

## Article

# Comparing Soil pH Mapping from Multi-Temporal PlanetScope and Sentinel-2 Data Across Land Use Types

Ziyu Wang<sup>1</sup>, Wei Wu<sup>2</sup> and Hongbin Liu<sup>1,\*</sup> 

<sup>1</sup> College of Resources and Environment, Southwest University, Beibei District, Chongqing 400716, China; forward521@email.swu.edu.cn

<sup>2</sup> College of Computer and Information Science, Southwest University, Beibei District, Chongqing 400716, China; ww@swu.edu.cn

\* Correspondence: lhbin@swu.edu.cn

**Abstract:** In vegetated areas, soil pH impacts plant growth, soil properties, and spectral characteristics. Remote sensing enables soil pH mapping by delivering detailed surface data, and while high-resolution satellite images show great potential in complex terrains, research in this area is still limited. This study evaluated PlanetScope (high-resolution) and Sentinel-2 (medium-resolution) images in estimating soil pH across diverse land use types in southwestern China's hilly areas. It examined how spectral variables from four seasonal images affect prediction accuracy. We integrated topographic and spectral variables at seven spatial resolutions (3 m, 10 m, 20 m, 30 m, 40 m, 50 m, and 60 m), using extreme gradient boosting (XGboost) for orchards, dry land, and paddy fields. We found that the models developed with PlanetScope images tended to achieve better prediction accuracy compared to those utilizing Sentinel-2 images. For each satellite, single-temporal images showed greater predictive power under each land use type. In particular, the spring spectral data showed desirable predictive performance for the orchards and the paddy fields, while the autumn spectral data contributed more effectively to the models for the dry land. Specifically, PlanetScope provided the best prediction accuracy for soil pH at 3 m resolution (orchard:  $R^2 = 0.72$ ,  $MAE = 0.24$ ,  $RMSE = 0.30$ ,  $RPD = 1.91$ ; dry land:  $R^2 = 0.77$ ,  $MAE = 0.37$ ,  $RMSE = 0.40$ ,  $RPD = 2.09$ ; paddy field:  $R^2 = 0.66$ ,  $MAE = 0.35$ ,  $RMSE = 0.41$ ,  $RPD = 1.71$ ), while Sentinel-2 performed better at 10 m resolution (orchard:  $R^2 = 0.67$ ,  $MAE = 0.29$ ,  $RMSE = 0.33$ ,  $RPD = 1.75$ ; dry land:  $R^2 = 0.70$ ,  $MAE = 0.39$ ,  $RMSE = 0.47$ ,  $RPD = 1.83$ ; paddy field:  $R^2 = 0.64$ ,  $MAE = 0.34$ ,  $RMSE = 0.42$ ,  $RPD = 1.66$ ). Our findings demonstrate that sensor selection, land use, temporal phases, and modeling resolution significantly impact outputs. High-resolution PlanetScope images prove effective for predicting soil pH in complex terrains.



Academic Editor: Dusan Gleich

Received: 28 November 2024

Revised: 31 December 2024

Accepted: 6 January 2025

Published: 7 January 2025

**Citation:** Wang, Z.; Wu, W.; Liu, H. Comparing Soil pH Mapping from Multi-Temporal PlanetScope and Sentinel-2 Data Across Land Use Types. *Remote Sens.* **2025**, *17*, 189. <https://doi.org/10.3390/rs17020189>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** hilly topography; extreme gradient boosting; digital soil mapping; feature importance; soil property prediction

## 1. Introduction

Soil pH plays a key role in determining soil fertility and evaluating acidification [1]. Soil acidity or alkalinity is integral to numerous chemical reactions that have a profound impact on soil structure, nutrient availability, elemental migration during soil formation, and microbial activity [2,3]. In recent years, there has been growing attention towards discussions on food security and environmental concerns. As an important parameter for evaluating land quality, soil pH influences the presence, transformation, and efficacy of soil

nutrients by affecting various soil properties [4]. Understanding soil pH regional distribution is essential for sensible and effective land management and sustainable agricultural development. Nevertheless, conventional approaches employed for field-scale soil pH measurement are characterized by their time-consuming nature, labor-intensive requirements, and substantial associated costs. They are also not suitable for large-scale monitoring and do not capture the spatial variability of soil pH [5]. Therefore, it is imperative to explore reliable and efficient approaches to monitoring soil pH.

At present, Digital Soil Mapping (DSM) is recognized as a highly effective method for accurately mapping different soil properties [6–8]. In contrast to traditional approaches, DSM provides numerous benefits, such as efficiency, cost-effectiveness, and the ability to capture detailed spatial variations in soil properties [9–15]. DSM is the process of creating models to predict soil properties by establishing quantitative correlations between these variables and pertinent environmental conditions. Powerful and accurate machine learning models have been created to achieve this goal, including Support Vector Machine (SVM), Random Forest (RF), Multivariate Adaptive Regression Splines (MARSs), the Cubist model, the Deep Learning Neural Network model, Gradient Boosting Regression Trees (GBRT), and Extreme Gradient Boosting (XGboost) [16–19]. Of these, the XGboost algorithm is preferred for its powerful processing abilities and strong resistance to overfitting, making it highly suitable for predicting soil properties [20–22].

Earlier research on DSM has largely depended on observed indicators and a limited set of environmental variables to predict soil pH levels. In regions with vegetation, pH levels affect plant growth and the physicochemical properties of the soil, leading to unique spectral patterns [23–25]. Leveraging remote sensing technology, numerous reliable spectral variables derived from optical satellite data have been utilized to model soil–landscape relationships and effectively predict soil pH. In recent years, the use of sensors such as Sentinel-2, MODIS, and Landsat for soil pH prediction has gained widespread application and achieved significant success [26–29]. However, the predictive performance of satellite data for soil pH varies substantially at different levels of terrain complexity. In flat regions, where geomorphological variability is minimal, optical sensors can accurately capture the surface textures and vegetation growth patterns that indirectly reflect the chemical properties of the soils in the area. For example, Asa et al. conducted studies in four agricultural zones in Prestavlky, using Sentinel-2 data. They demonstrated that incorporating spectral indices further enhanced the robustness of medium-resolution imagery modeling [30]. Similarly, Xia et al. utilized high-resolution Gaofen-2 data combined with vegetation indices, such as the Normalized Difference Vegetation Index (NDVI), to generate pH distribution maps in the North China Plain, achieving excellent model performance ( $R^2 = 0.82$ ) [26]. These studies indicate that in flat terrains, various satellite sensors can effectively model soil pH by capturing vegetation cover and bare surface features. However, conditions differ markedly in hilly or mountainous regions characterized by complex terrain. These areas often involve steep slopes, increased shadow effects, and irregular distributions of surface reflectance signals. Relying solely on single-source optical satellite data may not adequately describe soil properties under such conditions; additional parameters, such as the Digital Elevation Model (DEM), are often required, to improve predictions. For instance, Zhang et al. investigated sloping areas on the Qinghai–Tibet Plateau, using Landsat 8 data integrated with DEM [31]. They demonstrated that incorporating variables like slope aspect and elevation gradients significantly improved the ability of low-resolution imagery to characterize large-scale soil pH gradients. In addition to DEMs, radar data have increasingly emerged as a valuable supplement for addressing challenges posed by complex terrains. Wang et al. proposed a fusion method combining Sentinel-1 radar intensity with Sentinel-2 optical imagery to achieve more accurate DSM mapping in the hilly regions of

eastern China. This multi-source integration not only compensated for information loss from traditional optical data in forested areas but also captured micro-scale local variations, enhancing prediction accuracy ( $R^2 = 0.76$ ,  $RMSE = 0.45$ ) [21].

While traditional medium- and low-resolution sensors have played a significant role in DSM, their inherent data limitations often restrict their effectiveness. In regions characterized by complex terrain, rapid dynamic changes, or high small-scale heterogeneity, such sensors struggle to capture sufficient detail. This leads to reduced model-prediction accuracy and limits the ability to precisely identify and map soil properties. However, advances in remote sensing technologies—particularly high-resolution satellite imagery and Unmanned Aerial Vehicle (UAV)-based remote sensing—offer promising solutions to these challenges. For instance, Zheng et al. employed UAV-based multispectral imagery to map soil organic matter content in the hilly terrain of a national modern agricultural industrial park in northern Cixi City, Zhejiang Province, China [32]. Their results demonstrated that the high spatial resolution of UAV imagery significantly improves prediction accuracy in areas with small-scale heterogeneity, enhancing the sensitivity of models to spatial variability in soil properties. Similarly, Wang et al. investigated the synergistic use of Sentinel-2 and PlanetScope data for soil property mapping in mountainous regions. They found that PlanetScope's ultra-high spatial resolution effectively compensates for Sentinel-2's limited temporal resolution, facilitating comprehensive dynamic monitoring [33]. Furthermore, Webb et al. utilized hyperspectral data acquired from UAV-mounted cameras combined with parent material and slope aspect information to accurately delineate micro-topographic units and their associated soil properties on spring wheat slopes in southwestern Montana, USA [34]. These findings underscore the substantial potential of high-resolution imagery for predicting soil attributes in complex terrains. Notably, the recent emergence of PlanetScope satellites has garnered increasing attention from researchers, due to their superior temporal and spatial resolution [35–39]. Beyond generating ultra-fine-grained remote sensing imagery, PlanetScope satellites feature spectral response curves specifically optimized for agricultural land applications [40]. This design enables them to extract core effective bands critical for identifying detailed soil characteristics within farmland environments. In complex terrain settings, PlanetScope's ultra-high-resolution remote sensing imagery combined with auxiliary variables, such as DEM, may hold great potential for distinguishing discontinuities between adjacent small units more precisely. This capability could further enhance DSM model applicability under challenging environmental conditions by improving accuracy across diverse landscapes with intricate topography and variable landforms.

The PlanetScope land observation mini-satellite system is deployed by Planet, a renowned American commercial remote sensing satellite company. These advanced satellites offer stable data transmission capabilities and can capture remote sensing images daily, with spatial resolutions ranging from 3 to 5 m [40]. With their high spatial resolution, spectral resolution, and coverage frequency characteristics, PlanetScope satellites can monitor the entire globe once a day [40]. Compared to Sentinel-2, although PlanetScope satellites are not available for free, their higher temporal and spatial resolutions provide significant opportunities for analyzing soil information at different scales. Currently, PlanetScope satellite data have shown dependable outcomes in areas like agricultural monitoring, marine observation, and land cover classification [37–39,41,42]. However, the complete capabilities of PlanetScope satellite sensors in forecasting soil pH remain underutilized. Prior research has mainly focused on comparing different medium-resolution sensors' impacts on models predicting soil properties, seldom examining how high-spatial-resolution (<5 m) and medium-spatial-resolution (10–30 m) satellite sensors affect these models. Specifically, the widely applicable PlanetScope satellite's potential has not been thoroughly investigated.

Compared to other high-resolution sensors with sub-meter detail capabilities (such as Gaofen-2), PlanetScope stands out not only for its excellent technical performance but also for its broader acceptance, due to international market accessibility. On the other hand, Sentinel-2 has become a favored choice among users because of its free availability and strong adaptability. Although Sentinel-2 is slightly inferior in spatial resolution, with a medium resolution of 10–20 m, it achieves an optimal balance between coverage area and computational cost. Additionally, its 13 key spectral bands, particularly the red-edge bands, provide valuable support for soil property prediction. A comparison of PlanetScope and Sentinel-2 could help elucidate the specific roles and impacts of spatial resolution and spectral characteristics in soil pH prediction within hilly regions.

In vegetated regions, spectral imagery can estimate soil pH by capturing surface vegetation data. This ability may stem from the significant interactions between vegetation variability and soil attributes [23,43]. However, remote sensing images capture only the surface features at a specific moment, so features recorded at different times can vary significantly. Different land use practices, due to their variations in vegetation cover, farming methods, and water resource management, lead to dynamic changes in soil properties over time. These dynamic changes manifest as different spectral characteristics at different time points [21,36]. By analyzing these temporal differences, it is possible to more accurately capture features related to soil pH.

Selecting the appropriate time for satellite image acquisition could reduce complex interference caused by external environmental changes, better reflect the interactions between vegetation and soil, and enhance the accuracy of soil pH prediction by analyzing spectral characteristic differences at different times. Generally speaking, acquiring satellite images during spring, summer, autumn, and winter can enable assessment of how vegetation affects spectral features related to soil properties during stages of emergence, vigor, senescence, and dormancy [33]. For example, in springtime new leaves strongly absorb red light and reflect near-infrared light, while dense and active summer vegetation provides information about health status—both potentially linked to soil pH. During autumn, leaves change color due to environmental conditions, and their gradual yellowing or reddening alters reflection characteristics across different spectral bands; data from this period reveal potential changes in soil pH, due to nutrient retraction or degradation. In winter, with plants reduced or absent, exposed bare surfaces become primary observation targets; directly obtaining reflection data from the exposed ground aids precise analysis of inherent properties, including pH value. Moreover, comparing the predictive performance across imagery acquired in different seasons enables a more comprehensive understanding of how acquisition timing affects model accuracy. It also provides deeper insights into the year-round dynamics of vegetation–soil interactions in varying environmental conditions.

The rugged topography of Southwest China, characterized by its hilly and mountainous regions, poses challenges for accurate prediction of soil pH using remote sensing data, including significant terrain fluctuations, diverse land use types, and fragmented land parcels. Current research has shown that soil pH exhibits different variations depending on the specific land use types, which are affected by factors such as vegetation [44,45]. Nevertheless, accurate soil pH prediction across various land uses remains under-researched. Although Zhang et al. and Xia et al.'s models included land use as a variable, its impact on prediction accuracy was minimal [26,27]. Given the influence of land use on prediction outcomes, exploring remote sensing for soil pH estimation in specific land uses is highly valuable. We hypothesized that focusing remote sensing efforts on single land use categories could improve model precision.

Hence, this study aimed to assess PlanetScope satellite's capability in estimating soil pH by comparing it with the widely used Sentinel-2 satellite. The specific goals were

to (i) compare different models and determine the optimal one for mapping the spatial distribution of soil pH; (ii) explore the influence of optical image acquisition time on prediction accuracy; (iii) evaluate the prediction accuracy of soil pH under different land use types. These objectives were achieved by obtaining satellite images of four seasons, and using the XGboost algorithm to develop soil pH prediction models for different land use types.

## 2. Material and Methods

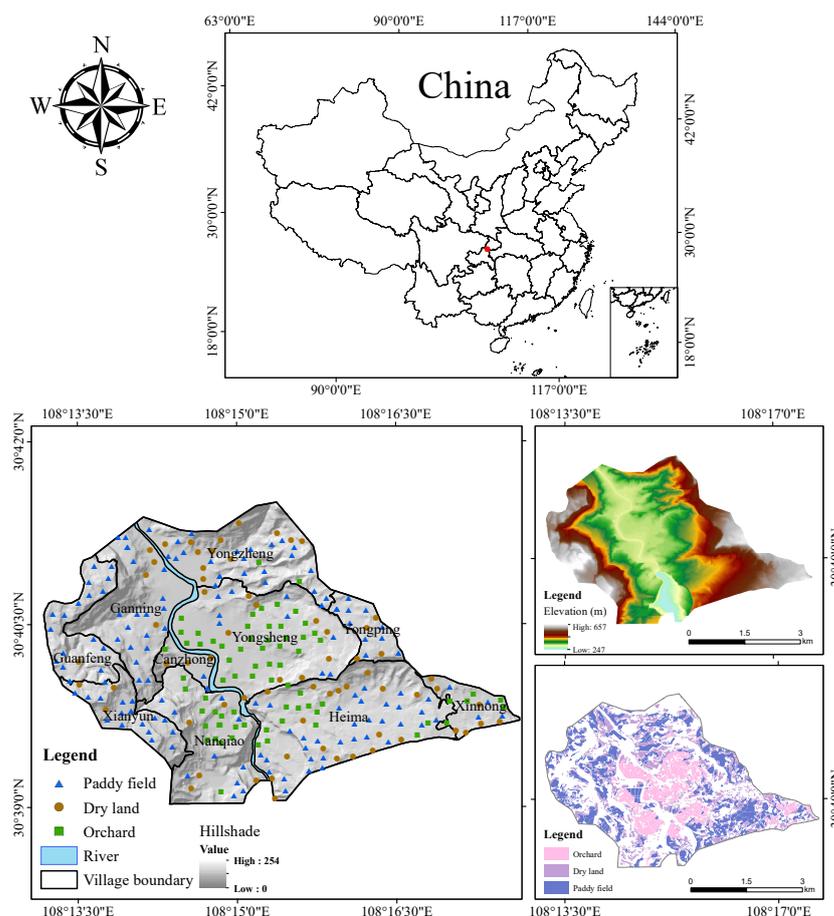
### 2.1. Study Area

The study area is located in Ganning Town, Wanzhou District, Chongqing, China, covering longitudes  $108^{\circ}13'10''$ – $108^{\circ}17'40''$ E and latitudes  $30^{\circ}39'1''$ – $30^{\circ}41'30''$ N. Covering a vast expanse of approximately 1800 hectares, the study area boasts a subtropical monsoon humid climate characterized by four distinct seasons. Based on observational data from the Wanzhou District Meteorological Station, this region experiences an average yearly temperature of  $17^{\circ}\text{C}$ , receives about 1209.7 h of sunshine annually, and has an annual precipitation of approximately 1293.3 mm. The relative humidity is around 75.4%. Regarding geomorphology, the study area features the East Sichuan Tectonic Denudation Hills. The Gannin River traverses through this region, contributing to its complex topographic conditions. The study area includes a notable natural slope, with elevations spanning from 248 to 657 m. Additionally, the topographic slope ranges between 0 and 79 degrees.

The study area exhibits a diverse range of land use types, with orchard, dry land, and paddy field standing out as the most significant ones. Spanning across 233.3 hectares (14.67% of the total land area), the orchards are primarily cultivated with rose oranges, mandarins, and loquats. Following closely is the expansive dry land covering 285.9 hectares (17.98% of the total land area), where crops such as maize, wheat, soya beans, and rapeseed flourish. Lastly, the paddy fields, covering a significant 394.11 hectares (24.78% of the total land area), serves primarily for single-season rice cultivation from April to September, while also functioning as an important water reservoir during the winter months.

### 2.2. Soil Sampling

In 2019, we conducted surveys within the study area, and we determined soil sampling locations, using a “grid combined with plot” method, focusing on surface soils. The core of this method lies in capturing the diversity and representativeness of surface soils through a well-designed grid layout and plot configuration. Initially, the study area was divided into several grids, to systematically cover the entire region. The grid size was determined based on a comprehensive consideration of the terrain complexity and land use types. For the agricultural areas, we selected a grid size with 4 to 6 sampling points per square kilometer. This setup effectively captured the spatial variability of the soil while providing sufficient data points for statistical analysis within a reasonable workload. At each designated sampling point, using the GPS location as the center, we identified 4 to 6 subsampling points within a 30-to-50-m radius. These sub-samples were then combined in equal parts, to form a composite sample. Each initial sample, weighing over 1.5 kg, was placed in specialized bags. After collection, the samples were air-dried at room temperature and passed through a 2 mm sieve before chemical analysis. Soil pH was measured with a pH meter, using a 1:2.5 soil-to-water suspension ratio [46]. A total of 290 topsoil samples were obtained from the study area, including 75 from the orchards, 67 from the dry land, and 148 from the paddy fields (Figure 1).



**Figure 1.** Overview of the study area. Notes: the red dot indicates the location of the study area within China.

### 2.3. Environmental Data for Modeling

Recognizing the link between environmental conditions and soil characteristics, we gathered remote sensing imagery and terrain data for use as environmental inputs in our modeling analysis. These inputs were converted into raster layers with spatial resolutions of 3 m, 10 m, 20 m, 30 m, 40 m, 50 m, and 60 m via bilinear interpolation using ArcGIS 10.8. We then extracted the attribute values from each soil sample point to serve as input for the model [47]. Below is a detailed explanation of how these environmental variables were obtained and processed.

#### 2.3.1. Satellite Images Collection and Processing

##### (1) PlanetScope Images

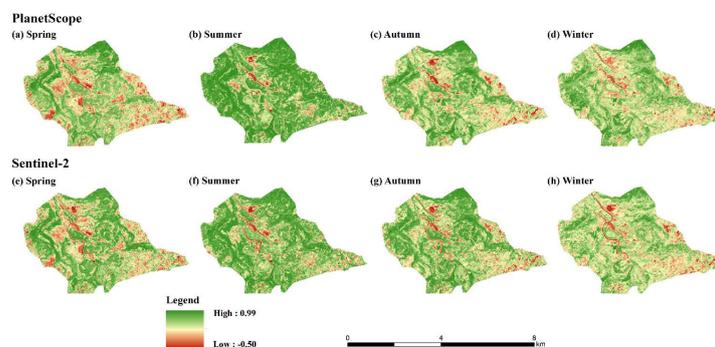
The PlanetScope satellite excels in ground monitoring, with its 3-meter resolution and near-daily revisit cycle. For this study, we utilized L3B-level PSB-SD products from PlanetScope, which feature four spectral bands: blue, green, red, and near-infrared (Table 1). Four images were taken when the cloud cover was less than 5%, covering the four distinct seasons within the study area: spring (17 April 2020), summer (3 August 2020), autumn (14 November 2020), and winter (12 January 2021). All PlanetScope images were subjected to geometric correction using 30 ground control points. This correction was performed by applying an affine transformation, which was determined by the utilization of the least squares registration characterization tool in ArcGIS 10.8. The FLAASH atmospheric model, based on the MODTRAN 4 radiative transfer code created in ENVI 5.1 software, was then used to radiometrically and atmospherically correct all the remote sensing datasets [48].

Following the use of radiometric correction, the pixel values of the image were transformed into spectral reflectance.

**Table 1.** Spectral bands and resolutions of the PlanetScope and Sentinel-2 sensors used in this study.

PlanetScope			Sentinel-2		
Band	Wavelength (nm)	Resolution (m)	Band	Wavelength (nm)	Resolution (m)
Blue (B)	457.5–522.5	3	Blue (B)	458–523	10
Green (G)	542–577.5	3	Green (G)	543–578	10
Red (R)	650–680	3	Red (R)	650–680	10
Near-infrared (NIR)	855–875	3	Red Edge 1 (RE1)	698–713	20
			Red Edge 2 (RE2)	733–748	20
			Red Edge 3 (RE3)	773–793	20
			Near-Infrared (NIR)	785–900	10
			Near-Infrared Narrow (NIRn)	855–875	20
			Shortwave Infrared 1 (SWIR1)	1565–1655	20
			Shortwave Infrared 2 (SWIR2)	2100–2280	20

Figure 2a–d illustrate the NDVI values for the four different time periods of PlanetScope, which serve as an indicator of vegetation distribution during these specific periods. In January, the study area exhibits a low vegetation cover, with many crops yet to be planted. As spring arrives in April, cultivation activities commence, including rice transplantation in the paddy fields and seedling planting in the dry land. Moreover, trees start to develop and thrive in the orchards and woodlands. By August, the crops flourish, resulting in the highest vegetation cover throughout the year. November marks the autumn harvest season, when rice is harvested and oranges ripen in orchards. Additionally, other crops, such as rapeseed, are being planted in dry land during this time period.



**Figure 2.** Images of Normalized Difference Vegetation Index (NDVI) taken by PlanetScope and Sentinel-2, respectively, at different temporal phases over the study area. Notes: for PlanetScope, the spring imagery is from 17 April 2020, the summer imagery is from 3 August 2020, the autumn imagery is from 24 November 2020, and the winter imagery is from 12 January 2021; for Sentinel-2, the spring imagery is from 28 April 2020, the summer imagery is from 26 August 2020, the autumn imagery is from 14 November 2020, and the winter imagery is from 13 January 2021.

## (2) Sentinel-2 Images

We obtained four views of L1C-level images with clouds below 5% from ESA's Sentinel Science Data Centre (<https://scihub.copernicus.eu>, accessed on 20 April 2024) for the years 2020 to 2021, all from Sentinel-2A satellite images, again covering four seasons: spring (28 April 2020), summer (26 August 2020), autumn (14 November 2020), and winter (13 January 2021). Each image independently covered the study area, and only the ten higher-resolution bands were selected as potential predictors for modeling. These bands included Blue (10 m), Green (10 m), Red (10 m), Red Edge 1 (20 m), Red Edge 2 (20 m), Red Edge 3 (20 m), Near-Infrared (10 m), Near-Infrared Narrow (20 m), Shortwave Infrared 1 (20 m), and Shortwave Infrared 2 (20 m) (Table 1). For all Sentinel-2 images, we used the same pre-processing methods as PlanetScope, including atmospheric correction and radiometric calibration, to match the PlanetScope data. In addition, we performed geometric correction on the Sentinel-2 images by utilizing geometrically corrected PlanetScope images, resulting in an RMSE below 0.5 pixels per control point. The study area was then clipped from these Sentinel-2 images.

Figure 2e–h illustrate the NDVI values captured by Sentinel-2 at four specific time points. Due to disparities in the ability of Sentinel-2 and PlanetScope to differentiate between spatial scales of target features, there exist slight variations in both the values and distribution of vegetation information obtained by these two instruments during similar temporal phases. The April Sentinel-2 image was taken towards the end of the month, coinciding with the full transplantation of rice into the fields. Consequently, the vegetation cover during this period was slightly higher compared to mid-April (as shown in Figure 2e). August represented the final month of summer, characterized by vigorous growth of diverse plant species, intensified photosynthesis, heightened chlorophyll concentration, and subsequently elevated NDVI values compared to other times of the year.

Since the conditions were similar during the PlanetScope and Sentinel-2 satellite image collections, they met the comparative soil pH prediction criteria.

Compared to Sentinel-2, PlanetScope, with its high spatial resolution of 3 m and daily imaging capability, provides more precise insights into local variations in soil pH at finer scales [35]. This is particularly advantageous for areas with complex topography, such as steep slopes, intertwined valleys, or small agricultural fields. Additionally, PlanetScope's spectral response curves are optimized for agricultural land use, and its narrower bandwidths help mitigate the mixed-pixel effect (i.e., a single pixel containing multiple distinct surface features), thereby offering more accurate and effective spectral information [49]. However, its limited spectral range may restrict the detection of certain critical soil properties. In contrast, while Sentinel-2 has a lower spatial resolution of 10 m, it compensates with broader spectral coverage across 13 bands, from visible to shortwave infrared regions. This extensive spectral information provides robust support for identifying and characterizing complex soil chemical attributes (e.g., organic matter content or mineral composition). In modeling applications, PlanetScope's high spatial resolution serves as the primary factor driving accuracy improvements, whereas Sentinel-2 relies on its diverse spectral information to enhance overall model performance.

### 2.3.2. Spectral Variables

At each time phase, the original satellite bands and their spectral indices were used as variables to predict soil pH. The spectral indices encompassed four soil radiometric indices, eight vegetation radiometric indices, and two soil salinity indices (Table 2). Based on the association that exists between soil pH and soil organic carbon, the soil radiometric indices and the vegetation radiometric indices were chosen as proxies for predicting the soil pH [50]. Numerous studies have shown a connection between soil salinity and pH levels [51,52].

Since the loss of salt ions increases soil acidity and decreases soil pH, the spectral index associated with soil salinity was selected as the spectral variable to predict the soil pH [53].

**Table 2.** Spectral indices selected as spectral variables for the soil pH prediction.

Type	Spectral Index	Calculation Formula	Reference
Soil radiometric indices	Brightness Index (BI)	$\sqrt{\frac{\rho_{\text{red}}^2 + \rho_{\text{green}}^2}{2}}$	[54]
	Second Brightness Index (BI2)	$\sqrt{\frac{\rho_{\text{red}}^2 + \rho_{\text{green}}^2 + \rho_{\text{nir}}^2}{3}}$	[54]
	Redness Index (RI)	$\frac{\rho_{\text{red}}^2}{\rho_{\text{green}}^3}$	[55]
	Color Index (CI)	$\frac{\rho_{\text{red}} - \rho_{\text{green}}}{\rho_{\text{red}} + \rho_{\text{green}}}$	[55]
Vegetation radiometric indices	Soil-Adjusted Vegetation Index (SAVI)	$\frac{(\rho_{\text{nir}} - \rho_{\text{red}}) \times (1 + L)}{\rho_{\text{nir}} + \rho_{\text{red}} + L}$	[56]
	Modified Soil-Adjusted Vegetation Index (MSAVI)	$\frac{(1 + M) \times (\rho_{\text{nir}} - \rho_{\text{red}})}{\rho_{\text{nir}} + \rho_{\text{red}} + L}$	[57]
	Second Modified Soil-Adjusted Vegetation Index (MSAVI2)	$\frac{2\rho_{\text{nir}} + 1 - \sqrt{(2\rho_{\text{nir}} + 1)^2 - 8(\rho_{\text{nir}} - \rho_{\text{red}})}}{2}$	[58]
	Difference Vegetation Index (DVI)	$\rho_{\text{nir}} - \rho_{\text{red}}$	[59]
	Ratio Vegetation Index (RVI)	$\frac{\rho_{\text{nir}}}{\rho_{\text{red}}}$	[60]
	Weighted Difference Vegetation Index (WDVI)	$\rho_{\text{nir}} - S \times \rho_{\text{red}}$	[61]
	Perpendicular Vegetation Index (PVI)	$\sin(b) \times \rho_{\text{nir}} - \cos(b) \times \rho_{\text{red}}$	[62]
Soil salinity indices	Normalized Difference Vegetation Index (NDVI)	$\frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}}}$	[63]
	Soil salinity index1 (SSII)	$\frac{\rho_{\text{blue}} \times \rho_{\text{red}}}{\rho_{\text{green}}}$	[51]
	Soil salinity index2 (SSI2)	$\frac{\rho_{\text{green}} + \rho_{\text{red}}}{2}$	[64]

Notes:  $L$  is a correction factor that varies from 0, indicating very high vegetation cover, to 1, indicating very low vegetation cover. The parameter  $s$  represents the slope of the soil line.  $M = 1 - 2 \times s \times NDVI \times WDVI$ ;  $b$  denotes the angle between the soil line and the NIR axis, measured in degrees. In this paper, the values assigned to  $L$ ,  $s$ , and  $b$  are 0.5, 0.5, and 45, respectively.

### 2.3.3. Topographic and Land Use Data

To estimate the soil pH, topographic variables were incorporated, due to the hilly and mountainous terrain having significant variations in elevation. Using the DEM data at 5-meter resolution provided by the Land Survey, we utilized ArcGIS 10.8 to derive three key variables: altitude, aspect, and slope. These were subsequently incorporated as indicators for predicting the soil pH.

The land use data included in this analysis were obtained from the Ministry of Natural Resources of the People's Republic of China (<http://www.mnr.gov.cn>, accessed on 6 January 2024). Specifically, the data mostly consisted of the 2019 edition of the land use map scaled to a 1:10,000 scale for the Wanzhou District. The initial compilation of this map relied on information from the Wanzhou District Land Use Database, originally established in 2008 and updated every year by the local Ministry of Natural Resources [65].

## 2.4. Methods

### 2.4.1. Statistical Analysis

The Kolmogorov-Smirnov (K-S) test was employed to evaluate whether the soil pH distribution followed a normal pattern. To investigate the differences in the soil pH across the orchards, dry land, and paddy fields, a one-way ANOVA was conducted, with post

hoc comparisons made using the Least Significant Difference (LSD) test. The significance threshold was set at 0.05. All statistical procedures were carried out using the SPSS V.25 software package.

#### 2.4.2. Extreme Gradient Boosting

In this study, we utilized the extreme gradient boosting (XGboost) method to analyze the association between soil pH and its influencing factors. XGboost, an optimized variant of the Gradient Boosting Decision Tree (GBDT), introduces improvements in both the algorithm's structure and the system's configuration [66]. This approach combines the predictions of multiple weak models to build a more accurate one, using advanced techniques. Unlike GBDT, XGboost leverages second-order derivatives to enhance the objective function's optimization, and it includes a regularization factor, which helps to precisely establish the objective while reducing the potential for overfitting [67]. On the algorithmic front, XGboost uses a weighted quantile sketching algorithm to pre-construct candidate split nodes for each feature during tree creation, thereby decreasing computational demands and speeding up training. From a system architecture standpoint, XGboost organizes the input features by sorting them into memory blocks to enable efficient re-use in later iterations. Additionally, it allows for parallel computations during the training phase, greatly enhancing performance [68]. For this research, we utilized Python 3.9 to implement the XGboost model.

In XGboost, feature importance is evaluated using three key metrics: 'gain', 'frequency', and 'coverage'. 'Gain' quantifies how effectively a feature splits tree nodes, while 'frequency' denotes the occurrence rate of features within the model's structure. Meanwhile, 'coverage' indicates the average impact of feature observations [69]. Features that are repeatedly selected for essential splitting decisions receive higher scores [70]. 'Gain' is particularly important for determining the relevance of features in branching. In our analysis, we focused on feature importance, primarily through the 'gain' metric. For example, the relative importance of feature  $j$  was calculated as follows:

$$\hat{I}_j^2(T) = \sum_{t=1}^{J-1} i_t^2 P(v_t = j) \quad (1)$$

$$\hat{I}_j^2 = \frac{1}{M} \sum_{m=1}^M \hat{I}_j^2(T_m) \quad (2)$$

Let  $T$  represent a tree with  $J$  branching nodes, where  $t$  refers to each node. The metric  $i_t^2$  represents the reduction in squared error at node  $t$ , and  $V_t$  denotes the feature linked to that node.  $M$  stands for the total number of trees in the forest.

The significance of feature  $j$  within a specific tree  $T$  was determined using Equation (1). To assess the overall importance of feature  $j$  across all  $M$  trees, we computed the average importance, using Equation (2).

As proxies for predicting the soil pH, we used raw bands, spectral indices, and topographic variables from each image over four temporal phases. Models were constructed with seven different resolutions for PlanetScope and six different resolutions for Sentinel-2 (Table 3). A flowchart of the soil pH mapping process using these experimental models is shown in Figure 3.

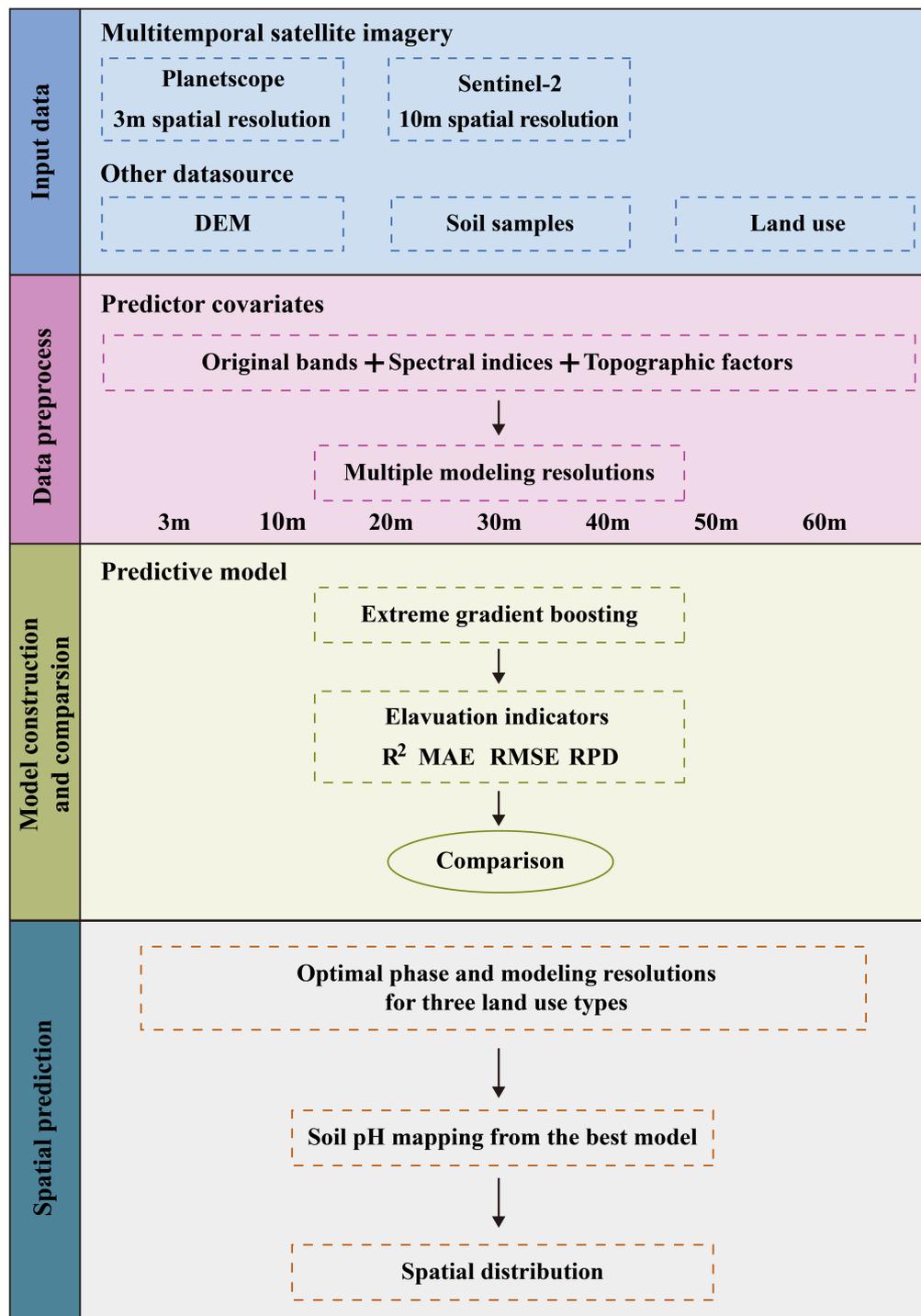


Figure 3. Technical workflow.

Table 3. Model construction across different resolutions.

Model	Modeling Resolution	
	PlanetScope	Sentinel-2
Model A	3 m	
Model B	10 m	10 m
Model C	20 m	20 m
Model D	30 m	30 m
Model E	40 m	40 m
Model F	50 m	50 m
Model G	60 m	60 m

### 2.4.3. Model Validation

In the XGboost model, we focused on four key hyperparameters: learning rate, maximum tree depth, minimum child weight, and lambda for regularization. These parameters were selected due to their significant influence on model performance: learning rate determines the optimization step size, maximum tree depth controls model complexity to balance underfitting and overfitting, minimum child weight prevents overfitting by setting a threshold for node splitting, and lambda enhances robustness through L2 regularization. To identify the optimal parameter combination within the predefined ranges (as detailed in Table 4), we employed a systematic four-level nested loop approach. This method ensures comprehensive exploration, exhaustively testing all potential parameter combinations without overlooking any possibilities, while maintaining simplicity and clarity, as it is both easy to implement and interpret with transparent parameter selection. After determining the optimal hyperparameter combination, 80% of randomly selected samples from each land use category were used as the training dataset, while the remaining 20% were reserved as the validation dataset. This partitioning was designed to evaluate the model's predictive accuracy and generalization capability effectively.

**Table 4.** Parameters setting to determine the optimal parameters for the XGboost algorithm.

Parameters	Threshold	Intervals
Eta	0.01, 0.05, 0.1	-
Max_Depth	1–11	1
Min_Child_Weight	0–21	1
Lambda	0–11	1

Notes: Eta = learning rate; Max\_Depth = maximum tree depth; Min\_Child\_Weight = minimum child weight; Lambda = regression lambda.

To assess the predictive capacity and reliability of the soil pH prediction model, four metrics were chosen to evaluate the model: Coefficient of Determination ( $R^2$ ), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Residual Prediction Deviation (RPD), which were calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (3)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (5)$$

$$RPD = \frac{1}{\sqrt{1 - R^2}} \quad (6)$$

where  $m$  is the sample size,  $y_i$  is the true value,  $\hat{y}_i$  is the predicted value, and  $\bar{y}$  is the mean of  $y_i$ .

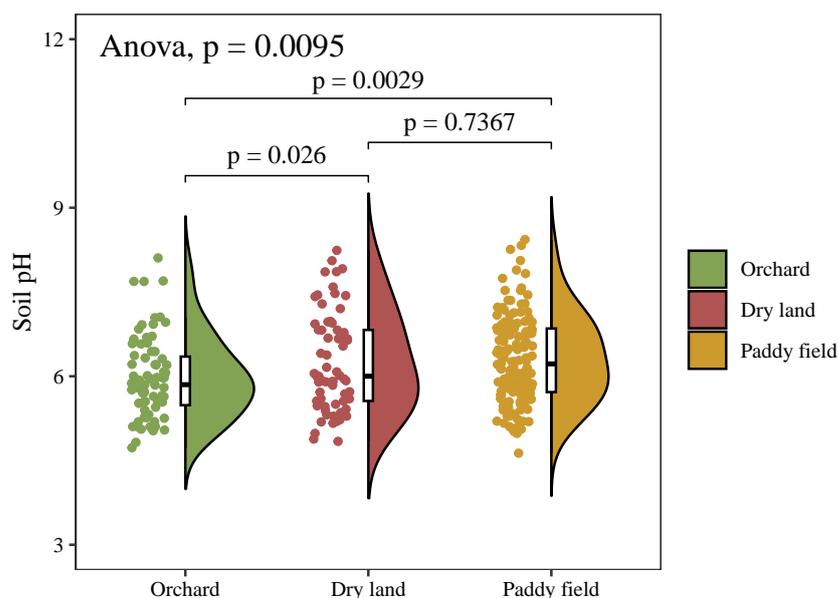
The  $R^2$  score measures how well the model fits the data, with values approaching 1 indicating a stronger fit [71]; MAE and RMSE evaluate the robustness of the model, and when the value is smaller, the lower the prediction error of the model [72]. The RPD evaluates the predictive ability of the model. A higher RPD value suggests the better effectiveness of the model. Specifically, when the RPD value exceeds 2, it implies a higher level of effectiveness [73].

### 3. Results

#### 3.1. Descriptive Statistics and Difference Test of Soil pH

The descriptive statistics for the total area, orchards, dry land, and paddy fields soil pH are displayed in Table 5. The soil pH values in the study area spanned a wide range, from 4.63 to 8.43. The average soil pH was calculated to be 6.20, with a Coefficient of Variation (CV) of 12.71%. These findings suggest a degree of variability in soil pH within the study area. The mean soil pH in the orchards was lower than in the dry land and paddy fields. This was due to the fact that the orchards sample sites were collected during the post-harvest fertilization period of blood oranges, which may have resulted in a low soil pH after fertilization. The Coefficient of Variation (CV) of the soil pH was higher in the dry land compared to the orchards and the paddy fields. This may have been due to the multi-crop rotation in the dry land, where different crops have different fertilizer requirements, resulting in greater variation in soil pH across fields.

The soil pH data from the total area followed a normal distribution (the K-S test). The ANalysis Of VAriance (ANOVA) and subsequent multiple comparisons revealed a statistically significant difference ( $p < 0.05$ ) in the soil pH across the orchards, dry land, and paddy fields (Figure 4). These findings suggest that soil pH is affected to some extent by land use types.



**Figure 4.** Soil pH Least Significant Difference (LSD) test results across orchards, dry land, and paddy fields.

**Table 5.** The statistics of soil pH values for total area, orchards, dry land, and paddy fields.

Land Use Type	N	Min	Max	Mean	SD	CV (%)	Skewness	Kurtosis
Total area	290	4.63	8.43	6.20	0.79	12.71	0.52	−0.25
Orchard	75	4.73	8.11	5.96b	0.72	12.06	0.80	0.54
Dry land	67	4.84	8.24	6.25a	0.87	13.89	0.47	−0.70
Paddy field	148	4.63	8.43	6.29a	0.76	12.13	0.43	−0.16

Notes: N = number; Min = minimum; Max = maximum; SD = Standard Deviation; CV = Coefficient of Variation. Different letters within the mean column indicate that the difference in soil pH among land use types is significant at  $p < 0.05$ .

### 3.2. Analysis of Soil pH Prediction Results

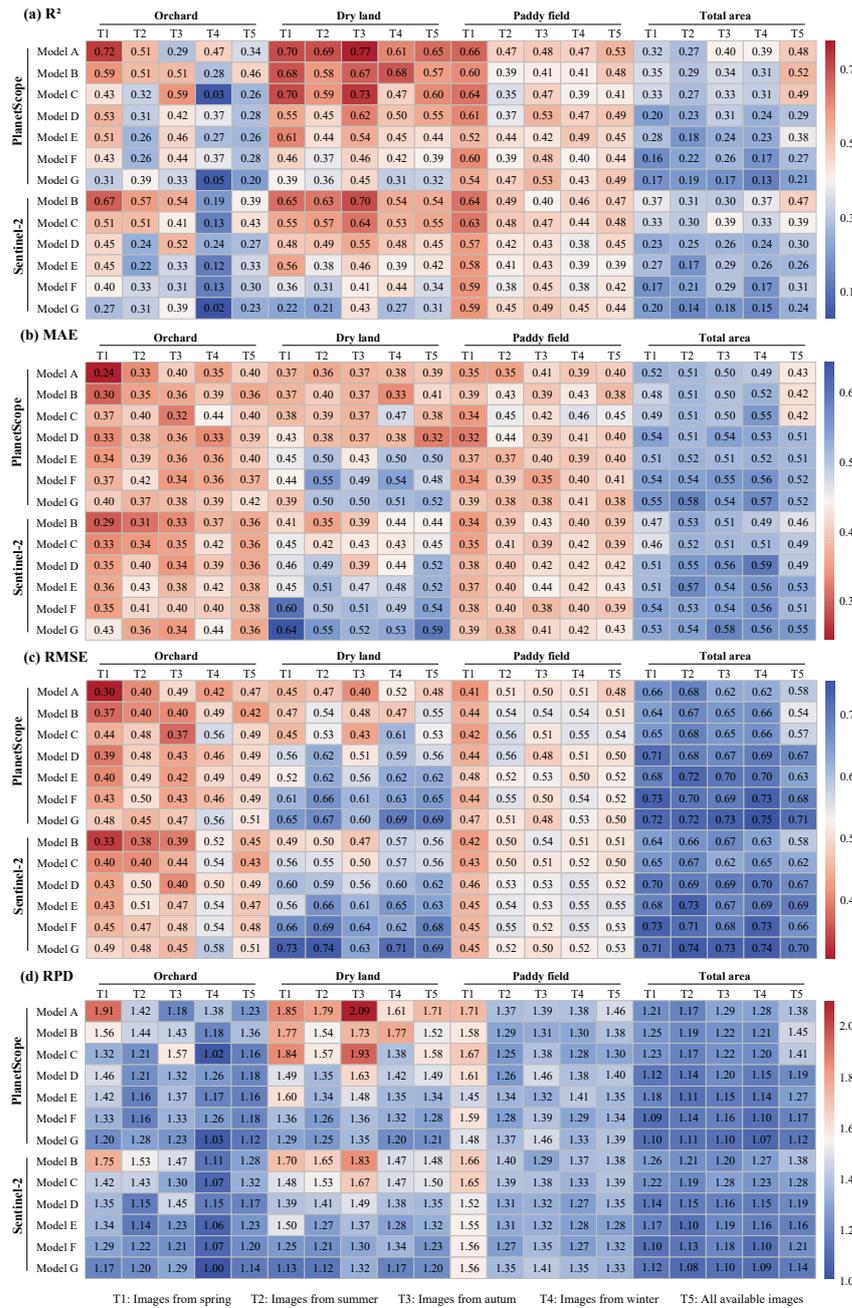
Seven soil pH prediction models were developed for the orchards, dry land, and paddy fields, using the XGboost algorithm based on four temporal phases of PlanetScope and Sentinel-2 images. The prediction accuracies are shown in Figure 5.

For each satellite, under the same modeling resolution, significant variations were observed in the performance of the predictive models built using spectral variables from different temporal phases. This indicates that the timing of spectral image acquisition can influence soil pH prediction. The contribution of images taken at different times to the models varied according to the type of land use (Figure 5). Specifically, for the orchards and paddy fields, images taken during spring from both PlanetScope and Sentinel-2 had a greater impact on the model, with lower error rates compared to other temporal phases. This was consistent with the findings of Wang et al. [33] and Forkuor et al. [74], who also found that satellite data acquired in spring made a significant contribution to soil property predictions. However, in dry land areas, autumn spectral data exhibited a higher contribution to the predictive model. Notably, single-temporal images played a more substantial role in modeling under specific land use types, while multi-temporal images dominated when considering the entire area. Generally, remote sensing images captured diverse information about features at different time points. Consequently, the temporal phase of the remote sensing images could moderately affect the predicted soil property results.

A comparison of the performance between high-spatial-resolution PlanetScope and medium-spatial-resolution Sentinel-2 images in predicting soil pH during the same temporal phase and modeling resolution revealed differences in accuracy. The prediction models based on PlanetScope images demonstrated higher overall prediction accuracy, greater robustness, and better predictive power compared to those based on Sentinel-2 images. However, there was no substantial increase in model accuracy with the high-resolution images. This was similar to the conclusions reached by Xia et al. [26] and Lu et al. [28] regarding soil pH prediction. In addition, the prediction accuracy for the three land use types showed a decreasing trend with different patterns as the modeling resolution decreased over the same time phases. It is worth highlighting that the change in prediction accuracy was relatively substantial in the orchards and dry land, but not in the paddy fields. The highest prediction accuracy was achieved at the original spatial resolution for all three land use types.

For the models based on PlanetScope images, the highest prediction accuracies were achieved for the orchards ( $R^2 = 0.72$ ,  $RPD = 1.91$ ) and paddy fields ( $R^2 = 0.66$ ,  $RPD = 1.71$ ) in the spring model A and for the dry land ( $R^2 = 0.77$ ,  $RPD = 2.09$ ) in the autumn model A. For the models based on Sentinel-2 images, the optimal prediction accuracies were achieved for the orchards ( $R^2 = 0.67$ ,  $RPD = 1.75$ ) and paddy fields ( $R^2 = 0.64$ ,  $RPD = 1.66$ ) in the spring model B and for dry land ( $R^2 = 0.70$ ,  $RPD = 1.83$ ) in the autumn model B.

Finally, we found that model performance differed with land use types at the optimal time phase. Both the Sentinel-2 and PlanetScope image-based prediction models achieved the highest prediction accuracy for the dry land (PlanetScope:  $R^2 = 0.77$ ; Sentinel-2:  $R^2 = 0.70$ ), higher than for the paddy fields and orchards. This was consistent with findings by Stenberg et al., who noted that model-prediction accuracy increases with higher Coefficients of Variation (in this study, the dry land exhibited a higher Coefficient of Variation compared to paddy fields and orchards) [75]. Additionally, modeling soil sample sites separately for the orchards, dry land, and paddy fields showed a substantial improvement in prediction accuracy compared to total area soil pH predictions, suggesting that a single land use approach may enhance the performance of soil pH predictions. This finding aligns with conclusions drawn by Samira et al. [76].

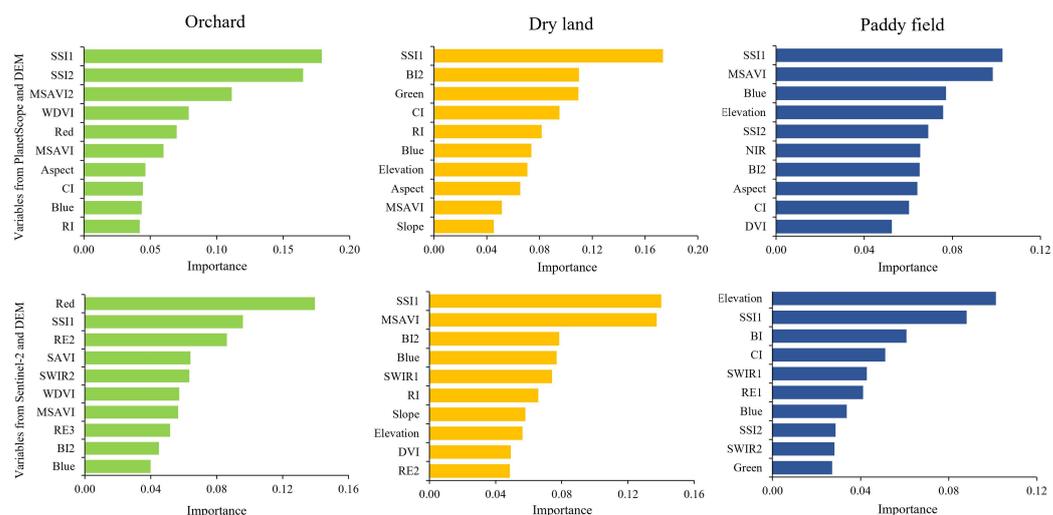


**Figure 5.** Prediction results of models with different combinations of inputs. Performance indicators include (a)  $R^2$ , (b) MAE, (c) RMSE, and (d) RPD. Notes: For PlanetScope, Models A, B, C, D, E, F, and G correspond to modeling resolutions of 3 m, 10 m, 20 m, 30 m, 40 m, 50 m, and 60 m, respectively. T1 uses spring imagery acquired on 17 April 2020; T2 uses summer imagery acquired on 3 August 2020; T3 uses autumn imagery acquired on 14 November 2020; T4 uses winter imagery acquired on 12 January 2021; T5 uses all available temporal phases. For Sentinel-2, Models B, C, D, E, F, and G correspond to modeling resolutions of 10 m, 20 m, 30 m, 40 m, 50 m, and 60 m, respectively. T1 uses spring imagery acquired on 28 April 2020; T2 uses summer imagery acquired on 26 August 2020; T3 uses autumn imagery acquired on 14 November 2020; T4 uses winter imagery acquired on 13 January 2021; T5 uses all available temporal phases.

### 3.3. Importance of Predictor Variables

For each satellite, a permutation method was applied to identify the significance of the top 10 spectral variables in the optimal XGboost model (Figure 6). In predicting soil pH with the two satellite images, the model utilized a blend of original spectral bands, derived

indices, and topographical variables as inputs. These factors showed differing degrees of significance throughout the prediction process. In the orchards, for the PlanetScope image, SSI1 (0.18), SSI2 (0.16), and MSAVI2 (0.11) were the three key variables; for the Sentinel-2 image, Red band (0.14), SSI1 (0.10), and Red Edge 2 band (0.09) were the three most influential variables. In the dry land, for the PlanetScope image, SSI1 (0.17) was a key variable; for the Sentinel-2 image, SSI1 (0.14) and MSAVI (0.14) were two important variables. In the paddy fields, for the PlanetScope image, SSI1 (0.10) and MSAVI (0.10) were two major variables; for the Sentinel-2 image, Elevation (0.10) and SSI1 (0.09) were the two principal variables. For all three land use types, it was evident that the Soil Salinity Index SSI1 was highly significant for predicting soil pH using both satellite images.

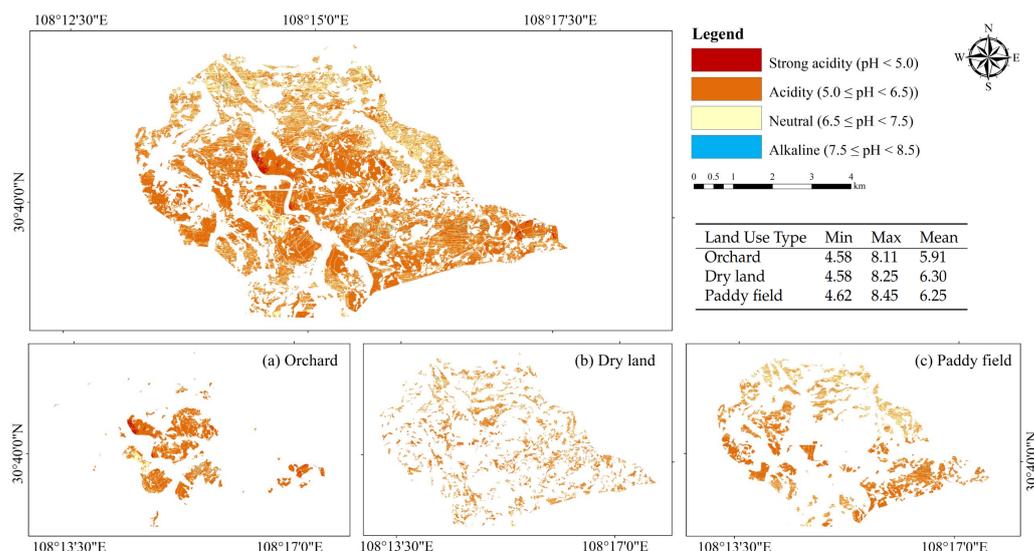


**Figure 6.** Variable importance rankings produced by extreme gradient boosting based on the most optimal model for each land use type (the abbreviations are listed in Table 2). Notes: For PlanetScope, the optimal model for the orchards and paddy fields was Model A, based on spring imagery acquired on 17 April 2020, while the optimal model for dry land was Model A, based on autumn imagery acquired on 14 November 2020; For Sentinel-2, the optimal model for the orchards and paddy fields was Model B, based on spring imagery acquired on 28 April, while the optimal model for dry land was Model B, based on autumn imagery acquired on 14 November 2020.

### 3.4. Mapping Soil pH in the Study Area

To obtain the best predictive results, we used the spring and autumn PlanetScope images with 3 m modeling resolution for mapping soil pH. Figure 7 shows the distribution of soil pH across three land use types. Based on the “Specification of Land Quality Geochemical Assessment (2016)”, soil pH was classified into five categories: strongly acidic ( $\text{pH} < 5.0$ ), acidic ( $5.0 \leq \text{pH} < 6.5$ ), neutral ( $6.5 \leq \text{pH} < 7.5$ ), alkaline ( $7.5 \leq \text{pH} < 8.5$ ), and strongly alkaline ( $\text{pH} \geq 8.5$ ). Additionally, descriptive statistical analyses of soil pH for the orchards, dry land, and paddy fields were conducted.

The prediction results indicate that no strongly alkaline areas existed within the study region; instead, the soils in the northern regions were mainly neutral-to-alkaline, while the other areas showed acidic or strongly acidic characteristics. For the orchards located primarily in the central and southern parts of the region, the average soil pH was 5.93, primarily due to the cultivation of blood oranges, which are well-suited to soils with a pH range of 5.5 to 7.0. Paddy fields and dry land were distributed around the periphery of the study area, with the soil pH characteristics showing an alkaline tendency in the northern regions, while acidic or strongly acidic conditions prevailed on the eastern and western sides.



**Figure 7.** Soil pH prediction map for cultivated land in the study area, generated from the best optimal model using PlanetScope data with 3 m spatial resolution (white color: non-cultivated land): (a) soil pH prediction map for the orchards; (b) soil pH prediction map for the dry land; (c) soil pH prediction map for the paddy fields.

## 4. Discussion

### 4.1. Effects of Spectral Image Acquisition Time

The effectiveness of the soil prediction models was significantly impacted by the choice of optical satellite image acquisition timing, as revealed through comparative analysis (Figure 5). For each satellite, we found that images from different temporal phases contributed differently to the prediction, depending on the land use types. Specifically, spring spectral data demonstrated desirable predictive performance for the orchards and the paddy fields, while autumn spectral data contributed more effectively to models for the dry land. The observed variability in the study area may have been strongly associated with factors such as crop residue cover, soil moisture, surface roughness, and cloud cover [43,77].

In this study, the spectral variables from spring and autumn were most effective for predicting the soil pH, which can be attributed to several factors. For the orchards, spring is characterized by a higher proportion of exposed soil, as it occurs after the pruning of spring buds but before the sprouting of summer buds. This period of increased soil exposure enhances the correlation between spectral variables and soil pH. For the paddy fields, spring coincides with the rice-sowing season, during which field preparations, such as plowing and leveling, reduce surface roughness and crop residues. This allows spectral signals to more directly reflect soil properties. For the dry land, autumn marks a critical growth stage for rapeseed crops, during which vegetation coverage increases significantly. Research has shown that satellite-derived spectral variables obtained during active vegetative growth stages tend to exhibit stronger relationships with soil properties [78]. These agricultural practices and management activities directly influence factors such as soil exposure, surface roughness, and vegetation coverage, making spectral variables in spring and autumn particularly effective for predicting soil pH. Studies from other regions further support the notion that “soil exposure” and “vegetation absorption dynamics” are key determinants of remote sensing signal accuracy. For instance, Bolorani et al., through their study of the Jajrud and Karaj river basins on the Iranian Plateau, demonstrated that the interaction between bare soil and other soil characteristics (such as soil type, geomorphology, and land use) significantly influences the interpretability of remote sensing data [79]. Particularly

during spring, when vegetation has not yet fully developed, the properties of bare soil provide clear information about the soil surface, thereby enhancing the correlation between remote sensing variables and soil attributes. Similarly, in paddy cultivation areas of Dehui City, Jilin Province, China, Li et al. observed a significant improvement in remote sensing signals following rice field tillage [80]. Additionally, limited rainfall during both spring and autumn results in relatively stable shallow-layer moisture conditions, thereby reducing water interference and providing reliable data for spectral analysis [81].

It is worth noting that single-temporal images of specific land use types achieve higher predictive accuracy. This may be due to the consistency of crop types and growth cycles within a single land use type, allowing images captured during specific periods to effectively highlight the most crucial information about these crops. Conversely, in larger regions with diverse crop types and varying growth conditions, multi-temporal satellite images are necessary, to comprehensively reveal all relevant information about crop growth [82]. Generally, the spectral characteristics of feature targets change over time. Since remote sensing images record instantaneous information, understanding the process and range of spectral changes through dynamic monitoring can help identify optimal times for target identification.

#### *4.2. Effects of Different Optical Satellites*

From an application standpoint, high-resolution satellite remote sensing can more clearly express the spatial structure characteristics and surface texture of the feature target and distinguish the finer composition inside the feature, which provides the conditions and basis for effective visual interpretation and has undoubted application prospects in agriculture, forestry, resource, and environment monitoring and management, etc. [83,84]. A key aim of this study was to investigate whether employing high-spectral-resolution satellite imagery could enhance the accuracy of soil pH predictions. The findings indicate that the prediction model of PlanetScope images has higher prediction accuracy compared to the popular Sentinel-2 images over the same temporal phases for the same land use type. This aligns with the research conducted by Xia et al. and Guo et al. [26,85]. The reason for this could possibly be that high-resolution remote sensing images serve as a viable data source for extracting fine land cover information, which enhances the prediction accuracy of soil pH, to some extent [83]. However, using satellite images with high spectral resolution did not make the soil pH predictions much more accurate. The reasons for this may be due to the different data characteristics of the two satellites. Although PlanetScope has a higher spatial resolution, it has a narrower spectral range and a smaller number of bands, while Sentinel-2 images have a relatively lower spatial resolution but a larger number of spectra. On the other hand, due to the hilly terrain and complex topography of the study area, the satellite images are affected by topographic factors and the validity of their data may have been reduced.

For each satellite, we also found that as the modeling resolution decreased, the prediction accuracy gradually decreased for the orchards and dry land, while it remained essentially unchanged for the paddy fields in the same time phase. The reasons for this may be related to topography and vegetation growth [86]. Even though the orchards were in the middle part of the study area at lower elevations, the natural slopes were more variable, and the dry land was in the higher areas around the study area with more topographic relief. These topographic factors made the topography and vegetation growth of the orchards and dry areas more variable in a small area, and it was difficult to capture such detailed information as the image elements became bigger. In the context of the paddy fields, the presence of water helped mitigate the errors in the spectral information of the optical data arising from the complex ground environment. Therefore, the variability in

prediction accuracy became less as the image elements became larger and the spectral information became less variable.

#### 4.3. Effects of a Single Land Use Type

The variability in soil pH can differ across various land use types, due to the influence of factors like vegetation cover and other environmental conditions [87]. These land use types significantly impact the accuracy of soil pH predictions, with noticeable differences in accuracy observed among dry land, orchard, and paddy field, where dry land shows higher prediction accuracy compared to the other two. The aforementioned statement aligns with the findings presented in the study conducted by Wang et al. [21]. Stenberg et al. have shown that a rise in the Coefficient of Variation corresponds to better model-prediction accuracy [75]. The study found that the Coefficient of Variation for pH in dry land soil was greater than that observed in orchard and paddy field soil. This disparity in variability may contribute to the superior predictive performance of the model in dry land areas. Additionally, errors in spectral information caused by complex crop types are avoided in the November dry season, when the ground is relatively flat, due to autumn harvesting, and most of the surface is bare.

It should be noted that the accuracy of soil pH prediction can be substantially enhanced by building prediction models under a single land use approach. Variations in soil and vegetation characteristics across different land use types might explain this occurrence. The accuracy of the prediction model is improved by its ability to limit the heterogeneity to a more precise regional scale through the use of sub-regions [88]. Given the limited scope of our research area, it is possible that these findings are specific to this particular region. To ensure more reliable findings, future studies may validate these observations across various land use categories and over a broader geographical extent.

#### 4.4. Effects of Predictor Variables

In this work, we used a variety of new proxies to estimate soil pH levels. These models were mostly composed of spectral variables and topographic variables, including the raw bands of two satellite images, different spectral indices derived from the raw bands, and elevation, slope, and aspect. The XGboost model combined soil pH predictions, to assess each proxy's contribution. The importance and contribution of various indices in soil pH prediction models differs among the three land use types. This variation arises mainly from the distinct ways in which these indices affect soil pH across varying environmental settings [89,90].

Each land use type exhibits unique environmental characteristics that determine the varying significance and contribution of each index in predicting soil pH. Among all land use types, Soil Salinity Index 1 (SSI1) consistently emerges as a key predictor, highlighting the strong correlation between soil salinity and pH. SSI1 reflects overall soil salinity levels, which directly influence the chemical, physical, and biological properties of soil—including cation exchange capacity and nutrient availability—thereby playing a crucial role in regulating soil pH [91,92]. This finding aligns with the study by Xia et al., which used GF-2 images to predict soil pH [26].

Specifically, in orchards SSI1 demonstrates significant predictive power, due to its critical role in modulating local chemical balance. Simultaneously, the red band may also contribute to pH prediction by indirectly capturing variations in vegetation cover and bare soil reflectance. The red band exhibits high spectral responsiveness, enabling it to detect differences between orchard vegetation health and surface physical characteristics, thereby providing supplementary insights into pH prediction.

For dry land, where evaporation rates are high and precipitation is limited, Brightness Index 2 (BI2) stands out as another important factor alongside SSI1. BI2 captures changes in surface brightness that are closely related to mineral weathering of exposed soils and organic matter content. In dry environments with limited biological activity, soil properties are more dominantly influenced by physical processes such as mineral weathering, salt accumulation, and organic matter decomposition. These processes directly govern the dynamic variations in soil acidity or alkalinity. Consequently, BI2 becomes vital for precisely assessing and predicting soil pH in arid regions by monitoring bare soil characteristics under such conditions.

In contrast, within paddy fields characterized by unique agricultural practices and prolonged waterlogging conditions, elevation emerges as another significant variable influencing pH beyond SSI1. Elevation directly determines localized water depth and drainage efficiency while further impacting oxygen availability and redox conditions [93,94]. Even minor elevation differences can lead to substantial variations in water depth, oxygen accessibility, and redox dynamics. For instance, higher elevations typically facilitate faster drainage under irrigation or rainfall scenarios. This promotes aerobic decomposition of organic matter under well-aerated conditions, to produce alkaline substances that raise pH levels. Conversely, lower elevations with poor drainage tend to remain waterlogged over extended periods; these anoxic (anaerobic) conditions enhance microbial fermentation or iron oxide reduction processes that increase acidity and lower pH values.

However, our study did not identify a statistically significant effect of NDVI on pH prediction, diverging from the findings reported by Qi et al. This discrepancy may be attributed to the use of imagery data from April and November, when vegetation coverage in the study area was relatively low [95]. Since NDVI primarily relies on vegetation spectral reflectance for its calculation, reduced vegetation coverage significantly diminishes the informational value related to plant–soil interactions. Consequently, its ability to capture indicators such as plant health status and canopy cover is directly limited. In summary, across all land use types, SSI1 emerges as a universal driving factor that links salinity with pH through its strong influence over fundamental soil chemical processes. Conversely, other indices—such as the red band in orchards, BI2 in dry lands, and elevation in paddy fields—reflect distinct environmental dynamics unique to their respective land use contexts.

#### *4.5. Application Potential and Limitations*

This study compared the performance of high-spatial-resolution (PlanetScope) and medium-spatial-resolution (Sentinel-2) remote sensing images in predicting soil pH across different temporal phases and land use types. The results highlight the potential of high-resolution satellite data for constructing soil pH prediction models in complex mountainous terrains. Additionally, this work provides valuable insights into the selection of remote sensing datasets, the timing of image acquisition, and the role of land use in spatial modeling of soil properties.

However, despite its significant contributions, this study had certain limitations regarding its applicability. The research was specifically focused on hilly and mountainous regions where the intricate topography, diverse vegetation cover, and varied land use patterns strongly influenced the soil pH variability. Consequently, the findings may only be applicable to areas with similar natural conditions and may not be directly extrapolated to other ecosystems, such as plains, wetlands, or arid regions. For example, in plain areas, soil heterogeneity is typically driven more by human activities than natural factors [96], whereas in wetlands hydrological conditions predominantly govern soil physicochemical properties, leading to mechanisms of pH variation that differ substantially from those observed in hilly regions [97,98]. Furthermore, because this study primarily relied on optical

remote sensing data from the PlanetScope and Sentinel-2 platforms, discrepancies between sensors—such as variations in spectral resolution, band configurations, and processing workflows—may have affected the result consistency and limited broader applicability.

To improve the generalizability of this approach, future efforts should focus on conducting cross-regional validation experiments covering a wider range of climatic zones, ecosystems, and diversified land use patterns. Incorporating additional data sources, such as radar imagery, hyperspectral data, or long-term time series monitoring could further enhance model robustness and support greater transferability across regions.

Despite these limitations, this study represents a significant step forward in exploring new methodologies for digital soil mapping using high-resolution remote sensing imagery. It addresses key knowledge gaps related to complex environmental conditions in mountainous areas while offering a foundation for future advancements aimed at expanding applicability through broader sampling frameworks and multi-scale validations.

## 5. Conclusions

This study conducted a comparison of the performance between high-spatial-resolution PlanetScope and medium-spatial-resolution Sentinel-2 for predicting soil pH under different temporal and land use types, and it selected the most suitable model for mapping soil pH in hilly mountainous areas. The main results are summarized below:

(1) The comparative analysis of the results suggests that by conducting a quantitative evaluation of satellite sensor selection, spectral image acquisition time, land use types, and modeling resolution, it is feasible to enhance the prediction accuracy of soil pH.

(2) Overall, the models built from PlanetScope images were more accurate than those built using Sentinel-2 images. The best model was obtained using PlanetScope images with a modeling resolution of 3 m. These findings indicate that high-spatial-resolution satellite images (PlanetScope) hold significant potential for improving the accuracy of digital soil pH mapping in mountainous and hilly regions.

(3) The effect of spectral image acquisition time on the accuracy of soil pH prediction differed by land use type. Spring images contributed more to the soil pH prediction for the orchards and paddy fields, and autumn images contributed more to the soil pH prediction for the dry land.

(4) The spatial heterogeneity of vegetation was limited to specific subregions, due to the presence of a single land use type. Constructing soil prediction models based on a single land use pattern may help to enhance the accuracy of predictions.

(5) The prediction map of the soil pH exhibited spatial heterogeneity. Overall, acidic soil was mainly concentrated in orchard areas with lower altitudes and steeper terrain.

(6) According to the feature importance ranking obtained from the XGboost algorithm, the Soil Salinity Index played a pivotal role in predicting the soil pH.

**Author Contributions:** Z.W.: methodology, software, conceptualization, writing—original draft preparation, validation, visualization, resources, formal analysis; W.W.: writing—review and editing, validation; H.L.: methodology, writing—review and editing, funding acquisition, supervision, resources, investigation. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The satellite imagery used in this study is publicly available and can be accessed through the specified website or repository. However, the specific sampling data and computational code used for analysis are proprietary and confidential, as they are part of ongoing research within Chongqing Key Laboratory of Land Quality Geological Survey. These resources are not publicly available, due to non-disclosure agreements. Requests for further information about

the data and code can be directed to the corresponding author, subject to the constraints of the confidentiality agreement.

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

## References

1. Tu, C.; He, T.; Lu, X.; Luo, Y.; Smith, P. Extent to which pH and topographic factors control soil organic carbon level in dry farming cropland soils of the mountainous region of Southwest China. *Catena* **2018**, *163*, 204–209. [[CrossRef](#)]
2. Lauber, C.L.; Hamady, M.; Knight, R.; Fierer, N. Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Appl. Environ. Microbiol.* **2009**, *75*, 5111–5120. [[CrossRef](#)] [[PubMed](#)]
3. Zhalnina, K.; Dias, R.; de Quadros, P.D.; Davis-Richardson, A.; Camargo, F.A.O.; Clark, I.M.; McGrath, S.P.; Hirsch, P.R.; Triplett, E.W. Soil pH Determines Microbial Diversity and Composition in the Park Grass Experiment. *Microb. Ecol.* **2015**, *69*, 395–406. [[CrossRef](#)]
4. Heggelund, L.R.; Diez-Ortiz, M.; Lofts, S.; Lahive, E.; Jurkschat, K.; Wojnarowicz, J.; Cedergreen, N.; Spurgeon, D.; Svendsen, C. Soil pH effects on the comparative toxicity of dissolved zinc, non-nano and nano ZnO to the earthworm *Eisenia fetida*. *Nanotoxicology* **2014**, *8*, 559–572. [[CrossRef](#)] [[PubMed](#)]
5. Minhoni, R.T.d.A.; Scudiero, E.; Zaccaria, D.; Saad, J.C.C. Multitemporal satellite imagery analysis for soil organic carbon assessment in an agricultural farm in southeastern Brazil. *Sci. Total Environ.* **2021**, *784*, 147216. [[CrossRef](#)]
6. Zeraatpisheh, M.; Ayoubi, S.; Jafari, A.; Finke, P. Comparing the efficiency of digital and conventional soil mapping to predict soil types in a semi-arid region in Iran. *Geomorphology* **2017**, *285*, 186–204. [[CrossRef](#)]
7. Tajik, S.; Ayoubi, S.; Shirani, H.; Zeraatpisheh, M. Digital mapping of soil invertebrates using environmental attributes in a deciduous forest ecosystem. *Geoderma* **2019**, *353*, 252–263. [[CrossRef](#)]
8. Zeraatpisheh, M.; Ayoubi, S.; Jafari, A.; Tajik, S.; Finke, P. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* **2019**, *338*, 445–452. [[CrossRef](#)]
9. Tajik, S.; Ayoubi, S.; Zeraatpisheh, M. Digital mapping of soil organic carbon using ensemble learning model in Mollisols of Hyrcanian forests, Northern Iran. *Geoderma Reg.* **2020**, *21*, e00256. [[CrossRef](#)]
10. Moradpour, S.; Entezari, M.; Ayoubi, S.; Karimi, A.; Naimi, S. Digital exploration of selected heavy metals using Random Forest and a set of environmental covariates at the watershed scale. *J. Hazard. Mater.* **2023**, *455*, 131609. [[CrossRef](#)]
11. Casa, R.; Castaldi, F.; Pascucci, S.; Palombo, A.; Pignatti, S. A comparison of sensor resolution and calibration strategies for soil texture estimation from hyperspectral remote sensing. *Geoderma* **2013**, *197–198*, 17–26. [[CrossRef](#)]
12. Chagas, C.d.S.; de Carvalho Junior, W.; Bhering, S.; Calderano Filho, B. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. *Catena* **2016**, *139*, 232–240. [[CrossRef](#)]
13. Gallo, B.; Demattê, J.; Rizzo, R.; Safanelli, J.; Mendes, W.; Lepsch, I.; Sato, M.; Romero, D.; Lacerda, M. Multi-Temporal Satellite Images on Topsoil Attribute Quantification and the Relationship with Soil Classes and Geology. *Remote Sens.* **2018**, *10*, 1571. [[CrossRef](#)]
14. Loiseau, T.; Chen, S.; Mulder, V.; Román Dobarco, M.; Richer-de Forges, A.; Lehmann, S.; Bourennane, H.; Saby, N.; Martin, M.; Vaudour, E.; et al. Satellite data integration for soil clay content modelling at a national scale. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101905. [[CrossRef](#)]
15. Zhai, Y.; Thomasson, J.; Boggess, J.; Sui, R. Soil texture classification with artificial neural networks operating on remote sensing data. *Comput. Electron. Agric.* **2006**, *54*, 53–68. [[CrossRef](#)]
16. Chen, S.; Liang, Z.; Webster, R.; Zhang, G.; Zhou, Y.; Teng, H.; Hu, B.; Arrouays, D.; Shi, Z. A high-resolution map of soil pH in China made by hybrid modelling of sparse soil data and environmental covariates and its implications for pollution. *Sci. Total Environ.* **2019**, *655*, 273–283. [[CrossRef](#)] [[PubMed](#)]
17. Ghazali, M.; Wikantika, K.; Harto, A.; Kondoh, A. Generating soil salinity, soil moisture, soil pH from satellite imagery and its analysis. *Inf. Process. Agric.* **2020**, *7*, 294–306. [[CrossRef](#)]
18. Jia, P.; Shang, T.; Zhang, J.; Sun, Y. Inversion of soil pH during the dry and wet seasons in the Yinbei region of Ningxia, China, based on multi-source remote sensing data. *Geoderma Reg.* **2021**, *25*, e00399. [[CrossRef](#)]
19. Tziachris, P.; Aschonitis, V.; Chatzistathis, T.; Papadopoulou, M.; Doukas, I.J.D. Comparing Machine Learning Models and Hybrid Geostatistical Methods Using Environmental and Soil Covariates for Soil pH Prediction. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 276. [[CrossRef](#)]
20. Emadi, M.; Taghizadeh-Mehrjardi, R.; Cherati, A.; Danesh, M.; Mosavi, A.; Scholten, T. Predicting and Mapping of Soil Organic Carbon Using Machine Learning Algorithms in Northern Iran. *Remote Sens.* **2020**, *12*, 2234. [[CrossRef](#)]

21. Wang, H.; Zhang, X.; Wu, W.; Liu, H. Prediction of Soil Organic Carbon under Different Land Use Types Using Sentinel-1/-2 Data in a Small Watershed. *Remote Sens.* **2021**, *13*, 1229. [CrossRef]
22. Zhang, X.; Xue, J.; Chen, S.; Wang, N.; Shi, Z.; Huang, Y.; Zhuo, Z. Digital mapping of soil organic carbon with machine learning in dryland of Northeast and North plain China. *Remote Sens.* **2022**, *14*, 2504. [CrossRef]
23. Ou, Y.; Rousseau, A.; Wang, L.; Yan, B. Spatio-temporal patterns of soil organic carbon and pH in relation to environmental factors—A case study of the Black Soil Region of Northeastern China. *Agric. Ecosyst. Environ.* **2017**, *245*, 22–31. [CrossRef]
24. Fu, Y.; Wan, L.; Xiao, D.; Le, B. Remote sensing inversion of saline and alkaline land based on reflectance spectroscopy and D-TELM algorithm in Wuyuan areas. *Infrared Phys. Technol.* **2020**, *109*, 103367. [CrossRef]
25. Hou, L.; Liu, Z.; Zhao, J.; Ma, P.; Xu, X. Comprehensive assessment of fertilization, spatial variability of soil chemical properties, and relationships among nutrients, apple yield and orchard age: A case study in Luochuan County, China. *Ecol. Indic.* **2021**, *122*, 107285. [CrossRef]
26. Xia, C.; Zhang, Y. Comparison of the use of Landsat 8, Sentinel-2, and Gaofen-2 images for mapping soil pH in Dehui, northeastern China. *Ecol. Inform.* **2022**, *70*, 101705. [CrossRef]
27. 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]
28. Lu, Q.; Tian, S.; Wei, L. Digital mapping of soil pH and carbonates at the European scale using environmental variables and machine learning. *Sci. Total Environ.* **2023**, *856*, 159171. [CrossRef] [PubMed]
29. Zhang, Y.; Shen, H.; Gao, Q.; Zhao, L. Estimating soil organic carbon and pH in Jilin Province using Landsat and ancillary data. *Soil Sci. Soc. Am. J.* **2020**, *84*, 556–567. [CrossRef]
30. Gholizadeh, A.; Žižala, D.; Saberioon, M.; Borůvka, L. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens. Environ.* **2018**, *218*, 89–103. [CrossRef]
31. Zhang, B.; Niu, Z.; Zhang, D.; Huo, X. Dynamic changes and driving forces of alpine wetlands on the Qinghai–Tibetan Plateau based on long-term time series satellite data: A case study in the Gansu Maqu wetlands. *Remote Sens.* **2022**, *14*, 4147. [CrossRef]
32. Zheng, G.; Chen, T.; Wang, Y.; Li, X.; Dai, W.; Xu, M.; Jiao, C.; Zhao, C. Rapid monitoring of the spatial distribution of soil organic matter using unmanned aerial vehicle imaging spectroscopy. *Ann. GIS* **2024**, *30*, 367–381. [CrossRef]
33. Wang, Z.; Wu, W.; Liu, H. Spatial Estimation of Soil Organic Carbon Content Utilizing PlanetScope, Sentinel-2, and Sentinel-1 Data. *Remote Sens.* **2024**, *16*, 3268. [CrossRef]
34. Webb, H.; Barnes, N.; Powell, S.; Jones, C. Does drone remote sensing accurately estimate soil pH in a spring wheat field in southwest Montana? *Precis. Agric.* **2021**, *22*, 1803–1815. [CrossRef]
35. Moon, M.; Richardson, A.; Friedl, M. Multiscale assessment of land surface phenology from harmonized Landsat 8 and Sentinel-2, PlanetScope, and PhenoCam imagery. *Remote Sens. Environ.* **2021**, *266*, 112716. [CrossRef]
36. Csillik, O.; Asner, G. Near-real time aboveground carbon emissions in Peru. *PLoS ONE* **2020**, *15*, e0241418. [CrossRef]
37. Cheng, Y.; Vrieling, A.; Fava, F.; Meroni, M.; Marshall, M.; Gachoki, S. Phenology of short vegetation cycles in a Kenyan rangeland from PlanetScope and Sentinel-2. *Remote Sens. Environ.* **2020**, *248*, 112004. [CrossRef]
38. Kimm, H.; Guan, K.; Jiang, C.; Peng, B.; Gentry, L.; Wilkin, S.; Wang, S.; Cai, Y.; Bernacchi, C.; Peng, J.; et al. Deriving high-spatiotemporal-resolution leaf area index for agroecosystems in the U.S. Corn Belt using Planet Labs CubeSat and STAIR fusion data. *Remote Sens. Environ.* **2020**, *239*, 111615. [CrossRef]
39. Qayyum, N.; Ghuffar, S.; Ahmad, H.; Yousaf, A.; Shahid, I. Glacial Lakes Mapping Using Multi Satellite PlanetScope Imagery and Deep Learning. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 560. [CrossRef]
40. Planet. Satellite Imagery and Archive, 2021. Available online: <https://planet.com/products/planet-imagery/> (accessed on 29 June 2024).
41. Breunig, F.; Galvão, L.; Dalagnol, R.; Dauve, C.; Parraga, A.; Santi, A.; Della Flora, D.; Chen, S. Delineation of management zones in agricultural fields using cover–crop biomass estimates from PlanetScope data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *85*, 102004. [CrossRef]
42. Wang, J.; Yang, D.; Detto, M.; Nelson, B.; Chen, M.; Guan, K.; Wu, S.; Yan, Z.; Wu, J. Multi-scale integration of satellite remote sensing improves characterization of dry-season green-up in an Amazon tropical evergreen forest. *Remote Sens. Environ.* **2020**, *246*, 111865. [CrossRef]
43. Zhang, Y.; Sui, B.; Shen, H.; Wang, Z. Estimating temporal changes in soil pH in the black soil region of Northeast China using remote sensing. *Comput. Electron. Agric.* **2018**, *154*, 204–212. [CrossRef]
44. Comte, I.; Colin, F.; Grünberger, O.; Follain, S.; Whalen, J.; Caliman, J.P. Landscape-scale assessment of soil response to long-term organic and mineral fertilizer application in an industrial oil palm plantation, Indonesia. *Agric. Ecosyst. Environ.* **2013**, *169*, 58–68. [CrossRef]
45. Li, S.; Yang, Y.; Li, Y.; Gao, B.; Tang, Y.; Xie, J.; Zhao, H. Remediation of saline-sodic soil using organic and inorganic amendments: Physical, chemical, and enzyme activity properties. *J. Soils Sediments* **2020**, *20*, 1454–1467. [CrossRef]

46. Bao, S. *Soil and Agricultural Chemistry Analysis*; China Agriculture Press: Beijing, China, 2000.
47. Chen, D.; Chang, N.; Xiao, J.; Zhou, Q.; Wu, W. Mapping dynamics of soil organic matter in croplands with MODIS data and machine learning algorithms. *Sci. Total Environ.* **2019**, *669*, 844–855. [[CrossRef](#)] [[PubMed](#)]
48. Lin, C.; Zhu, A.X.; Wang, Z.; Wang, X.; Ma, R. The refined spatiotemporal representation of soil organic matter based on remote images fusion of Sentinel-2 and Sentinel-3. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *89*, 102094. [[CrossRef](#)]
49. Swain, K.C.; Singha, C.; Pradhan, B. Estimating Total Rice Biomass and Crop Yield at Field Scale Using PlanetScope Imagery Through Hybrid Machine Learning Models. *Earth Syst. Environ.* **2024**, *8*, 1713–1731. [[CrossRef](#)]
50. Zhang, J.; Wu, X.; Shi, Y.; Jin, C.; Yang, Y.; Wei, X.; Mu, C.; Wang, J. A slight increase in soil pH benefits soil organic carbon and nitrogen storage in a semi-arid grassland. *Ecol. Indic.* **2021**, *130*, 108037. [[CrossRef](#)]
51. Abbas, A.; Khan, S. Using Remote Sensing Techniques for Appraisal of Irrigated Soil Salinity: International Congress on Modelling and Simulation (MODSIM). In Proceedings of the International Congress on Modelling and Simulation (MODSIM), Christchurch, New Zealand, 10–13 December 2007; pp. 2632–2638.
52. Yue, J.; Tian, J.; Tian, Q.; Xu, K.; Xu, N. Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. *ISPRS J. Photogramm. Remote Sens.* **2019**, *154*, 216–230. [[CrossRef](#)]
53. Omar, M.M.; Shitindi, M.; Massawe, B.; Fue, K.; Meliyo, J.; Pedersen, O. Salt-affected soils in Tanzanian agricultural lands: Type of soils and extent of the problem. *Sustain. Environ.* **2023**, *9*, 2205731. [[CrossRef](#)]
54. Escadafal, R.; Girard, M.C.; Courault, D. Munsell soil color and soil reflectance in the visible spectral bands of landsat MSS and TM data. *Remote Sens. Environ.* **1989**, *27*, 37–46. [[CrossRef](#)]
55. Pouget, M.; Madeira, J.; Le Floch, E.; Kamal, S. Caracteristiques Spectrales des Surfaces Sableuses de La Region Cotiere Nord-Ouest de L’Egypte: Application Aux Donnees Satellitaires SPOT. In Proceedings of the 2eme Journées de T&Detection: Caracterisation et Suivi des Milieux Terrestres En Regions Arides et Tropicales, Bondy, France, 6 December 1990; pp. 27–38.
56. Huete, A. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [[CrossRef](#)]
57. Qi, J.; Chehbouni, A.; Huete, A.; Kerr, Y.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [[CrossRef](#)]
58. Qi, J.; Kerr, Y.; Chehbouni, A. External factor consideration in vegetation index development. *Remote Sens. Environ.* **1994**, *723*, 730.
59. Moreau, L.; Li, Z. A new approach for remote sensing of canopy absorbed photosynthetically active radiation. II: Proportion of canopy absorption. *Remote Sens. Environ.* **1996**, *55*, 192–204. [[CrossRef](#)]
60. Gupta, R. Comparative study of AVHRR ratio vegetation index and normalized difference vegetation index in district level agricultural monitoring. *Int. J. Remote Sens.* **1993**, *14*, 53–73. [[CrossRef](#)]
61. Richardson, A.; Wiegand, C. Distinguishing vegetation from soil background information. *Photogramm. Eng. Remote Sens.* **1977**, *43*, 1541–1552.
62. Clevers, J. The derivation of a simplified reflectance model for the estimation of leaf area index. *Remote Sens. Environ.* **1988**, *25*, 53–69. [[CrossRef](#)]
63. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.; Gao, X.; Ferreira, L. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
64. Douaoui, A.; Nicolas, H.; Walter, C. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma* **2006**, *134*, 217–230. [[CrossRef](#)]
65. Liu, W.; Henneberry, S.; Ni, J.; Radmehr, R.; Wei, C. Socio-cultural roots of rural settlement dispersion in Sichuan Basin: The perspective of Chinese lineage. *Land Use Policy* **2019**, *88*, 104162. [[CrossRef](#)]
66. Zhou, J.; Li, E.; Wang, M.; Chen, X.; Shi, X.; Jiang, L. Feasibility of Stochastic Gradient Boosting Approach for Evaluating Seismic Liquefaction Potential Based on SPT and CPT Case Histories. *J. Perform. Constr. Facil.* **2019**, *33*, 04019024. [[CrossRef](#)]
67. Yu, X.; Wang, Y.; Wu, L.; Chen, G.; Wang, L.; Qin, H. Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting. *J. Hydrol.* **2020**, *582*, 124293. [[CrossRef](#)]
68. Fan, J.; Yue, W.; Wu, L.; Zhang, F.; Cai, H.; Wang, X.; Lu, X.; Xiang, Y. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agric. For. Meteorol.* **2018**, *263*, 225–241. [[CrossRef](#)]
69. Jia, Y.; Jin, S.; Savi, P.; Gao, Y.; Tang, J.; Chen, Y.; Li, W. GNSS-R Soil Moisture Retrieval Based on a XGboost Machine Learning Aided Method: Performance and Validation. *Remote Sens.* **2019**, *11*, 1655. [[CrossRef](#)]
70. Zheng, H.; Yuan, J.; Chen, L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* **2017**, *10*, 1168. [[CrossRef](#)]
71. Taghizadeh-Mehrjardi, R.; Schmidt, K.; Amirian-Chakan, A.; Rentschler, T.; Zeraatpisheh, M.; Sarmadian, F.; Valavi, R.; Davatgar, N.; Behrens, T.; Scholten, T. Improving the Spatial Prediction of Soil Organic Carbon Content in Two Contrasting Climatic Regions by Stacking Machine Learning Models and Rescanning Covariate Space. *Remote Sens.* **2020**, *12*, 1095. [[CrossRef](#)]

72. Wei, L.; Yuan, Z.; Wang, Z.; Zhao, L.; Zhang, Y.; Lu, X.; Cao, L. Hyperspectral Inversion of Soil Organic Matter Content Based on a Combined Spectral Index Model. *Sensors* **2020**, *20*, 2777. [[CrossRef](#)] [[PubMed](#)]
73. Bellon-Maurel, V.; Fernandez-Ahumada, E.; Palagos, B.; Roger, J.M.; McBratney, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends Anal. Chem.* **2010**, *29*, 1073–1081. [[CrossRef](#)]
74. Forkuor, G.; Hounkpatin, O.K.L.; Welp, G.; Thiel, M. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLoS ONE* **2017**, *12*, e0170478. [[CrossRef](#)]
75. Stenberg, B.; Viscarra Rossel, R.; Mouazen, A.; Wetterlind, J. Visible and Near Infrared Spectroscopy in Soil Science. *Adv. Agron.* **2010**, *107*, 163–215. [[CrossRef](#)]
76. Safaee, S.; Libohova, Z.; Kladvik, E.J.; Brown, A.; Winzeler, E.; Read, Q.; Rahmani, S.; Adhikari, K. Influence of sample size, model selection, and land use on prediction accuracy of soil properties. *Geoderma Reg.* **2024**, *36*, e00766. [[CrossRef](#)]
77. Vaudour, E.; Gomez, C.; Loiseau, T.; Baghdadi, N.; Loubet, B.; Arrouays, D.; Ali, L.; Lagacherie, P. The Impact of Acquisition Date on the Prediction Performance of Topsoil Organic Carbon from Sentinel-2 for Croplands. *Remote Sens.* **2019**, *11*, 2143. [[CrossRef](#)]
78. Yang, R.M.; Guo, W.W. Using time-series Sentinel-1 data for soil prediction on invaded coastal wetlands. *Environ. Monit. Assess.* **2019**, *191*, 462. [[CrossRef](#)] [[PubMed](#)]
79. Bolorani, A.D.; Bakhtiari, M.; Samany, N.N.; Papi, R.; Soleimani, M.; Mirzaei, S.; Bahrami, H.A. Land degradability mapping using remote sensing data and soil chemical properties. *Remote Sens. Appl. Soc. Environ.* **2023**, *32*, 101027. [[CrossRef](#)]
80. Li, J.; Lu, J.; Fu, H.; Zou, W.; Zhang, W.; Yu, W.; Feng, Y. Research on the Inversion of Key Growth Parameters of Rice Based on Multisource Remote Sensing Data and Deep Learning. *Agriculture* **2024**, *14*, 2326. [[CrossRef](#)]
81. Tekin, Y.; Tumsavas, Z.; Mouazen, A. Effect of Moisture Content on Prediction of Organic Carbon and pH Using Visible and Near-Infrared Spectroscopy. *Soil Sci. Soc. Am. J.* **2012**, *76*, 188–198. [[CrossRef](#)]
82. Shafizadeh-Moghadam, H.; Minaei, F.; Talebi-khiyavi, H.; Xu, T.; Homaei, M. Synergetic use of multi-temporal Sentinel-1, Sentinel-2, NDVI, and topographic factors for estimating soil organic carbon. *Catena* **2022**, *212*, 106077. [[CrossRef](#)]
83. Khanal, S.; Fulton, J.; Klopfenstein, A.; Douridas, N.; Shearer, S. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Comput. Electron. Agric.* **2018**, *153*, 213–225. [[CrossRef](#)]
84. Žižala, D.; Minařík, R.; Zádorová, T. Soil Organic Carbon Mapping Using Multispectral Remote Sensing Data: Prediction Ability of Data with Different Spatial and Spectral Resolutions. *Remote Sens.* **2019**, *11*, 2947. [[CrossRef](#)]
85. Guo, L.; Fu, P.; Shi, T.; Chen, Y.; Zhang, H.; Meng, R.; Wang, S. Mapping field-scale soil organic carbon with unmanned aircraft system-acquired time series multispectral images. *Soil Tillage Res.* **2020**, *196*, 104477. [[CrossRef](#)]
86. Kokulan, V.; Akinremi, O.; Moulin, A.; Kumaragamage, D. Importance of terrain attributes in relation to the spatial distribution of soil properties at the micro scale: A case study. *Can. J. Soil Sci.* **2018**, *98*, 292–305. [[CrossRef](#)]
87. Odhiambo, B.; Kenduiwo, B.; Were, K. Spatial prediction and mapping of soil pH across a tropical afro-montane landscape. *Appl. Geogr.* **2020**, *114*, 102129. [[CrossRef](#)]
88. Brungard, C.; Nauman, T.; Duniway, M.; Veblen, K.; Nehring, K.; White, D.; Salley, S.; Anchang, J. Regional ensemble modeling reduces uncertainty for digital soil mapping. *Geoderma* **2021**, *397*, 114998. [[CrossRef](#)]
89. Dedeoğlu, M.; Başayığıt, L.; Yüksel, M.; Kaya, F. Assessment of the vegetation indices on Sentinel-2A images for predicting the soil productivity potential in Bursa, Turkey. *Environ. Monit. Assess.* **2020**, *192*, 16. [[CrossRef](#)] [[PubMed](#)]
90. Abbaszad, P.; Asadzadeh, F.; Rezapour, S.; Khosravi, Aqdam, K.; Shabani, F. Evaluation of Landsat 8 and Sentinel-2 vegetation indices to predict soil organic carbon using machine learning models. *Model. Earth Syst. Environ.* **2023**, *10*, 2581–2592. [[CrossRef](#)]
91. Chamaki, S.; Taghvaeian, S.; Kisling, L.J. Effect of alternate irrigation with fresh and saline waters on soil profile salinity and chemical properties. In Proceedings of the World Environmental and Water Resources Congress 2020, Henderson, NV, USA, 17–21 May 2020. [[CrossRef](#)]
92. Abbas, J. Soil salinity assessment by using spectral salinity indices in Al-Sweira project middle of the Iraqi alluvial plain. *Int. J. Environ. Sci. Technol.* **2022**, *20*, 10847–10860. [[CrossRef](#)]
93. Sheng, H.; Yin, Z.; Zhou, P.; Thompson, M.L. Soil C:N:P ratio in subtropical paddy fields: Variation and correlation with environmental controls. *J. Soils Sediments* **2021**, *22*, 21–31. [[CrossRef](#)]
94. Suzuki, K.; Katashima, K.; Miki, T.; Igarashi, H.; Xu, Q.; Ohkubo, S.; Iwaishi, S.; Harada, N. Bacterial community composition under paddy conditions is more strongly affected by the difference in soil type than by field management. *Microb. Ecol. Int. J.* **2023**, *86*, 2552–2559. [[CrossRef](#)] [[PubMed](#)]
95. Qi, L.; Wang, S.; Zhuang, Q.; Yang, Z.; Bai, S.; Jin, X.; Lei, G. Spatial-Temporal Changes in Soil Organic Carbon and pH in the Liaoning Province of China: A Modeling Analysis Based on Observational Data. *Sustainability* **2019**, *11*, 3569. [[CrossRef](#)]
96. Nyengere, J.; Okamoto, Y.; Funakawa, S.; Shinjo, H. Analysis of spatial heterogeneity of soil physicochemical properties in Northern Malawi. *Geoderma Reg.* **2023**, *35*, e00733. [[CrossRef](#)]

97. Wang, S.; Wang, M.; Gao, X.; Zhao, W.; Miao, P.; Liu, Y.; Zhang, R.; Wang, X.; Sui, X.; Li, M.H. The Diversity and Composition of Soil Microbial Communities Differ in Three Land Use Types of the Sanjiang Plain, Northeastern China. *Microorganisms* **2024**, *12*, 780. [[CrossRef](#)]
98. Patel, K.F.; Fansler, S.J.; Campbell, T.P.; Bond-Lamberty, B.; Smith, A.P.; RoyChowdhury, T.; McCue, L.A.; Varga, T.; Bailey, V.L. Soil texture and environmental conditions influence the biogeochemical responses of soils to drought and flooding. *Commun. Earth Environ.* **2021**, *2*, 127. [[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.