

Article

Unveiling the Effects of Crop Rotation on Cropland Soil pH Mapping: A Remote Sensing-Based Soil Sample Grouping Strategy

Yuan Liu ^{1,2,3,4} , Songchao Chen ^{5,6} , Ge Shen ^{7,*},[†], Cheng Chen ⁴ , Zejiang Cai ³, Ji Zhu ^{1,2}, Xia Zhang ^{1,2}, Guofei Shang ^{1,2}, Qingbo Zhou ⁸, Sonoko Dorothea Bellingsrath-Kimura ^{4,9}, Qiangyi Yu ^{3,†} and Wenbin Wu ³

¹ School of Land Science and Space Planning, Hebei GEO University, Shijiazhuang 050031, China; liuyuan05@hgu.edu.cn (Y.L.); zhujj@hgu.edu.cn (J.Z.); zhangxia396@hgu.edu.cn (X.Z.); shangguofei@hgu.edu.cn (G.S.)

² Hebei International Joint Research Center for Remote Sensing of Agricultural Drought Monitoring, Hebei GEO University, Shijiazhuang 050031, China

³ State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China; caizejiang@caas.cn (Z.C.); yuqiangyi@caas.cn (Q.Y.); wuwenbin@caas.cn (W.W.)

⁴ Leibniz Centre for Agricultural Landscape Research (ZALF), 15374 Müncheberg, Germany; cheng.chen@zalf.de (C.C.); sonoko.bellingsrath-kimura@zalf.de (S.D.B.-K.)

⁵ Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 311200, China; chensongchao@zju.edu.cn

⁶ College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China

⁷ School of Public Administration, Zhejiang University of Finance and Economics, Hangzhou 310018, China

⁸ Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China; zhouqingbo@caas.cn

⁹ Faculty of Life Science, Humboldt University of Berlin, 14195 Berlin, Germany

* Correspondence: shenge@zufe.edu.cn

[†] These authors contributed equally to this work.



Academic Editors: Chong Luo, Xiaoling Wu, Liujun Zhu and Xiaoji Shen

Received: 21 March 2025

Revised: 27 April 2025

Accepted: 1 May 2025

Published: 6 May 2025

Citation: Liu, Y.; Chen, S.; Shen, G.; Chen, C.; Cai, Z.; Zhu, J.; Zhang, X.; Shang, G.; Zhou, Q.; Bellingsrath-Kimura, S.D.; et al. Unveiling the Effects of Crop Rotation on Cropland Soil pH Mapping: A Remote Sensing-Based Soil Sample Grouping Strategy. *Remote Sens.* **2025**, *17*, 1643. <https://doi.org/10.3390/rs17091643>

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/>).

Abstract: Crop rotation affects soil pH by disturbing H⁺ production and consumption within soil–crop systems, primarily through fertilization, irrigation, cropping, and harvest. Studies have shown that crop rotation improves soil organic matter prediction. However, simply incorporating crop rotation may not significantly improve soil pH prediction, because the spatial variability in soil pH is lower and the way crop rotation influences pH is different. To quantify the extent to which crop rotation improves soil pH mapping, we introduced the strategy of grouping soil samples by crop rotation and modeling separately. We chose a typical multiple-cropping region suffering soil acidification in Southern China, where the complex crop rotation was mapped by Sentinel-1/2 time series and a legend featuring three main systems (i.e., paddy, vegetable, and orchard) and nine subsystems. This crop rotation map was then combined with other variables to derive multiple combinations and predict soil pH. Based on the best combination, we further assessed the grouping strategy. The results showed that simply incorporating crop rotation in one joint model was useful but could not obtain the expected accuracy, with a root mean squared error (RMSE) of 0.66 and an R² of 0.36. The individual statistical accuracies were quite low for the vegetable and orchard rotations, with an RMSE of 0.77/0.70 and an R² of 0.30/−0.04. Grouping soil samples by crop rotation significantly enhanced soil pH predictability with a decrease in the RMSE of 15% and an increase in the R² of 53%. The results proved that grouping by crop rotation can fit and optimize the sub-models after learning the characteristics of the rotation subsamples, offering a way for improving digital mapping of soil pH over heterogeneous agricultural landscapes.

Keywords: soil pH; Sentinel-1/2 images; cropland soil; crop rotation; soil sample grouping; machine learning

1. Introduction

Soil pH, often referred to as the “master soil variable”, has an enormous influence on soil biogeochemical processes, thereby affecting nutrient cycling, crop growth and yield, and environmental health [1]. For example, soil biodiversity–function relationships and the mobility and bioavailability of toxic elements in soils and crops are all closely related to soil pH [2,3]. Characterized by a decrease in soil pH, soil acidification has been a major global issue [4–6]. Soils with a pH of less than 5.5 could cause stunted plant growth due to factors such as aluminum toxicity [7,8], while root systems can be damaged in highly acidic soils with a pH of less than 4.5, resulting in heavy crop yield losses [9]. Soil acidification is particularly severe in soils of intensive Chinese agricultural systems, reflected by an average pH decline between 0.13 and 0.80 units in major croplands between 1980s and 2000s [10,11]. Without sufficient mitigation measures, soil pH could decline by about 1 unit with relative yield losses increasing from 4% to 24% during 2010–2050 in China [12]. To support site-specific decision making for alleviating soil acidification and promoting environmental health, detailed and accurate spatial information on soil pH is urgently required. Cropland soil pH mapping is especially important for the sustainability of both food production and agricultural ecosystems.

Currently, digital soil mapping (DSM) is the most widely used method, which infers spatial soil information from soil observation and related environmental variables based on quantitative soil–landscape relationships [13,14]. Because cropland soils are actively managed by humans, human interventions must be considered in the modeling process [15]. For Chinese croplands, nitrogen (N) fertilizer use is a key influencing factor of soil pH and has led to severe soil acidification [11,12,16]. Directly mapping N fertilization is rather challenging due to the intricate nature of farmer activities [17]. Given that crops’ nutritional needs vary considerably and that crops may be under different fertilization regimes, maps on cropping systems somehow indicate the spatial and temporal variability of N fertilization [18]. It was shown that Chinese vegetable production involves more intensive N fertilization than cereal crops [19]. Orchards have also been severely overfertilized in China, and farmland-to-orchard conversion causes soil pH decline [20]. Cropping systems can also reflect other management practices related to soil pH change, such as irrigation which leads to temporally flooded fields and a larger acid buffering capacity [21]. The conversion from upland to paddy systems modifies N-cycling microbes and increases soil pH [22]. Moreover, soil pH is strongly affected by cropping and harvesting which removes base cations by crop growth and residue removal [11,23]. Maize–wheat rotations have been shown to have higher acidification rates than rice–wheat and rice–fallow, with crop harvesting being the main driver [24]. These imply that spatial data on cropping systems could provide information on fertilization, cropping frequency, and other management practices relevant to the spatial heterogeneity of soil pH.

The great advancement in the satellite-based mapping of crop rotation offers a unique opportunity for integrating cropping systems and DSM [25–29]. For example, crop rotation in the northern Hunan Province was mapped by a hybrid deep learning model that synergized Sentinel-1 and Sentinel-2 time series [28]. As a result, cropping system maps are increasingly used as inputs for predicting soil properties, especially in agricultural regions with intense human activity and homogenous climate and parent material [30]. For example, incorporating crop rotation in soil organic matter (SOM) prediction led to a decrease

in the RMSE of 7% and an increase in the R^2 of 24% [31]. However, simply incorporating cropping systems may not bring significant improvements in soil pH prediction, because the way crop rotation influences pH is different, and soil pH has low levels of spatial variability that make its prediction more challenging than SOM [32]. Grouping soil samples by cropping systems and modeling them separately may achieve better soil pH predictions. The grouping strategy builds an individual prediction model for each sample group such that each model (sub-model) is customized and optimal, creating conditions under which the explanatory variables work more reasonably [33]. In contrast, joint modeling for all soil samples may ignore the specificity of different groups and limit soil pH predictability. The fact that soil pH and its influencing factors vary greatly among different cropping systems indicates the rationality and potential of grouping by cropping systems for soil pH mapping. An analysis based on 13 long-term experiments in Southern China revealed that natural bicarbonate leaching was the dominant driver of soil pH decline in paddy soils while it was N fertilizer application in upland soils [34]. Recently, the validity of separately modeling SOM for drylands and paddy fields in Northern China was proven [35]. In the case of a more heterogeneous environment, the grouping strategy appears more necessary. However, in the literature, the performance of separately modeling soil pH for cropping systems has not yet been investigated. Therefore, to what extent the incorporation of cropping systems could contribute to improving soil pH mapping remains unknown.

Therefore, cropping systems (e.g., crop rotation) affect soil pH considerably in both direct (e.g., removal of base cations by harvest) and indirect (e.g., fertilization, irrigation, tillage) ways, yet the digital mapping of soil pH has rarely considered such factors. Furthermore, the extent to which cropping systems affect soil pH mapping accuracy remains unclear, leaving important research gaps. In this research, we introduced a soil sample grouping strategy and explored the effects of incorporating crop rotation in soil pH mapping. We hypothesized that grouping soil samples by crop rotation could build optimized sub-models for different rotations and achieve a higher accuracy than simply incorporating crop rotation. To test this hypothesis, we selected a typical acidic soil region in Southern China where the agricultural landscape is highly heterogeneous due to intensive and diversified cropping systems.

Specifically, the objectives of this study were (1) to explore the effectiveness of simply incorporating crop rotation in mapping soil pH and (2) verify whether and to what extent the soil sample grouping strategy and separate modeling of different crop rotations could further improve the digital mapping of soil pH.

2. Materials and Methods

2.1. Study Region

We focused on a heterogeneous agricultural region in Southern China with quite low soil pH (Figure 1). Soil pH in the Zengcheng District is low mainly due to soil type and climate, which deteriorates due to long-term intensive cropping with excessive chemical fertilization [36]. The soil parent materials consist predominately of granite-weathered materials, and the soil types of agricultural land are mainly red, lateritic red, and paddy soils [31]. Zengcheng is a typical multi-cropping region, where the subtropical monsoon climate enables crops to be grown all year round. Vegetables and fruits occupy more than 90% of agricultural land, while the rest is mainly covered by paddy rice [27]. The monitoring of cultivated land's quality in the Guangdong Province shows that between 1984–1986 and 2020–2022 the average pH decreased from 5.74 to 5.61, with a decrease of 0.13 units.

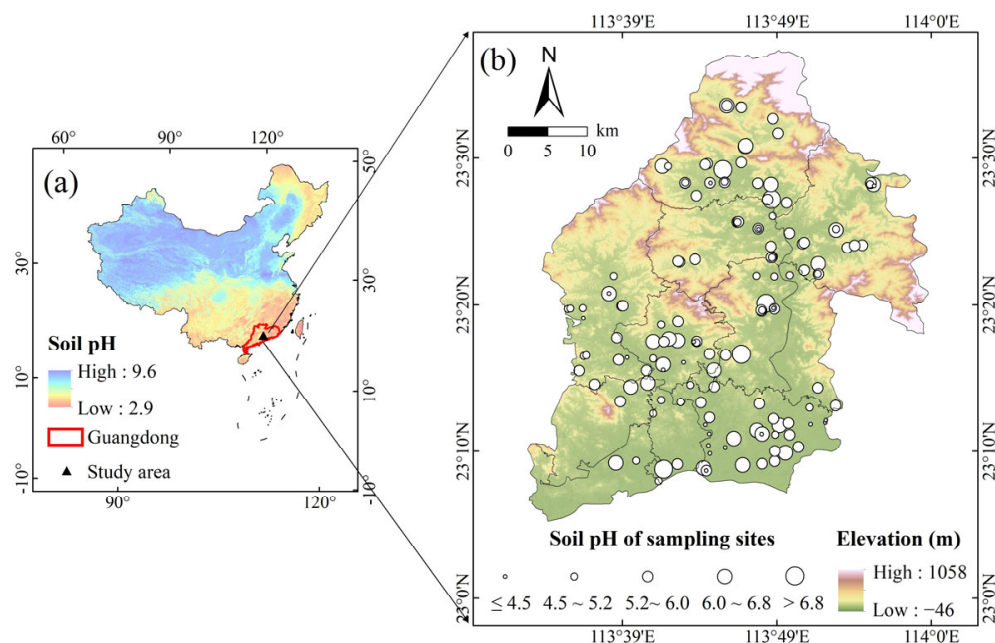


Figure 1. The study region and soil samples: (a) location of the Guangdong Province and Zengcheng District in China, overlaid with the soil pH map from [37]; (b) elevation, distribution of soil samples in Zengcheng, and measured pH values.

2.2. Data Collection

2.2.1. Soil Samples

To train and validate the soil pH prediction models, we collected a total of 150 topsoil samples (0–20 cm) across the study area in October 2021 by the grid sampling scheme (Figure 1b). The measured soil pH varied from 3.62 to 9.44, with a mean of 5.46 and a standard deviation of 0.83. The samples were collected according to the following procedure: First, we created a 2×2 km fishnet to cover the entire study area. Each grid was regarded as a sampling unit. Second, we adopted a land-cover product, i.e., the global land-cover product with a fine classification system at 30 m (GLC_FCS30), in 2020 to screen the grids outside the cropland [38,39]. Third, we randomly took ten soil cores from each unit using a 5 cm diameter auger and mixed the soils thoroughly for one composite sample. We recorded their geographic locations with a global positioning system. Fourth, we air-dried and crushed the collected soil samples, filtered them with a 1.0 mm sieve, and stored the preprocessed soils in sealed glass jars for further analysis. Finally, we measured the soil pH (in water, 1:2) using a pH detector.

2.2.2. Environmental Variables

According to the soil spatial prediction function with spatially autocorrelated errors, the reference (Scorpan-SSPFe), and the characteristics of the soil pH, we selected a suite of factors to represent soil-forming environments. These factors were closely related to organisms, soil, climate, and relief. The factor of “organisms, vegetation, fauna or human activities” consisted of Sentinel-1’s radar backscatter and Sentinel-2’s spectral bands and indices, cropland type, and crop rotation systems. All variables in raster format were resampled to 10 m using the bilinear method and were used as the candidate independent variables in predicting the soil pH (Table 1).

Table 1. The collected environmental variables used to predict the soil pH.

Environmental Variables		Year	Resolution	Source
Organisms, vegetation, fauna, or human activities	Vertical–horizontal (VH) and vertical–vertical (VV) polarization backscattering coefficients, cross-polarization ratio (VH/VV, CR)	May 2015–December 2021	10 m	Bimonthly median composites of Sentinel-1
	Spectral bands (2–8, 8A, 11, and 12), normalized difference vegetation index (NDVI), soil-adjusted vegetation index (SAVI), land surface water index (LSWI), and three soil salinity indices (SSI1, SSI2, and SSI3)	July 2017–March 2022	10 m	Seasonal median composites of Sentinel-2
	Cropland type	1985–2021	30 m	[39]
	Crop rotation systems	2020–2021	10 m	[27]
Soil	Soil organic carbon (SOC), pH, texture (sand, silt, clay), bulk density, and thickness	—	1 km	[40]
Climate	Mean annual temperature (MAT) and mean annual precipitation (MAP)	2021	1 km	[41]
	Mean annual daytime and nighttime land surface temperature (LST)	2021	1 km	MOD11A2
	Minimum, mean and maximum temperature, precipitation	1950–2000	~1 km	WorldClim Climatology V1
Relief	Elevation, aspect, slope, topographic wetness index (TWI), plan curvature (PLC), profile curvature (PRC), slope-length factor (LS), valley depth (VD), convergence index (CI), channel network base level (CNBL), catchment network distance (CND), relative slope position (RSP), and vertical distance to channel network (VDCN)	-	30 m	DEM from SRTM

2.2.3. Crop Rotation

We conducted multiple field investigations in Zengcheng and learned that the crop rotations are extremely complex due to fragmented cropland, multi-cropping, and diverse crops. There are many kinds of vegetables that could be harvested three or more times per year, and rice could be harvested twice a year (double paddy rice). Usually, vegetables are rotated with rice in a year to balance economic output and grain production, e.g., early rice–late rice–winter vegetable and spring vegetable–late rice.

To represent the complex crop rotations, we first provided a classification scheme based on the field investigation, farmer interviews, expert knowledge, and agricultural statistics. The classification included three main systems, i.e., paddy, vegetable, and orchard systems, which were divided into nine subsystems with distinct seasonal dynamics. The paddy system had single and double rice, which were further distinguished by the presence of vegetables. The vegetable system was classified into low- and high-intensity (includes low- and high-diversity) vegetables. Finally, the orchard system was classified into short- (e.g., banana, papaya) and long-term (e.g., litchi, longan) orchards. Based on the analyses of their spectral and temporal signatures, we then proposed a hierarchical rule-based algorithm to identify the crop rotation systems in 2020 and 2021 (Figure 2). We initially obtained four key indicators: flooding frequency, cropping intensity, cropping diversity, and coefficient of variation. Then, we mapped the three main crop rotation systems according to Sentinel-2 images and the random forest (RF) algorithm. Finally, the nine crop rotation subsystems were identified by the four indicators. For example, to identify the low-diversity and high-diversity vegetables, we designed and calculated an indicator called “cropping diversity”.

It was defined as crop diversity over time and could be measured by quantifying differences among crops with phenological metrics. More details on crop rotation mapping can be found in [27].

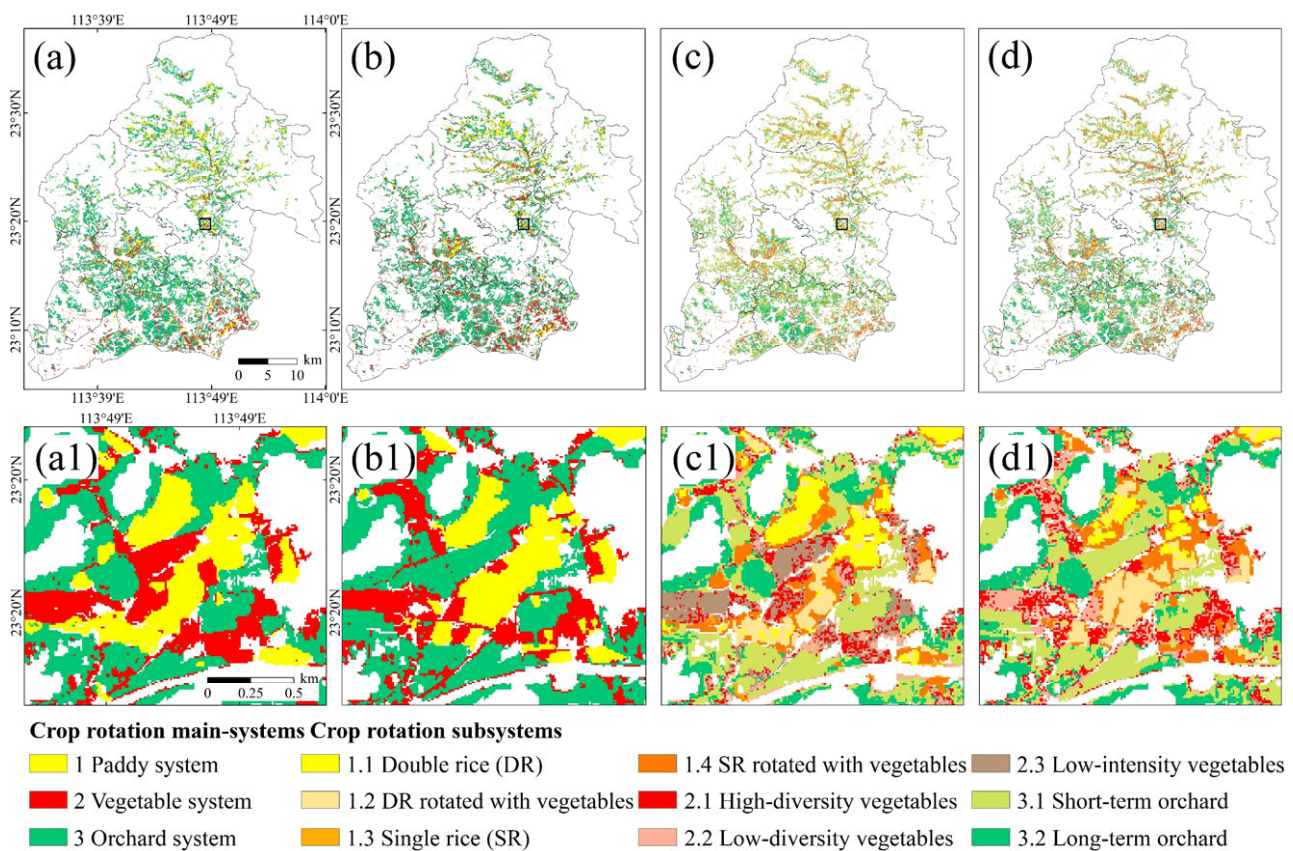


Figure 2. Crop rotation main systems in 2020 (a) and 2021 (b) and the subsystems in 2020 (c) and 2021 (d). Figures (a1–d1) correspond to the zoomed-in maps for the black squares in (a–d).

2.2.4. Sentinel Images

To characterize the factor “o”—organisms—we also used long-term Sentinel-1/2 time series and compared their effectiveness with crop rotation. Sentinel-2 carries a multispectral imager (MSI) and has been increasingly used in DSM. Besides optical images, radar data such as synthetic aperture radar (SAR) sensitive to soil moisture and ground-surface conditions may also be useful for soil pH mapping, yet it remains unexplored [42,43]. In this study, we took advantage of the full collection of Sentinel-1/2 since the satellite constellations consistently provided dense time series. The time frames for accessing Sentinel-1 and Sentinel-2 were May 2015–December 2021 and July 2017–February 2022, respectively, and they finally produced 40 bimonthly radar and 14 seasonal optical composites. The Sentinel-2 images were seasonally composited because they were frequently blocked by clouds.

For Sentinel-1, the Level-1 ground range-detected (GRD) product was used. Besides the vertical–horizontal (VH) and vertical–vertical (VV) polarization backscattering coefficients, we also computed the cross-ratio index (VH/VV), which was more sensitive to the dynamics of land surface and the phenological development of crops [44]. For Sentinel-2, the Level-2A surface reflectance product was used. Multiple vegetation and water indices associated with soil pH were calculated, including the normalized difference vegetation index (NDVI), the soil-adjusted vegetation index (SAVI), and the land surface water index (LSWI) [45]. We also incorporated three soil salinity indices (SSI1, SSI2, and SSI3), considering that they can reflect the soil salinity which is closely related to the soil pH [46]. The

Google Earth Engine (GEE) was exploited to access and process the Sentinel images. The calculation formulas can be found in Table 2.

Table 2. Spectral indices calculated from seasonal composites of Sentinel-2.

Spectral Index	Expression	Properties	References
Normalized difference vegetation index (NDVI)	$(\text{NIR} - \text{R})/(\text{NIR} + \text{R})$	Vegetation	[47]
Soil-adjusted vegetation index (SAVI)	$(\text{NIR} - \text{R}) \times (1 + 0.5)/(\text{NIR} + \text{R} + 0.5)$	Vegetation	[48]
Land surface water index (LSWI)	$\text{NIR} - \text{SWIR1}/\text{NIR} + \text{SWIR1}$	Leaf water and soil moisture	[49]
Soil salinity index1 (SSI1)	$(\text{B} \times \text{R})/\text{G}$	Soil salinity	[46]
Soil salinity index2 (SSI2)	$(\text{G} + \text{R})/2$	Soil salinity	[46]
Soil salinity index3 (SSI3)	$(\text{B} \times \text{R})^{1/2}$	Soil salinity	[50]

Note: B, G, R, NIR, and SWIR1 are the surface reflectance of bands 2, 3, 4, 8, and 11 in the Sentinel-2 multispectral instrument (MSI) sensor.

2.2.5. Other Variables

The following datasets were also used as environmental covariates: (1) The cropland layers were extracted from GLC_FCS30 for the period from 1985 to 2021, which were also used to represent organisms [39]. We selected this product because of the fine spatial details, high spatial resolution, and long time span (1985–2022) that captured historical changes in cropland. The maps were updated every 5 years before 2000 and annually after 2000, and the category “cropland” comprised rainfed and irrigated cropland. The rainfed cropland included herbaceous cover and tree or shrub cover (orchard). (2) The high-resolution National Soil Information Grids of China (NSIGC) was used to indicate soil information [40]. It was produced by advanced ensemble machine learning and 5000~ representative soil profiles, leading to more detailed and accurate results.

(3) The topographical variables were calculated using the terrain analysis tool of QGIS based on the Shuttle Radar Topographic Mission (SRTM) Digital Elevation product. (4) The climatic variables included the mean annual temperature (MAT) and mean annual precipitation (MAP) in 2021 from the National Tibetan Plateau Data Center available at <https://data.tpdc.ac.cn/zh-hans/data/71ab4677-b66c-4fd1-a004-b2a541c4d5bf> (accessed on 20 May 2024) [41]. In addition, we used the mean annual daytime and nighttime land surface temperature (LST) in 2021 from the MOD11A2 V6 product and the historical climate patterns supplied by WorldClim Climatology version 1 [51].

2.3. Soil pH Mapping Framework

To explore the effectiveness of incorporating crop rotation in mapping soil pH and compare it with cropland type and Sentinel images, we designed seven scenarios with different combinations of environmental variables. Each scenario was used to predict soil pH combining the soil samples and the random forest (RF) model. Based on the scenario with the highest accuracy, we then verified whether and to what extent grouping soil samples by crop rotation and separate modeling could further improve the digital mapping of soil pH. The flowchart of the method of this study is illustrated in Figure 3.

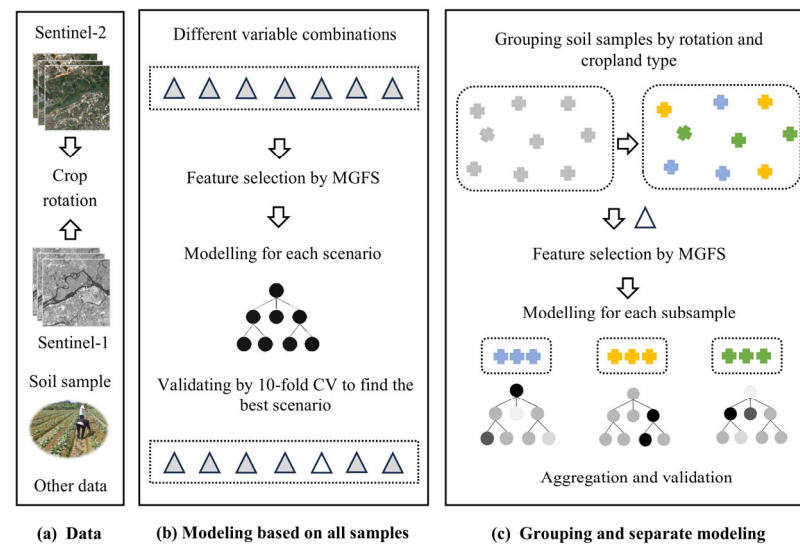


Figure 3. Flowchart of the soil pH prediction in this study. MGFS represents the modified greedy feature selection algorithm (Section 2.3.1). The color of soil samples is used to refer to different groups.

2.3.1. Variable Scenario and Variable Selection

We designed seven scenarios: scenarios 1 and 2 only included the temporal composites of Sentinel-1 and Sentinel-2, respectively, to compare their behavior in soil pH prediction, whereas scenario 3 combined radar and optical satellites (Figure 3, Table 3). Scenario 4 added the soil, climatic, and topographic variables to the previous variable pool. Instead of using natural variables, scenarios 5 and 6 introduced the cropland type and crop rotation systems, respectively. For scenario 7, all the candidate variables were fed as inputs to the modeling process.

Table 3. Different combinations of environmental variables.

Combinations of Environmental Variables		Number of Variables
1	Bimonthly median composites of Sentinel-1	120 (3 bands $\times$ 40 intervals)
2	Seasonal median composites of Sentinel-2	224 (16 bands and indices $\times$ 14 intervals)
3	Median composites of Sentinel-1/2	344 (120 + 224)
4	Median composites of Sentinel-1/2, soil, climatic, and topographic variables	372 (344 + 28)
5	Median composites of Sentinel-1/2 and cropland type	369 (344 + 25 periods)
6	Median composites of Sentinel-1/2 and crop rotation systems	348 (344 + 4)
7	Median composites of Sentinel-1/2, soil, climatic, topographic variables, cropland type, and crop rotation systems	401 (344 + 28 + 25 + 4)

Since the scenarios had a different number of variables and the satellite data dramatically increased the variables' numbers, variable selection was necessary to deal with redundancy and collinearity and improve model accuracy. Here, we selected the modified greedy feature selection (MGFS) algorithm, owing to its advantages in model parsimony and accuracy over Boruta, recursive feature elimination (RFE), and variance inflation factor analysis [52]. Opposite to RFE, MGFS adopts a forward selection strategy including five steps: (1) fit a model using all the variables and calculate the variables' importance; (2) select the most important variable (only one) to fit an initial model and calculate its performance by k-fold cross-validation; (3) fit a list of models using two variables (the combinations of variable(s) in the pool and one of the remaining variables), calculate their

model performances, and record the model with the best performance; (4) update the pool by taking the variables from the best model in the last step; (5) repeat steps 3 and 4 by increasing the number of variables from three to n . The variables in the model with the best performance are selected for the final model. To improve computational efficiency, the “early stop” function was activated when the model performance started to decrease. More details on MGFS can be found in [53].

2.3.2. Modeling for Each Scenario

We leveraged the RF regression algorithm to predict soil pH employing the R packages of “randomForest”, “raster”, and “caret” by the open-source statistical language R v4.4.2 (R Studio, Boston, MA, USA). RF has been widely used in DSM and usually provides a high accuracy. It is easy to execute and is also insensitive to noise and overtraining. RF is an ensemble learning algorithm by combining multiple independent decision trees [54]. RF first randomly selects bootstrap samples of observations, and a decision tree is fitted to each subsample. Then, the decision tree randomly selects explanatory variables with $mtry$ features ($mtry$: number of variables randomly sampled as candidates at each split). All splits of the tree are examined with explanatory variables, and the best split at each step is determined to build the decision tree. The final modeled value is the average across all decision trees.

For each scenario, variable selection was first completed by MGFS based on the whole soil sample sets to improve computing efficiency and ensure consistency across sample folds. Then, based on selected variables, 10-fold cross-validation was performed to evaluate soil pH prediction performance. At each fold, the number of variables randomly sampled as candidates at each split ($mtry$) was fine-tuned by grid-search and cross-validation. The number of trees ($ntree$) was held constant at the default value of 500. Our experiment indicated that, when the number of trees increased to 500, the accuracy tended to stabilize. The variables’ relative importance was assessed by the increase in the mean square error (%IncMSE) as a result of randomly shuffling variable values. Finally, the soil pH map was produced by the RF model fitted by all the soil samples under each scenario. To feed categorical variables such as crop rotation to the models, the one for hot encoding was used by converting each categorical value into a new categorical column and assigning a binary value of 1 or 0 to those columns.

2.3.3. Grouping by Cropland and Crop Rotation

Grouping soil samples appropriately and separate modeling hold the potential to improve the DSM of soil pH. Two different strategies were adopted and compared, i.e., grouping by cropland type and crop rotation systems. Specifically, the cropland layer in 2021 led to rainfed and irrigated subsamples, while the map of crop rotation systems in 2021 (Figure 2b) led to paddy, vegetable, and orchard subsamples. To ensure sufficient samples for the subsample, here we only used the main crop rotation systems. For each subsample, an RF regression model was fitted based on the variable scenario with the highest accuracy. Finally, the cropland and rotation strategy built two and three sub-models, respectively, and the soil pH predictions from multiple sub-models were summed to calculate the prediction accuracy for each strategy.

2.3.4. Model Validation

The 10-fold cross-validation was performed to evaluate the performances of the models. This procedure resulted in 135 training and 15 validation samples at each iteration. For the soil sample grouping strategy, the resulting validation points from all sub-models were combined to determine the overall accuracy. Three indices were calculated for quantifying model accuracy: mean error (ME), root mean squared error (RMSE), and coefficient of

determination (R^2). The soil pH maps were produced by the RF model fitted by all the soil samples for each scenario.

$$ME = \sum_{i=1}^n \frac{p_i - o_i}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (o_i - p_i)^2}{n}} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n (o_i - \bar{o})^2} \quad (3)$$

3. Results

3.1. Modeling for Each Scenario

Based on the 10-fold cross-validation method, we validated the soil pH prediction models for all scenarios (Table 4, Figure 4). The predictive results of scenarios 1 and 2, using radar or optical satellite composites alone, were quite similar, while scenario 3, using both sensors, evidently enhanced the predictability of soil pH, with an RMSE of 0.69 and an R^2 of 0.29. As shown by the results of scenarios 4 and 5, combining Sentinel data with the natural variables (soil, climate, and terrain) or cropland data did not bring further improvement but instead reduced the prediction accuracy. However, combining Sentinel data with the maps of crop rotation systems significantly improved the prediction accuracy, with a decrease in the RMSE of 4% and an increase in the R^2 of 24% compared to scenario 3, which combined Sentinel-1 and Sentinel-2 composites. As for scenario 7, which incorporated all the environmental variables in soil pH prediction, a similar accuracy was observed to that of scenario 3.

Table 4. Soil pH mapping accuracies for different scenarios.

Scenarios	Number of Selected Variables	ME	RMSE	R^2
1	9	0.00	0.72	0.25
2	7	0.01	0.72	0.24
3	11	0.01	0.69	0.29
4	3	−0.01	0.72	0.23
5	3	−0.01	0.72	0.23
6	9	0.02	0.66	0.36
7	8	0.02	0.70	0.29

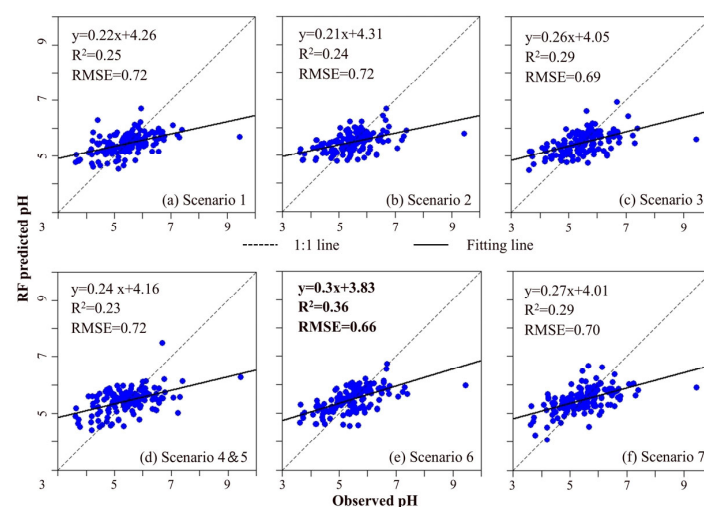


Figure 4. Scatter plots of predicted versus observed soil pH from the 10-fold cross-validation for scenarios 1 (a), 2 (b), 3 (c), 4 and 5 (d), 6 (e), and 7 (f). The dashed line represents the 1:1 line, and the solid line represents the fitted line.

3.2. Grouping by Cropland and Crop Rotation

Using the environmental variables of scenario 6 which had the highest prediction accuracy, we fitted the soil pH prediction models for different grouping strategies, i.e., grouping by cropland type and crop rotation systems. The models were also validated by the 10-fold cross-validation method, and the results are shown in Table 5 and Figure 5. The two different strategies led to quite distinct predictions. The strategy based on cropland type obtained similar results to scenario 6, with an RMSE of 0.67 and an R^2 of 0.35. On the other hand, grouping by crop rotation systems further enhanced the predictability of soil pH compared to scenario 6, with a decrease in the RMSE of 15% and an increase in the R^2 of 53%. The individual statistical accuracy of the sub-models fitted by subsamples also varied markedly. For cropland type, the irrigated subsample had a higher accuracy, with an RMSE of 0.65 and an R^2 of 0.41, while the rainfed one only had an R^2 of 0.28. For the crop rotation systems, the orchard subsample performed the best out of the three systems, with an RMSE of 0.46 and an R^2 of 0.55, followed by the vegetable subsample, with an RMSE of 0.68 and an R^2 of 0.45. Despite having the lowest R^2 , the paddy subsample performed the best regarding the RMSE (0.44).

Table 5. Soil pH prediction accuracies for different soil sample grouping strategies.

Subsample Determination		Number of Selected Variables	ME	RMSE	R^2
Cropland type	Rainfed	13	0.01	0.68	0.28
	Irrigated	5	0.00	0.65	0.41
	Total	-	0.00	0.67	0.35
Crop rotation systems	Paddy	4	−0.05	0.44	0.25
	Vegetable	8	0.00	0.68	0.45
	Orchard	5	0.02	0.46	0.55
	Total	-	−0.01	0.56	0.55

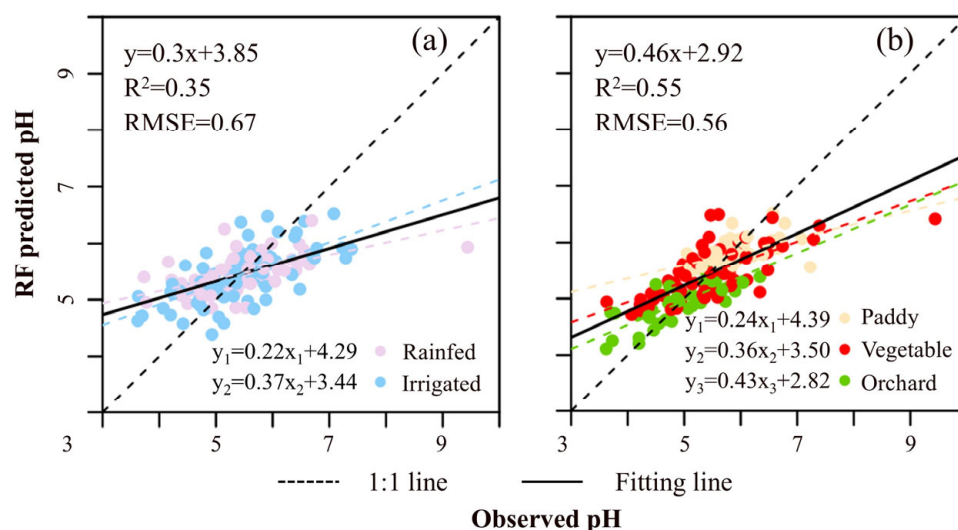


Figure 5. Scatter plots of predicted versus observed soil pH from the 10-fold cross-validation for the two different grouping strategies, namely cropland type (a) and crop rotation systems (b). The fitting lines with colors are fitted individually by different subsamples. The dashed line represents the 1:1 line, and the solid line represents the fitted line.

3.3. Relative Variable Importance

3.3.1. Modeling for Each Scenario

Figure 6 displays the relative importance of the variables selected by MGFS for each scenario. For scenarios 1–3, using Sentinel-1/2 alone or together led to different variables deemed important. When using them together, the VV backscatter composite 37 (September to October 2021) and the composites 12 of the red edge band (B8A) and NDVI (May to August 2021) were identified as the most important variables. Furthermore, there were almost the same number of Sentinel-1 and Sentinel-2 composites in scenario 3. For scenarios 4–5, only three variables were selected by MGFS from 370~ variables. For scenarios 6–7, the main crop rotation systems in 2021 were identified as the most critical variables, especially the third, which represented the orchard system, outperforming the Sentinel composites. The crop rotation subsystems in 2021 contributed to scenario 6, while the historical cropland maps contributed to scenario 7. Additionally, more optical images, including B2, B8A, B11, and SAVI, than radar images (VH backscatter) were selected by scenario 6. Surprisingly, the natural variables, i.e., soil, climate, and topography, were deemed insignificant by all the scenarios, except for soil thickness, which was chosen by scenario 7.

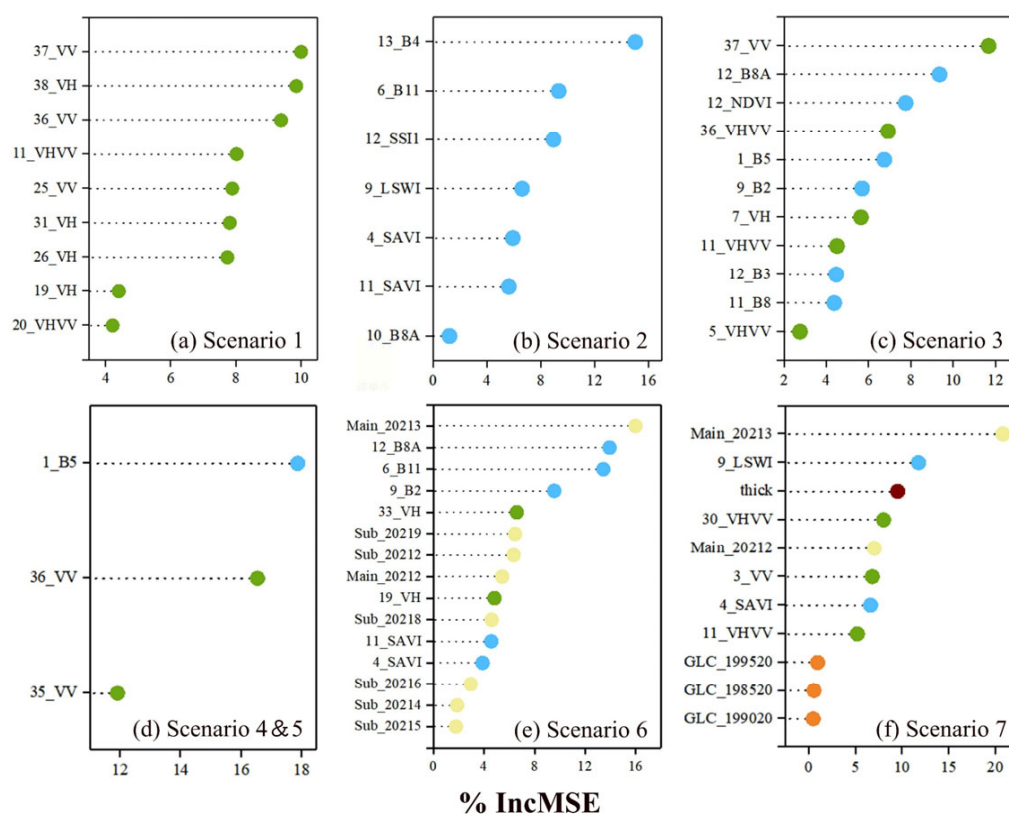


Figure 6. The relative variable importance for scenarios 1 (a), 2 (b), 3 (c), 4 and 5 (d), 6 (e), and 7 (f). The prefixes of the satellite composites represent the order in the compositing sequence, ranging from 0 to 39 for radar and 0 to 13 for optical imagery. The suffixes of crop rotation maps indicate the production year and the systems, and for the main crop rotation, 1, 2, and 3 indicate the paddy, vegetable, and orchard systems, respectively. The dots of green, blue, yellow, brown, and dark brown represent radar, optical, crop rotation system, cropland type, and soil variables, respectively.

3.3.2. Grouping by Cropland and Crop Rotation

Figure 7 displays the relative importance of the variables selected by MGFS from the sub-models built on different subsamples. There were significant differences in the selected variables and their rankings between the two grouping strategies and also between the

models built on different subsamples. For grouping by crop rotation, only Sentinel composites participated in predicting the soil pH without the crop rotation systems themselves, and the optical composites contributed more. All the important variables for the paddy and vegetable subsample were Sentinel-2 composites, including many historical images, while both sensors were used for the orchard subsample. As for grouping by cropland type, the crop rotation systems were selected by MGFS alongside the Sentinel composites. The VHVV and VH backscatters were listed as the most important variables. The rainfed subsample used more variables, which were also more diverse than those in the irrigated subsample.

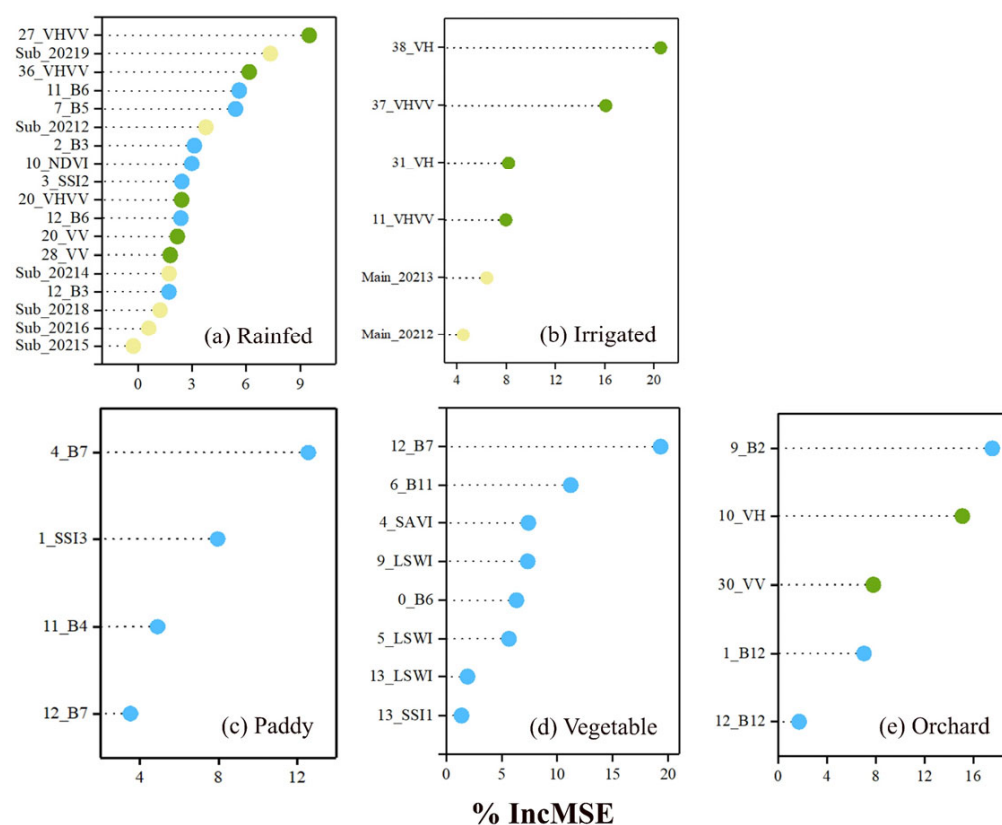


Figure 7. The relative variable importance in the sub-models built on different subsamples: rainfed (a), irrigated (b), paddy (c), vegetable (d), and orchard (e). The dots of green, blue, and yellow represent radar, optical, crop rotation system, cropland type, and soil variables, respectively.

3.4. Spatial Distribution of Soil pH

Figure 8 presents the soil pH maps with significant differences predicted by the seven scenarios. Overall, the study region was dominated by acid soils, and the figure illustrates the increasing soil pH, from southwest to northeast. Averagely 20% of the agricultural soils were predicted to be very acid soils ($\text{pH} < 5$) and concentrated in the southwestern plain area. Scenarios 2, 3, and 7 assigned around 15% of the pixels to very acid soils, while this category occupied nearly 30% of the pixels for the scenarios 4, 5, and 6. Scenario 3 (Figure 8c), which only used Sentinel-1/2 composites, and scenario 6 (Figure 8e), which combined Sentinel data and crop rotation, had the highest (5.38) and lowest (5.25) mean soil pH value, respectively. Relatively acid soils ($6 < \text{pH} < 6.5$) were mainly distributed in the northern mountainous area. Notably, there were considerable differences in small-scale spatial characteristics, and introducing crop rotation maps could help reduce the salt-and-pepper effect.

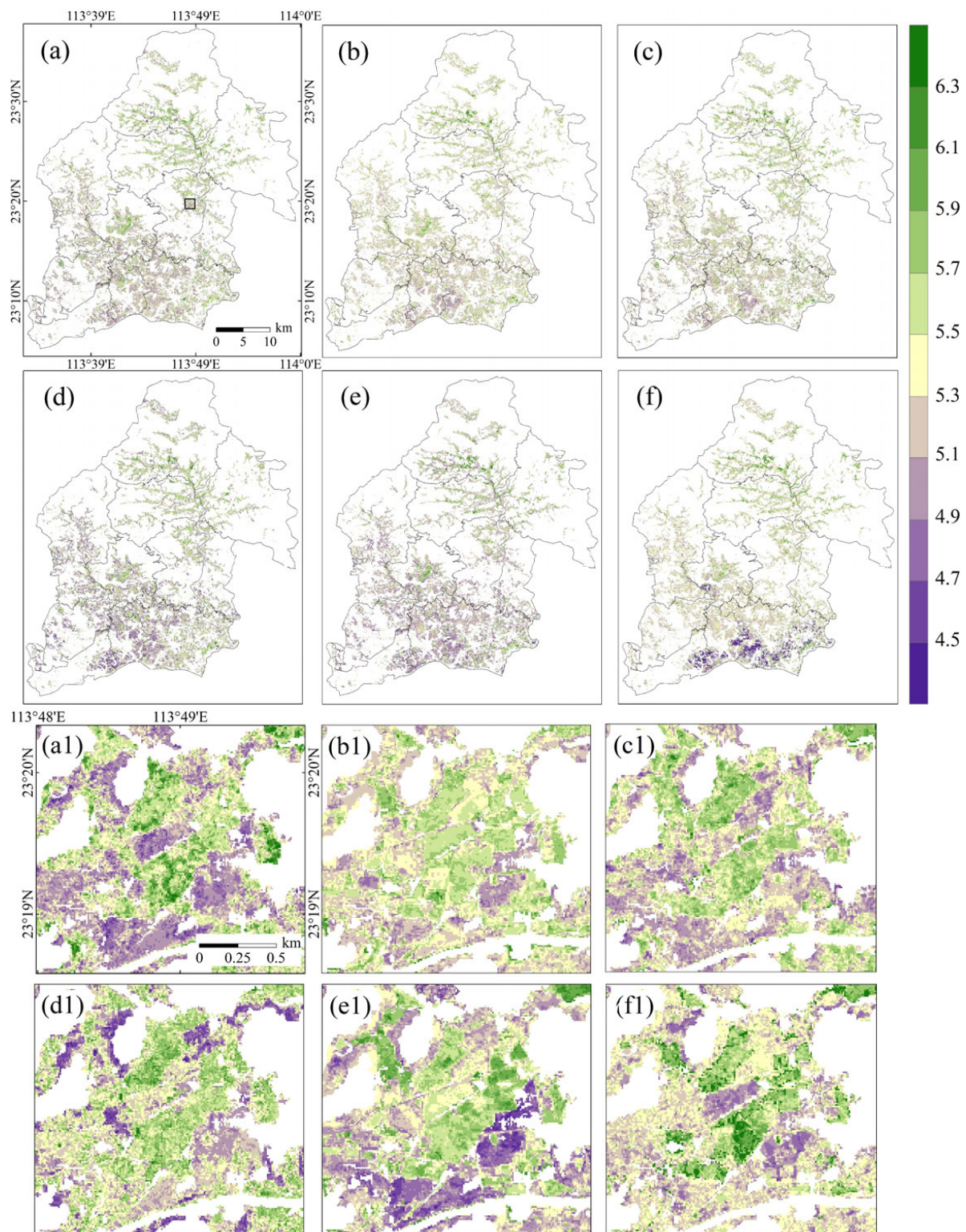


Figure 8. Predicted soil pH distribution in 2021 from scenarios 1 (a), 2 (b), 3 (c), 4 and 5 (d), 6 (e), and 7 (f). Figures (a1–f1) correspond to the zoomed-in maps for the black squares in (a–f).

Figure 9 visualizes the soil pH maps predicted by the soil sample grouping strategies. The spatial patterns of soil pH are generally similar to those in Figure 8e. However, grouping by crop rotation generated a map with more pixels assigned to very acid soils (Figure 9b), and the predicted mean pH was the lowest (5.24).

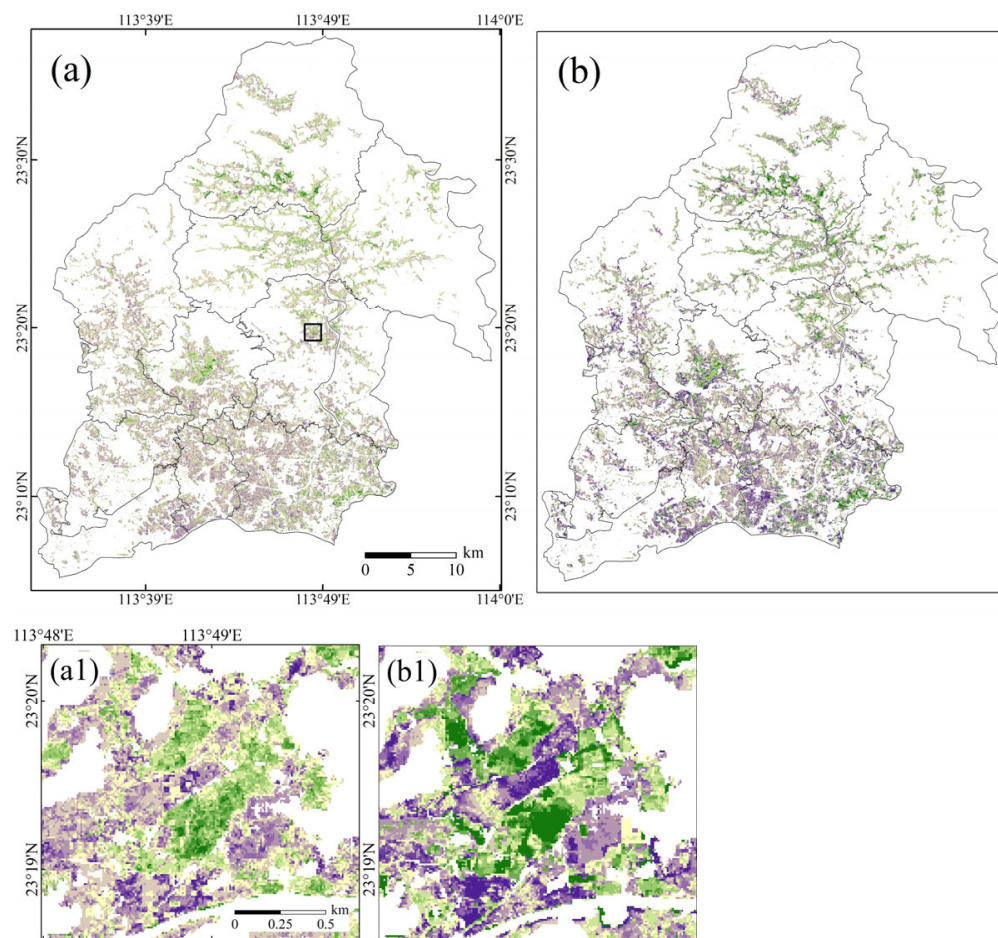


Figure 9. Predicted soil pH distribution in 2021 derived from the two soil sample grouping strategies: grouping by cropland type (a) and crop rotation systems (b). Figures (a1,b1) correspond to the zoomed-in maps for the black squares in (a,b). The legend is identical to that of Figure 8.

4. Discussion

4.1. Modeling for Each Scenario

We evaluated the performance of Sentinel-1/2 images, commonly used natural variables, cropland, and crop rotation maps in predicting soil pH by different data configurations (Table 4, Figure 4). Firstly, the results indicated that Sentinel-1 radar images played an important role in predicting soil pH (Figure 6). Prediction efforts benefit from the uniqueness of radar sensors, which operate in the microwave spectrum and respond to different physical processes [42,55]. Sentinel-1 thus supplemented the optical perceptions of Sentinel-2 through the retrieval of geophysical variables related to soil pH. Long-term radar data have been shown to also be important in mapping the soil pH of Europe, and SAR-optical fusion has been previously shown to achieve better results [56]. Another advantage of active radar sensors is that they can obtain good-quality observations under all weather conditions, allowing the usage of dense time series. Secondly, as indicated by Figure 6, all the natural variables except for soil thickness were excluded from the optimal variable sets selected by MGFS, contrary to most previous findings [31,37,57]. This could be attributed to the site being a small agricultural region with relatively homogenous natural conditions and the fact that we only focused on cropland soils that are predominantly influenced by human activities. The limited spatial resolution of the data used may also help explain this.

Thirdly, in agreement with the soil organic carbon modeling work by [31], we found that the cropland map had no positive effects, while the crop rotation systems remarkably

enhanced the prediction accuracy of soil pH (Figure 4e). This was likely because the soil pH in this area was largely influenced by the intensive and diversified cropping systems, and the crop rotation maps managed to capture the spatio-temporal heterogeneity of cropping systems. Several studies have also reported the effectiveness of cropping-related factors in DSM [58–60]. On the other hand, the cropland map could hardly provide valuable information related to the heterogeneity of soil pH, and there could have been large intra-class variations for rainfed and irrigated cropland. This hypothesis was supported by the statistical analyses. The one-way analysis of variance (ANOVA) suggested that the differences in soil pH between the rainfed and irrigated cropland were not significant, whereas significant differences were observed among the crop rotation systems ($p < 0.05$). The least significant difference test (LSD) revealed that the differences were primarily from the main systems, especially for the crop rotation map produced in 2021 (Figure 10c). The crop rotation map in 2021 distinguished soil pH differences more effectively than that of 2020. This may indicate that the soil pH values kept changing under high-intensity cropping and the soil sampling date influenced the relationship between the soil pH and the crop rotation system containing dynamic crop information.

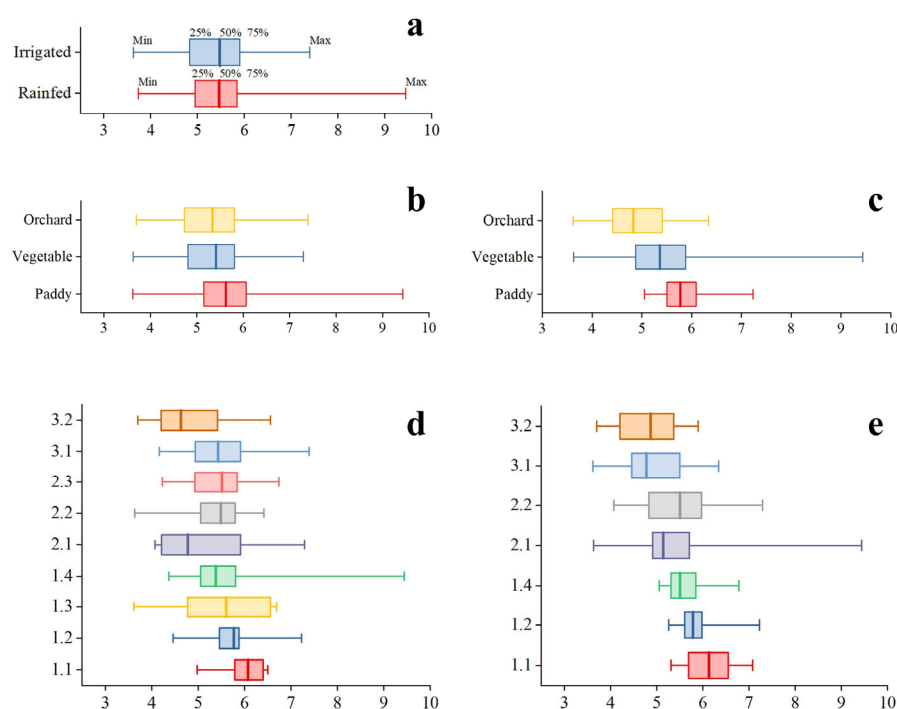


Figure 10. Boxplots of soil pH for cropland type in 2021 (a), main systems of crop rotation in 2020 (b) and 2021 (c), and subsystems of crop rotation in 2020 (d) and 2021 (e). The numbers from 1.1 to 3.2 on the y-axis of (d,e) represent double rice, double rice rotated with vegetables, single rice, single rice rotated with vegetables, high-diversity vegetables, low-diversity vegetables, low-intensity vegetables, short-term orchard, and long-term orchard systems, respectively.

As shown in Figure 10b,c, the paddy system had the highest soil pH, followed by the vegetable and orchard systems. Fertilization, water, and tillage management changing the soil moisture, redox reactions, microbial characteristics, and C/N-cycling processes should explain the differences [23]. The paddy system had a higher soil pH mainly because of less chemical fertilizer application and a larger acid buffering capacity due to flooded fields [21]. On the contrary, vegetables and orchards were severely overfertilized, resulting in a lower soil pH. In Zengcheng, vegetables consumed huge amounts of fertilizers, over three times those of grain crops. In addition, the vegetable system had a high cropping frequency (>3), which could have led to the excess uptake and removal of soil base cations. For the orchard

system, some of the fruit trees tended to be planted in hilly areas with relatively higher altitudes, e.g., litchi and longan, which increased the leaching loss of base cations and thus caused the soil pH to decrease [61].

4.2. Grouping by Cropland and Crop Rotation

We evaluated the performances of grouping soil samples by cropland type and crop rotation systems for predicting soil pH (Table 5, Figure 5). On the one hand, grouping by cropland type did not increase soil pH prediction accuracy (Table 5, Figure 5a). This was inconsistent with the finding of [35], which conducted a separate prediction of SOM in drylands and paddy fields and received positive results. This inconsistency may arise from the two study regions, characterized by distinct soil and cropping environments. The study region of [35] is the Sanjiang Plain in Northeast China, where the agricultural landscapes are more homogenous and the croplands can be clearly divided into drylands and paddy fields. In addition, crops there can only be harvested once per year. The difference in SOM between paddy and dryland fields thus enabled the separate prediction in [35], reducing the influence of spectral differences on predicting the SOM [35]. Compared to the previous context, Zengcheng in Southern China is characterized by more complex agricultural landscapes and cropping systems. The soil pH in Zengcheng thus has a higher spatial heterogeneity that cannot be easily explained by cropland type (Figure 10a), limiting the strategy of grouping by cropland type.

On the other hand, we found that grouping by crop rotation and carrying out separate predictions for the three main rotations gave rise to significant improvements in soil pH estimations (Table 5, Figure 5b). This was because the crop rotation systems managed to capture the spatial heterogeneity of the soil pH (Figure 10c). The proper clustering of soil samples made the models more capable of capturing the complex soil property–environment relationships. In other words, highly personalized RF models could be fitted, fully accounting for the characteristics of the subsamples. In this study, the important variables selected by the sub-models and their rankings noticeably differed from scenario 6 (Figure 6e vs. Figure 7) and each other (Figure 7a vs. Figure 7b vs. Figure 7c). As indicated in Figure 7c, the orchard sub-model with the highest accuracy used both Sentinel-1 and Sentinel-2 composites, which could have impacted performance as radar imagery is more sensitive to the structure of orchards. The sub-models for the paddy and vegetable systems determined band 7 (red edge) as the most influential variable, with the vegetable sub-model using more variables, including SAVI, the water index LSWI, and the soil salinity index SSI1 (Figure 7a,b). We directly compared grouping by crop rotation to the joint modeling of scenario 6 and found that the grouping strategy significantly improved the prediction performance for the vegetable and orchard rotations, probably benefiting from optimized RF models (Tables 5 and 6).

Table 6. Individual statistical soil pH prediction accuracies for different crop rotations based on scenario 6.

Crop Rotation	Number of Selected Variables	ME	RMSE	R ²
Paddy	9	−0.07	0.47	0.15
Vegetable		0.02	0.77	0.30
Orchard		0.15	0.70	−0.04
Total		0.02	0.66	0.36

4.3. Limitations

Some limitations still need to be dealt with in future efforts. Firstly, because of frequent clouds in Southern China, the Sentinel-2 images were seasonally composited, meaning that they may have missed key stages of the soil surface. Denser optical time series (e.g., by the fusion of multisource satellite data) are worth looking into for improving soil pH prediction accuracy [62,63]. Secondly, the current soil pH values are related to cropping history, suggesting that long-term cropping system data could be more effective in DSM [64]. However, limited by crop rotation samples and remote sensing images (the number and quality of Sentinel images in previous years cannot be guaranteed), we only produced and used crop rotation maps for two years. This failed to cover the long-term dynamics of the cropping systems. Thirdly, some other variables associated with farming management could be considered to characterize human activities in soil pH prediction, especially N fertilization, which is closely related to soil pH change. In a different study mapping soil pH for the Netherlands, agricultural management type, ammonia and nitrogen emissions, manure application, and water drainage classes were included, obtaining good results [60]. Unfortunately, we were unable to obtain spatial data on those farming practices. Statistical data cannot reveal spatial differences in small regions, while satellite remote sensing is quite challenging to use for accurately mapping N fertilization and other practices. Finally, our experiment was conducted in a small region, with special characteristics in terms of its climate and cropping practices. Cropping practices in other regions and their influence on the soil pH are different, which may change the effectiveness of grouping by crop rotation in predicting the soil pH.

4.4. Implications

Even though a few studies have proven the effectiveness of crop rotation in DSM, they mostly focused on SOM or SOC and hardly mentioned soil pH [31,58,65]. Our study implies that cropping information should be included in the DSM of soil pH in agricultural landscapes. Moreover, the results highlight the great potential of the soil sample grouping strategy, considering differences in crop rotation, for obtaining more accurate soil pH estimates (Table 5, Figure 5b), which may also be valid for other soil properties and similar regions. Testing this grouping approach in other regions with different soil types and cropping systems is meaningful and necessary but also challenging in terms of gathering remote sensing data of cropping systems. Previous studies have explored the effects of the grouping strategy for DSM. Yet, they mainly adopted remote sensing, soil spectra, or soil type to group soil samples, while grouping by crop rotation has not yet been investigated over complex agricultural landscapes. In addition, we proved the usefulness of Sentinel-1 SAR images in predicting soil pH. Therefore, when mapping soil pH over cloudy regions (such as Southern China) where optical sensors are limited, radar data are particularly valuable and should be considered. A few historical satellite composites were selected by the models (Figures 6 and 7), suggesting the usefulness of using long-term satellite images in predicting soil pH.

Our study also has substantial implications for agricultural and environmental management. For example, it can support decision making for remediating and preventing soil acidification by redistributing cropping systems [8]. The statistical analysis revealed average soil pH values for different crop rotation systems, according to which the paddy system had the highest pH value (Figure 10c). Therefore, farmers could consider converting very acid soils which grow vegetables or orchards into paddy systems. Alternatively, inter-annual rotations between paddy and vegetable/orchard systems and the intra-annual rotations between vegetables and paddy rice, which balance economic output and soil conservation, could be promoted. Soil pH maps could also be used to support soil heavy

metal pollution control and remediation, because soil pH closely relates to the activity and availability of heavy metals. For example, maps of the tolerable threshold concentration of heavy metals in arable land have been previously derived based on produced soil pH maps and national standards [30,37].

5. Conclusions

Crop rotation affects soil pH considerably in both direct (e.g., removal of base cations by harvest) and indirect (e.g., fertilization, irrigation, tillage) ways, yet its usefulness and effectiveness for soil pH mapping remains unclear. This study assessed whether and to what extent incorporating crop rotation could improve the digital mapping of soil pH. The innovation was the incorporation of satellite crop rotation maps in predicting cropland soil pH. To unveil the effects of crop rotation on soil pH mapping, we proposed a new soil sample grouping and separate modeling strategy for paddy, vegetable, and orchard rotation systems. Our results showed the following: (1) simply incorporating crop rotation was useful for predicting soil pH and contributed the most to the spatial predictive model, followed by long-term Sentinel-1/2 composites; (2) grouping soil samples by crop rotation significantly increased the prediction accuracy of soil pH by fitting and optimizing the RF sub-models based on different crop rotation subsamples. This likely benefited from the fact that soil pH and its influencing factors varied greatly among the three main crop rotations. We also found that radar images could be considered in the digital mapping of soil pH to supplement optical perceptions through the retrieval of geophysical variables. Our study thus provides a new way of characterizing human interventions in soil pH mapping and further introduces a new soil sample grouping strategy for improving model performance over heterogeneous agricultural landscapes.

Author Contributions: Conceptualization, Q.Y., Q.Z., G.S. (Ge Shen) and W.W.; methodology, Y.L. and S.C.; software, Y.L.; validation, Y.L. and Z.C.; formal analysis, X.Z.; investigation, Y.L., Z.C. and Q.Y.; resources, W.W., Q.Y. and G.S. (Guofei Shang); data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, Q.Y., G.S. (Ge Shen), S.C., C.C. and S.D.B.-K.; visualization, Y.L.; supervision, Q.Y. and S.C.; project administration, W.W.; funding acquisition, Y.L., G.S. (Ge Shen), and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (Nos. 42171271 and 42401331), the open project of State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China, the Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences (No. EUAL-2024-03), Hebei Natural Science Foundation (No. D2024403075), the National Statistical Science Research Project (2024LY051), the National Key Research and Development Program of China (2023YFD150130), the Guiding Funds of Central Government for Supporting the Development of the Local Science and Technology (236Z4201G), and the Doctoral Research Start-up Fund Project of Hebei GEO University (No. BQ2024042).

Data Availability Statement: The data used in this study included freely available satellite images available on GEE. Field survey data are available upon requests to the corresponding author (shenge@zufe.edu.cn).

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

Abbreviations

The following abbreviations are used in this manuscript:

DSM	Digital soil mapping
RMSE	Root mean squared error
N	Nitrogen
GLC_FCS30	Global land-cover product with fine classification system at 30 m

GPS	Global positioning system
Scorpan-SSPFe	Soil spatial prediction function with spatially autocorrelated errors and reference
VH	Vertical–horizontal
VV	Vertical–vertical
CR	Cross-polarization ratio
NDVI	Normalized difference vegetation index
SAVI	Soil-adjusted vegetation index
LSWI	Land surface water index
SSI1	Soil salinity index 1
SSI2	Soil salinity index 2
SSI3	Soil salinity index 3
SOC	Soil organic carbon
SOM	Soil organic matter
MAT	Mean annual temperature
MAP	Mean annual precipitation
LST	Land surface temperature
TWI	Topographic wetness index
PLC	Plan curvature
PRC	Profile curvature
LS	Slope-length factor
VD	Valley depth
CI	Convergence index
CNBL	Channel network base level
RSP	Relative slope position
CND	Catchment network distance
VDCN	Vertical distance to channel network
RF	Random forest
MSI	Multispectral imager
SAR	Synthetic aperture radar
GRD	Ground range-detected
GEE	Google Earth Engine
NSIGC	High-resolution National Soil Information Grids of China
SRTM	Shuttle Radar Topographic Mission
RFE	Recursive feature elimination
MGFS	Modified greedy feature selection
mtry	Number of variables randomly sampled as candidates at each split
ntree	Number of trees
ME	Mean error
R ²	Coefficient of determination
%IncMSE	Increase in mean square error

References

1. Neina, D. The Role of Soil PH in Plant Nutrition and Soil Remediation. *Appl. Environ. Soil Sci.* **2019**, *2019*, 1–9. [[CrossRef](#)]
2. Tan, K.; Wang, H.; Chen, L.; Du, Q.; Du, P.; Pan, C. Estimation of the Spatial Distribution of Heavy Metal in Agricultural Soils Using Airborne Hyperspectral Imaging and Random Forest. *J. Hazard. Mater.* **2020**, *382*, 120987. [[CrossRef](#)] [[PubMed](#)]
3. Xu, D.; Shen, Z.; Dou, C.; Dou, Z.; Li, Y.; Gao, Y.; Sun, Q. Effects of Soil Properties on Heavy Metal Bioavailability and Accumulation in Crop Grains under Different Farmland Use Patterns. *Sci. Rep.* **2022**, *12*, 9211. [[CrossRef](#)] [[PubMed](#)]
4. Breemen, V.N.; Mulder, J.; Driscoll, C.T. Acidification and Alkalinization of Soils. *Plant Soil* **1983**, *75*, 283–308. [[CrossRef](#)]
5. Breemen, V.N.; Driscoll, C.T.; Mulder, J. Acidic Deposition and Internal Proton Sources in Acidification of Soils and Waters. *Nature* **1984**, *307*, 599–604. [[CrossRef](#)]
6. von Uexküll, H.R.; Mutert, E. Global Extent, Development and Economic Impact of Acid Soils. *Plant Soil* **1995**, *171*, 1–15. [[CrossRef](#)]
7. Naz, M.; Dai, Z.; Hussain, S.; Tariq, M.; Danish, S.; Khan, I.U.; Qi, S.; Du, D. The Soil PH and Heavy Metals Revealed Their Impact on Soil Microbial Community. *J. Environ. Manage.* **2022**, *321*, 115770. [[CrossRef](#)] [[PubMed](#)]

8. Zhu, X.F.; Shen, R.F. Towards Sustainable Use of Acidic Soils: Deciphering Aluminum-Resistant Mechanisms in Plants. *Fundam. Res.* **2023**, *4*, 1533–1541. [[CrossRef](#)]
9. Kochian, L.V.; Piñeros, M.A.; Liu, J.; Magalhaes, J.V. Plant Adaptation to Acid Soils: The Molecular Basis for Crop Aluminum Resistance. *Annu. Rev. Plant Biol.* **2015**, *66*, 571–598. [[CrossRef](#)]
10. Guo, J.H.; Liu, X.J.; Zhang, Y.; Shen, J.L.; Han, W.X.; Zhang, W.F.; Christie, P.; Goulding, K.W.T.; Vitousek, P.M.; Zhang, F.S. Significant Acidification in Major Chinese Croplands. *Science* **2010**, *327*, 1008–1010. [[CrossRef](#)]
11. Lu, X.; Zhang, X.; Zhan, N.; Wang, Z.; Li, S. Factors Contributing to Soil Acidification in the Past Two Decades in China. *Environ. Earth Sci.* **2023**, *82*, 74. [[CrossRef](#)]
12. Zhu, Q.; Liu, X.; Hao, T.; Zeng, M.; Shen, J.; Zhang, F.; de Vries, W. Cropland Acidification Increases Risk of Yield Losses and Food Insecurity in China. *Environ. Pollut.* **2020**, *256*, 113145. [[CrossRef](#)]
13. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On Digital Soil Mapping. *Geoderma* **2003**, *117*, 3–52. [[CrossRef](#)]
14. Minasny, B.; McBratney, A.B. Digital Soil Mapping: A Brief History and Some Lessons. *Geoderma* **2016**, *264*, 301–311. [[CrossRef](#)]
15. Sünneemann, M.; Beugnon, R.; Breikreuz, C.; Buscot, F.; Cesarz, S.; Jones, A.; Lehmann, A.; Lochner, A.; Orgiazzi, A.; Reitz, T.; et al. Climate Change and Cropland Management Compromise Soil Integrity and Multifunctionality. *Commun. Earth Environ.* **2023**, *4*, 394. [[CrossRef](#)]
16. Meng, H.; Xu, M.; Lv, J.; He, X.H.; Wang, B.; Cai, Z. Quantification of Anthropogenic Acidification under Long-Term Fertilization in the Upland Red Soil of South China. *Soil Sci.* **2014**, *179*, 486–494. [[CrossRef](#)]
17. Shin, J.; Won, J.; Kim, S.M.; Kim, D.C.; Cho, Y. Fertilization Mapping Based on the Soil Properties of Paddy Fields in Korea. *Agric.* **2023**, *13*, 2049. [[CrossRef](#)]
18. Adalibieke, W.; Cui, X.; Cai, H.; You, L.; Zhou, F. Global Crop-Specific Nitrogen Fertilization Dataset in 1961–2020. *Sci. Data* **2023**, *10*, 617. [[CrossRef](#)] [[PubMed](#)]
19. Wang, X.; Dou, Z.; Shi, X.; Zou, C.; Liu, D.; Wang, Z.; Guan, X.; Sun, Y.; Wu, G.; Zhang, B.; et al. Innovative Management Programme Reduces Environmental Impacts in Chinese Vegetable Production. *Nat. Food* **2021**, *2*, 47–53. [[CrossRef](#)]
20. Zhao, J.; Liu, Z.; Zhai, B.; Jin, H.; Xu, X.; Zhu, Y. Long-Term Changes in Soil Chemical Properties with Cropland-to-Orchard Conversion on the Loess Plateau, China: Regulatory Factors and Relations with Apple Yield. *Agric. Syst.* **2023**, *204*, 103562. [[CrossRef](#)]
21. Lu, H.-L.; Li, K.-W.; Nkoh, J.N.; He, X.; Xu, R.-K.; Qian, W.; Shi, R.-Y.; Hong, Z.-N. Effects of PH Variations Caused by Redox Reactions and PH Buffering Capacity on Cd(II) Speciation in Paddy Soils during Submerging/Draining Alternation. *Ecotoxicol. Environ. Saf.* **2022**, *234*, 113409. [[CrossRef](#)] [[PubMed](#)]
22. Li, B.; Zhu, D.; Li, J.; Liu, X.; Yan, B.; Mao, L.; Zhang, M.; Wang, Y.; Li, X. Converting Upland to Paddy Fields Alters Soil Nitrogen Microbial Functions at Different Depths in Black Soil Region. *Agric. Ecosyst. Environ.* **2024**, *372*, 109089. [[CrossRef](#)]
23. Wu, Z.; Sun, X.; Sun, Y.; Yan, J.; Zhao, Y.; Chen, J. Soil Acidification and Factors Controlling Topsoil PH Shift of Cropland in Central China from 2008 to 2018. *Geoderma* **2022**, *408*, 115586. [[CrossRef](#)]
24. Hao, T.; Liu, X.; Zhu, Q.; Zeng, M.; Chen, X.; Yang, L.; Shen, J.; Shi, X.; Zhang, F.; de Vries, W. Quantifying Drivers of Soil Acidification in Three Chinese Cropping Systems. *Soil Tillage Res.* **2022**, *215*, 105230. [[CrossRef](#)]
25. Waldhoff, G.; Lussem, U.; Bareth, G. Multi-Data Approach for Remote Sensing-Based Regional Crop Rotation Mapping: A Case Study for the Rur Catchment, Germany. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *61*, 55–69. [[CrossRef](#)]
26. Li, R.; Xu, M.; Chen, Z.; Gao, B.; Cai, J.; Shen, F.; He, X.; Zhuang, Y.; Chen, D. Phenology-Based Classification of Crop Species and Rotation Types Using Fused MODIS and Landsat Data: The Comparison of a Random-Forest-Based Model and a Decision-Rule-Based Model. *Soil Tillage Res.* **2021**, *206*, 104838. [[CrossRef](#)]
27. Liu, Y.; Yu, Q.; Zhou, Q.; Wang, C.; Bellingrath-Kimura, S.D.; Wu, W. Mapping the Complex Crop Rotation Systems in Southern China Considering Cropping Intensity, Crop Diversity and Their Seasonal Dynamics. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9584–9598. [[CrossRef](#)]
28. Liu, Y.; Zhao, W.; Chen, S.; Ye, T. Mapping Crop Rotation by Using Deeply Synergistic Optical and Sar Time Series. *Remote Sens.* **2021**, *13*, 4160. [[CrossRef](#)]
29. Ren, T.; Xu, H.; Cai, X.; Yu, S.; Qi, J. Smallholder Crop Type Mapping and Rotation Monitoring in Mountainous Areas with Sentinel-1/2 Imagery. *Remote Sens.* **2022**, *14*, 566. [[CrossRef](#)]
30. Hu, B.; Xie, M.; Shi, Z.; Li, H.; Chen, S.; Wang, Z.; Zhou, Y.; Ni, H.; Geng, Y.; Zhu, Q.; et al. Fine-Resolution Mapping of Cropland Topsoil PH of Southern China and Its Environmental Application. *Geoderma* **2024**, *442*, 116798. [[CrossRef](#)]
31. Liu, Y.; Chen, S.; Yu, Q.; Cai, Z.; Zhou, Q.; Bellingrath-Kimura, S.D.; Wu, W. Improving Digital Mapping of Soil Organic Matter in Cropland by Incorporating Crop Rotation. *Geoderma* **2023**, *438*, 116620. [[CrossRef](#)]
32. Bouslihim, Y.; John, K.; Miftah, A.; Azmi, R.; Aboutayeb, R.; Bouasria, A.; Razouk, R.; Hssaini, L. The Effect of Covariates on Soil Organic Matter and PH Variability: A Digital Soil Mapping Approach Using Random Forest Model. *Ann. GIS* **2024**, *30*, 215–232. [[CrossRef](#)]

33. Bao, Y.; Meng, X.; Ustin, S.; Wang, X.; Zhang, X.; Liu, H.; Tang, H. Vis-SWIR Spectral Prediction Model for Soil Organic Matter with Different Grouping Strategies. *Catena* **2020**, *195*, 104703. [\[CrossRef\]](#)
34. Zhu, X.; Ros, G.H.; Xu, M.; Xu, D.; Cai, Z.; Sun, N.; Duan, Y.; de Vries, W. The Contribution of Natural and Anthropogenic Causes to Soil Acidification Rates under Different Fertilization Practices and Site Conditions in Southern China. *Sci. Total Environ.* **2024**, *934*, 172986. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Ma, H.; Wang, C.; Liu, J.; Yuan, Z.; Yao, C.; Wang, X.; Pan, X. Separate Prediction of Soil Organic Matter in Drylands and Paddy Fields Based on Optimal Image Synthesis Method in the Sanjiang Plain, Northeast China. *Geoderma* **2024**, *447*, 116929. [\[CrossRef\]](#)
36. Duan, D.; Sun, X.; Wang, C.; Zha, Y.; Yu, Q.; Yang, P. A Remote Sensing Approach to Estimating Cropland Sustainability in the Lateritic Red Soil Region of China. *Remote Sens.* **2024**, *16*, 1069. [\[CrossRef\]](#)
37. 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. [\[CrossRef\]](#)
38. Zhang, X.; Liu, L.; Chen, X.; Gao, Y.; Xie, S.; Mi, J. GLC_FCS30: Global Land-Cover Product with Fine Classification System at 30 m Using Time-Series Landsat Imagery. *Earth Syst. Sci. Data Discuss.* **2020**, *13*, 2753–2776. [\[CrossRef\]](#)
39. Zhang, X.; Zhao, T.; Xu, H.; Liu, W.; Wang, J.; Chen, X.; Liu, L. GLC_FCS30D: The First Global 30m Land-Cover Dynamics Monitoring Product with a Fine Classification System for the Period from 1985 to 2022 Generated Using Dense-Time-Series Landsat Imagery and the Continuous Change-Detection Method. *Earth Syst. Sci. Data* **2024**, *16*, 1353–1381. [\[CrossRef\]](#)
40. Liu, F.; Wu, H.; Zhao, Y.; Li, D.; Yang, J.; Song, X.; Shi, Z. Mapping High Resolution National Soil Information Grids of China. *Sci. Bull.* **2022**, *67*, 328–340. [\[CrossRef\]](#)
41. Peng, S.; Ding, Y.; Liu, W.; Li, Z. 1 Km Monthly Temperature and Precipitation Dataset for China from 1901 to 2017. *Earth Syst. Sci. Data* **2019**, *11*, 1931–1946. [\[CrossRef\]](#)
42. Domenech, M.B.; Amioti, N.M.; Costa, J.L.; Castro-Franco, M. Prediction of Topsoil Properties at Field-Scale by Using C-Band SAR Data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *93*, 102197. [\[CrossRef\]](#)
43. van Hateren, T.C.; Chini, M.; Matgen, P.; Pulvirenti, L.; Pierdicca, N.; Teuling, A.J. On the Potential of Sentinel-1 for Sub-Field Scale Soil Moisture Monitoring. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *120*, 103342. [\[CrossRef\]](#)
44. Meroni, M.; d'Andrimont, R.; Vrieling, A.; Fasbender, D.; Lemoine, G.; Rembold, F.; Seguí, L.; Verhegghen, A. Comparing Land Surface Phenology of Major European Crops as Derived from SAR and Multispectral Data of Sentinel-1 and -2. *Remote Sens. Environ.* **2021**, *253*, 112232. [\[CrossRef\]](#)
45. Yescas-Coronado, P.; Segura-Castruita, M.Á.; Chávez-Rodríguez, A.M.; Gómez-Leyva, J.F.; Martínez-Sifuentes, A.R.; Amador-Camacho, O.; González-Medina, R. Covariables of Soil-Forming Factors and Their Influence on PH Distribution and Spatial Variability. *Agriculture* **2022**, *12*, 2132. [\[CrossRef\]](#)
46. 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. [\[CrossRef\]](#)
47. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [\[CrossRef\]](#)
48. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [\[CrossRef\]](#)
49. Xiao, X.; Boles, S.; Froking, S.; Salas, W.; Moore, I.; Li, C.; He, L.; Zhao, R. Observation of Flooding and Rice Transplanting of Paddy Rice Fields at the Site to Landscape Scales in China Using VEGETATION Sensor Data. *Int. J. Remote Sens.* **2002**, *23*, 3009–3022. [\[CrossRef\]](#)
50. Khan, N.M.; Rastokuev, V.V.; Sato, Y.; Shiozawa, S. Assessment of Hydrosaline Land Degradation by Using a Simple Approach of Remote Sensing Indicators. *Agric. Water Manag.* **2005**, *77*, 96–109. [\[CrossRef\]](#)
51. Hijmans, R.J.; Cameron, S.E.; Parra, J.L.; Jones, P.G.; Jarvis, A. Very High Resolution Interpolated Climate Surfaces for Global Land Areas. *Int. J. Climatol.* **2005**, *25*, 1965–1978. [\[CrossRef\]](#)
52. Xiao, Y.; Xue, J.; Zhang, X.; Wang, N.; Hong, Y.; Jiang, Y.; Zhou, Y.; Teng, H.; Hu, B.; Lugato, E.; et al. Improving Pedotransfer Functions for Predicting Soil Mineral Associated Organic Carbon by Ensemble Machine Learning. *Geoderma* **2022**, *428*, 116208. [\[CrossRef\]](#)
53. Zhang, X.; Chen, S.; Xue, J.; Wang, N.; Xiao, Y.; Chen, Q.; Hong, Y.; Zhou, Y.; Teng, H.; Hu, B.; et al. Improving Model Parsimony and Accuracy by Modified Greedy Feature Selection in Digital Soil Mapping. *Geoderma* **2023**, *432*, 116383. [\[CrossRef\]](#)
54. Wadoux, A.M.J.C.; Minasny, B.; McBratney, A.B. Machine Learning for Digital Soil Mapping: Applications, Challenges and Suggested Solutions. *Earth-Sci. Rev.* **2020**, *210*, 103359. [\[CrossRef\]](#)
55. Han, D.; Vahedifard, F.; Aanstoos, J.V. Investigating the Correlation between Radar Backscatter and in Situ Soil Property Measurements. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *57*, 136–144. [\[CrossRef\]](#)
56. Geng, Y.; Zhou, T.; Zhang, Z.; Cui, B.; Sun, J.; Zeng, L.; Yang, R.; Wu, N.; Liu, T.; Pan, J.; et al. Continental-Scale Mapping of Soil pH with SAR-Optical Fusion Based on Long-Term Earth Observation Data in Google Earth Engine. *Ecol. Indic.* **2024**, *165*, 112246. [\[CrossRef\]](#)

57. Guo, J.; Wang, K.; Jin, S. Mapping of Soil PH Based on SVM-RFE Feature Selection Algorithm. *Agronomy* **2022**, *12*, 2742. [[CrossRef](#)]
58. Wu, Z.; Liu, Y.; Han, Y.; Zhou, J.; Liu, J.; Wu, J. Mapping Farmland Soil Organic Carbon Density in Plains with Combined Cropping System Extracted from NDVI Time-Series Data. *Sci. Total Environ.* **2021**, *754*, 142120. [[CrossRef](#)]
59. Yang, L.; He, X.; Shen, F.; Zhou, C.; Zhu, A.X.; Gao, B.; Chen, Z.; Li, M. Improving Prediction of Soil Organic Carbon Content in Croplands Using Phenological Parameters Extracted from NDVI Time Series Data. *Soil Tillage Res.* **2020**, *196*, 104465. [[CrossRef](#)]
60. Helfenstein, A.; Mulder, V.L.; Heuvelink, G.B.M.; Okx, J.P. Tier 4 Maps of Soil PH at 25 m Resolution for the Netherlands. *Geoderma* **2022**, *410*, 115659. [[CrossRef](#)]
61. Zhang, J.E.; Ouyang, Y.; Ling, D.J. Impacts of Simulated Acid Rain on Cation Leaching from the Latosol in South China. *Chemosphere* **2007**, *67*, 2131–2137. [[CrossRef](#)] [[PubMed](#)]
62. Claverie, M.; Ju, J.; Masek, J.G.; Dungan, J.L.; Vermote, E.F.; Roger, J.C.; Skakun, S.V.; Justice, C. The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set. *Remote Sens. Environ.* **2018**, *219*, 145–161. [[CrossRef](#)]
63. 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](#)]
64. Wang, Y.; Wang, S.; Adhikari, K.; Wang, Q.; Sui, Y.; Xin, G. Effect of Cultivation History on Soil Organic Carbon Status of Arable Land in Northeastern China. *Geoderma* **2019**, *342*, 55–64. [[CrossRef](#)]
65. Yang, L.; Song, M.; Zhu, A.X.; Qin, C.; Zhou, C.; Qi, F.; Li, X.; Chen, Z.; Gao, B. Predicting Soil Organic Carbon Content in Croplands Using Crop Rotation and Fourier Transform Decomposed Variables. *Geoderma* **2019**, *340*, 289–302. [[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.