risks. POP had relatively weak effects on O3 and NPP overall, but fluctuated more noticeably in low-population-density areas. These findings indicate that the ecological

and environmental pattern in 2018 was jointly shaped by topography, climate, vegetation, and human activities, with marked differences in the response direction and intensity of different indicators to the same factor.

The nonlinear response patterns in 2023 further show that ecological and environmental indicators still exhibited significant nonlinear responses, threshold effects, and interactions with topographic, climatic, vegetation, and human activity factors. Overall, DEM mainly had negative effects on LST, PM_{2.5}, and O₃, with SHAP values generally decreasing as elevation increased, whereas its effect on NPP increased rapidly at first and then became stable. PRE had clear inhibitory effects on PM_{2.5} and O₃, but generally promoted LST and NPP, suggesting that increased precipitation helps reduce pollutant concentrations and improve vegetation productivity. NDVI showed overall negative contributions to LST, PM_{2.5}, and O₃, indicating that enhanced vegetation cover can mitigate thermal environmental pressure and air pollution, while it had a significant positive effect on NPP. PET mostly had positive effects on LST and NPP, whereas its impacts on PM_{2.5} and O₃ showed segmented changes and local fluctuations. NTL had a clear positive contribution to LST, reflecting that intensified human activities may aggravate the surface thermal effect. POP had relatively weak effects on O₃ and NPP overall, with greater fluctuations mainly in low-population-density ranges. These results suggest that regional ecological and environmental changes in 2023 were jointly driven by natural conditions and human activities, and that the direction and intensity of each factor varied significantly across different target variables.

4. Discussion

Based on the results of Pearson correlation analysis, multiple linear regression, and XGBoost-SHAP interpretable machine learning, this study identified consistent relationships between the environmental response variables and their driving factors in Henan Province. The spatial variation in LST, PM_{2.5}, O₃, and NPP was jointly associated with topography, hydrothermal conditions, vegetation cover, and human activities. DEM, PRE, PET, and NDVI generally showed relatively large regression coefficients and high SHAP importance across different years and response variables, indicating that natural background and ecological factors played important roles in explaining spatial differences in the thermal environment, air pollution, and carbon sequestration capacity [9,12,37].

Anthropogenic variables, including NTL, POP, BLA, BV, and RD, generally showed lower SHAP importance and smaller standardized regression coefficients than the dominant natural factors. Nevertheless, several consistent associations were identified. In particular, NTL was positively associated with LST and negatively associated with NPP across the three years, indicating that intensified human activity was statistically associated with greater thermal stress and lower ecological productivity. The directions and magnitudes of the coefficients for POP, BLA, BV, and RD varied across response variables and years, suggesting that their environmental associations were less temporally stable and may depend on local land-use and ecological conditions.

4.1. Integrated Roles of Natural, Ecological, and Anthropogenic Factors

The results highlight the important roles of natural background factors in the spatial variation in the thermal environment, air pollution, and carbon sequestration capacity in Henan Province. DEM showed relatively large regression coefficients and high SHAP importance in the PM_{2.5} and O₃ models and also contributed substantially to the LST and NPP models in some years. Low-elevation areas are generally characterized by denser populations and more intensive construction activities, which may be associated with greater pollutant accumulation and surface thermal pressure. By contrast, high-elevation areas often have lower development intensity and more favorable pollutant dispersion

conditions. PRE and PET represent the regional hydrothermal background. Their associations with LST, PM_{2.5}, O₃, and NPP varied across response variables and exhibited nonlinear patterns in the SHAP dependence plots. Precipitation may be related to air pollution, surface moisture, and vegetation productivity through wet deposition and water availability, whereas PET is closely associated with thermal conditions and atmospheric evaporative demand [9,23,41].

NDVI was an important ecological factor across the three analytical approaches. Pearson correlation analysis showed that NDVI was generally negatively correlated with LST, PM_{2.5}, and O₃ but positively correlated with NPP. After the other explanatory variables were controlled for, the multiple linear regression results showed the same overall association pattern. The SHAP results further indicated that increasing NDVI was generally associated with negative contributions to the predictions of LST, PM_{2.5}, and O₃ and positive contributions to NPP. These consistent findings suggest that vegetation cover is closely related to surface cooling, pollution mitigation, and ecological productivity [1,8,13,18,19]. Possible mechanisms include shading, evapotranspiration, surface-energy regulation, particulate interception, and improved habitat conditions. Therefore, vegetation cover represents an important ecological factor for the coordinated management of heat, air pollution, and carbon sequestration in Henan Province.

The dominant roles of topography, hydrothermal conditions, and vegetation identified in Henan Province are broadly consistent with findings reported for other regions. Previous studies conducted in different climatic and urban contexts have similarly shown that higher vegetation cover is generally associated with lower land surface temperature and improved ecosystem productivity, mainly through shading, evapotranspiration, and surface-energy regulation [5,10,20,24,25]. Studies from other regions have also reported that elevation, precipitation, and atmospheric dispersion conditions are important determinants of the spatial distributions of PM_{2.5} and O₃ [11,12,15,41,42]. However, the relative importance and response direction of these factors vary among regions because of differences in climate, topography, urban form, emission structure, and vegetation type. Compared with highly urbanized regions where traffic and industrial emissions may dominate environmental variation, the results for Henan Province indicate a relatively strong contribution from topographic and hydrothermal background conditions. These similarities and differences suggest that the environmental effects of individual driving factors are context-dependent and that conclusions derived from one region should not be directly generalized to other geographical settings.

The anthropogenic variables were used to characterize urbanization intensity, land development, and traffic-related activities. The regression results showed that their association directions and magnitudes varied across environmental response variables and years. NTL was consistently positively associated with LST and negatively associated with NPP, suggesting that areas with stronger human activity tended to experience greater thermal stress and lower ecological productivity. BV was generally positively associated with LST and negatively associated with NPP, whereas the coefficients of POP, BLA, and RD were smaller or changed direction across years and response variables. These results indicate that anthropogenic factors remain relevant to regional environmental variation, although their linear associations were generally less stable than those of the dominant natural factors [20,33,38–40].

The observed variations in PM_{2.5} and O₃ may also be influenced by emission- and meteorology-related processes that were not fully represented by the explanatory variables used in this study. Changes in emission-control policies, industrial structure, energy consumption, traffic activity, and precursor emissions can alter the formation and spatial distribution of atmospheric pollutants. PM_{2.5} concentrations are affected by primary emissions, secondary aerosol formation, wet and dry deposition, and regional transport [6,28,36,42]. O₃ is a secondary pollutant produced through complex photochemical

reactions involving nitrogen oxides (NO_x) and volatile organic compounds (VOCs) under favorable meteorological conditions. Solar radiation, temperature, humidity, wind conditions, and boundary-layer height can therefore influence O₃ formation, pollutant accumulation, dispersion, and transport [11,15,23,34]. In addition, regional transport may introduce pollutants and precursors from neighboring areas. Consequently, the observed changes in PM_{2.5} and O₃ should not be attributed solely to the spatial factors included in the present models but should be interpreted as the combined outcome of local emissions, meteorological variability, atmospheric chemical processes, and regional transport.

4.2. Implications for Coordinated Ecological-Environmental Governance in Henan Province

The results of this study provide implications for coordinated ecological and environmental governance in Henan Province. First, thermal environment mitigation, air pollution control, and carbon sequestration enhancement should be incorporated into an integrated governance framework. Traditional governance often treats the urban thermal environment, PM_{2.5} pollution, O₃ pollution, and ecological carbon sinks as relatively independent issues. However, this study shows that these environmental responses share several key driving factors, especially NDVI, DEM, PRE, and PET, which exert cross-target effects. Therefore, future governance should shift from single-objective control to multi-objective coordination, so as to avoid situations in which a policy is effective for one environmental target but offsets another [3,35,43–45].

Second, the regulatory role of vegetation cover and ecological space should be strengthened. NDVI showed relatively consistent associations with LST, PM_{2.5}, O₃, and NPP, indicating that increasing high-quality vegetation cover is an important approach to coordinated cooling, pollution reduction, and carbon sequestration enhancement [1,2,8,13,19]. In urban renewal, territorial spatial planning, and ecological restoration, priority should be given to protecting and improving urban green spaces, agricultural ecological spaces, riparian vegetation, and mountainous ecological barriers. In areas where high thermal risk, severe pollution, and low carbon sequestration capacity overlap, continuous green spaces should be expanded to enhance ecosystem regulation capacity.

Third, differentiated governance should be implemented according to topographic and hydrothermal conditions. The effects of DEM, PRE, and PET on each dependent variable showed clear spatial differences, indicating that the causes of environmental problems vary across regions in Henan Province. In low-elevation areas with high construction intensity and weak pollution dispersion conditions, priority should be given to controlling construction expansion, traffic emissions, and anthropogenic heat emissions. In areas sensitive to hydrothermal conditions, blue–green space construction and ecological water replenishment capacity should be strengthened. In areas with high ecological productivity, protection should be enhanced to prevent built-up land expansion from weakening regional carbon sequestration functions.

Finally, greater attention should be paid to anthropogenic factors associated with urban development. Although most human activity variables showed lower SHAP importance and smaller standardized regression coefficients than the dominant natural factors, their associations with the environmental responses were not negligible. In particular, NTL and BV showed relatively consistent relationships with increased LST and reduced NPP. Therefore, the promotion of new-type urbanization should be accompanied by controls on disorderly built-up land expansion, improvements in land-use efficiency, and optimization of transportation and industrial layouts. These measures may help reduce the combined environmental pressure associated with high-intensity development [46–48].

4.3. Limitations and Future Research

Several limitations remain. First, this study selected three time points, 2013, 2018, and 2023, which can reflect the phased changes in the thermal environment, air pollution, and

carbon sequestration capacity in Henan Province. However, it cannot fully reveal dynamic processes at continuous annual or seasonal scales. Thermal conditions and air pollution show obvious seasonal variations, especially O₃, which is highly sensitive to temperature, radiation, and precursor conditions [11,15,23]. Future studies could introduce monthly or seasonal data to improve the explanation of short-term mechanisms.

Second, this study mainly examined natural, ecological, and human activity factors at the grid scale and did not fully incorporate detailed emission- and meteorology-related variables. Specifically, emission inventories, industrial structure, energy consumption, traffic emissions, NO_x and VOC emissions, solar radiation, wind conditions, and boundary-layer height were not directly included. In addition, the present study considered road density but did not separately incorporate railway density or railway-related transport intensity. However, PM_{2.5} and O₃ are influenced by primary emissions, secondary formation, photochemical reactions, atmospheric dispersion, and regional transport [6,34,36,42]. The omission of these variables and processes may limit the mechanistic interpretation of the observed pollutant variations. Future studies should integrate high-resolution emission inventories, detailed meteorological observations, road and railway indicators, atmospheric chemistry models, and regional transport analyses to more comprehensively explain the formation and spatial variation in PM_{2.5} and O₃.

Third, multiple linear regression and XGBoost-SHAP provide complementary perspectives but also have methodological limitations. Multiple linear regression provides interpretable estimates of independent linear associations but assumes linear relationships and independent residuals [22,23,36]. Because the observations were derived from spatial raster data, residual spatial autocorrelation may affect coefficient uncertainty and statistical significance. Moreover, the large sample size may result in statistically significant coefficients with limited practical magnitude. XGBoost-SHAP can capture nonlinear response patterns and improve model interpretability, but its results also represent statistical associations rather than strict causal evidence. Future studies could combine spatial regression, geographically weighted regression, spatial panel models, causal inference methods, and scenario simulations to improve the robustness of the analysis [33,35,39,41]. In addition, future research may incorporate explicit multiscale spatial-complexity descriptors to better characterize the heterogeneous organization of environmental patterns across scales. Recent studies in human-settlement research have shown that reproducible, scale-explicit metrics can provide a quantitative basis for describing complex spatial organization, although such metrics should be interpreted as comparative descriptors rather than direct causal mechanisms [49–51].

The XGBoost-SHAP model used in this study has clear strengths. It can analyze nonlinear driving effects. It can also measure the importance of different variables for the thermal environment, air pollution, and carbon sequestration capacity. However, it still has some limits in geospatial modeling. First, XGBoost is a tree-based ensemble learning model. It usually treats observations as independent during modeling. This setting may ignore part of the spatial autocorrelation and spatial differences in geographical data. Second, the model depends on high-quality spatial data. Spatial resolution can affect the relationships between driving factors and response variables. Different resolutions may therefore bring uncertainty when the model is used at different scales. Future studies could add spatial weight matrices to machine learning models. They could also combine XGBoost-SHAP with spatial methods such as geographically weighted regression (GWR) [30,38,39,41]. These methods may help the model better describe spatial interactions and the driving links in complex surface environmental systems.

Overall, this study examined the relationships between the environmental response variables and their driving factors from four perspectives: bivariate correlations, independent linear associations, factor importance, and nonlinear responses. The results indicate

that ecological and environmental management in Henan Province should simultaneously consider natural background constraints, ecological regulation, and anthropogenic disturbance. Future management should move beyond single-target interventions and integrate multiple environmental objectives with region-specific strategies [10,14]. Such an integrated approach may contribute to surface-heat mitigation, air pollution control, and carbon sequestration enhancement.

5. Conclusions

Taking Henan Province as the study area, this study selected 2013, 2018, and 2023 as representative years. Land surface temperature (LST), particulate matter concentration (PM_{2.5}), ozone concentration (O₃), and net primary productivity (NPP) were used to characterize the thermal environment, air pollution, and carbon sequestration capacity. Pearson correlation analysis, multiple linear regression, and XGBoost-SHAP were combined to examine the relationships between natural, ecological, and anthropogenic factors and the four environmental response variables. The analysis focused on bivariate associations, independent linear associations, factor importance, nonlinear responses, and potential threshold characteristics. The main conclusions are as follows.

First, LST, PM_{2.5}, O₃, and NPP showed marked spatial heterogeneity in Henan Province. LST and air pollution were generally higher in areas characterized by more intensive construction and stronger human activities, whereas NPP was generally higher in areas with better vegetation cover and ecological conditions. Pearson correlation analysis showed that NDVI, DEM, PRE, and PET were closely related to the four environmental response variables. In particular, NDVI was generally negatively correlated with LST, PM_{2.5}, and O₃ but positively correlated with NPP, indicating that greater vegetation cover was associated with lower thermal and pollution levels and higher ecological productivity.

Second, the multiple linear regression models showed varying explanatory power across the four environmental response variables. The R^2 values ranged from 0.574 to 0.690 for LST, from 0.238 to 0.720 for O₃, from 0.810 to 0.848 for PM_{2.5}, and from 0.574 to 0.625 for NPP. Natural background factors generally showed relatively large standardized regression coefficients. PRE and PET were consistently positively associated with LST, whereas DEM and PRE were negatively associated with PM_{2.5} and O₃. DEM, PRE, PET, and NDVI were positively associated with NPP. NDVI showed an environmentally favorable association pattern, being negatively associated with LST, PM_{2.5}, and O₃ but positively associated with NPP. In comparison, most anthropogenic variables showed smaller and less temporally stable coefficients.

Third, the XGBoost models showed good predictive performance for the four environmental response variables, particularly for PM_{2.5}, with R^2 values of 0.945, 0.920, and 0.905 in 2013, 2018, and 2023, respectively. The models also showed relatively stable performance for LST and O₃, whereas the R^2 values for NPP ranged from 0.752 to 0.780. The SHAP importance results identified DEM, PRE, PET, and NDVI as the main contributors to model predictions. Across the three years, the most important variable for LST shifted from PRE in 2013 to DEM in 2018 and 2023; that for PM_{2.5} shifted from DEM to PRE; DEM remained the most important variable for O₃; and the most important variable for NPP shifted from DEM in 2013 to PRE in 2018 and 2023.

Fourth, the SHAP dependence plots revealed clear nonlinear responses and potential threshold characteristics. DEM generally showed negative SHAP contributions to PM_{2.5} and O₃ predictions, indicating that higher elevations were associated with lower predicted pollution levels. PRE showed negative contributions to PM_{2.5} and O₃ but generally positive contributions to NPP. PET showed predominantly positive contributions to LST and NPP, although its relationships with the air pollution variables varied across different value ranges. NDVI generally showed negative contributions to LST, PM_{2.5},

and O₃ but positive contributions to NPP, further demonstrating the close relationship of vegetation cover with cooling, pollution mitigation, and ecological productivity.

Overall, the thermal environment, air pollution, and carbon sequestration capacity in Henan Province were associated with a shared set of natural, ecological, and anthropogenic factors. Future ecological and environmental governance should move beyond single-objective interventions and place greater emphasis on coordinated management. Particular attention should be given to vegetation cover, topographic and hydrothermal constraints, and anthropogenic disturbance. Improving blue-green space patterns, controlling disorderly built-up land expansion, enhancing vegetation quality, and adopting region-specific management strategies may contribute to coordinated improvements in heat mitigation, air pollution control, and carbon sequestration capacity across Henan Province.

This study has several limitations. The analysis was based on three representative years rather than a continuous temporal series, and detailed emission inventories, atmospheric chemical processes, meteorological variables, railway-related indicators, and regional pollutant transport were not fully incorporated. In addition, the use of spatial raster observations may introduce spatial autocorrelation and scale-related uncertainty. Therefore, the results should be interpreted as statistical associations rather than causal effects. Future studies should integrate continuous temporal observations, high-resolution emission and meteorological data, railway and other transportation indicators, spatial modeling approaches, and atmospheric chemistry models to further validate and extend the present findings.

Author Contributions: Writing—original draft, Methodology, Data curation and Visualization, C.L.; Conceptualization, Supervision, Writing—review and editing, S.Z. (Shaowei Zhang); Methodology, Validation, L.S.; Data curation, Formal analysis, Writing—review and editing, C.Z.; Methodology, Validation, Data curation, P.J. Methodology, Validation, S.Z. (Shennian Zhang); All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Henan Provincial Soft Science Research Program (Grant No. 252400410525).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The author sincerely appreciates this valuable support, which provided an important foundation for the successful completion of this study and the continued development of related research.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Abbreviations

The following abbreviations are used in this manuscript:

LST	Land Surface Temperature
PM _{2.5}	Particulate Matter with an aerodynamic diameter of 2.5 µm or less
O ₃	Ozone
NPP	Net Primary Productivity
WATER	Water area
BLA	Built-up land area
PRE	Precipitation
PET	Potential evapotranspiration

NDVI	Normalized Difference Vegetation Index
NTL	Nighttime Light
DEM	Digital elevation model
BV	Building Volume
RD	Road Density
POP	Population Density

Appendix A

Multiple Linear Regression Results and Multicollinearity Diagnostics

Table A1. Multicollinearity Test from 2013 to 2023.

Variable	2013 VIF	2018 VIF	2023 VIF
BV	2.196	2.203	2.251
DEM	2.713	4.610	4.432
NDVI	1.908	2.902	3.165
NTL	1.907	1.962	1.991
POP	1.116	1.130	2.082
PET	2.765	4.165	4.104
WATER	1.092	1.109	1.132
BLA	2.169	2.453	2.944
RD	1.167	1.468	1.634
PRE	1.185	1.410	1.577

Note: All VIF values for the three years were below 5, indicating that the models did not exhibit serious multicollinearity.

Table A2. 2013 Multiple Linear Regression Results.

Variable	LST β (t-Value)	O3 β (t-Value)	PM2.5 β (t-Value)	NPP β (t-Value)
Intercept	−0.007 (−4.189 ***)	0.008 (3.654 ***)	0.000 (0.416)	−0.046 (−27.642 ***)
BV	0.093 (25.648 ***)	−0.032 (−6.566 ***)	0.023 (10.421 ***)	−0.072 (−19.478 ***)
DEM	0.056 (21.661 ***)	−0.293 (−83.601 ***)	−0.864 (−541.507 ***)	0.788 (295.285 ***)
NDVI	−0.306 (−128.553 ***)	−0.196 (−61.032 ***)	−0.126 (−86.398 ***)	0.213 (87.208 ***)
NTL	0.139 (45.190 ***)	0.141 (34.039 ***)	0.013 (6.728 ***)	−0.073 (−23.310 ***)
POP	0.004 (1.611)	−0.013 (−3.728 ***)	0.016 (9.793 ***)	−0.007 (−2.513 *)
PET	0.388 (146.985 ***)	−0.123 (−34.519 ***)	−0.249 (−153.457 ***)	0.496 (182.786 ***)
WATER	−0.117 (−43.082 ***)	−0.018 (−4.844 ***)	−0.051 (−30.736 ***)	−0.041 (−14.669 ***)
BLA	−0.103 (−37.626 ***)	0.045 (12.302 ***)	−0.015 (−8.986 ***)	−0.059 (−20.911 ***)
RD	0.016 (8.162 ***)	−0.029 (−11.190 ***)	−0.014 (−11.891 ***)	0.001 (0.266)
PRE	0.553 (315.616 ***)	−0.265 (−112.229 ***)	−0.566 (−526.253 ***)	0.349 (193.776 ***)
R ²	0.574	0.238	0.848	0.574
Adjusted R ²	0.574	0.238	0.848	0.574
F-statistic	21,100.124 ***	4905.423 ***	87,122.081 ***	21,086.135 ***
N	156,640	156,640	156,640	156,640

Note: Standardized regression coefficients are reported, with t-values in parentheses. * $p < 0.05$ and *** $p < 0.001$.

Table A3. 2018 Multiple Linear Regression Results.

Variable	LST β (t-Value)	O3 β (t-Value)	PM2.5 β (t-Value)	NPP β (t-Value)
Intercept	−0.005 (−3.454 ***)	−0.001 (−0.801)	−0.002 (−1.437)	−0.048 (−30.312 ***)
BV	0.124 (36.053 ***)	−0.041 (−14.039 ***)	−0.019 (−8.446 ***)	−0.045 (−13.272 ***)
DEM	−0.074 (−22.404 ***)	−1.138 (−402.526 ***)	−0.803 (−361.452 ***)	0.785 (238.997 ***)
NDVI	−0.141 (−50.118 ***)	−0.018 (−7.327 ***)	−0.125 (−66.166 ***)	0.394 (140.876 ***)
NTL	0.163 (57.194 ***)	−0.018 (−7.357 ***)	0.011 (5.990 ***)	−0.069 (−24.511 ***)
POP	−0.032 (−12.503 ***)	0.024 (10.791 ***)	0.019 (10.919 ***)	−0.004 (−1.487)
PET	0.441 (139.226 ***)	−0.502 (−185.651 ***)	−0.095 (−44.522 ***)	0.506 (161.135 ***)
WATER	0.053 (20.773 ***)	−0.111 (−50.921 ***)	−0.058 (−33.951 ***)	−0.027 (−10.865 ***)
BLA	−0.044 (−15.968 ***)	0.091 (38.197 ***)	0.018 (9.380 ***)	−0.009 (−3.136 **)
RD	0.026 (12.019 ***)	−0.027 (−14.198 ***)	−0.020 (−13.425 ***)	0.001 (0.575)
PRE	0.431 (231.532 ***)	−0.593 (−372.992 ***)	−0.573 (−458.802 ***)	0.451 (244.455 ***)
R ²	0.596	0.720	0.828	0.621
Adjusted R ²	0.596	0.720	0.828	0.621
F-statistic	23,087.362 ***	40,208.522 ***	75,314.798 ***	25,715.578 ***
N	156,742	156,742	156,742	156,742

Note: Standardized regression coefficients are reported, with t-values in parentheses. ** $p < 0.01$, and *** $p < 0.001$.

Table A4. 2023 Multiple Linear Regression Results.

Variable	LST β (t-Value)	O3 β (t-Value)	PM2.5 β (t-Value)	NPP β (t-Value)
Intercept	−0.003 (−2.398 *)	−0.002 (−1.114)	−0.001 (−0.714)	−0.044 (−27.887 ***)
BV	0.099 (33.235 ***)	−0.029 (−9.375 ***)	0.000 (0.197)	−0.090 (−26.453 ***)
DEM	−0.248 (−88.262 ***)	−0.757 (−258.160 ***)	−0.511 (−222.751 ***)	0.695 (216.994 ***)
NDVI	−0.023 (−9.111 ***)	−0.125 (−47.442 ***)	−0.267 (−129.054 ***)	0.321 (111.188 ***)
NTL	0.211 (88.506 ***)	0.012 (4.723 ***)	0.003 (1.438)	−0.078 (−28.867 ***)
POP	0.042 (12.149 ***)	−0.022 (−6.185 ***)	0.033 (11.502 ***)	0.065 (16.433 ***)
PET	0.364 (133.514 ***)	−0.244 (−86.057 ***)	0.017 (7.630 ***)	0.488 (157.416 ***)
WATER	0.051 (23.243 ***)	−0.090 (−39.642 ***)	−0.093 (−52.285 ***)	−0.046 (−18.452 ***)
BLA	0.063 (24.108 ***)	0.066 (24.274 ***)	−0.049 (−22.856 ***)	−0.022 (−7.300 ***)
RD	0.027 (14.055 ***)	−0.012 (−5.952 ***)	−0.024 (−15.198 ***)	−0.036 (−16.388 ***)
PRE	0.477 (279.208 ***)	−0.555 (−311.776 ***)	−0.540 (−387.517 ***)	0.523 (268.816 ***)
R ²	0.690	0.686	0.810	0.625
Adjusted R ²	0.690	0.686	0.810	0.625
F-statistic	34,865.458 ***	34,239.407 ***	66,552.943 ***	26,118.767 ***
N	156,596	156,596	156,596	156,596

Note: Standardized regression coefficients are reported, with t-values in parentheses. * $p < 0.05$ and *** $p < 0.001$.

References

- Feng, R.; Wang, F.; Liu, S.; Qi, W.; Zhengchen, R.; Wang, D. Synergistic effects of urban forest on urban heat island-air pollution-carbon stock in mega-urban agglomeration. *Urban For. Urban Green.* **2025**, *103*, 128590. [\[CrossRef\]](#)
- Wang, K.; Du, J.; Li, C.; Li, M.; Chen, C.; Wang, Z.; Chen, J. Uncovering the coupling mechanisms among urban form, surface thermal environments and air pollution. *Sustain. Cities Soc.* **2025**, *130*, 106537. [\[CrossRef\]](#)
- Wu, Q.; Huang, Y.; Irga, P.; Kumar, P.; Li, W.; Wei, W.; Shon, H.K.; Lei, C.; Zhou, J.L. Synergistic control of urban heat island and urban pollution island effects using green infrastructure. *J. Environ. Manag.* **2024**, *370*, 122985. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wu, T.; Qiao, Z. Synergizing climate, energy, and air quality: Uncovering urban heat island, energy consumption, carbon emission, and air pollution nexus linkages. *Environ. Impact Assess. Rev.* **2026**, *117*, 108202. [\[CrossRef\]](#)
- Zhou, D.; Zhao, S.; Zhang, L.; Sun, G.; Liu, Y. The footprint of urban heat island effect in China. *Sci. Rep.* **2015**, *5*, 11160. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, Y.; Gao, W.; Wang, S.; Song, T.; Gong, Z.; Ji, D.; Wang, L.; Liu, Z.; Tang, G.; Huo, Y.; et al. Contrasting trends of PM_{2.5} and surface-ozone concentrations in China from 2013 to 2017. *Natl. Sci. Rev.* **2020**, *7*, 1331–1339. [\[CrossRef\]](#) [\[PubMed\]](#)
- Running, S.; Zhao, M. MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500m SIN Grid V061; Dataset; NASA Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA, 2021. [\[CrossRef\]](#)
- Cao, W.; Zhou, W.; Yu, W.; Wu, T. Combined effects of urban forests on land surface temperature and PM_{2.5} pollution in the winter and summer. *Sustain. Cities Soc.* **2024**, *104*, 105309. [\[CrossRef\]](#)
- Ji, Y.; Zhou, G.; Luo, T.; Dan, Y.; Zhou, L.; Lv, X. Variation of net primary productivity and its drivers in China's forests during 2000–2018. *For. Ecosyst.* **2020**, *7*, 15. [\[CrossRef\]](#)
- Yuan, Y.; Li, C.; Geng, X.; Yu, Z.; Fan, Z.; Wang, X. Natural-anthropogenic environment interactively causes the surface urban heat island intensity variations in global climate zones. *Environ. Int.* **2022**, *170*, 107574. [\[CrossRef\]](#) [\[PubMed\]](#)
- Guo, B.; Wu, H.; Pei, L.; Zhu, X.; Zhang, D.; Wang, Y.; Luo, P. Study on the spatiotemporal dynamic of ground-level ozone concentrations on multiple scales across China during the blue sky protection campaign. *Environ. Int.* **2022**, *170*, 107606. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, G.; Peng, W.; Zhang, L.; Zhang, J. Quantifying the impacts of natural and human factors on changes in NPP using an optimal parameters-based geographical detector. *Ecol. Indic.* **2023**, *155*, 111018. [\[CrossRef\]](#)
- Han, L.; Zhang, R.; Wang, J.; Cao, S.-J. Spatial synergistic effect of urban green space ecosystem on air pollution and heat island effect. *Urban Clim.* **2024**, *55*, 101940. [\[CrossRef\]](#)
- Qiao, Z.; Wang, N.; Chen, J.; Xu, X.; Liu, L.; Han, D. Understanding the differences in the contribution and impact of urbanization on urban warming during heatwave and non-heatwave periods in China. *J. Clean. Prod.* **2024**, *474*, 143626. [\[CrossRef\]](#)
- Lyu, Y.; Xu, H.; Wu, H.; Han, F.; Lv, F.; Kang, A.; Pang, X. Spatiotemporal variations of PM_{2.5} and ozone in urban agglomerations of China and meteorological drivers for ozone using explainable machine learning. *Environ. Pollut.* **2025**, *365*, 125380. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wei, X.; Yang, J.; Luo, P.; Lin, L.; Lin, K.; Guan, J. Assessment of the variation and influencing factors of vegetation NPP and carbon sink capacity under different natural conditions. *Ecol. Indic.* **2022**, *138*, 108834. [\[CrossRef\]](#)
- Zeng, J.; Zhou, T.; Wang, Q.; Xu, Y.; Lin, Q.; Zhang, Y.; Wu, X.; Zhang, J.; Liu, X. Spatial patterns of China's carbon sinks estimated from the fusion of remote sensing and field-observed net primary productivity and heterotrophic respiration. *Ecol. Inform.* **2023**, *76*, 102152. [\[CrossRef\]](#)
- Yu, Z.; Yang, G.; Zuo, S.; Jørgensen, G.; Koga, M.; Vejre, H. Critical review on the cooling effect of urban blue-green space: A threshold-size perspective. *Urban For. Urban Green.* **2020**, *49*, 126630. [\[CrossRef\]](#)
- Lei, Y.; Duan, Y.; He, D.; Zhang, X.; Chen, L.; Li, Y.; Gao, Y.G.; Tian, G.; Zheng, J. Effects of Urban Greenspace Patterns on Particulate Matter Pollution in Metropolitan Zhengzhou in Henan, China. *Atmosphere* **2018**, *9*, 199. [\[CrossRef\]](#)
- Chen, X.; Wei, F. Combined effects of urban morphology on land surface temperature and PM_{2.5} concentration across fine-scale urban blocks in Hangzhou, China. *Build. Environ.* **2025**, *278*, 112979. [\[CrossRef\]](#)
- Wang, A.; Zhang, M.; Chen, E.; Zhang, C.; Han, Y. Impact of seasonal global land surface temperature change on gross primary production in the early 21st century. *Sustain. Cities Soc.* **2024**, *110*, 105572. [\[CrossRef\]](#)
- Wicki, A.; Parlow, E. Multiple Regression Analysis for Unmixing of Surface Temperature Data in an Urban Environment. *Remote Sens.* **2017**, *9*, 684. [\[CrossRef\]](#)
- Li, M.; Geng, C.; Li, L.; Zheng, Z.; Xu, B.; Yang, W.; Wang, X. Meteorological impacts on surface ozone: A case study based on Kolmogorov–Zurbenko filtering and multiple linear regression. *Front. Environ. Sci.* **2023**, *10*, 1081453. [\[CrossRef\]](#)
- Tian, Q.; Qiu, Q.; Wang, Z.; Cai, Z.; Hu, L.; Liu, H.; Feng, Y.; Li, X. Spatial Configuration of Urban Greenspace Affects Summer Air Temperature: Diurnal Variations and Scale Effects. *Atmosphere* **2023**, *14*, 1433. [\[CrossRef\]](#)
- Uhrin, A.; Onačillová, K. Spatiotemporal analysis of land surface temperature and land cover changes in Prešov city using downscaling approach and machine learning algorithms. *Environ. Monit. Assess.* **2025**, *197*, 126. [\[CrossRef\]](#) [\[PubMed\]](#)

26. Chang, S.; Gao, X.; Li, Q.; Song, X.; Lin, J. Threshold response of net primary productivity to climate change in the Qilian Mountains (2001–2022). *Ecol. Indic.* **2025**, *178*, 113900. [\[CrossRef\]](#)
27. Calatayud, V.; Diéguez, J.J.; Agathokleous, E.; Sicard, P. Machine learning model to predict vehicle electrification impacts on urban air quality and related human health effects. *Environ. Res.* **2023**, *228*, 115835. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Wei, J.; Li, Z.; Lyapustin, A.; Sun, L.; Peng, Y.; Xue, W.; Su, T.; Cribb, M. Reconstructing 1-Km-Resolution High-Quality PM_{2.5} Data Records from 2000 to 2018 in China: Spatiotemporal Variations and Policy Implications. *Remote Sens. Environ.* **2021**, *252*, 112136. [\[CrossRef\]](#)
29. Wang, S.; Ren, Y.; Xia, B. PM_{2.5} and O₃ concentration estimation based on interpretable machine learning. *Atmos. Pollut. Res.* **2023**, *14*, 101866. [\[CrossRef\]](#)
30. Yuan, X.; Hong, X.; Huang, Z.; Sheng, L.; Zhang, J.; Chen, D.; Zhong, Z.; Wang, B.; Zheng, J. Uncovering key sources of regional ozone simulation biases using machine learning and SHAP analysis. *Environ. Pollut.* **2025**, *372*, 126012. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Zhang, S.; Guo, X.; Zhang, S.; Li, C.; Zhang, C. Urban–Rural Spatial Patterns, Landscape Configuration, and Carbon Emission Performance: A County-Level Analysis in Henan Province, China. *Land* **2026**, *15*, 1021. [\[CrossRef\]](#)
32. Jiang, W.; Gao, W.; Gao, X.; Ma, M.; Zhou, M.; Du, K.; Ma, X. Spatio-temporal heterogeneity of air pollution and its key influencing factors in the Yellow River Economic Belt of China from 2014 to 2019. *J. Environ. Manag.* **2021**, *296*, 113172. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Li, M.; Li, C.; Zhang, M. Exploring the spatial spillover effects of industrialization and urbanization factors on pollutant emissions in China’s Huang-Huai-Hai region. *J. Clean. Prod.* **2018**, *195*, 154–162. [\[CrossRef\]](#)
34. Yang, Z.; Li, Z.; Cheng, F.; Lv, Q.; Li, K.; Zhang, T.; Zhou, Y.; Zhao, B.; Xue, W.; Wei, J. Two-decade surface ozone (O₃) pollution in China: Enhanced fine-scale estimations and environmental health implications. *Remote Sens. Environ.* **2025**, *317*, 114459. [\[CrossRef\]](#)
35. Wu, T.; Qiao, Z. Synergistic governance of urban heat islands, energy consumption, carbon emissions, and air pollution in China: Evidence from a Spatial Durbin Model. *Environ. Pollut.* **2025**, *372*, 126025. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Zheng, H.; Kong, S.; Zhai, S.; Sun, X.; Cheng, Y.; Yao, L.; Song, C.; Zheng, Z.; Shi, Z.; Harrison, R.M. An intercomparison of weather normalization of PM_{2.5} concentration using traditional statistical methods, machine learning, and chemistry transport models. *npj Clim. Atmos. Sci.* **2023**, *6*, 214. [\[CrossRef\]](#)
37. Yi, Z.; Wu, L. Identification of factors influencing net primary productivity of terrestrial ecosystems based on interpretable machine learning—Evidence from county-level administrative districts in China. *J. Environ. Manag.* **2023**, *326*, 116798. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Li, H.; Yang, J.; Xin, J.; Yu, W.; Ren, J.; Yu, H.; Xiao, X.; Xia, J. Investigating the effect of urban form on land surface temperature at block and grid scales based on XGBoost-SHAP. *Environ. Model. Softw.* **2026**, *195*, 106738. [\[CrossRef\]](#)
39. Liu, X.; Wang, S.; Tang, G. Understanding nonlinear and spatially heterogeneous effects of urban residential morphology on land surface temperature: Integrating SOM, XGBoost-SHAP, and GWR models. *Sustain. Cities Soc.* **2026**, *136*, 107100. [\[CrossRef\]](#)
40. Wang, Z.; Zhou, R.; Rui, J.; Yu, Y. Revealing the impact of urban spatial morphology on land surface temperature in plain and plateau cities using explainable machine learning. *Sustain. Cities Soc.* **2025**, *118*, 106046. [\[CrossRef\]](#)
41. Liu, Q.; Wu, R.; Zhang, W.; Li, W.; Wang, S. The varying driving forces of PM_{2.5} concentrations in Chinese cities: Insights from a geographically and temporally weighted regression model. *Environ. Int.* **2020**, *145*, 106168. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Wang, H.L.; Qiao, L.P.; Lou, S.R.; Zhou, M.; Ding, A.J.; Huang, H.Y.; Chen, J.M.; Wang, Q.; Tao, S.K.; Chen, C.H.; et al. Chemical composition of PM_{2.5} and meteorological impact among three years in urban Shanghai, China. *J. Clean. Prod.* **2016**, *112*, 1302–1311. [\[CrossRef\]](#)
43. Cheng, J.; Tong, D.; Liu, Y.; Geng, G.; Davis, S.J.; He, K.; Zhang, Q. A synergistic approach to air pollution control and carbon neutrality in China can avoid millions of premature deaths annually by 2060. *One Earth* **2023**, *6*, 978–989. [\[CrossRef\]](#)
44. Gong, J.; Yin, Z.; Lei, Y.; Lu, X.; Zhang, Q.; Cai, C.; Chai, Q.; Chen, H.; Chen, R.; Chen, W.; et al. The 2023 report of the synergetic roadmap on carbon neutrality and clean air for China: Carbon reduction, pollution mitigation, greening, and growth. *Environ. Sci. Ecotechnol.* **2025**, *23*, 100517. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Yang, Y.; Xu, H.; Yang, X.; Zhang, Y.; Liu, T. Exploring synergistic efficiency of air pollution and carbon reduction and its influencing factors: Insights from China. *Resour. Conserv. Recycl.* **2025**, *212*, 107973. [\[CrossRef\]](#)
46. Lopez-Aparicio, S.; Grythe, H.; Drabicki, A.; Chwastek, K.; Toboła, K.; Górka-Niemas, L.; Kierpiec, U.; Markelj, M.; Strużewska, J.; Kud, B.; et al. Environmental sustainability of urban expansion: Implications for transport emissions, air pollution, and city growth. *Environ. Int.* **2025**, *196*, 109310. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Zhang, M.; Liu, X.; Ding, Y. Assessing the influence of urban transportation infrastructure construction on haze pollution in China: A case study of the Beijing-Tianjin-Hebei region. *Environ. Impact Assess. Rev.* **2021**, *87*, 106547. [\[CrossRef\]](#)
48. Wang, Y.; Wang, Z.; Wang, J.; Chen, T.; Wang, G. Same source, different control: Threshold-driven divergence of PM_{2.5} and CO₂ responses to urban built environment features. *Sustain. Cities Soc.* **2026**, *139*, 107230. [\[CrossRef\]](#)
49. Zhang, C.; Li, W.; Zhang, S.; Wang, X.; Hu, J.; Fan, Q.; Wang, Q. Multifractal Analysis of Heritage Architecture: Quantifying Complexity in 30 Iconic Historic Buildings. *npj Herit. Sci.* **2026**. [\[CrossRef\]](#)

50. Katona, V. The Hidden Dimension of Façades: Fractal Analysis Reveals Composition Rules in Classical and Renaissance Architecture. *Fractal Fract.* **2023**, *7*, 257. [[CrossRef](#)]
51. Lorenz, W.E.; Kulcke, M. Multilayered Complexity Analysis in Architectural Design: Two Measurement Methods Evaluating Self-Similarity and Complexity. *Fractal Fract.* **2021**, *5*, 244. [[CrossRef](#)]

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