

Article

Multi-Scale Driving Mechanisms of the Spatiotemporal Distribution of Soil pH in High-Standard Farmland Within the Hilly Regions of Southern China

Zhengxiang Zeng ¹, Fanghao Li ², Meiqiu Chen ^{1,*} , Yujun Cai ¹ and Yefeng Jiang ¹ 

¹ College of Land Resources and Environment, Jiangxi Agricultural University, Nanchang 330045, China; zzx@bjsiwei.com (Z.Z.); caiyujun809@gmail.com (Y.C.); jiangyefeng@jxau.edu.cn (Y.J.)

² Jiangxi Provincial Institute of Land and Space Survey and Planning, Nanchang 330025, China; lfh@bjsiwei.com

* Correspondence: cmq12@jxau.edu.cn; Tel.: +86-13677001278

Abstract

Elucidating the multi-scale driving mechanisms of the spatiotemporal distribution of soil pH is critical for regional soil ecological regulation and the preservation of cultivated land fertility. However, systematic elucidation of the multi-scale driving mechanisms of the spatiotemporal distribution of soil pH remains limited. Here, we integrated random forest, XGBoost, and geographically interpretable SHAP (GeoShapley) to systematically investigate the spatiotemporal changes in soil pH and their driving mechanisms across three analytical scales (global, raster, and point-level) before and after high-standard farmland construction, based on 149 paired soil sampling points collected in 2008 and 2023 in Taihe County, Jiangxi Province. The mean soil pH in the study area increased from 5.12 in 2008 to 5.53 in 2023, suggesting mitigation of soil acidification. XGBoost outperformed random forest in both periods, with R^2 values of 0.71 and 0.67 for the two periods, respectively, and the spatial predictions indicated that soil pH in the western and southeastern regions increased following high-standard farmland construction. At the global scale, feature importance based on XGBoost indicated that climate contributed most to soil pH variation in both periods, followed by anthropogenic activities. At the raster scale, GeoShapley revealed a shift from a combination of anthropogenic activities and climate in 2008 to an interplay among climate, vegetation, and topography in 2023, suggesting a decreasing influence of anthropogenic disturbance. At the point level, the factors primarily influencing soil pH shifted from topographic and anthropogenic ones in 2008 to topographic and climatic ones in 2023, regardless of whether the sampling points exhibited low or high pH values, further corroborating the diminishing role of anthropogenic influence. Interaction analyses indicated that topography and vegetation consistently modulated the effects of anthropogenic and climatic factors throughout the study period, suggesting that these two factors were consistently associated with soil pH dynamics across the study area. In summary, this study provides a scientific basis for the targeted improvement of cultivated land fertility and the formulation of differentiated soil management strategies for high-standard farmland in the hilly regions of southern China.



Academic Editors: Eugenio Cavallo and Marcella Biddoccu

Received: 13 July 2026

Revised: 4 August 2026

Accepted: 4 August 2026

Published: 5 August 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](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: soil pH; high-standard farmland; interpretable machine learning; GeoShapley; multi-scale driver analysis

1. Introduction

Soil pH is a fundamental property governing nutrient availability, microbial activity, and crop growth, and plays a pivotal role in agricultural production and ecosystem functioning [1]. However, accelerating industrialization and agricultural intensification have intensified soil acidification globally, driven by acid deposition from fossil fuel combustion and the excessive application of nitrogen fertilizers [2]. This problem is particularly severe in the hilly regions of southern China, where the combined effects of excessive nitrogen fertilizer inputs and acid deposition have caused a sustained decline in farmland soil pH since the 1980s [1], severely constraining agricultural sustainability. Soil acidification impairs farmland productivity through multiple pathways. In terms of ion toxicity, acidification promotes the hydrolysis of aluminum oxides, elevating soluble aluminum concentrations that inhibit root tip cell division and elongation [3,4], while simultaneously exacerbating manganese toxicity, which disrupts crop photosynthesis and enzymatic activity [5]. With respect to nutrient loss, acidification accelerates the leaching of base cations such as calcium and magnesium, reducing soil nutrient availability and buffering capacity. Furthermore, acidification mobilizes soil heavy metals, undermining soil ecological structure and function [6], and ultimately depresses crop yields, threatening regional food security [7]. To mitigate farmland soil acidification and safeguard national food security, China has implemented high-standard farmland construction as a key policy initiative [8], which systematically modifies farmland physicochemical conditions through measures such as land leveling, irrigation and drainage improvement, the application of calcareous soil amendments, and straw incorporation [9], thereby profoundly shaping the spatiotemporal patterns and dynamics of soil pH. Therefore, elucidating the spatiotemporal patterns and dynamics of soil pH and their underlying driving mechanisms before and after high-standard farmland construction is of both theoretical and practical significance.

The spatial distribution of soil pH is jointly shaped by a range of natural and anthropogenic factors, including climate [10,11], topography [12,13], vegetation [14], atmospheric pollution [15], and human activities [16,17]. A thorough understanding of the relative contributions and mechanisms of these driving factors is an essential prerequisite for the precise management of soil pH. With respect to natural factors, Sun et al. [18] demonstrated at the global scale that temperature influences the soil proton balance by regulating rock weathering rates and microbial activity, whereas precipitation accelerates the loss of base cations through leaching, with both factors collectively driving the spatial patterns of soil pH. Similarly, Zhang et al. [19] found at the global scale that topographic factors exert a compound influence on the spatial differentiation of soil pH by regulating water redistribution and material translocation. Furthermore, Zhang et al. [20] revealed at the global scale a significant spatiotemporal coupling between NPP and soil pH, whereby vegetation productivity influences the soil acid–base balance through organic matter inputs and root exudates. With respect to anthropogenic factors, Yan et al. [17] demonstrated at the point scale that land use type is the primary factor driving spatial variability in soil pH across the red soil region of southern China, with significant pH differences among land use types. Li et al. [21] further confirmed at the global scale that atmospheric PM_{2.5} continuously introduces acidic substances into the soil via dry and wet deposition, significantly affecting the spatial distribution of farmland pH. However, most existing studies have focused on point-level or global-scale analyses of driving factors, and a systematic characterization of drivers across multiple scales remains lacking—particularly the pixel-level spatial heterogeneity of driving mechanisms at the raster scale—thus limiting the formulation of differentiated precision soil management strategies.

Geostatistics is a widely used approach for mapping soil properties and can effectively characterize their spatial variability structures [22]; however, its underlying principle relies

on spatial autocorrelation-based interpolation, which makes it difficult to quantitatively decompose the relative contributions of multiple environmental factors [14]. Machine learning models, represented by random forest and extreme gradient boosting (XGBoost), have been widely adopted in recent years for the spatial prediction of soil pH and the identification of its driving factors, owing to their advantages in capturing nonlinear relationships and handling high-dimensional interaction effects [23]. For example, Hengl et al. [24] leveraged 272,000 global soil pH samples and applied spatiotemporal machine learning to produce global dynamic soil pH maps at 30 m resolution, demonstrating the considerable applicability of machine learning for large-scale spatiotemporal mapping of soil properties. In addition, spatializing machine learning prediction residuals to characterize prediction uncertainty has increasingly become an important component in evaluating the reliability of digital soil mapping [25]. Nevertheless, the inherent “black-box” nature of machine learning models makes it difficult to quantify the absolute contributions of individual factors, thereby constraining a deeper understanding of the mechanisms underlying the spatiotemporal distribution of soil pH. The introduction of the SHapley Additive exPlanations (SHAP) method, however, offers an effective solution to this challenge. Grounded in cooperative game theory, SHAP quantifies the contributions of driving factors by computing the marginal contribution of each feature to the predicted outcome [26]. This method is capable of ranking feature importance at the global level, attributing predictions at the point level, and effectively capturing nonlinear interaction effects among multiple factors. For instance, Huang et al. [14] combined random forest with SHAP in Lanxi City, Zhejiang Province, to achieve quantitative decomposition of the driving factors of farmland soil pH and to reveal the mechanisms underlying its spatial differentiation. However, SHAP analyses at both the global and point scales assume spatially homogeneous contributions of driving factors, making it difficult to capture the spatial heterogeneity of driving mechanisms as they vary across geographic locations. To address this limitation, Li [27] proposed the GeoShapley method, which incorporates geographic location as a player within the Shapley value framework, enabling location-specific quantification of the spatial effects of driving factors and extending the analysis of driving mechanisms from the global and point scales to the raster spatial scale. In summary, integrating machine learning with GeoShapley provides a scientifically sound and effective framework for elucidating the drivers of soil pH spatial heterogeneity from a multi-scale perspective. However, how the relative contributions of driving factors change in farmland undergoing systematic construction, such as high-standard farmland, has rarely been examined.

In this study, we integrated random forest, XGBoost, and the GeoShapley interpretability framework to systematically investigate the spatiotemporal patterns and driving mechanisms of soil pH before and after high-standard farmland construction across three analytical scales—global, raster, and point-level—based on 149 paired soil sampling points collected in Taihe County, Jiangxi Province, in 2008 and 2023. The specific objectives were to (1) characterize the changes in spatiotemporal distribution patterns of soil pH before and after high-standard farmland construction; (2) identify the multi-scale driving factors influencing the spatial distribution of soil pH and compare their relative contributions before and after high-standard farmland construction; and (3) explore the nonlinear response relationships and interaction effects among dominant driving factors. These findings provide a scientific basis for mitigating soil acidification and implementing precision management in high-standard farmland across the hilly regions of southern China.

2. Materials and Methods

2.1. Study Area

The study area is located in Taihe County, Ji'an City, Jiangxi Province ($114^{\circ}17'–115^{\circ}20'$ E, $26^{\circ}38'–27^{\circ}06'$ N), at the center of the Ji-Tai Basin, and covers a total area of 2495 km². The terrain is predominantly hilly, exhibiting a basin configuration with higher elevations in the southeastern and western portions and lower elevations in the central zone. The county lies within the mid-subtropical humid monsoon climate zone, characterized by abundant rainfall and distinct seasons, with a mean annual temperature of 19 °C and mean annual precipitation of 1457.9 mm. The soils in the study area are dominated by red soil and paddy soil, with a small distribution of fluvo-aquic soil, and the red soils are generally acidic under the prolonged subtropical climate. Owing to these favorable conditions, the study area is endowed with abundant cultivated land resources, with double-season rice as the dominant crop. However, under long-term intensive cultivation, the sustained application of nitrogen fertilizers has continuously introduced acidic substances into the soil, resulting in pronounced soil acidification in the region [28]. In recent years, the construction of high-standard farmland in Taihe County has advanced steadily, with a cumulative area of approximately 361.1 km² established by 2023, predominantly concentrated in flat farmland areas in the western and southeastern parts of the study area (Figure 1).

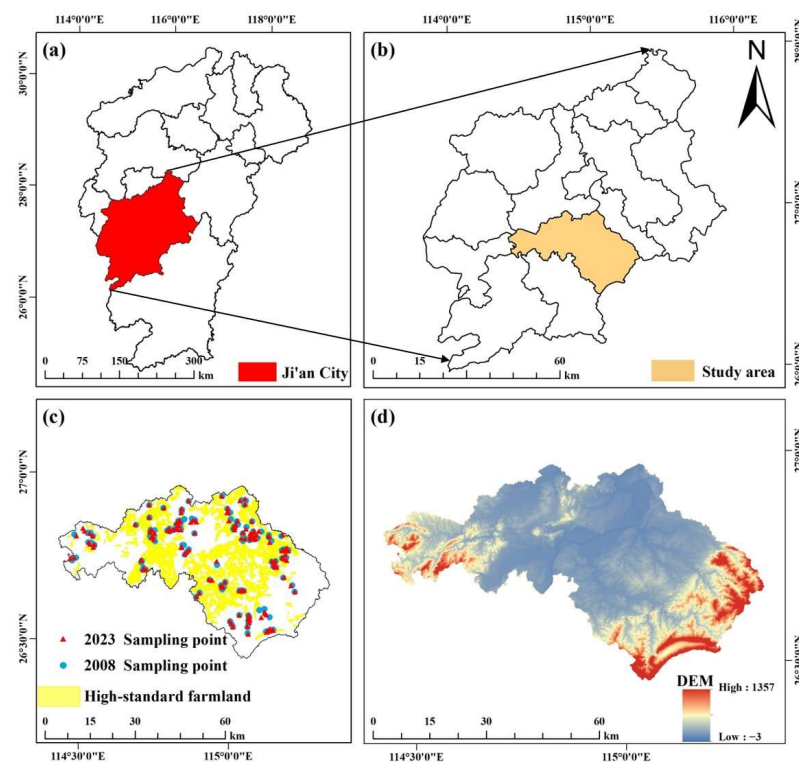


Figure 1. Geographical location, topography and sampling point distribution of the study area. (a) Location of Ji'an City within Jiangxi Province, China. (b) Location of the study area (Taihe County) within Ji'an City. (c) Distribution of high-standard farmland (as of 2023), together with the 2008 sampling points (blue circles) and 2023 sampling points (red triangles) within the study area. The high-standard farmland boundaries represent the current (2023) state of construction and are overlaid on the 2008 sampling points solely for spatial comparison; high-standard farmland had not yet been constructed at the time of the 2008 sampling. (d) Digital elevation model of the study area.

Specifically, high-standard farmland construction involves a set of engineering and agronomic measures, including field consolidation, soil improvement, irrigation and drainage improvement, field road construction, and farmland protection. Among these,

the soil improvement component is most directly relevant to soil pH: in the acidic red soil regions of southern China, it typically involves the application of calcareous amendments such as lime to raise soil pH, increased application of organic fertilizers, straw incorporation, and green manure planting to enhance soil organic matter and acid-buffering capacity. In addition, the improved irrigation and drainage facilities alter the soil water regime and leaching intensity, while the standardization of cultivation and fertilization practices reduces the excessive use of acidifying nitrogen fertilizers. These measures jointly modify the physicochemical conditions of the farmland and may therefore influence soil pH.

2.2. Soil Sample Collection

Surface soil samples (0–20 cm) were collected from 149 points across the cultivated land of the study area between July and August 2008 using a random sampling strategy. After natural air-drying, grinding, and sieving, soil pH was determined in the laboratory using the potentiometric method with a water-to-soil ratio of 2.5:1. Resampling was conducted between July and August 2023 at locations adjacent to the 2008 sampling points within high-standard farmland, yielding another 149 sampling points. The mean distance between paired points from the two sampling periods was 386 m, ranging from 17 to 2592 m. The high-standard farmland in the study area consists of contiguous and concentrated paddy fields with uniform soil types, within which topography and cultivation regimes are highly homogeneous. Consequently, even the maximum paired distance remained within the same continuous farmland system. The paired points therefore share essentially uniform soil-forming conditions, and the observed pH differences reflect temporal change. To ensure comparability between the two datasets, the sampling strategy and laboratory measurement procedures adopted in 2023 were both kept consistent with those of 2008. The spatial distribution of sampling points from both periods is shown in Figure 1.

2.3. Data Source and Processing

In this study, we selected 34 environmental variables spanning four categories—topography, climate, remote sensing, and human activities—for spatiotemporal mapping of soil pH (Table 1). These variables were selected based on the five soil-forming factors (climate, organisms, relief, parent material and time), together with human activities, and all of them are available as spatially continuous raster layers covering the study area. Topographic variables included the digital elevation model (DEM), slope, aspect, multiresolution index of ridge top flatness (MRRTF), the multiresolution index of valley bottom flatness (MRVBF), terrain ruggedness index (TRI), and topographic position index (TPI), all derived from NASA SRTM data (<https://earthdata.nasa.gov>) at a spatial resolution of 30 m. Climatic variables included mean annual temperature (MAT), mean annual precipitation (MAP), annual maximum temperature (MMAT), annual minimum temperature (LMAT), potential evapotranspiration (PET), and aridity index (AI), all obtained from the National Earth System Science Data Center (<https://www.geodata.cn>) at a spatial resolution of 1 km, covering the period from 2008 to 2023. Remote sensing variables included 15 variables encompassing three aspects: vegetation, moisture, and land cover. Vegetation-related variables included the normalized difference vegetation index (NDVI), fractional vegetation cover (FVC), net primary productivity (NPP), gross primary productivity (GPP), net ecosystem productivity (NEP), tree cover (TC), soil-adjusted vegetation index (SAVI), and enhanced vegetation index (EVI). Among these, NDVI, FVC, NPP, GPP, NEP, and TC were obtained from the National Ecosystem Science Data Center (<https://www.nesdc.org.cn>), whereas SAVI and EVI were derived from Landsat imagery using the Google Earth Engine (GEE) platform (<https://earthengine.google.com>), all at a resolution of 30 m. Moisture-related variables included the normalized difference water index (NDWI), modified normalized difference water index (MNDWI), normalized difference moisture index (NDMI), and tasseled cap wetness (Wet), all computed from Landsat imagery on

the GEE platform at a resolution of 30 m. Land cover variables included the bare soil index (BSI), normalized difference built-up index (NDBI), and soil conservation services (SC). BSI and NDBI were derived from Landsat imagery on the Google Earth Engine platform at a resolution of 30 m, whereas SC data were sourced from Li et al. [29], covering the period 1992–2019 at a resolution of 300 m. All Landsat-derived remote sensing indices were computed as annual mean composites for the corresponding sampling years (2008 and 2023) to minimize the influence of phenological stage, cloud contamination, and imaging time associated with individual scenes. Human activity variables included real GDP, nighttime light (NL), population density (PD), and PM_{2.5}, PM₁₀, and ozone (O₃) concentrations. GDP data were obtained from the Resource and Environment Science Data Platform (<http://www.resdc.cn>), covering 1992–2019 at a spatial resolution of 1 km. NL data were sourced from Wu et al. [30], spanning 1992–2023 at a spatial resolution of 1 km. PD data were derived from the LandScan dataset (<https://landscan.ornl.gov>), covering 2000–2022 at a spatial resolution of 1 km. PM_{2.5} and PM₁₀ data were obtained from Wei et al. [31] and Wei et al. [32], respectively, at a spatial resolution of 1 km covering 2008–2023. O₃ data were sourced from Yang et al. [33], at a spatial resolution of 1 km covering 2008–2023. All variables were resampled to a uniform spatial resolution of 30 m to ensure spatial consistency across multi-source datasets. For environmental covariates with missing data in certain years, we substituted values from the nearest available year.

Table 1. Environmental variables used for the spatiotemporal mapping of soil pH.

Category	Variable Name	Abbreviation	Period	Resolution	Data Source
Terrain	Digital elevation model	DEM	/	30 m	NASA SRTM (https://earthdata.nasa.gov)
	Slope	Slope	/	30 m	
	Aspect	Aspect	/	30 m	
	Multiresolution index of ridge top flatness	MRRTF	/	30 m	
	Multiresolution index of valley bottom flatness	MRVBF	/	30 m	
	Terrain ruggedness index	TRI	/	30 m	
	Topographic position index	TPI	/	30 m	
Climate	Mean annual temperature	MAT	2008–2022	1 km	National Earth System Science Data Center (https://www.geodata.cn)
	Mean annual precipitation	MAP	2008–2022	1 km	
	Annual maximum temperature	MMAT	2008–2022	1 km	
	Annual minimum temperature	LMAT	2008–2022	1 km	
	Potential evapotranspiration	PET	2008–2023	1 km	
	Aridity index	AI	2008–2023	1 km	
Remote Sensing	Normalized Difference Vegetation Index	NDVI	2008–2022	30 m	National Ecosystem Science Data Center (https://www.nesdc.org.cn)
	Fractional vegetation cover	FVC	2008–2023	250 m	
	Net primary productivity	NPP	2008–2023	1 km	
	Gross primary productivity	GPP	2008–2023	1 km	
	Net ecosystem productivity	NEP	2008–2023	1 km	
	Tree cover	TC	2008–2023	30 m	
	Soil-adjusted vegetation index	SAVI	2008/2023	30 m	Google Earth Engine (https://earthengine.google.com)
	Enhanced vegetation index	EVI	2008/2023	30 m	
	Normalized Difference Water Index	NDWI	2008/2023	30 m	
	Modified Normalized Difference Water Index	MNDWI	2008/2023	30 m	
	Normalized Difference Moisture Index	NDMI	2008/2023	30 m	
	Tasseled cap wetness	Wet	2008/2023	30 m	
	Bare soil index	BSI	2008/2023	30 m	
	Normalized Difference Built-up Index	NDBI	2008/2023	30 m	
	Soil conservation services	SC	2008–2019	300 m	
					[29]

Table 1. Cont.

Category	Variable Name	Abbreviation	Period	Resolution	Data Source
Anthropogenic	Real GDP	GDP	2008–2019	1 km	Resource and Environment Science Data Platform (http://www.resdc.cn)
	Nighttime light	NL	2008–2023	1 km	[30]
	Population density	PD	2008–2022	1 km	LandScan (https://landscan.ornl.gov)
	PM2.5 concentration	PM2.5	2008–2023	1 km	[31]
	PM10 concentration	PM10	2008–2023	1 km	[32]
	Ozone concentration	O3	2008–2023	1 km	[33]

2.4. Machine Learning Models

We employed two machine learning models—random forest and extreme gradient boosting (XGBoost)—to spatially predict soil pH for both sampling periods. Random forest is an ensemble learning algorithm based on decision trees that constructs mutually independent trees through bootstrap sampling and averages the predictions of all trees as the final output, which reduces the risk of overfitting and makes the model robust to multicollinearity among predictors [34]. Its prediction formula is as follows:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

where $\hat{y}$ denotes the final predicted value, B is the number of decision trees, and $T_b(x)$ is the prediction of the b -th decision tree for input x . During the construction of each tree, m features are randomly selected ($m < p$, where p is the total number of features) for node splitting, thereby enhancing diversity among trees.

XGBoost is based on the gradient boosting framework, which iteratively accumulates multiple weak learners into a strong learner and incorporates regularization terms into the objective function to control model complexity, offering marked advantages in predictive accuracy, computational efficiency, and robustness to multicollinearity [35]. Its objective function is as follows:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

where $\mathcal{L}^{(t)}$ denotes the objective function, n is the number of samples, x_i is the input variable for the i -th sample, g_i and h_i are the first- and second-order gradients of the loss function with respect to the current prediction, f_t is the decision tree added at iteration t ($t = 1, 2, \dots, T$), T is the number of leaf nodes, w is the vector of leaf node weights, and γ and λ are regularization coefficients.

Both models were systematically tuned using grid search for hyperparameter optimization. The search ranges for each model are provided in Table S1, and the optimal parameter combination was determined based on the minimum RMSE under cross-validation. Model accuracy was subsequently evaluated using leave-one-out cross-validation, which was preferred over k -fold cross-validation because, given the limited sample size (149 points), it maximizes the use of the available data for training in each iteration and provides a more stable and less biased performance estimate for a dataset of this size. The coefficient of determination (R^2) and root mean square error (RMSE) were used as evaluation metrics to comprehensively assess predictive performance and generalization capacity. Furthermore, to evaluate the spatial reliability of model predictions, a non-parametric bootstrap proce-

dures were applied to both random forest and XGBoost models, each resampled 100 times. For each pixel, the upper and lower bounds of the 95% confidence interval and the mean predicted value were computed, and the ratio of the 95% confidence interval width to the mean was used to characterize model prediction uncertainty, yielding spatially explicit uncertainty maps for both sampling periods [36].

2.5. GeoShapley Interpretability Framework

SHAP is grounded in cooperative game theory, treating each environmental variable as a player in a cooperative game. By enumerating all possible feature subset combinations, it computes the expected marginal contribution of each feature upon its inclusion in the model prediction, thereby deriving the Shapley value for that feature [37]. This method supports both global-level feature importance ranking and local-level attribution of individual sample contributions, enabling quantitative characterization of the magnitude and direction of each driving factor's influence on soil pH. The calculation formula is as follows:

$$\phi_i(f, x) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f_x(S \cup \{i\}) - f_x(S)]$$

where ϕ_i denotes the Shapley value of feature i , quantifying its marginal contribution to the predicted outcome for the current sample; F is the complete feature set; S is an arbitrary feature subset excluding feature i ; $f_x(S)$ is the model output based solely on subset S ; and $f_x(S \cup \{i\})$ is the model output after adding feature i to subset S .

However, SHAP assumes spatially homogeneous contributions of driving factors, making it difficult to capture the spatial heterogeneity of driving mechanisms as they vary across geographic locations. To address this limitation, GeoShapley was further introduced to enable location-specific decomposition of the spatial heterogeneity of driving factors. Building upon the SHAP framework, GeoShapley incorporates geographic location as an additional player in the cooperative game, decomposing the contribution of each driving factor into two components: spatial and non-spatial effects [27]. Compared with conventional SHAP, which yields only a single global importance ranking, this enables GeoShapley to reveal how the influence of each driving factor varies across space, thereby extending the interpretation of driving mechanisms from the global scale to the raster spatial scale. Based on the spatial effect values of each feature derived from GeoShapley, the feature with the largest absolute spatial effect value was designated as the dominant driving factor for each pixel, enabling quantitative mapping of the spatial distribution of dominant drivers and comparative analysis across the two sampling periods.

2.6. Data Analysis and Tools

All descriptive statistical analyses, machine learning model construction, interpretability analyses, and graphical visualization in this study were performed in a Python 3.10.11 programming environment. Specifically, *pandas* version 2.3.3 and *NumPy* version 1.26.4 were used for data processing and organization, *SciPy* version 1.15.3 for descriptive statistics and Kolmogorov–Smirnov normality testing, the *scikit-learn* version 1.7.2 library for random forest and XGBoost version 3.1.2 model construction and hyperparameter tuning, the *SHAP* version 0.49.1 and *GeoShapley* version 0.2.0.0 libraries for geospatial interpretability analysis and spatial effect quantification, and *Matplotlib* version 3.10.9 and *Seaborn* version 0.13.2 for statistical visualization.

3. Results

3.1. Descriptive Statistics of Soil pH

Descriptive statistical results (Table 2 and Figure 2) showed that soil pH values across the 149 sampling points in 2008 ranged from 3.90 to 6.80, with a mean of 5.12 and a coefficient of variation of 10.44%, indicating relatively weak overall variability. The minimum value was as low as 3.90, indicating severe acidification at certain sampling points. By 2023, the mean soil pH had increased by 0.41 units to 5.53, the minimum value had risen to 4.34, and the CV was 11.39%. Kolmogorov–Smirnov tests further revealed that the distribution of soil pH shifted from non-normal in 2008 ($p < 0.001$) to normal in 2023 ($p > 0.05$). A paired-sample t -test further confirmed that the increase in soil pH from 2008 to 2023 was statistically significant ($t = 6.58$, $p < 0.001$), with 69.1% of the paired points showing a pH increase. In summary, these results indicate that soil pH increased following high-standard farmland construction, and soil acidification was ameliorated.

Table 2. Descriptive statistics of soil pH in 2008 and 2023.

Periods	Number	Maximum	Minimum	Mean	Median	SD	CV (%)	K-S p -Value
2008	149	6.8	3.9	5.12 ^b	5	0.53	10.44	<0.001
2023	149	7.14	4.34	5.53 ^a	5.56	0.63	11.39	0.772

SD = Standard Deviation; CV = Coefficient of Variation; K-S = Kolmogorov–Smirnov test. A paired-sample t -test indicated that the difference in soil pH between 2008 and 2023 was statistically significant ($t = 6.58$, $p < 0.001$; mean difference = 0.41, 95% CI: 0.28–0.53). Different lowercase letters following the mean values indicate a significant difference between the two periods ($p < 0.001$).

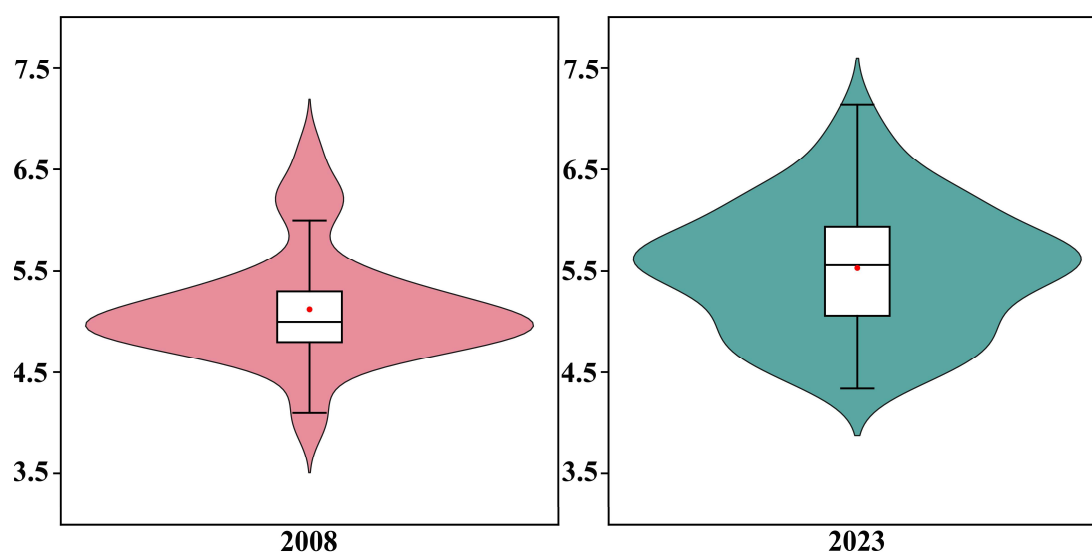


Figure 2. Violin plots of soil pH distribution in 2008 and 2023. The box plots within show the median and interquartile range (IQR), and the red dots represent the mean value.

3.2. Spatiotemporal Distribution and Prediction Uncertainty of Soil pH

The results shown in Figure 3 indicate that random forest yielded relatively lower predictive accuracy for soil pH ($R^2 = 0.58$ and 0.51 , RMSE = 0.38 and 0.48 for 2008 and 2023, respectively), whereas XGBoost achieved relatively higher accuracy ($R^2 = 0.71$ and 0.67 , RMSE = 0.37 and 0.47 for 2008 and 2023, respectively). Figure 4 shows that the spatial patterns of soil pH predicted by the two models were broadly consistent within each period but differed between periods. Specifically, soil pH in the study area was generally low in 2008, with areas of low soil pH (<5.0) predominantly concentrated in the central zone and areas of high soil pH (>5.5) sparsely distributed in localized parts of the southeastern

region. By contrast, areas of high soil pH expanded markedly in 2023, forming contiguous patches in the western and southeastern parts of the study area, whereas areas of low soil pH decreased considerably in extent. From 2008 to 2023, soil pH increased overall following high-standard farmland construction, with acidification improvement concentrated primarily in the western and southeastern regions. Figure 4 also shows the spatial distribution of the prediction uncertainty. In 2008, the uncertainty was generally low across the study area, with the lowest-uncertainty areas (<0.05) concentrated in the central and northern parts and only scattered high-uncertainty patches (>0.10) elsewhere. In 2023, a small high-uncertainty area (>0.10) emerged in the northern part of the study area, corresponding to a locality where soil pH remained relatively lower than in the surrounding areas that had risen to higher levels, whereas the rest of the region showed generally moderate and spatially more uniform uncertainty. Overall, the prediction uncertainty remained low to moderate across most of the study area in both periods, indicating that the predicted spatiotemporal patterns of soil pH are generally reliable.

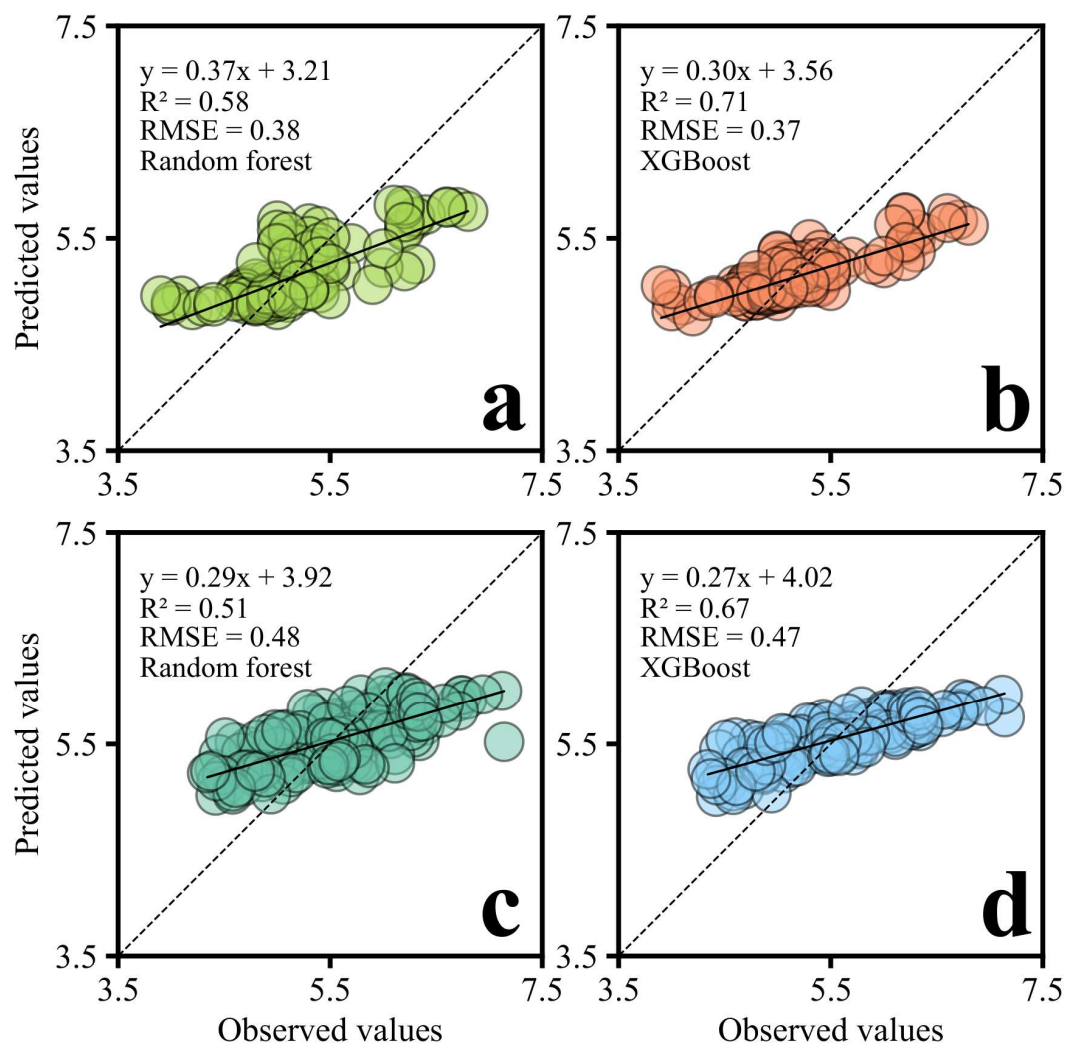


Figure 3. Comparison of predictive performance of Random Forest and XGBoost models for soil pH in 2008 and 2023. Panels (a,b) show the observed vs. predicted scatterplots for Random Forest and XGBoost in 2008, respectively; panels (c,d) show the corresponding results for 2023. The dashed line represents the 1:1 reference line ($y = x$) and the solid line represents the linear fit between predicted and observed values. R^2 and RMSE are annotated in each panel.

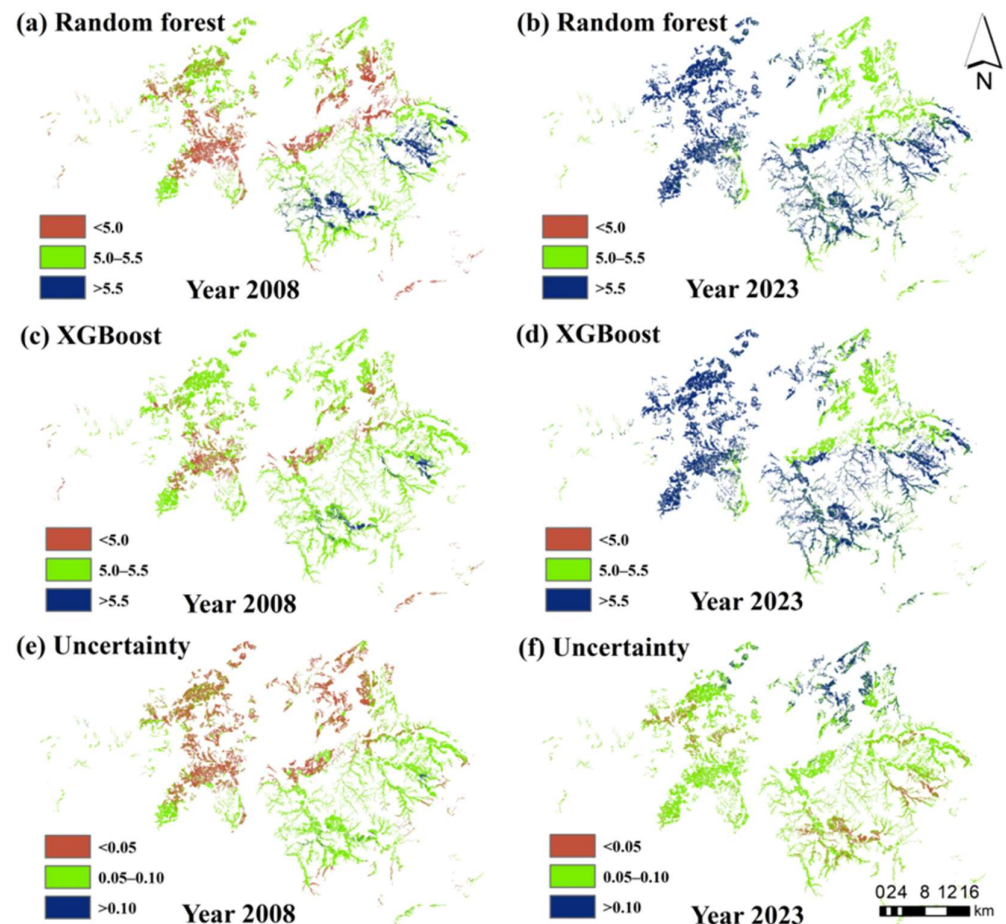


Figure 4. Spatial distribution of predicted soil pH and prediction uncertainty in 2008 and 2023. Panels (a,b) show the Random Forest predicted soil pH maps for 2008 and 2023, respectively; panels (c,d) show the corresponding XGBoost predictions; panels (e,f) show the spatial distribution of prediction uncertainty for 2008 and 2023, respectively.

3.3. Multi-Scale Driving Factors of the Spatiotemporal Distribution of Soil pH

Figure 5 presents the global feature importance for soil pH predictions in both sampling periods. In 2008, MAT was the most influential variable affecting soil pH variation, followed by PM2.5, PM10, MMAT, and NDMI, indicating that climate and human activities were the dominant driving factors. In 2023, LMAT surged to become the most influential variable, followed by PM2.5, NPP, MRVBF, and MAT, suggesting that soil pH variation was jointly influenced by climate, human activities, vegetation, and topography. Overall, from 2008 to 2023, climate consistently remained the most important factor influencing the spatial differentiation of soil pH, followed by human activities.

The results shown in Figure 6 indicate that, at the raster scale, PM2.5 was the dominant variable with the largest spatial coverage in 2008, predominantly distributed in the northern part of the study area, followed by MAT, which was mainly distributed in the southeastern part; NDMI and other variables exerted relatively minor influence. Specifically, the SHAP values of PM2.5 exhibited pronounced north–south differentiation, with positive values predominating in the north and negative values in the south. The SHAP values of MAT approached zero in the central zone of the study area, whereas positive values prevailed in the surrounding regions. The SHAP values of NDMI were predominantly negative in the western part and approached zero in the eastern part, reflecting its limited overall influence. In 2023, LMAT became the dominant variable with the largest spatial coverage, predominantly distributed in the northeastern part of the study area, followed by NPP and

MRVBF, distributed in the northwestern and southeastern parts, respectively. The SHAP values of LMAT were predominantly positive in the northwestern part and predominantly negative in the northeastern part, while exerting limited influence in the southern region. The SHAP values of NPP were predominantly positive in the northwestern part and negative in localized areas of the east-central zone, while exerting relatively minor influence elsewhere. The SHAP values of NPP were predominantly positive in the northwestern part, negative in localized areas of the east-central zone, and exerted relatively minor influence elsewhere. The SHAP values of MRVBF approached zero in the central zone, whereas positive values prevailed in the surrounding areas. Overall, the spatial variation in soil pH shifted from being primarily influenced by anthropogenic factors represented by PM2.5 and climatic factors represented by MAT in 2008 to being jointly influenced by climatic factors represented by LMAT, vegetation factors represented by NPP, and topographic factors represented by MRVBF in 2023, indicating a declining influence of human activities over time.

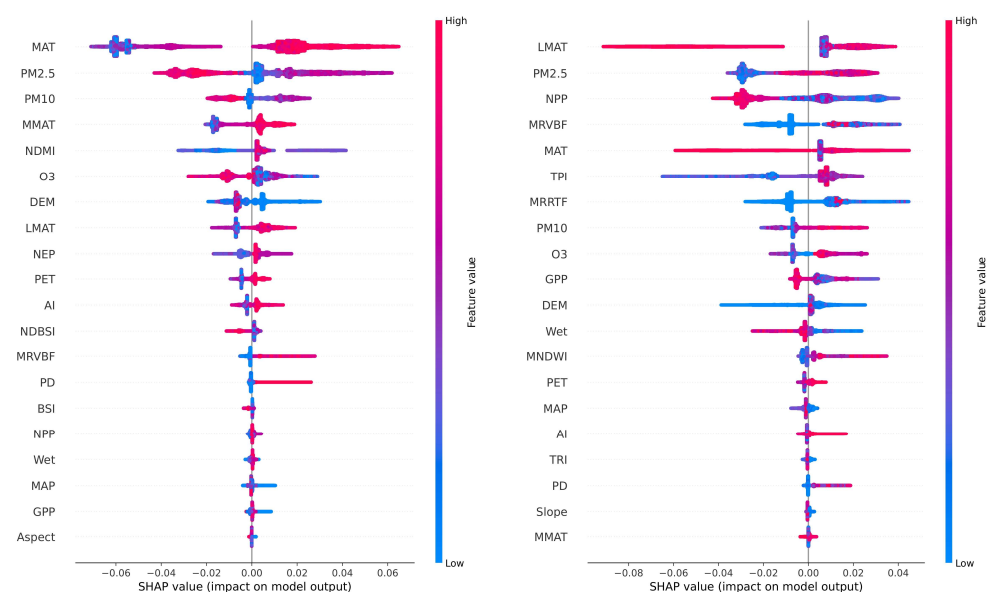


Figure 5. Global SHAP feature importance for soil pH prediction in 2008 (**left**) and 2023 (**right**). Each point represents a soil sample, and the color indicates the feature value (red = high, blue = low). The horizontal position of each point reflects its SHAP value, indicating the direction and magnitude of the feature's contribution to the model output. Features are ranked in descending order of mean absolute SHAP value. The definitions of the feature (predictor variable) abbreviations are provided in Table 1.

At the point scale, one low-pH and one high-pH sample were selected from each of the two sampling periods to construct SHAP waterfall plots (Figure 7). The results indicate that, for the low-pH samples, soil pH variation in 2008 was primarily influenced by negative contributions from anthropogenic variables such as PM10 and PM2.5 (SHAP values of -0.09 and -0.05 , respectively) and topographic variables such as MRVBF (SHAP value of -0.03); by 2023, this shifted to negative contributions dominated by climatic variables such as LMAT (SHAP value of -0.16) and topographic variables such as DEM and TPI (SHAP values of -0.05 and -0.04 , respectively). For the high-pH samples, soil pH variation in 2008 was primarily influenced by positive contributions from topographic variables such as MRVBF (SHAP value of $+0.15$) and anthropogenic variables such as PM10 (SHAP value of $+0.11$); by 2023, this shifted to positive contributions dominated by climatic variables such as LMAT (SHAP value of $+0.14$) and topographic variables such as TPI (SHAP value of $+0.08$). Taken together, regardless of whether the samples exhibited low or high pH values,

soil pH variation in 2008 was primarily influenced by human activities and topography, whereas by 2023 this had transitioned to topography and climate, further corroborating the sustained attenuation of anthropogenic influence from 2008 to 2023.

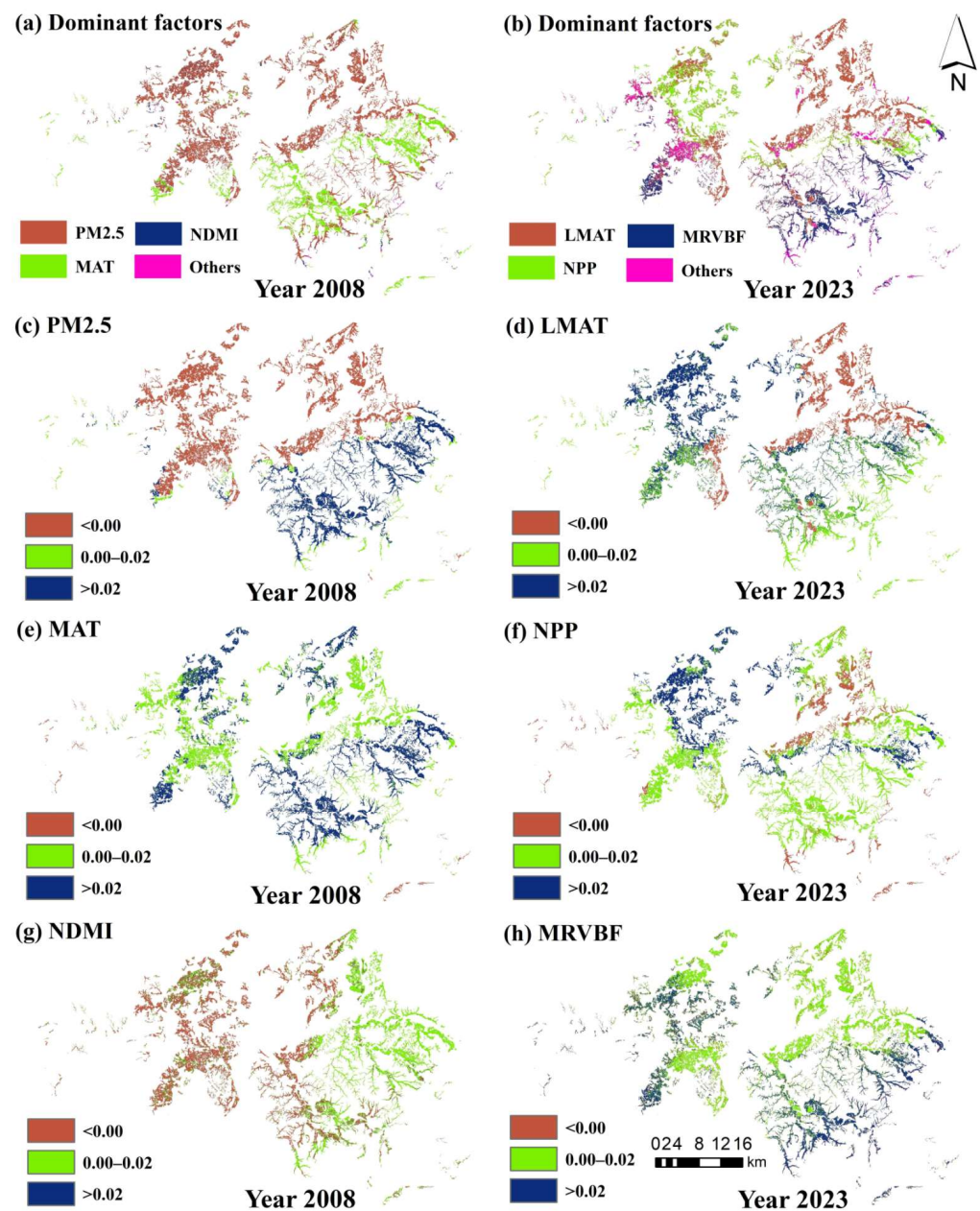


Figure 6. Spatial distribution of dominant driving factors and their GeoShapley values for soil pH in 2008 and 2023. Panels (a,b) show the spatially dominant driving factors identified by GeoShapley for 2008 and 2023, respectively. Panels (c,e) show the spatial distribution of GeoShapley values for PM2.5 and MAT in 2008, and panel (g) shows the corresponding results for NDMI. Panels (d,f) show the spatial distribution of GeoShapley values for LMAT and NPP in 2023, and panel (h) shows the corresponding results for MRVBF. Positive GeoShapley values indicate a positive contribution to soil pH, while negative values indicate a negative contribution. The definitions of the predictor variable abbreviations are provided in Table 1.

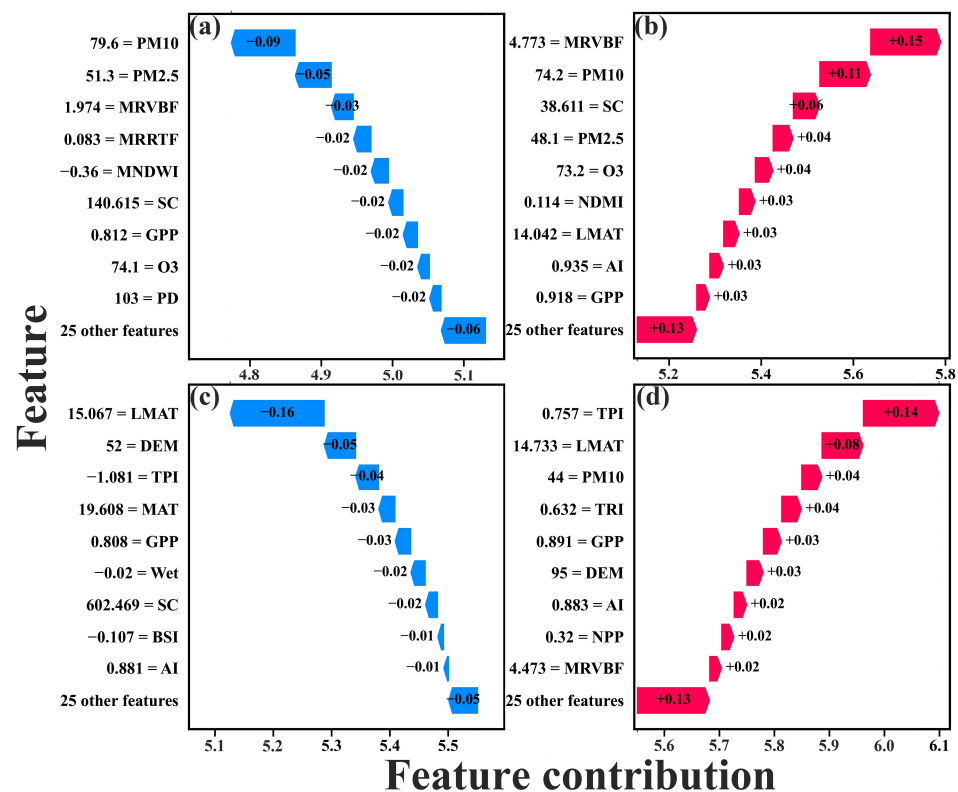


Figure 7. SHAP waterfall plots for representative soil samples in 2008 and 2023. Panels (a,b) show a low-pH sample (predicted pH = 4.77) and a high-pH sample (predicted pH = 5.79) from 2008, respectively. Panels (c,d) show a low-pH sample (predicted pH = 5.13) and a high-pH sample (predicted pH = 6.06) from 2023, respectively. The base value represents the mean model prediction, and each bar indicates the SHAP contribution of an individual feature to the final prediction, with positive values indicating an increase and negative values indicating a decrease in soil pH. The definitions of the predictor variable abbreviations are provided in Table 1.

3.4. Nonlinear Interaction Effects of Soil pH Driving Factors

Figure 8 further illustrates the nonlinear response relationships and interaction effects between the dominant variables and soil pH in both sampling periods. In 2008, PM2.5 exhibited a nonlinear relationship with soil pH: SHAP values were predominantly positive when PM2.5 concentrations were below approximately $49.5 \mu\text{g}/\text{m}^3$, shifting to negative values beyond this threshold, with this negative influence being attenuated in areas with higher MRVBF values. MAT showed a nonlinear positive relationship with soil pH, with SHAP values predominantly negative below approximately 18.5°C and transitioning to positive values beyond this threshold; lower NEP levels further enhanced the positive contribution of MAT to soil pH. NDMI exhibited a piecewise nonlinear relationship with soil pH, with SHAP values shifting from negative to positive at approximately 0.15, and higher PET values helped enhance the positive influence of NDMI on soil pH. In 2023, LMAT showed an overall nonlinear negative relationship with soil pH, with SHAP values predominantly positive below approximately 15°C and transitioning to negative values beyond this threshold; in areas with higher TPI values, this negative influence was partially suppressed. NPP exhibited a nonlinear relationship with soil pH characterized by an initial increase followed by a decrease, with SHAP values shifting from negative to positive at approximately 0.32 and returning to negative at approximately 0.36; in areas with higher DEM values, the positive contribution of NPP to soil pH was more stable. MRVBF similarly exhibited a nonlinear relationship characterized by an initial increase followed by a decrease, with SHAP values shifting from negative to positive at approximately 1.5 and

from positive to negative at approximately 4.5; higher LMAT values tended to attenuate the positive influence of MRVBF on soil pH. Overall, the dominant variables in both periods exhibited pronounced nonlinear relationships with soil pH, and their influence was modulated interactively by other variables.

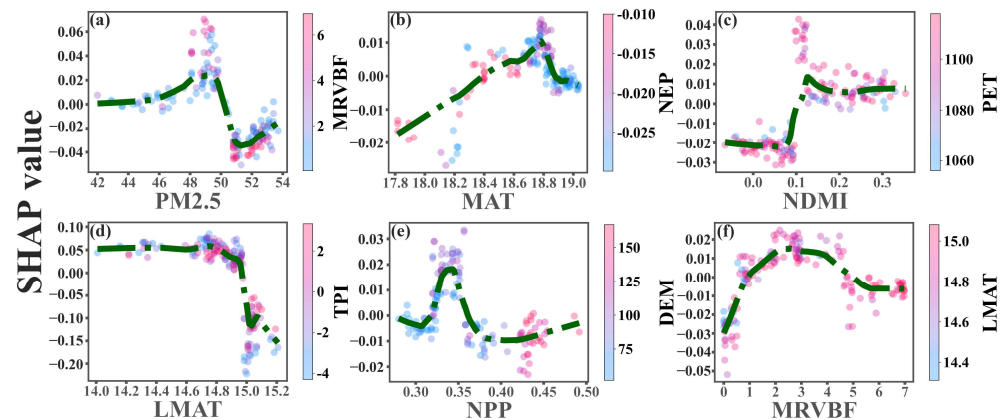


Figure 8. SHAP interaction dependence plots for the dominant driving factors of soil pH in 2008 and 2023. Panels (a–c) show the dependence plots for PM2.5, MAT, and NDMI in 2008, and panels (d–f) show the corresponding plots for LMAT, NPP, and MRVBF in 2023. Each point represents a soil sample, and the color indicates the value of the interacting feature with the strongest interaction effect. The dark green line represents the smoothed mean trend of SHAP values. The definitions of the predictor variable abbreviations are provided in Table 1.

4. Discussion

This study integrated random forest and XGBoost machine learning models with the GeoShapley interpretability framework to systematically elucidate the spatiotemporal changes in soil pH and its multi-scale driving mechanisms before and after high-standard farmland construction in the hilly regions of southern China. Our results showed that the mean soil pH in the study area increased from 5.12 in 2008 to 5.53 in 2023, a rise of 0.41 units. This finding is consistent with Shen et al. [38], who reported that the progressive implementation of acid rain control policies has led to a year-on-year reduction in acid deposition, contributing to an overall increase in soil pH in Yujiang County. Previous research [28] has demonstrated that crop cultivation patterns are an important driver of spatial heterogeneity in soil pH across the red soil region of southern China. Building on this, high-standard farmland construction may have contributed to elevated soil pH by standardizing cultivation practices, promoting the rational application of calcareous soil amendments [39], and advancing straw incorporation and other soil fertility improvement measures, thereby targeting the amelioration of originally scattered, strongly acidic points. We also found that XGBoost outperformed random forest in predictions for both periods, and the spatial prediction patterns of the two models were broadly consistent, indicating good stability of the predictive results. The slightly higher accuracy of XGBoost may reflect its greater capacity to capture the nonlinear interactions among the driving factors of soil pH. Nonetheless, the R^2 values of 0.67–0.71 indicate that a considerable proportion of the variance in soil pH remained unexplained, which may relate to the limited number of sampling points available for county-scale prediction [40]. With respect to spatial patterns, low-pH areas in 2008 were predominantly concentrated in the central part of the study area; by 2023, high-pH areas had expanded into contiguous patches in the western and southeastern regions and low-pH areas had contracted markedly, while the prediction uncertainty remained generally low to moderate across most of the study area. Notably, the expansion of high-pH areas in 2023 was predominantly concentrated in relatively flat

farmland in the western and southeastern regions, broadly coinciding with the areas where high-standard farmland construction had been most intensively implemented. This is consistent with the findings of Xu et al. [41], who reported that the improvement effects of high-standard farmland construction on farmland quality are more pronounced in topographically flat areas and comparatively limited in steeper hilly terrain.

Global feature importance results indicated that climatic factors such as MAT and LMAT were the most important variables associated with soil pH variation in both sampling periods, a pattern consistent with the reported role of temperature in soil acid–base balance through its effects on soil microbial activity and organic matter decomposition rates [42,43]. Meanwhile, anthropogenic factors such as PM_{2.5} ranked second only to climatic factors throughout the study period, indicating that human activities are a key driver of farmland soil pH. Zhang et al. [44], drawing on a national dataset of 7024 regional survey records, similarly found that changes in anthropogenic factors such as nitrogen fertilizer application patterns were a major factor in the spatiotemporal variation in farmland soil pH across China. However, global feature importance reflects the overall contribution of each factor and is insufficient to reveal relative importance at the raster and point scales [36]. Specifically, GeoShapley analyses revealed that in 2008, anthropogenic activities exerted greater influence in the northern region while climatic factors were more influential in the southern region. By 2023, this had shifted to a pattern of combined influence, with climatic factors more influential in the northeastern region, vegetation factors in the northwestern region, and topographic factors in the southwestern region. This shift is likely attributable to the changes in farmland conditions brought about by high-standard farmland construction through measures such as irrigation improvement and land leveling, which in turn altered the dominant driving factors across different regions. This finding is consistent with the results of Yang et al. [45], whose large-scale study across 31 provinces in China demonstrated that high-standard farmland construction modified farmland conditions and thereby altered the response mechanisms of driving factors. Notably, the influence of anthropogenic factors declined markedly over this period, which may be associated with optimized fertilization management and the standardization of cropping systems in the context of high-standard farmland construction; improvements in fertilization practices have been shown to significantly reduce the contribution of human activities to soil acidification rates [46]. The spatially explicit effects of individual variables further revealed pronounced differences in the direction and magnitude of their influence across different periods and regions. This may be related to the compound spatial heterogeneity in soil properties, climatic conditions, topographic characteristics, and cropping systems across different regions, which gives rise to divergent directions and intensities of influence for each variable. Hu et al. [47] similarly found in their study of farmland in the Poyang Lake Plain that the dominant drivers of SOC varied by region, with certain variables even exhibiting opposing directions of influence in different areas. Point-scale analyses further revealed differences in the contributions of driving factors to soils with contrasting pH levels. For instance, in 2008, low-pH samples were primarily influenced by negative contributions from anthropogenic factors such as PM₁₀ and PM_{2.5}, whereas high-pH samples were predominantly influenced by positive contributions from topographic factors such as MRVBF. This contrast may be related to differences in soil pH buffering capacity between the two groups—low-pH soils have weaker acid buffering capacity and are more sensitive to acidic inputs, whereas high-pH soils possess relatively stronger buffering capacity and greater resistance to acidic inputs [48]. From a temporal perspective, regardless of whether the points exhibited low or high pH values, the dominant drivers shifted from human activities and topography in 2008 to climate and topography in 2023. It is noteworthy that while climatic factors consistently ranked highest at the global scale throughout the study

period, climate only emerged as the most influential factor at the point scale by 2023. This may be because anthropogenic factors tended to mask the influence of climate at the point level; as anthropogenic influences weakened, the effect of climate gradually became more apparent [49].

SHAP interaction dependence plots further elucidated the nonlinear response relationships between each dominant factor and soil pH, as well as their interaction effects. In 2008, the SHAP values of PM_{2.5} were predominantly positive below approximately 49.5 µg/m³ and shifted to negative values beyond this threshold. This negative influence was attenuated in areas with higher MRVBF values, a pattern that may be related to the role of topographic factors in moderating the acidifying influence of anthropogenic factors through enhanced local soil acid–base buffering capacity [50]. The SHAP values of MAT were predominantly negative below approximately 18.5 °C and transitioned to positive values beyond this threshold, with positive contributions more pronounced in areas with lower NEP, consistent with the understanding of Khan et al. [51] that vegetation conditions constrain the capacity of temperature to regulate soil microbial activity and organic matter decomposition processes. In 2023, the SHAP values of LMAT were predominantly positive below approximately 15 °C and shifted to negative values beyond this threshold. Higher TPI values were associated with a weakening of this negative influence, which may reflect the role of topographic factors in moderating the effect of climatic factors on enzyme-mediated organic matter decomposition through drainage conditions [52]. The SHAP values of NPP shifted from negative to positive at approximately 0.32 and returned to negative at approximately 0.36. Positive contributions were more stable at higher DEM values, which may be explained by low-temperature conditions at higher elevations limiting excessive organic matter decomposition, thereby stabilizing the positive contribution of vegetation factors to soil pH [53]. The SHAP values of MRVBF shifted from negative to positive at approximately 1.5 and from positive to negative at approximately 4.5; higher LMAT values attenuated the positive effect of MRVBF, which may indicate a counteracting influence of climate warming on the base cation deposition associated with topographic factors [54]. Overall, the moderating roles of topographic and vegetation factors on both anthropogenic and climatic factors persisted across both sampling periods, indicating that the influence of natural factors is long-standing in the study area, and that changes in their apparent importance reflect shifts in the relative strength among driving factors.

In summary, this study showed that soil pH in the study area increased overall following high-standard farmland construction, with a marked reduction in the proportion of sampling points exhibiting extremely low pH values, suggesting that the observed improvement in soil acidity coincided with the implementation of high-standard farmland construction. Furthermore, this study systematically elucidated the driving mechanisms of soil pH variation across three analytical scales—global, raster, and point-level—showing that the relative contribution of anthropogenic factors declined between 2008 and 2023, and demonstrating the methodological value of a multi-scale framework for characterizing the spatial differentiation of soil pH driving mechanisms. Nevertheless, several limitations remain. First, soil chemical properties and specific agricultural management practices were not incorporated into the model, as spatially continuous data on these variables were unavailable, which may have caused part of the anthropogenic intervention effects to be attributed to correlated environmental variables. Second, because this is an observational study without a spatial control (i.e., cultivated land outside high-standard farmland), both the observed pH changes and the factor contributions revealed by the GeoShapley framework reflect statistical associations rather than causal relationships [55]. Third, as the study was conducted in a single county, the findings require further validation before being extended to other regions. Future research could increase sampling density over

larger areas, introduce multi-period continuous monitoring, and explicitly incorporate management factors into the feature system. Extending this framework to other key soil properties such as soil organic carbon and total nitrogen would further advance the understanding of soil acidification processes and the effects of high-standard farmland construction in the hilly regions of southern China.

5. Conclusions

This study focused on Taihe County, Jiangxi Province, and integrated random forest and XGBoost machine learning models with the GeoShapley interpretability framework to systematically investigate the spatiotemporal changes in soil pH and its multi-scale driving mechanisms across three analytical scales—global, raster, and point-level—before and after high-standard farmland construction. The results showed that soil acidification in the study area was alleviated following high-standard farmland construction, with mean soil pH increasing from 5.12 to 5.53 and a marked reduction in the proportion of sampling points exhibiting extremely low pH values. XGBoost outperformed random forest in predictions for both periods, exhibiting superior nonlinear fitting and generalization capacity. Spatial prediction results indicated that soil pH in the western and southeastern regions of the study area increased overall following high-standard farmland construction. With respect to driving mechanisms, global-scale analysis revealed that soil pH spatial variation in both periods was predominantly influenced by climatic factors, followed by anthropogenic factors. At the raster scale, the dominant driving mechanism shifted from a combined influence of anthropogenic and climatic factors in 2008 to the joint influence of climate, vegetation, and topography, with the influence of anthropogenic factors declining markedly. Furthermore, the point-scale analyses corroborated this trend of diminishing anthropogenic influence. Interaction effect analyses indicated that the moderating roles of topographic and vegetation factors on anthropogenic and climatic factors persisted across both sampling periods, suggesting that the influence of natural factors is long-standing in the study area and that changes in their apparent importance reflect shifts in the relative strength among driving factors. Collectively, this study provides a scientific basis for the targeted improvement of cultivated land fertility in high-standard farmland in the hilly regions of southern China while also offering a methodological reference for the systematic application of interpretable machine learning in multi-scale digital soil mapping.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/agriculture16151685/s1>, Table S1. Hyperparameter search ranges for the Random Forest and XGBoost models.

Author Contributions: Conceptualization, Z.Z. and Y.J.; methodology, Z.Z.; software, Z.Z.; validation, Z.Z., F.L. and Y.J.; formal analysis, Y.C.; investigation, F.L.; resources, Z.Z.; data curation, Z.Z.; writing—original draft preparation, Z.Z.; writing—review and editing, Z.Z., Y.C. and Y.J.; visualization, F.L.; supervision, M.C.; project administration, M.C.; funding acquisition, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 42461041.

Data Availability Statement: The data presented in the study are available upon request from the corresponding author.

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

References

- Guo, J.H.; Liu, X.J.; Zhang, Y.; Shen, J.; Han, W.; Zhang, W.; Christie, P.; Goulding, K.; Vitousek, P.; Zhang, F. Significant acidification in major Chinese croplands. *Science* **2010**, *327*, 1008–1010. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bünemann, E.K.; Bongiorno, G.; Bai, Z.; Creamer, R.E.; De Deyn, G.; De Goede, R.; Flesskens, L.; Geissen, V.; Kuiper, T.W.; Mäder, P. Soil quality—A critical review. *Soil Biol. Biochem.* **2018**, *120*, 105–125. [\[CrossRef\]](#)
- Haynes, R.; Swift, R. Effects of soil acidification on the chemical extractability of Fe, Mn, Zn and Cu and the growth and micronutrient uptake of highbush blueberry plants. *Plant Soil* **1985**, *84*, 201–212. [\[CrossRef\]](#)
- Xu, X.-L.; Zhang, J.-B. Research progress on aluminum toxicity and its control in forest soil–plant system. *Chin. J. Ecol.* **2017**, *36*, 1106–1116. [\[CrossRef\]](#)
- Schmidt, S.B.; Husted, S. The biochemical properties of manganese in plants. *Plants* **2019**, *8*, 381. [\[CrossRef\]](#) [\[PubMed\]](#)
- De Vries, W.; Breeuwsma, A. The relation between soil acidification and element cycling. *Water Air Soil Pollut.* **1987**, *35*, 293–310. [\[CrossRef\]](#)
- Tao, L.; Li, F.-B.; Liu, C.-S.; Feng, X.-H.; Gu, L.-L.; Wang, B.-R.; Wen, S.-L.; Xu, M.-G. Mitigation of soil acidification through changes in soil mineralogy due to long-term fertilization in southern China. *Catena* **2019**, *174*, 227–234. [\[CrossRef\]](#)
- Hao, S.; Wang, G.; Yang, Y.; Zhao, S.; Huang, S.; Liu, L.; Zhang, H. Promoting grain production through high-standard farmland construction: Evidence in China. *J. Integr. Agric.* **2024**, *23*, 324–335. [\[CrossRef\]](#)
- Wu, Z.; Wang, B.; Feng, X.; Liang, J.; Zhong, Y.; Yan, Q.; Wu, Z. Mapping Soil Organic Carbon in High-standard Farmland Construction Areas Using Machine Learning Algorithms: Threshold and Interaction Effects of Environmental Variables. *Land Degrad. Dev.* **2025**, *36*, 4770–4782. [\[CrossRef\]](#)
- Xie, E.; Zhao, Y.; Li, H.; Shi, X.; Lu, F.; Zhang, X.; Peng, Y. Spatio-temporal changes of cropland soil pH in a rapidly industrializing region in the Yangtze River Delta of China, 1980–2015. *Agric. Ecosyst. Environ.* **2019**, *272*, 95–104. [\[CrossRef\]](#)
- Slessarev, E.W.; Lin, Y.; Bingham, N.L.; Johnson, J.E.; Dai, Y.; Schimel, J.P.; Chadwick, O.A. Water balance creates a threshold in soil pH at the global scale. *Nature* **2016**, *540*, 567–569. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, Q.; Gu, F.; Zhou, Y.; Xu, T.; Wang, L.; Zuo, Q.; Xiao, L.; Liu, J.; Tian, Y. Changes in the impacts of topographic factors, soil texture, and cropping systems on topsoil chemical properties in the mountainous areas of the subtropical monsoon region from 2007 to 2017: A case study in Hefeng, China. *Int. J. Environ. Res. Public Health* **2021**, *18*, 832. [\[PubMed\]](#)
- Zhang, Y.-Y.; Wu, W.; Liu, H. Factors affecting variations of soil pH in different horizons in hilly regions. *PLoS ONE* **2019**, *14*, e0218563. [\[CrossRef\]](#) [\[PubMed\]](#)
- Huang, H.; Liu, Y.; Liu, Y.; Tong, Z.; Ren, Z.; Xie, Y. Digital mapping of soil pH and driving factor analysis based on environmental variable screening. *Sustainability* **2025**, *17*, 3173.
- Zhu, Q.; De Vries, W.; Liu, X.; Zeng, M.; Hao, T.; Du, E.; Zhang, F.; Shen, J. The contribution of atmospheric deposition and forest harvesting to forest soil acidification in China since 1980. *Atmos. Environ.* **2016**, *146*, 215–222. [\[CrossRef\]](#)
- Wu, Z.; Sun, X.; Sun, Y.; Yan, J.; Zhao, Y.; Chen, J. Soil acidification and factors controlling topsoil pH shift of cropland in central China from 2008 to 2018. *Geoderma* **2022**, *408*, 115586. [\[CrossRef\]](#)
- Yan, P.; Peng, H.; Yan, L.; Zhang, S.; Chen, A.; Lin, K. Spatial variability in soil pH and land use as the main influential factor in the red beds of the Nanxiong Basin, China. *PeerJ* **2019**, *7*, e6342. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sun, W.; Li, S.; Zhang, G.; Fu, G.; Qi, H.; Li, T. Effects of climate change and anthropogenic activities on soil pH in grassland regions on the Tibetan Plateau. *Glob. Ecol. Conserv.* **2023**, *45*, e02532. [\[CrossRef\]](#)
- Zhang, X.; Xiang, D.-Q.; Yang, C.; Wu, W.; Liu, H.-B. The spatial variability of temporal changes in soil pH affected by topography and fertilization. *Catena* **2022**, *218*, 106586. [\[CrossRef\]](#)
- Zhang, X.; Nian, L.; Liu, X.; Li, X.; Adingo, S.; Liu, X.; Wang, Q.; Yang, Y.; Zhang, M.; Hui, C. Spatial–Temporal Correlations between Soil pH and NPP of Grassland Ecosystems in the Yellow River Source Area, China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 8852. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, D.; Xiao, E.; Xia, Y.; Liang, X.; Guo, M.; Ning, L.; Yan, J. Improving soil pH prediction and mapping using anthropogenic variables and machine learning models. *Geocarto Int.* **2025**, *40*, 2482699. [\[CrossRef\]](#)
- Wang, L.; Wu, W.; Liu, H.-B. Digital mapping of topsoil pH by random forest with residual kriging (RFRK) in a hilly region. *Soil Res.* **2019**, *57*, 387–396. [\[CrossRef\]](#)
- Huang, P.; Huang, Q.; Wang, J.; Shi, Y. Predicting surface soil pH spatial distribution based on three machine learning methods: A case study of Heilongjiang Province. *Environ. Monit. Assess.* **2025**, *197*, 367. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hengl, T.; Consoli, D.; Tian, X.; Nauman, T.W.; Nussbaum, M.; Isik, M.S.; Parente, L.; Ho, Y.-F.; Simoes, R.; Gupta, S. OpenLandMap-soildb: Global soil information at 30 m Spatial resolution for 2000–2022+ based on Spatiotemporal machine learning and harmonized legacy soil samples and observations. *Earth Syst. Sci. Data* **2026**, *18*, 989–1036. [\[CrossRef\]](#)
- Rohmer, J.; Belbeze, S.; Guyonnet, D. Insights into the prediction uncertainty of machine-learning-based digital soil mapping through a local attribution approach. *Soil* **2024**, *10*, 679–697. [\[CrossRef\]](#)
- Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4777.

27. Li, Z. GeoShapley: A game theory approach to measuring spatial effects in machine learning models. *Ann. Am. Assoc. Geogr.* **2024**, *114*, 1365–1385. [\[CrossRef\]](#)
28. Zou, H.; Li, W.; Guo, X.; Jiang, Y.; Cai, Y.; Wang, H.; Zhu, Q. Spatial heterogeneity of soil acidification driven by cropping patterns and soil types in red soil dryland of Southern China. *Eur. J. Agron.* **2025**, *170*, 127783. [\[CrossRef\]](#)
29. Li, J.; He, H.; Zeng, Q.; Chen, L.; Sun, R. A Chinese soil conservation dataset preventing soil water erosion from 1992 to 2019. *Sci. Data* **2023**, *10*, 319. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Wu, Y.; Shi, K.; Chen, Z.; Liu, S.; Chang, Z. An Improved Time-Series DMSP-OLS-Like Data (1992–2024) in China by Integrating DMSP-OLS and SNPP-VIIRS. Harvard Dataverse Dataset. 2021. Available online: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GIYGJU> (accessed on 15 March 2025).
31. Wei, J.; Li, Z.; Lyapustin, A.; Sun, L.; Peng, Y.; Xue, W.; Su, T.; Cribb, M. Reconstructing 1-km-resolution high-quality PM2.5 data records from 2000 to 2018 in China: Spatiotemporal variations and policy implications. *Remote Sens. Environ.* **2021**, *252*, 112136. [\[CrossRef\]](#)
32. Wei, J.; Li, Z.; Xue, W.; Sun, L.; Fan, T.; Liu, L.; Su, T.; Cribb, M. The ChinaHighPM10 dataset: Generation, validation, and spatiotemporal variations from 2015 to 2019 across China. *Environ. Int.* **2021**, *146*, 106290. [\[CrossRef\]](#) [\[PubMed\]](#)
33. 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. [\[CrossRef\]](#)
34. Cai, Y.; Zhang, Y.; Jiang, Y.; Guo, X. Spatiotemporal assessment and driving factors of ecosystem health: A case study of two provinces in Southern China. *Biology* **2025**, *14*, 671. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Jiang, Y.; Jiang, Y.; Guo, X.; Guo, Z.; Ye, Y.; Huang, J.; Liu, J. Monitoring of Farmland Abandonment Based on Google Earth Engine and Interpretable Machine Learning. *Agriculture* **2025**, *15*, 2090. [\[CrossRef\]](#)
36. Schmidinger, J.; Heuvelink, G.B. Validation of uncertainty predictions in digital soil mapping. *Geoderma* **2023**, *437*, 116585. [\[CrossRef\]](#)
37. Liao, Y.; Yu, Z.; Kuang, L.; Jiang, Y.; Yu, C.; Li, W.; Liu, M.; Guo, X.; Ye, Y. Analysis of cultivated land degradation in southern China: Diagnostics, drivers, and restoration solutions. *Front. Plant Sci.* **2025**, *16*, 1533855. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Shen, Y.; Zhang, Z.; Xue, Y. Study on the new dynamics and driving factors of soil pH in the red soil, hilly region of South China. *Environ. Monit. Assess.* **2021**, *193*, 304. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Zhang, S.; Zhu, Q.; de Vries, W.; Ros, G.H.; Chen, X.; Muneer, M.A.; Zhang, F.; Wu, L. Effects of soil amendments on soil acidity and crop yields in acidic soils: A world-wide meta-analysis. *J. Environ. Manag.* **2023**, *345*, 118531. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Dash, P.K.; Miller, B.A.; Panigrahi, N.; Mishra, A. Exploring the effect of sampling density on spatial prediction with spatial interpolation of multiple soil nutrients at a regional scale. *Land* **2024**, *13*, 1615. [\[CrossRef\]](#)
41. Xu, D.; Chen, M.; Li, R. How high-standard farmland construction can affect farmers' grain cultivation behavior: Evidence from rural China. *Habitat Int.* **2025**, *160*, 103384. [\[CrossRef\]](#)
42. Ma, C.; Tu, Q.; Zheng, S.; Deng, S.; Xia, Y.; Mao, W.; Gao, W.; Hu, L.; Kuzyakov, Y.; Hu, Y. Soil acidification induced by intensive agricultural use depending on climate. *J. Soils Sediments* **2022**, *22*, 2604–2607. [\[CrossRef\]](#)
43. Herreid, A.M.; Fazekas, H.M.; Nelson, S.J.; Wymore, A.S.; Murray, D.; Varner, R.K.; McDowell, W.H. Climate displaces deposition as dominant driver of dissolved organic carbon concentrations in historically acidified lakes. *Biogeochemistry* **2024**, *168*, 5. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Zhang, W.; Wei, C.; Li, J.; Zhou, Q.; Ma, J.; Li, Y.; He, Y.; Zou, Y.; Luo, Y.; Yao, Y. Stabilization of acidification in China's cropland soils. *Nat. Geosci.* **2025**, *18*, 1125–1132. [\[CrossRef\]](#)
45. Yang, B.; Guo, W.; Tan, Q.; Chen, F.; Liu, X.; Jin, G. Impacts of high-standard farmland construction policy on cultivated land system resilience. *Trans. Chin. Soc. Agric. Eng.* **2025**, *41*, 269–278. [\[CrossRef\]](#)
46. Zhu, X.; Ros, G.H.; Xu, M.; Xu, D.; Cai, Z.; Sun, N.; Duan, Y.; de Vries, W. The contribution of natural and anthropogenic causes to soil acidification rates under different fertilization practices and site conditions in southern China. *Sci. Total Environ.* **2024**, *934*, 172986. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Hu, B.; Geng, Y.; Lin, Y.; Ni, H.; Xie, M.; Wang, N.; Hu, J.; Zou, Q.; Chen, S.; Zhou, Y. Comparison of Machine Learning and Geostatistical Methods on Mapping Soil Organic Carbon Density in Regional Croplands and Visualizing Its Location-Specific Dominators via Interpretable Model. *Land Degrad. Dev.* **2025**, *36*, 3405–3427. [\[CrossRef\]](#)
48. Jiang, J.; Wang, Y.-P.; Yu, M.; Li, K.; Shao, Y.; Yan, J. Responses of soil buffering capacity to acid treatment in three typical subtropical forests. *Sci. Total Environ.* **2016**, *563*, 1068–1077. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Jorge, F.; Mutwale-Mutale, N.; Sandhage-Hofmann, A.; Braun, M.; Cambule, A.; Nhantumbo, A.; Chabala, L.; Shepande, C.; Chishala, B.; Lisboa, S. Anthropogenic disturbances superimpose climate effects on soil organic carbon in savanna woodlands of sub-Saharan Africa. *Glob. Biogeochem. Cycles* **2025**, *39*, e2023GB008086. [\[CrossRef\]](#)
50. Bowman, W.D.; Cleveland, C.C.; Halada, L.; Hreško, J.; Baron, J.S. Negative impact of nitrogen deposition on soil buffering capacity. *Nat. Geosci.* **2008**, *1*, 767–770. [\[CrossRef\]](#)

51. Khan, M.T.; Supronienė, S.; Žvirdauskienė, R.; Aleinikovienė, J. Climate, soil, and microbes: Interactions shaping organic matter decomposition in croplands. *Agronomy* **2025**, *15*, 1928. [[CrossRef](#)]
52. Razavi, B.S.; Blagodatskaya, E.; Kuzyakov, Y. Nonlinear temperature sensitivity of enzyme kinetics explains canceling effect—A case study on loamy haplic Luvisol. *Front. Microbiol.* **2015**, *6*, 1126. [[CrossRef](#)] [[PubMed](#)]
53. Filimonenko, E.; Kuzyakov, Y. Activation energy of organic matter decomposition in soil and consequences of global warming. *Glob. Change Biol.* **2025**, *31*, e70472. [[CrossRef](#)] [[PubMed](#)]
54. Husson, O. Redox potential (Eh) and pH as drivers of soil/plant/microorganism systems: A transdisciplinary overview pointing to integrative opportunities for agronomy. *Plant Soil* **2013**, *362*, 389–417. [[CrossRef](#)]
55. Rentschler, T.; Scholten, T. A note on spurious correlations and explainable machine learning in digital soil mapping. *Eur. J. Soil Sci.* **2025**, *76*, e70172. [[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.