Soil Use and Management

WILEY

RESEARCH ARTICLE

Advancing Provincial Cropland Soil Mapping With Temporal Satellite Data Integration

Xi Wang[1,2] | Furong Zhou[1,2] | Songchao Chen[1,2,3] | Xunfei Deng[4] | Zhouqiao Ren[4] | Si-Bo Duan[5] | Mengru Wang[6] | Zhou Shi[1,2]

[1]State Key Laboratory of Soil Pollution Control and Safety, Zhejiang University, Hangzhou, China | [2]College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China | [3]ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou, China | [4]Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou, China | [5]State Key Laboratory of Efficient Utilization of Arable Land in China/Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China | [6]Earth Systems and Global Change Group, Wageningen University & Research, Wageningen, the Netherlands

Correspondence: Songchao Chen (chensongchao@zju.edu.cn) | Xunfei Deng (dengxf@zaas.ac.cn)

Received: 11 April 2025 | **Revised:** 13 June 2025 | **Accepted:** 19 June 2025

Funding: This work is supported by the National Key Research and Development Program of China (2023YFD1900102) and the National Natural Science Foundation of China (42201054).

Keywords: digital soil mapping | management practices | satellite gap-filling | temporal information | time series satellite images

ABSTRACT

Understanding the spatiotemporal distribution of soil constraints, such as soil acidification and degradation, is critical for evaluating soil productivity and safeguarding ecosystem health. Digital soil mapping (DSM), which integrates environmental covariates, offers significant advantages in providing detailed soil information across scales. However, existing DSM approaches often fail to adequately represent agricultural management practices on cropland, limiting their ability to accurately characterise soil properties. In this study, we employed random forest (RF), Cubist, XGBoost and Artificial Neural Network (ANN) models within a DSM framework, incorporating crop growth dynamics, planting patterns and satellite time series spectral data to predict soil organic matter (SOM) and pH. The results showed that the RF model, which integrated temporal information, outperformed the other models for SOM prediction, achieving an R^2 of 0.42 and an RMSE of 9.67 g kg^{-1}, representing a 13.51% improvement in R^2 and a 3.21% reduction in RMSE compared to the baseline model without temporal information. For soil pH prediction, the XGBoost model, which incorporated temporal information, delivered the best performance with an R^2 of 0.61 and an RMSE of 0.63, resulting in a 12.96% improvement in R^2 and a 5.97% reduction in RMSE relative to the baseline model. This study demonstrates that integrating temporal information, such as crop growth and management dynamics, substantially enhances the accuracy of soil property predictions. These findings provide a promising approach for high-resolution soil mapping with broad applicability in agricultural and environmental management.

1 | Introduction

Cropland is fundamental to human survival and development (Gao et al. 2024; Guo, Yue, et al. 2023). In recent years, rapid global population growth and the intensifying impacts of climate change have exacerbated risks to cropland soils, posing significant threats to soil health and food security (Jin et al. 2022; Wang, Zhen, et al. 2023). In China, decades of intensive agricultural practices have contributed to diverse cropland soil challenges, including degradation, acidification, salinization, wind erosion, structural barriers, waterlogging, soil erosion and drought. These issues exhibit pronounced

regional and spatial variations (Wu et al. 2022; Zhang, Xiang, et al. 2022; Dai et al. 2024). For instance, in northeast China, waterlogging and salinization are predominant concerns (Guo, Liu, et al. 2023; Li, Li, et al. 2022). In contrast, southern regions of China face challenges such as soil acidification and structural constraints, including compacted layers and gleyed horizons (Hao et al. 2022; Zhu et al. 2024). To safeguard national food security, it is essential to obtain spatially detailed information on soil-related constraints—such as organic matter depletion and acidification—to inform land quality management and guide evidence-based policymaking (Das et al. 2024; Ma et al. 2023; Chen et al. 2023).

Digital soil mapping (DSM) utilises environmental covariates derived from soil-forming factors to predict soil properties at unsampled locations by analysing the relationships between observed soil data and associated environmental covariates (McBratney et al. 2003). This method effectively represents the spatiotemporal distribution of soils and their properties and has been widely applied to map various soil properties (Chen et al. 2022; Huang et al. 2022). Recent advances in DSM have focused on the integration of novel environmental covariates, such as agricultural practices and three-dimensional spatiotemporal (3D + T) variables (Helfenstein et al. 2024; Nussbaum et al. 2023; Temme et al. 2022). For instance, Liu et al. (2023) leveraged remote sensing data to identify crop rotation patterns in agricultural fields and employed this information to map soil organic matter (SOM) in cropland, highlighting the beneficial effects of sustainable agricultural practices like crop rotation on SOM accumulation. Similarly, He et al. (2021) extracted phenological parameters and spectral indices from Sentinel-2 imagery to explore the relationship between crop growth and soil organic carbon (SOC), demonstrating the efficacy of this approach for SOC mapping in cropland. Yang et al. (2020) applied phenological parameters extracted from HJ-1A/B imagery to assess the impact of agricultural activities on cropland. Additionally, Chen et al. (2024) incorporated soil spatial neighbour information to improve predictions of soil properties. These innovative approaches significantly enhance the DSM framework, offering new avenues for more accurate and comprehensive soil mapping.

Cropland as a highly anthropogenically influenced natural-artificial complex ecosystem is profoundly shaped by human agricultural activities, which in turn impact soil dynamics (Liu et al. 2023). However, detailed incorporation of crop growth and human activities in DSM is often constrained by data availability. In regions with fragmented cropland and a cloudy, rainy climate, such as southern China, single-source remote sensing imagery frequently lacks sufficient temporal resolution to capture dynamic changes in soil properties and the effects of agricultural practices (Babaeian et al. 2019; Li, Xu, et al. 2021). Integrating multi-source remote sensing data emerges as a promising solution to overcome these limitations. Temporal vegetation indices derived from complete and comprehensive imagery can effectively represent crop growth status, indirectly reflecting soil property changes (Pôças et al. 2020; Zeng et al. 2022). For instance, healthy vegetation often correlates with higher SOM levels, while yellowing leaves may indicate soil acidification (Guy et al. 2022; Hu et al. 2023, 2024; Shafizadeh-Moghadam et al. 2022). Temporal vegetation indices can also reveal crop rotation cycles, providing insights into agricultural management practices (Wu et al. 2023; Yang et al. 2023). Spectral data further enhance DSM by detecting changes in soil composition, offering valuable information on SOM, moisture, nutrients, and pH levels (Rossel et al. 2016). For example, functional groups like hydroxyl and carboxyl in SOM exhibit characteristic absorption in the infrared spectrum. Similarly, soil acidification alters the redox states of metal elements, such as iron, affecting their absorption characteristics in the visible and near-infrared spectra (Weng et al. 2022; Rasooli et al. 2023; Chen et al. 2024). To accurately represent the spatial distribution of complex agricultural soil properties, it is essential to leverage complete temporal imagery. This approach enables capturing dynamic information about crop growth, soil properties and human activities, offering new opportunities for advancing DSM in agroecosystems.

This study focuses on Zhejiang Province, China, a region characterised by a cloudy and rainy climate, using 7284 soil sample points collected in 2019 to predict soil organic matter (SOM) and pH in agricultural fields. The primary objective of this study is to evaluate whether the inclusion of temporal information (the vegetation index and multispectral data from time-series imagery) enhances model accuracy. Additionally, we quantify the contribution of each environmental covariate to soil property prediction and compare the SOM and pH maps generated with temporal information to those derived from the high-resolution National Soil Information Grids of China (NSIGC) and SoilGrids 2.0.

2 | Data and Methodology

2.1 | Study Area

The study area is located in Zhejiang Province (118°01′-123°10′ E, 27°06′-31°11′ N), southeastern China, encompassing an area of approximately 105,500 km^2 (Zhejiang Government's Official Web, https://www.zj.gov.cn/) (Figure 1). This region spans an elevation range of 1–1796 m and extends from the southwest to the northeast, incorporating both marine and terrestrial environments, with a terrain predominantly characterised by mountainous and hilly landscapes. Located within the central subtropical zone, the region is characterised by a monsoon humid climate that features high cloud cover, typically ranging from 50% to 70% between June and August (Li, Min, et al. 2022). The annual average temperature varies from 15°C to 18°C, while annual sunshine hours fluctuate between 1100 and 2200 h, and average annual precipitation ranges from 1100 to 2000 mm (Mao 2019). Influenced by the ocean and the Southeast Asian monsoon, significant seasonal variations in precipitation are experienced in the area, with peak rainfall being recorded in May and June, which accounts for approximately 30% of the total annual precipitation (Teng et al. 2014).

The parent materials for soil formation in Zhejiang Province comprise a diverse range of both acidic and basic rocks, with Quaternary red clay and sandstone predominating, along with siliceous and aluminosilicate weathering crusts. This geological

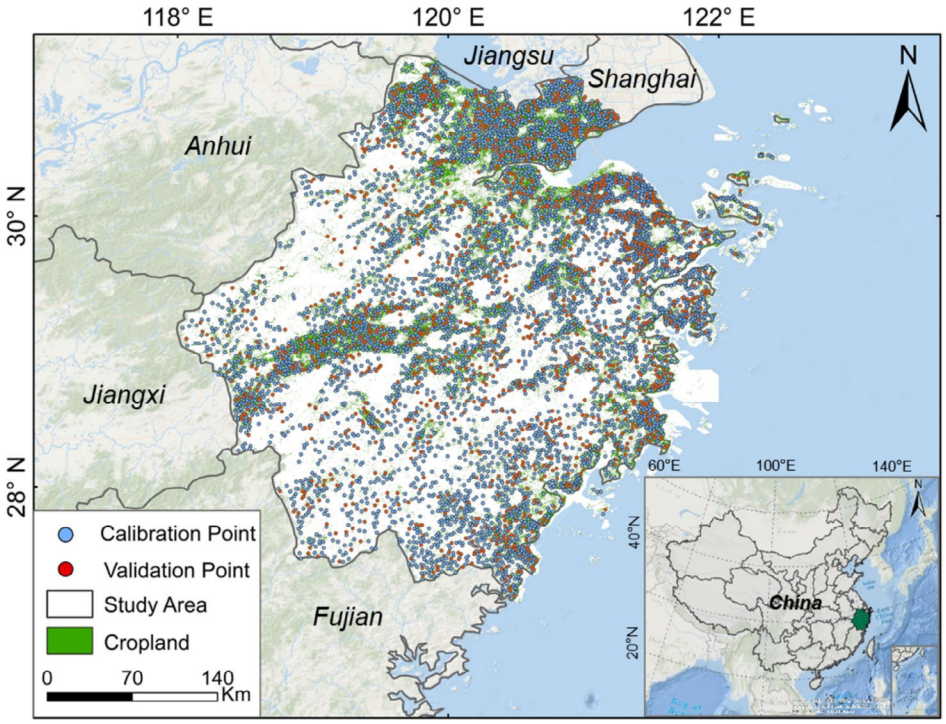

FIGURE 1 | Study area location (right-bottom subplot) and the distribution of soil sampling sites.

diversity supports a variety of 10 distinct soil types and 21 subtypes, such as red soils, yellow soils, purple soils, limestone soils, sandy soils, basic rock soils, mountain meadow soils, marsh soils, coastal saline soils, as well as paddy soils (Shi and Song 2016). Covering approximately 1281 km^2 of cropland, the province sustains productive cropping pattern that include both triple cropping (such as oilseed-rice-rice) and double cropping (with combinations like rice-rice, oilseed-rice and wheat-rice).

Through stratified sampling, a total of 7284 soil samples from the 0 to 20 cm topsoil layer were collected from cropland in 2019, ensuring representation of the major soil types, clay content, and agricultural systems. After all composite soil samples were air-dried and sieved to a particle size of < 2 mm, SOM (in g kg^{-1}) content was determined by the dichromate oxidation–external heating method, and soil pH was determined potentiometrically at a 5:1 water-to-soil ratio. Further details on sampling design and laboratory analyses refer to Zhuo et al. (2022) and Fornasier et al. (2018).

2.2 | Environmental Covariates

This study is based on the SCORPAN model to integrate diverse multi-source data products across Zhejiang Province, including soil, climate, topography, management practices, temporal biology and spectral data to support the analysis of the spatial variability of SOM and pH (Chen et al. 2022; McBratney et al. 2003). Detailed information on the environmental covariates is provided in Table 1. All selected environmental covariates were reprojected to the Mercator Projection coordinate system and uniformly converted to the World Geodetic System 1984 (WGS 1984).

2.2.1 | Topography

Topography plays a critical role in soil formation and development, influencing soil parent materials, hydrothermal conditions, and vegetation cover. In this study, we used 30 m elevation data from the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM). Using SAGA GIS 9.6 software, we calculated various topographic attributes, including elevation (ELE), slope, aspect (ASP), curvature (CUR), topographic wetness index (TWI), convergence index (CI), terrain ruggedness index (TRI), multiresolution index of valley bottom flatness (MRVBF), multiresolution index of ridge top flatness (MRRTF), topographic position index (TPI), analytical hillshading (AH) and stream power index (SPI).

2.2.2 | Climate

Climate plays a fundamental role in soil formation and its spatial distribution by shaping hydrothermal conditions and influencing the synthesis and decomposition of organic materials through factors such as temperature, humidity, evaporation and precipitation. In this study, 1 km resolution data on mean average precipitation (MAP), mean average temperature (MAT), and total potential evapotranspiration (PET) were obtained from the National Earth System Science Data Center (Peng et al. 2019). Solar radiation data (SR) were sourced from the 1 km annual land surface solar radiation dataset for China, spanning from 1960 to 2022, available through the Geographic Data Sharing Infrastructure and the Global Resources Data Cloud. Average annual daytime and nighttime land surface temperatures (LSTD and LSTN) were derived from the MOD11A2.061 data product (Wan et al. 2021), which offers a spatial resolution of 1 km and an 8-day temporal resolution. Additionally, since SOM and pH

TABLE 1 | Environmental covariates used in this study.

Environmental covariate		Resolution	Source
Topography	Elevation, slope, aspect, curvature, topographic wetness index, convergence index, terrain ruggedness index, multiresolution index of valley bottom flatness, multiresolution index of ridge top flatness, topographic position index, analytical hillshading, stream power index	30 m	DEM from SRTM
Climate	Mean average precipitation, mean average temperature, total potential evapotranspiration, solar radiation, daytime land surface temperature, nighttime land surface temperature	1000 m	Peng et al. (2019) Wan et al. (2021) Geographic Data Sharing Infrastructure, global resources data cloud (www.gis5g.com)
	Nitrogen deposition, sulfur deposition	0.25°	Zhou et al. (2023)
Soil	Total nitrogen, bulk density, total nitrogen density, soil organic carbon, coarse fragments, total phosphorus, total phosphorus density, total potassium, total Potassium density, clay, silt, sand, cation exchange capacity, thickness, texture classes	250 m	Liu et al. (2022)
	Depth-to-bedrock	100 m	Yan et al. (2020)
	Soil moisture	1000 m	Song et al. (2021)
	Spatial neighbour information	30 m	Calculated in this study
Temporal information	NDVI-TS*, EVI-TS, B2-TS, B3-TS, B4-TS, B5-TS, B6-TS, B7-TS, B8-TS, B11-TS, B12-TS	30 m	Wang et al. (2025)
Management measures	Cropping pattern	30 m	Extract in this study
	Nitrogen fertilise	5000 m	Yu et al. (2021)

*TS refers to the monthly time-series spectral bands.

can be influenced by acid rain, nitrogen (N) and sulfur (S) deposition data were also included in the analysis (Zhou et al. 2023).

2.2.3 | Soil

The physical properties and chemical composition of the soil parent material are critical in determining soil attributes. In this study, data on total nitrogen (TN), bulk density (BD), total nitrogen density (TND), soil organic carbon (SOC), coarse fragments (CF), total phosphorus (TP), total phosphorus density (TPD), total potassium (TK), total potassium density (TKD), clay content, silt content, sand content, cation exchange capacity (CEC), soil thickness and soil texture classes were obtained from the 250 m high-resolution National Soil Information Grids of China, as published by Liu et al. (2022). Soil moisture (SM) data were sourced from the Daily All-Weather Surface Soil Moisture dataset, with 1 km resolution, from the National Tibetan Plateau Data Center (Song et al. 2021). Depth-to-bedrock data were acquired from the Depth-to-Bedrock map of China, with a spatial resolution of 100 m, developed by Yan et al. (2020). Additionally, soil spatial neighbour information (SSNI) was incorporated into the environmental covariates, with the calculation method detailed in Chen et al. (2024).

2.2.4 | Temporal Spectrum and Vegetation Indices

The Normalised Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) were selected as primary indicators of vegetation cover. Due to the frequent cloudy and rainy weather conditions in Zhejiang Province, time-series optical remote sensing imagery often exhibits data gaps caused by cloud contamination. To mitigate this problem, Wang et al. (2025) developed a data reconstruction model named MSCAE-CNN, which utilised Landsat-8 imagery to reconstruct and impute missing Sentinel-2, thereby enabling the generation of temporally complete remote sensing time series. In this study, we acquired time-series Landsat-8 and Sentinel-2 imagery for 2019, which underwent preprocessing steps including radiometric calibration, atmospheric correction, bidirectional reflectance distribution function (BRDF) correction, terrain correction, resampling, and cloud masking for both datasets. The MSCAE-CNN model was used to fill cloud-induced gaps in Sentinel-2 imagery, generating a temporally continuous, monthly median-composited remote sensing dataset for 2019, hereafter referred to as HLS-MAC. We extracted the monthly time series of NDVI (NDVI-TS) and EVI (EVI-TS) from HLS-MAC for each sampling point, which were used as temporal biophysical variables.

Additionally, previous studies have shown that organic matter in various soil types responds significantly within the 600–800 nm range, while purple soils exhibit notable responses across a broader spectral range from 400 to 2450 nm (Ji et al. 2012). Based on the direct response of SOM to specific remote sensing spectra and the indirect response of pH to certain spectra, we extracted bands from HLS-MAC, including blue (B2), green (B3), red (B4), vegetation red edge (B5/6/7), near-infrared (B8) and shortwave infrared (B11/12). These bands were used to more comprehensively analyse and characterise the spatial variability of soil properties.

2.2.5 | Management Practices

SOM content and pH in cropland are significantly influenced by agricultural practices, such as tillage, crop rotation and fertilisation (Hou et al. 2018; Liu et al. 2023). In this study, we incorporated key management practices, including cropping pattern and nitrogen fertiliser application rates, into the prediction models for SOM and pH.

Nitrogen fertiliser (NF) with a spatial resolution of 5 km was obtained from the National Ecosystem Science Data Center's dataset on historical nitrogen fertiliser use in China from 1952 to 2018 (Yu et al. 2021). The extraction method for cropping pattern (CP) information is as follows: First, monthly NDVI values for each sampling point were extracted from HLS-MAC to generate a comprehensive NDVI time series curve. Then, based on survey data and statistical yearbook information, the primary cropping patterns in Zhejiang Province were determined, including wheat-rice rotation, oilseed-rice rotation, rice-rice rotation, oilseed-rice rotation and other planting patterns. Specific rules were established to classify the NDVI curves of each sampling point using unsupervised classification. The clustering results were identified by observing the variations in the NDVI curves, as well as the positions of the peak and valley values. The classification criteria for NDVI were established as follows: (1) The crop pattern is classified as 'Wheat-Rice' when the NDVI curve shows peak values exceeding 0.5 in May, June and either September or October. (2) If peak values of 0.5 are recorded in February or March, along with a specific threshold in June or July, it is categorised as 'Oilseed-Rice-Rice'. (3) A classification of 'Rice-Rice' is applied when the NDVI curve exceeds a peak value of 0.5 in both June or July and September or October. (4) The NDVI curve is classified as 'Oilseed-Rice' if it surpasses a peak value of 0.5 in February or March and again in July, August, or September. (5) Any instance not satisfying the above criteria is assigned to the 'Other' category. For analytical purposes, the classifications were encoded with values of 1, 2, 3, 4 and 5, respectively.

2.3 | Predictive Models

2.3.1 | Random Forest

Random Forest (RF) is an ensemble learning method that enhances model accuracy and stability by constructing multiple decision trees and aggregating their predictions. Each tree is trained on a random subset of the original data, and at each split, features are randomly selected for decision-making (Liu et al. 2012). This randomness helps mitigate overfitting and fosters greater model diversity. Ultimately, RF generates its final prediction by voting on the outcomes of all trees (for classification tasks) or averaging the results (for regression tasks). The RF model was implemented in Python 3.11. A grid search using GridSearchCV with 10-fold cross-validation and R^2 as the evaluation metric was applied to optimise model parameters. The search covered a wide range of hyperparameters, including tree numbers (50–500 in steps of 50), mtry values (1–137 in steps of 1, plus 'sqrt' and 'log2'), minimum node size (1–15 with varying steps), and maximum depths (10–40 in steps of 10). The final selected configuration included 500 trees, mtry values of 143 and 44, maximum depths of 20 and 25, and a minimum node size of 10 for SOM and pH, respectively.

2.3.2 | Cubist

Cubist is a regression model based on decision rules, which makes predictions by constructing a set of simple linear regression models, each based on specific conditions or decision rules (Kuhn et al. 2012). Cubist divides the data into multiple regions, where data points within each region satisfy certain conditions, and then uses linear regression to make predictions within each region. The model continuously optimises the rules and regression coefficients to minimise prediction error, while balancing the complexity of the rules with the simplicity of the model. To optimise the hyperparameters of the Cubist model, we conducted a grid search with 10-fold cross-validation and R^2 as the evaluation metric. The search covered committee sizes ranging from 10 to 30, neighbour numbers from 1 to 10, extrapolation levels including 0, 0.05 and 0.1, and maximum rule sizes between 500 and 1500. Optimal parameter combinations were selected based on cross-validated performance. For SOM and pH, the final settings included a committee size of 20 and 25, neighbour numbers of 8 and 9, an extrapolation level of 0.05, and a maximum rule size of 1000, respectively.

2.3.3 | XGBoost

XGBoost (Extreme Gradient Boosting) is an efficient machine learning algorithm based on Gradient Boosting, primarily used for regression and classification tasks (Chen 2015). The fundamental principle is to construct a strong learner by integrating multiple weak learners, typically decision trees. During training, XGBoost sequentially trains new trees to correct the errors of previous trees, with each tree optimised based on the residuals of the preceding tree. To mitigate the risk of overfitting, the XGBoost model incorporates L1 (Lasso) and L2 (Ridge) regularisation terms within its loss function, and it also effectively manages missing values by utilising sophisticated splitting strategies to handle incomplete data. The learning rate was tested at 0.01, 0.05, and 0.1, while the number of trees was searched between 100 and 500. The maximum tree depth was searched within the range of 5–20, while the minimum loss reduction for node splitting was tested between 0 and 1. The feature sampling ratio was varied from 0.1 to 1.0, the minimum child weight was tuned between 1 and 5

TABLE 2 | Descriptive statistics for SOM ($g\,kg^{-1}$) and pH in the all, calibration and validation datasets.

Soil	Dataset	Count	Min	Median	Mean	Max	SD	Skewness	Kurtosis
SOM	Total	7284	1.5	29.2	30.58	127	12.63	0.82	1.52
	Calibration	5827	1.5	29.2	30.66	127	12.68	0.84	1.68
	Validation	1457	2.7	28.95	30.39	107	12.50	0.74	1.14
pH	Total	7284	3.47	5.51	5.78	9.12	1.01	0.97	0.44
	Calibration	5827	3.47	5.50	5.78	9.12	1.01	0.98	0.45
	Validation	1457	3.50	5.51	5.78	9.12	0.99	0.94	0.41

to control node division rigorously, and the training sample subsampling rate was explored from 0.6 to 1.0. Optimal hyperparameters were selected through grid search combined with cross-validation. The final configuration included a learning rate of 0.05, 500 trees, a maximum depth of 15, a minimum loss reduction of 0, a feature sampling ratio of 0.8, a minimum child weight of 1, and a subsampling rate of 1.

2.3.4 | ANN

The Artificial Neural Network (ANN) emulates the human brain's information processing to uncover complex nonlinear relationships within data by iteratively optimising its weights through forward and backward propagation, thereby minimising prediction error (Prieto et al. 2016). To mitigate overfitting risk, ANN incorporate dropout regularisation during training and dynamically adjust network structure and parameters, while effective data handling strategies enable them to adapt to incomplete datasets by managing missing values robustly. The optimal hyperparameters for the ANN were obtained through Bayesian optimization. The search space included an initial learning rate ranging from 0.01 to 0.1, the number of hidden layers from 6 to 15, and the number of neurons per layer from 32 to 128. Dropout rates were explored between 0.1 and 0.5, and learning rate decay values ranged from 0 to 0.05. To balance sufficient training with computational efficiency, the number of epochs was fixed at 150. The final model configuration included a learning rate of 0.06, 12 hidden layers with 53 neurons per layer, a dropout rate of 0.22 and a learning rate decay of 0.01, ensuring both stable training and sufficient model capacity.

2.4 | Model Evaluation

The whole soil sample dataset, comprising 7284 samples, was randomly partitioned into calibration and validation sets at a ratio of 4:1, yielding a calibration set of 5827 samples and a validation set of 1457 samples. Model performance was evaluated using the coefficient of determination (R^2) and root mean square error (RMSE). The calculation formulas are as follows:

$$R^2 = 1 - \frac{\sum_i^n \left(\hat{y}_i - y_i\right)^2}{\sum_i^n \left(y_i - \bar{y}\right)^2} \quad (1)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_i^n \left(\hat{y}_i - y_i\right)^2}{n}} \quad (2)$$

where y_i and $\hat{y}_i$ represent the observed values and predicted values for sample i, $\bar{y}$ is the mean of all observations, and n is the total number of samples.

3 | Results

3.1 | Description of the SOM Content and pH

The statistical characteristics of the calibration and validation datasets of SOM content and pH are presented in Table 2. Quartile analysis indicated the absence of extreme outliers, and after applying a natural logarithmic transformation, the overall dataset approximated a normal distribution. The SOM content ranged from 1.5 to $127\,g\,kg^{-1}$, with a mean of $30.58\,g\,kg^{-1}$, a standard deviation of $12.63\,g\,kg^{-1}$, and skewness and kurtosis values of 0.82 and 1.52, respectively. For pH, the values ranged between 3.47 and 9.12, with a mean of 5.78, a standard deviation of 1.01 and skewness and kurtosis values of 0.97 and 0.44, respectively. The comparable standard deviations between the calibration and validation datasets (differing by only 0.12 for SOM and 0.02 for pH) indicate that the random partitioning successfully captured the full range of the original datasets, meeting the criteria for fair model validation.

3.2 | Cropping Pattern Distribution

Based on the established classification rules, NDVI curves corresponding to various cropping patterns were generated (Figure 2a). The variations in the NDVI curves for each category align with the growth trends of the respective crops.

In the oilseed-rice-rice cropping pattern, rapeseed typically enters the bolting stage from mid-February to early March, during which leaf thickness gradually increases and photosynthesis intensifies, leading to the first peak in the NDVI curve. Medium rice is usually sown and transplanted in early May, reaching the booting stage in early to mid-July. During this period, vigorous leaf growth occurs, causing a rise in NDVI values, which correspond to the second peak of the

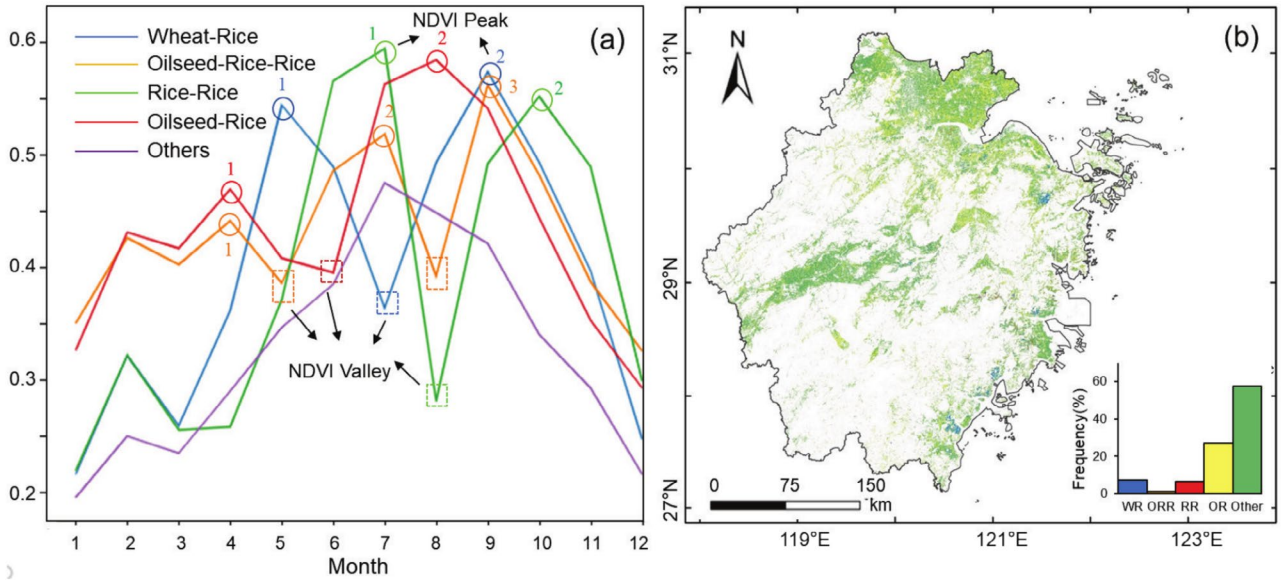

FIGURE 2 | Clustering of NDVI temporal curves (a). The spatial distribution map of the cropping pattern (b). WR represents wheat-rice, ORR represents oilseed-rice-rice, RR represents rice-rice, OR represents oilseed-rice.

NDVI curve. By mid-August, the harvest period for medium rice arrives, causing NDVI values to decline to their lowest levels. Late rice typically reaches the heading stage by mid-September, resulting in a third peak in the NDVI curve, with harvest completed by late October.

In the rice-rice rotation pattern, early rice is typically sown and transplanted between late March and early April. After transplantation, the crop enters the tillering stage, during which NDVI values rise, reaching the first peak of the curve in late June to early July. As the crop progresses into the filling and maturation stages, NDVI values decrease due to leaf senescence and reduced physiological activity, with harvest occurring by late July. Late rice is sown from late July to early August, initially exhibiting slow growth before transplantation, resulting in consistently low NDVI values. From mid-August to late September, as the crop advances through the tillering and jointing stages, NDVI values increase sharply, reaching their second peak at the heading stage. As the crop enters the filling and maturation stages from October to November, NDVI values gradually decline until harvest is completed by the end of October.

In the wheat-rice rotation pattern, wheat is sown between October and November. As temperatures rise, the crop enters the jointing stage in February, characterised by significant growth and an expansion in leaf area, which leads to an increase in NDVI values. During the heading and flowering stages in April and May, vegetation cover reaches its maximum, producing the first peak in NDVI due to intensified photosynthesis. Following maturity, NDVI declines, reaching its lowest point after the wheat harvest in late May. Rice seedlings are transplanted from late May to early June, progressing through the heading and booting stages, with NDVI peaking in August. As rice enters the filling and maturation stages, NDVI values gradually decrease until harvest in October.

In the oilseed-rice rotation pattern, rapeseed is typically sown in October, showing low NDVI values during the seedling stage. With rising temperatures in February and March, rapid growth during the budding stage leads to a significant increase in NDVI, which gradually peaks before declining toward harvest in late April to early May. After rapeseed is harvested, rice seedlings are transplanted from mid-May to June, entering the tillering and jointing stages, which result in a marked increase in NDVI. By mid-July to early August, rice reaches the heading stage and attains its growth peak, with NDVI reaching a second peak before gradually decreasing as the crop matures and approaches harvest.

Based on the pixel-by-pixel NDVI temporal curve clustering, the spatial distribution map of cropping patterns in Zhejiang Province was generated. As shown in Figure 2b, the cropping patterns across the province are influenced by factors such as topography, climate and soil properties, resulting in a diverse range of cropping structures with distinct regional characteristics. Specifically, the 'other cropping pattern', which covers the largest area, primarily represents single-season grain crops, vegetable fields, orchards or tea plantations in central and northern Zhejiang. The double-cropping patterns, including rice-wheat, rapeseed-rice and rice-rice rotations, occupy the second-largest area and are prevalent in the northwestern hilly regions, central hilly and basin areas, southwestern mountainous regions, southeastern coastal areas and eastern island zones. The triple-cropping pattern, mainly represented by the rapeseed-rice-rice rotation, is concentrated in the northern plains of Zhejiang. The spatial variability of cropping patterns aligns closely with local environmental and agricultural conditions.

3.3 | Evaluation of Model Performance

The accuracy evaluation results, presented in Figure 3, indicate that the RF model incorporating temporal information achieved the highest predictive performance, with an R^2 of 0.42 and an RMSE of 9.67 g kg^{-1} for predicting SOM content. This reflects a 0.05 improvement in R^2 and a 0.31 g kg^{-1}

reduction in RMSE compared to the model without temporal information. The accuracy of XGBoost and Cubist is nearly comparable, with XGBoost achieving an R^2 of 0.37 and an RMSE of 10.19 g kg^{-1}, while Cubist yields an R^2 of 0.36 and an RMSE of 10.07 g kg^{-1}. Compared to the baseline model without temporal information, XGBoost shows a 0.01 increase in R^2 and a 0.09 g kg^{-1}reduction in RMSE, while Cubist demonstrates a 0.04 increase in R^2 and a 0.31 g kg^{-1} reduction in RMSE. Although the accuracy of the ANN model incorporating temporal information is the lowest, its predictive performance improved compared to the baseline model without temporal data.

For soil pH prediction, the XGBoost model incorporating temporal data achieved the highest accuracy, with R^2 of 0.61 and RMSE of 0.63, reflecting improvements of 12.96% and 5.94%, respectively, compared to the baseline model. The RF and Cubist models with temporal data showed comparable predictive

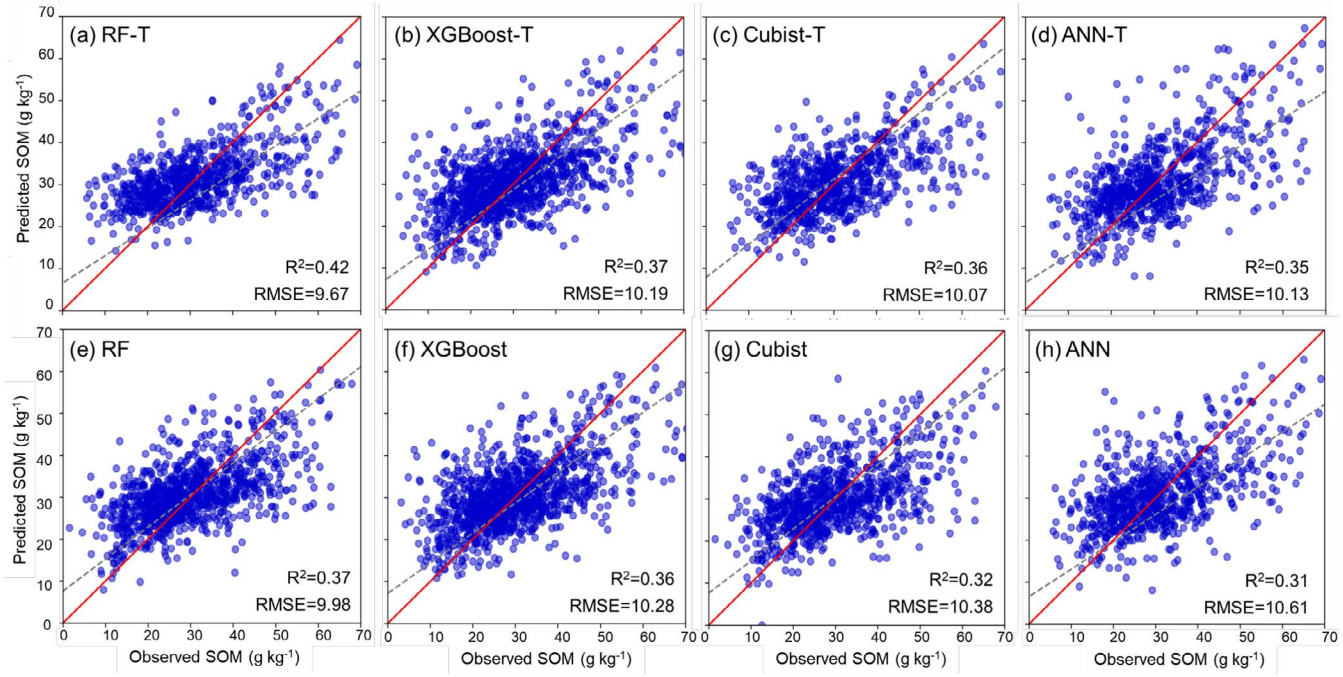

FIGURE 3 | Scatter plots depicting the model performance for validation set on SOM mapping including (a–d) and excluding (e–h) temporal information across four predictive models.

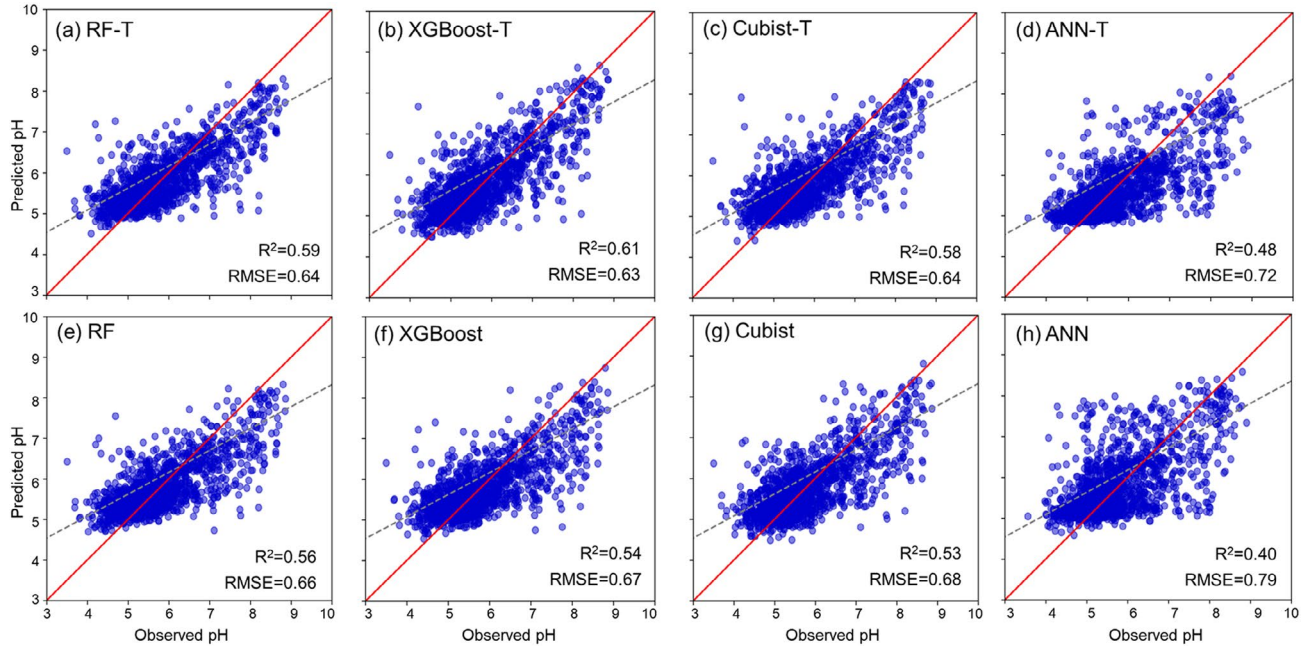

FIGURE 4 | Scatter plots depicting the model performance for the validation set on pH mapping including (a–d) and excluding (e–h) temporal information across four predictive models.

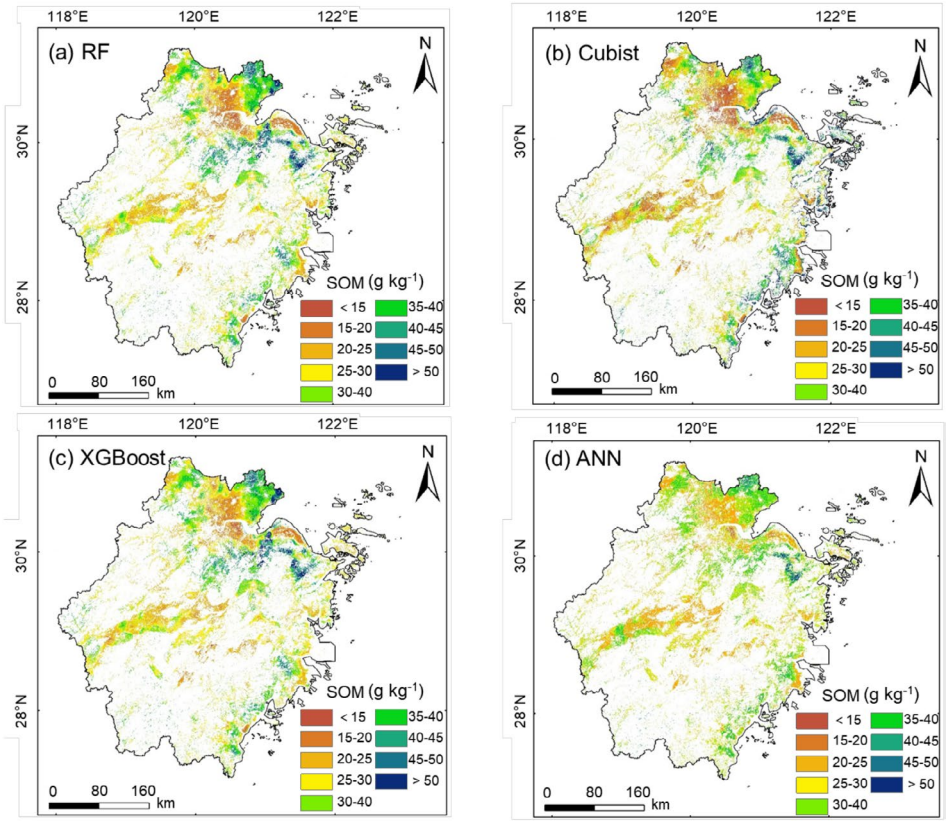

FIGURE 5 | Spatial distribution maps of SOM predicted by RF (a), Cubist (b), XGBoost (c) and ANN (d).

accuracies, with R^2 values of 0.59 and 0.58, and RMSE values of 0.64 for both models. Although the ANN model shows limited performance in predicting soil pH, the inclusion of temporal data resulted in a moderate improvement, achieving an R^2 of 0.48 and an RMSE of 0.72 (Figure 4).

Overall, these results underscore the significance of incorporating temporal information, which captures data across various time intervals. This approach provides deeper insights into soil property dynamics and trends that static analyses may overlook, thereby significantly enhancing the accuracy of soil property predictions.

3.4 | The Spatial Distributions of SOM and pH

The Figures 5 and 6 show the maps of SOM and pH at 30 m resolution. The spatial distribution patterns predicted by the four models demonstrates overall consistency. In the northeastern region, where paddy soils predominate and rice-rice or rice-oilseed rotation patterns are common, SOM content is relatively high. In contrast, the northern Zhejiang Plain and Ning-Shao Plain, which serve as major agricultural hubs, are characterised by rice-wheat, rice-rice or rice-oilseed rotations. These intensive cultivation patterns are associated with lower SOM levels. In the central mountainous areas, cropland is primarily located in hilly regions dominated by red-yellow soils. These soils typically have poor moisture retention, and steep slopes exacerbate soil erosion, limiting SOM accumulation and resulting in moderate to low SOM. Additionally, the saline tidal soils along the eastern coast exhibit comparatively low SOM levels.

The spatial distribution of cropland soil pH in Zhejiang Province exhibits a clear gradient, with higher values in the northeastern region and lower values in the southwestern region. Areas with soil pH exceeding 7, primarily located in the low-altitude eastern coastal zones, are dominated by coastal saline soils. These soils exhibit strong alkaline buffering capacity due to their high concentrations of base cations. Soils with pH values between 6 and 7 are mainly found in the northern Zhejiang Plain, characterised by fluvo-aquic soils and paddy soils, where rice-wheat or rice-oilseed rotations are the prevalent cropping patterns. In contrast, red soils, yellow soils, or paddy soils dominate the relatively high-altitude farmland in central and southwestern Zhejiang. In these areas, soil pH generally ranges from 5 to 6, but in some locations, it drops below 5 or even 4.5, indicating severe acidification.

3.5 | Variable Importance of Environmental Covariates

The contribution of each variable to the predictive model was quantified in the model with the highest prediction accuracy (Figure 7). For SOM prediction using the RF model, soil and temporal information emerged as the most influential factors, contributing 40.73% and 37.57% to the total importance, respectively. Climate factors also played a significant role, contributing 11.41%, while terrain and management practices had similar contributions, at 5.22% and 5.08%, respectively. In soil pH prediction using XGBoost, soil factors were the most influential, contributing 43.06%, followed closely by temporal information at 41.97%. Climate factors accounted for 8.15%, ranking third

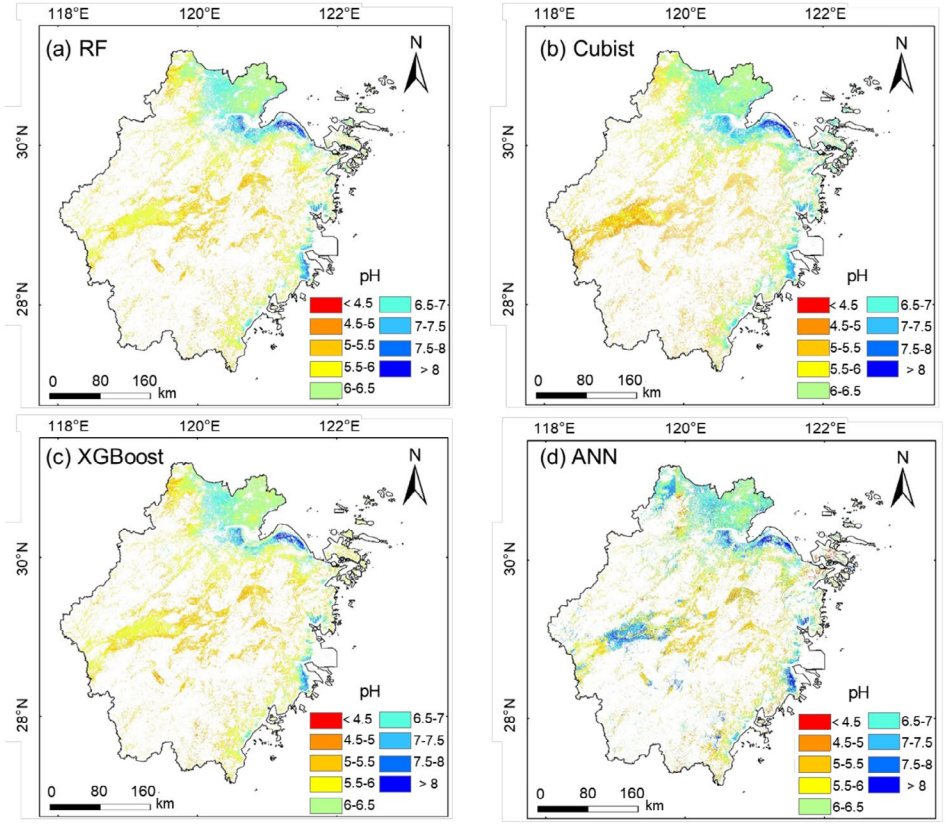

FIGURE 6 | Spatial distribution maps of soil pH predicted by RF (a), Cubist (b), XGboost (c) and ANN (d).

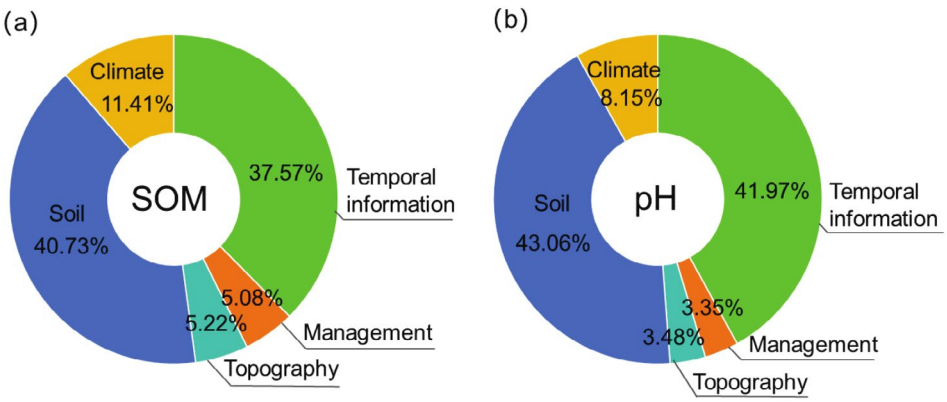

FIGURE 7 | Importance of environmental covariates (by major categories) in spatial predictive models for SOM (a) and pH (b).

in importance, while the contributions from management practices and terrain were relatively small. These results underscore the critical role of temporal information in improving the predictive accuracy of SOM and pH models.

4 | Discussion

4.1 | Relationships Between Environmental Covariates and SOM/pH

According to the contributions of various environmental covariates in the prediction, the top 20 most important variables were identified, and their relationships with pH or SOM were further analysed (Figure 8).

For SOM prediction, soil variables contribute the most, explaining 40.73% of the variance. Notably, SSNI plays a key role, supporting Tobler's First Law of Geography, which states that the soil properties at a given location are influenced by not only the surrounding environmental covariates but also the conditions of neighbouring sites (Chen et al. 2024). TN is another significant contributor to SOM prediction, as nitrogen supports plant growth, increases plant residue return to the soil and facilitates organic matter accumulation (Xu et al. 2021; Zheng et al. 2024). Nitrogen also stimulates

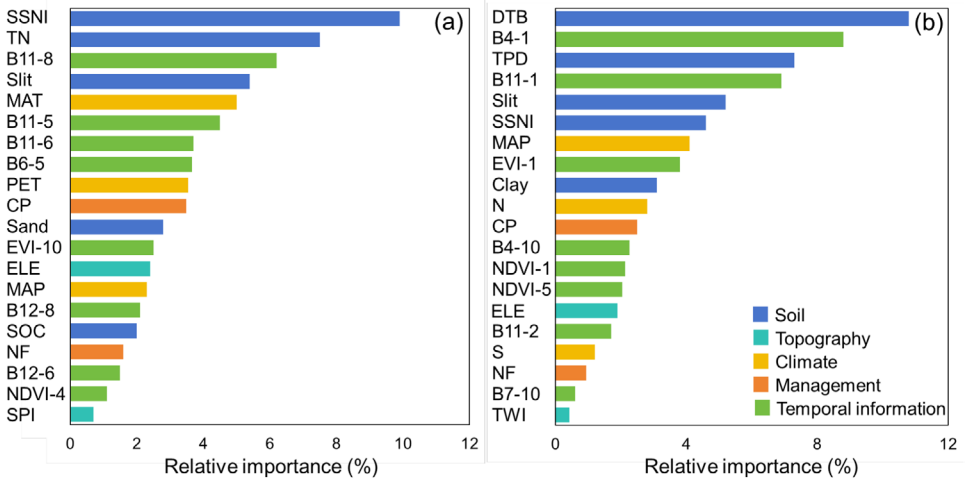

FIGURE 8 | Relative importance of environmental covariates in SOM (a) and pH (b) spatial predictive modelling (SSIN, spatial-neighbour-information; TN, total nitrogen; MAT, mean average temperature; PET, total potential evapotranspiration; CP, cropping pattern; ELE, elevation; DTB, depth-to-bedrock; MAP, mean average precipitation; SOC, soil organic matter; NF, nitrogen fertiliser; SPI, stream power index; TPD, total phosphorus density; N, nitrogen deposition; S, sulfur deposition; TWI, topographic wetness index).

microbial activity, accelerating the decomposition of plant residues, thus maintaining a dynamic equilibrium between SOM decomposition and accumulation (Jin et al. 2024; Xie et al. 2022). Additionally, silt and sand also contribute to modelling SOM, although their importance is relatively low. Temporal spectrum and vegetation indices contribute 37.57% to SOM prediction, with 8 variables among the top 20 most important factors. This is primarily due to the strong absorption of spectral light by polar covalent bonds (such as -COOH, -OH, -COOR and -NH_2) in the visible to shortwave infrared range (1000–2500 nm), which leads to a significant correlation between SOM and reflectance in this spectral range (Chenu et al. 2024; Rasooli et al. 2023; Weng et al. 2022). In this study, B11/12 (shortwave infrared) and B6 (visible and near-infrared) bands in spring and summer were found to be particularly important, consistent with the findings of Liu et al. (2023). Moreover, as plants fix carbon through photosynthesis and provide SOM to the soil through root exudates and the decomposition of plant residues, temporal NDVI and EVI are also key indicators for predicting SOM. Climate factors contribute 11.41% to SOM prediction, as climate conditions determine key parameters such as temperature (MAT) and moisture, which directly affect organic matter accumulation and decomposition rates. Higher temperatures accelerate microbial activity, promoting organic matter decomposition, while increased humidity supports its accumulation. Additionally, factors such as precipitation (MAP), seasonal variation and light intensity significantly influence plant growth and SOM return patterns (Bhattacharyya et al. 2022; Davidson and Janssens 2006; Maslov and Maslova 2021). Topography influences soil properties by controlling local soil moisture and temperature, and is considered a significant factor in determining soil characteristics in specific areas (Dai et al. 2022; Li, Liang, et al. 2021). Among topographic factors, elevation and SPI play a dominant role. In high-altitude areas, low temperatures, slow plant growth and slow organic matter decomposition hinder SOM accumulation (Ma et al. 2024; Zhang and Xu 2023). Meanwhile, areas with higher SPI may experience SOM loss due to erosion processes. Management practices also impact soil nutrient levels. For instance, rice-dryland rotation can enhance soil carbon sequestration, while changes in tillage methods (e.g., deep ploughing, no-till, or ridging) can alter soil structure, aeration and water retention capacity, which in turn affect SOM decomposition and accumulation (Blickensdörfer et al. 2022). Furthermore, different fertilisation methods can influence crop growth and microbial activity, thereby affecting SOM content (Chowdhury et al. 2021; Tian et al. 2017).

The spatial variation of soil pH in Zhejiang Province is influenced by both natural and anthropogenic factors, with natural factors playing a more dominant role. The relationship between soil pH and various environmental covariates is as follows: (1) Soil exert the most significant influence on topsoil pH, contributing 43.06% to its variation. DTB is the most pivotal variable, as it directly influences the mineral composition and calcium carbonate content of the soil, which subsequently determines the soil's capacity to neutralise acids. Notably, TPD, slit and SSNI also play important roles in predicting soil pH. Soils with higher TPD generally contain more phosphate compounds, which can interact with hydrogen or hydroxide ions in the soil, thereby influencing its acidity or alkalinity (Agegnehu et al. 2021; Ng et al. 2022). Additionally, fine soil particles can adsorb ions from the soil solution, disrupting the ion exchange balance and influencing soil pH (Rengasamy et al. 2016). (2) Temporal spectrum and vegetation indices (NDVI and EVI) contribute 41.97% to soil pH prediction, with 8 of the top 20 important variables related to temporal factors. In this study, B4 (red band) appeared particularly important, likely due to its sensitivity to elements such as iron and manganese, which influence soil pH through their interactions with the soil's mineral and chemical composition (Naz et al. 2022). Other soil properties, such as SOM, also affect pH values, with the shortwave infrared band (B11) showing an indirect relationship with soil pH. Additionally, the importance of NDVI-1/5 and EVI-1 highlights the correlation between the crop

growth cycle and soil pH, reflecting the dynamic influence of pH on nutrient availability, microbial activity, root health and physiological processes that change in response to crop demands over time (Das et al. 2022; Fageria and Moreira 2011). For instance, under acidic soil conditions (pH < 5.5), phosphorus readily binds with iron and aluminium to form insoluble compounds, significantly reducing its availability and causing deficiency symptoms during critical growth stages such as tillering and flowering (Osman 2013). Low pH levels increase the solubility of toxic ions such as aluminium and manganese, whose bioavailable forms inhibit cell division within the root apical meristem, thereby restricting root growth, particularly during the seedling and jointing stages (Rahman et al. 2024). Moreover, soil acidification profoundly alters bacterial community composition, diminishes the soil's ability to suppress fungal pathogens and reduces the activity of nitrifying bacteria, which limiting nitrogen availability and potentially delaying crop maturation (Li et al. 2023; Beeckman et al. 2018). In contrast, soils with a pH of 6.0–7.0 support optimal nitrogen mineralization, ensuring a steady nitrogen supply from seedling to grain filling (Brady et al. 2008; Kennedy 2022). Slightly alkaline conditions (pH 6.5–7.5) also stimulate root exudation of organic acids and enhance beneficial microbial activity in the rhizosphere, improving nutrient availability, uptake efficiency, and overall soil resource utilisation by crops (Cai et al. 2023). (3) Climate factors contribute 8.15% to soil pH variation, with precipitation, total nitrogen deposition, and total sulfur deposition being the most influential. Under the high-temperature and high-humidity conditions of Zhejiang Province, precipitation significantly leaches soil silica and base cations, leading to the accumulation of iron, manganese and aluminium oxides. This reduces base saturation, decreases acid buffering capacity and increases H^+ saturation, accelerating soil acidification (Wang, Tao, et al. 2023). Additionally, human activities, such as the burning of fuels that produce SO_2 and NO_X, along with acidic compounds from natural processes, contribute to acid deposition in the form of gases or particulate matter (Yu et al. 2021). These compounds enter the soil through dry deposition or acid rain, causing soil acidification. Acidic anions such as sulfate and nitrate, present in acid rain, promote leaching of base cations and increase hydrogen ion concentration, further lowering soil pH (Dong et al. 2022; Lie et al. 2023). (4) Topography holds a relative importance of 3.48%, with ELE and TWI being the primary contributors. Local climate and topography are inherently connected, as elevation influences temperature and precipitation patterns, making altitude a potential factor in predicting soil pH. Topographic variations significantly affect water flow and material transport. In areas with steeper slopes, water movement is faster, potentially reducing the accumulation of acidic substances, but it may also accelerate soil acidification due to water and soil erosion. On the other hand, flatter or gentler slopes facilitate water retention, promoting the infiltration of acidic substances into the soil and increasing the risk of acidification. (5) Management practices contribute 3.35% to soil pH changes, primarily driven by crop rotation, while fertilisation shows a smaller effect, potentially due to the low resolution of the data (5000 m). Crop rotation, which increases crop diversity and adjusts fertilisation practices, helps stabilise soil pH. However, the use of ammonium-based nitrogen fertilisers (e.g., ammonium bicarbonate, urea) leads to the nitrification of NH_4^+ into NO_3^-, releasing H^+ ions and increasing soil acidity, thus exacerbating soil acidification. Additionally, ammonium-based fertilisers can replace and leach base cations like calcium, magnesium and potassium, further depleting the soil's alkaline components and accelerating acidification (Hao et al. 2022; Xu et al. 2024).

4.2 | Comparison to Other Global and National Products

The accuracy of our SOM and pH predictions was compared to the National Soil Information Grids of China (NSIGC) and SoilGrids 2.0 by evaluating the agreement between measured and predicted SOM and pH values using a consistent validation dataset in this study. As shown in Figure 9–10, the SOM map derived from NSIGC demonstrates an R^2 of 0 and an RMSE of 16.78 g kg^{-1}, while the corresponding map from Soil Grid 2 reports an R^2 of 0 and an RMSE of 33.66 g kg^{-1}. For pH, the map from NSIGC yields an R^2 of 0.20 and an RMSE of 0.89, whereas the pH map from Soil Grid 2.0 has an R^2 of 0.11 and an RMSE of 0.90. In contrast, our products exhibit a notable improvement in accuracy for both SOM and pH, with an R^2 of 0.42 and an RMSE of 9.67 g kg^{-1} for SOM, and an R^2 of 0.61 and an RMSE of 0.63 for pH.

We calculated the spatial distribution differences by subtracting our maps from those of NSIGC and SoilGrids 2.0 (Figures 9 and 10). The difference frequency plot with NSIGC reveals that nearly 70% of the SOM predictions in this study are overestimated, especially in the northeastern and western regions of Zhejiang Province, while the northern plains are characterised by an underestimation of SOM. For soil pH, approximately 80% of the areas show overestimation, with the western and northern parts of the province being most affected, while the northeastern and southeastern coastal areas tend to have underestimates. Regarding SoilGrids 2.0, SOM predictions are generally underestimated in the northern Zhejiang Plain and Ning-Shao Plain, while overestimation is more prevalent in the western hilly agricultural zones. For pH, the overestimation and underestimation are more evenly distributed, with the greatest underestimation observed in the northeastern, southeastern coastal, and central regions.

The poor prediction accuracy of NSIGC and SoilGrids 2.0 in Zhejiang Province is primarily attributed to the sparse distribution of sampling points used during the development of these datasets. This limited sampling density restricts the ability to capture the spatial heterogeneity of soil properties, thus compromising the reliability of the predictions. To improve soil mapping accuracy at the regional scale, it is crucial to prioritise the expansion of high-quality soil data, which will provide a more comprehensive foundation for accurately representing the spatial distribution of soil properties. Additionally, adopting region-specific modelling approaches tailored to local conditions is essential for enhancing and optimising soil property maps at both national and global scales. For the regions with limited soil data, updating regional soil maps by model averaging of national and global soil products would provide a suitable solution (Chen et al. 2020).

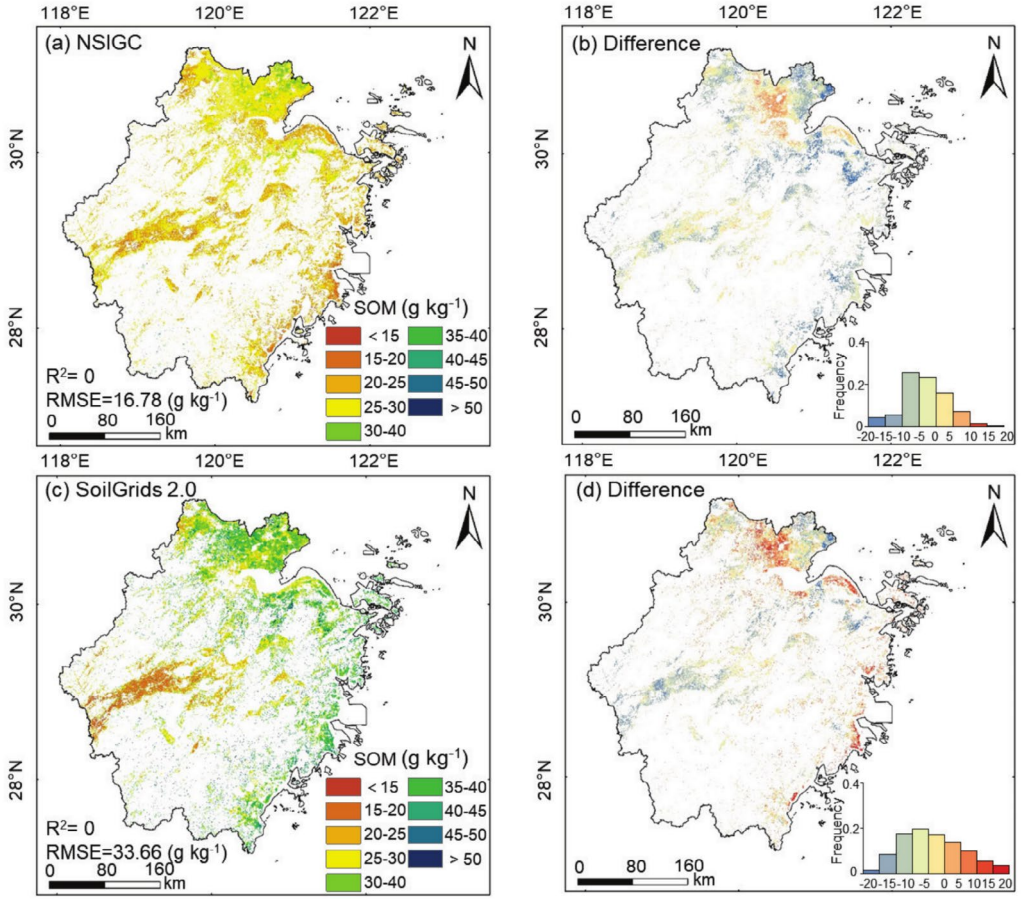

FIGURE 9 | Spatial distribution of SOM derived from SoilGrids 2.0 (a) and National Soil Information Grids of China (NSIGC, c), and their differences compared to the SOM spatial distribution generated in this study (b, d).

4.3 | Applicability of Temporal Information

As shown in Figures 3 and 4, incorporating temporal features as covariates improved the model's predictive accuracy for SOM and pH. This enhancement is likely due to the distinct temporal dynamics and sensitivities of SOM and pH to environmental covariates across different time scales.

The variation in SOM and pH is a continuous, dynamic process driven by factors such as vegetation growth, agricultural activities and seasonal changes, which lead to distinct temporal patterns in the accumulation and transformation of SOM, as well as fluctuations in pH (Bhattacharyya et al. 2022; Zhang, Hou, et al. 2022). Incorporating temporal spectral data, such as NDVI and EVI, can effectively capture the dynamic relationship between plant growth and soil properties. By integrating this temporal information into the model, its ability to detect changes in soil physical and chemical properties is enhanced, allowing for better assessment of soil conditions and their response to environmental changes. Furthermore, the use of temporal time series data increases the volume of available information, enabling the model to train on a more extensive dataset, identify more patterns and trends, and ultimately improve its generalisation capacity.

However, the inclusion of temporal information also presents certain challenges. Firstly, temporal data often require substantial computational resources and processing time, especially with large datasets. As the number of time steps increases, the volume of data that the model must handle grows significantly, which can lead to reduced computational efficiency and longer processing times (Zhang et al. 2023).

4.4 | Limitations and Perspectives

This study utilises environmental covariates, such as topography, vegetation, and climate, and further incorporates temporal and management factors to enhance the prediction of SOM content and pH. Given the complexity of the cropping patterns in Zhejiang Province, which include rotations involving orchards, fruit tree cultivation, medicinal plants, and vegetables, categorising management practices based solely on monthly time-series data and corresponding NDVI phenological curves proves to be challenging. As a result, this study focused on extracting rotation patterns for the major crops. Future research should aim to develop high-quality sample sets and apply supervised classification methods, such as convolutional neural networks (CNN), to achieve precise extraction of crop types. This will enable a more detailed categorization of agricultural cropping patterns and ultimately improve the predictive accuracy of SOM content and pH.

Furthermore, machine learning (ML) models such as random forest (RF) and XGBoost have inherent limitations when it

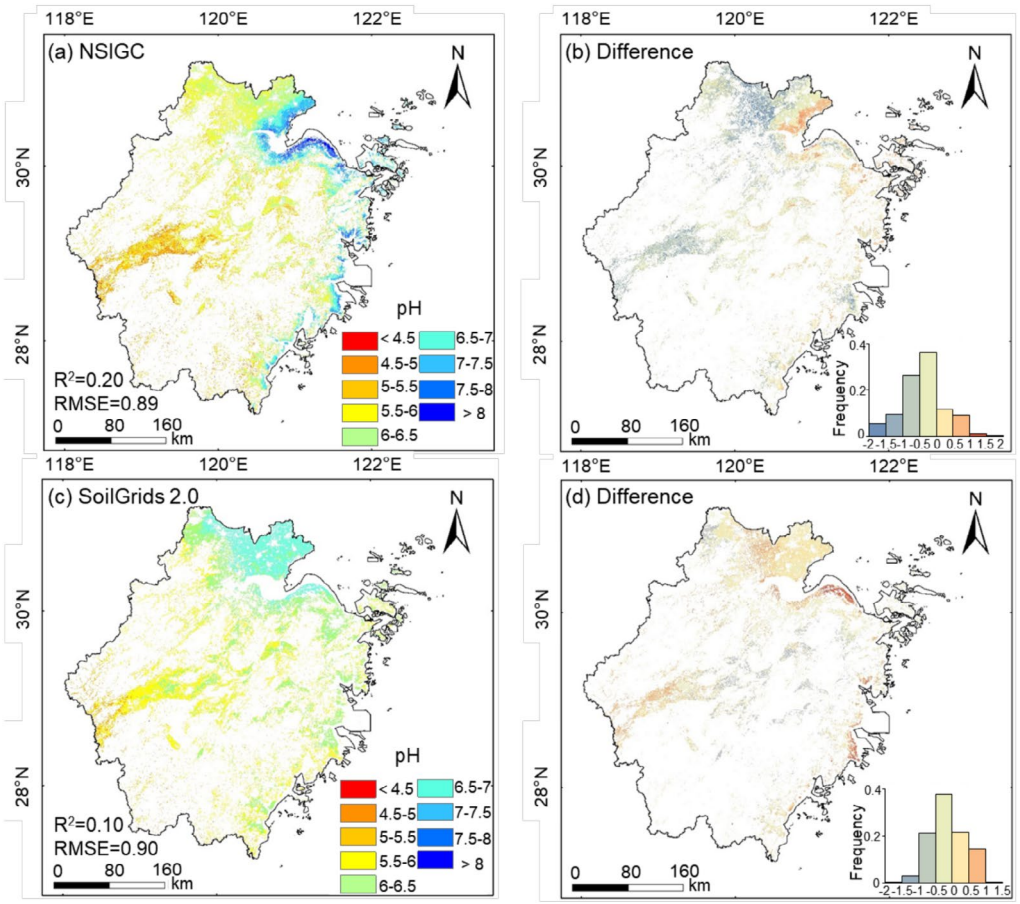

FIGURE 10 | Spatial distribution of pH derived from SoilGrids 2.0 (a) and National Soil Information Grids of China (NSIGC, c), and their differences compared to the pH spatial distribution generated in this study (b, d).

comes to modelling soil properties. In this study, the prediction accuracy for SOM did not exceed an R^2 of 0.5, indicating that current models may struggle to fully capture the spatial heterogeneity of soil and the complex interactions between environmental factors, particularly in relation to the nonlinear nature of SOM. Deep learning techniques, such as CNN and Long Short-Term Memory (LSTM), are expected to emerge as dominant approaches in soil property prediction (Meng et al. 2022; Wang et al. 2022). Compared to traditional ML methods, deep learning models can automatically extract features from large-scale, multidimensional datasets and adaptively adjust to capture complex spatial and temporal patterns in soil, which can significantly improve prediction accuracy. Additionally, integrating techniques like transfer learning and Generative Adversarial Networks (GANs) shows promise in addressing data scarcity issues and enhancing the model's generalisation capability across different regions or under varying conditions.

5 | Conclusion

In this study, we employed four models, including RF, Cubist, ANN and XGBoost, along with the incorporation of temporal data to demonstrate their effectiveness in enhancing the prediction of SOM and pH. The key findings are summarised as follows:

1. The RF model outperformed Cubist, ANN, and XGBoost in predicting SOM, achieving an R^2 of 0.42 and an RMSE of 9.67 g kg^{-1}. For soil pH, the XGBoost model performed best, with an R^2 of 0.61 and an RMSE of 0.63. These results indicate that the suitability of models depends on the type of soil attribute and the nature of the data, with each model having distinct strengths for specific predictive tasks.
2. The inclusion of temporal data significantly enhanced model performance for both SOM and pH, highlighting the importance of temporal dynamics in improving the accuracy of soil property predictions.
3. By extracting planting patterns from temporal data, additional information on agricultural management practices was incorporated into the models. This integration strengthened the prediction framework, enabling it to better capture the effects of different agronomic practices on soil properties, thus providing a more comprehensive understanding of the factors influencing soil characteristics.

Acknowledgements

This work is supported by the National Key Research and Development Program of China (2023YFD1900102) and the National Natural Science Foundation of China (42201054).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The produced maps for SOM and pH from different models are available from https://doi.org/10.5281/zenodo.15099076.

References

Agegnehu, G., T. Amede, T. Erkossa, et al. 2021. "Extent and Management of Acid Soils for Sustainable Crop Production System in the Tropical Agroecosystems: A Review." *Acta Agriculturae Scandinavica Section B Soil and Plant Science* 71, no. 9: 852–869.

Babaeian, E., M. Sadeghi, S. B. Jones, C. Montzka, H. Vereecken, and M. Tuller. 2019. "Ground, Proximal, and Satellite Remote Sensing of Soil Moisture." *Reviews of Geophysics* 57, no. 2: 530–616.

Beeckman, F., H. Motte, and T. Beeckman. 2018. "Nitrification in Agricultural Soils: Impact, Actors and Mitigation." *Current Opinion in Biotechnology* 50: 166–173.

Bhattacharyya, S., G. Ros, K. Furtak, H. Iqbal, and R. Parra-Saldívar. 2022. "Soil Carbon Sequestration–An Interplay Between Soil Microbial Community and Soil Organic Matter Dynamics." *Science of the Total Environment* 815: 152928.

Blickensdörfer, L., M. Schwieder, D. Pflugmacher, C. Nendel, S. Erasmi, and P. Hostert. 2022. "Mapping of Crop Types and Crop Sequences With Combined Time Series of Sentinel-1, Sentinel-2 and Landsat 8 Data for Germany." *Remote Sensing of Environment* 269: 112831.

Brady, N. C., R. R. Weil, and R. R. Weil. 2008. *The Nature and Properties of Soils, 13: 662–710*. Prentice Hall.

Cai, A., S. Tang, M. A. Waqas, et al. 2023. "Magnitude, Direction, and Drivers of Rhizosphere Effect on Soil Nitrogen and Phosphorus in Global Agroecosystem." *International Soil and Water Conservation Research* 11, no. 3: 482–493.

Chen, S., D. Arrouays, V. L. Mulder, et al. 2022. "Digital Mapping of GlobalSoilMap Soil Properties at a Broad Scale: A Review." *Geoderma* 409: 115567.

Chen, S., V. L. Mulder, G. B. Heuvelink, et al. 2020. "Model Averaging for Mapping Topsoil Organic Carbon in France." *Geoderma* 366: 114237.

Chen, T. 2015. "Xgboost: Extreme Gradient Boosting. R Package Version 0.4-2, 1(4)."

Chen, Z., Q. Shuai, Z. Shi, D. Arrouays, A. C. Richer-de-Forges, and S. Chen. 2023. "National-Scale Mapping of Soil Organic Carbon Stock in France: New Insights and Lessons Learned by Direct and Indirect Approaches." *Soil & Environmental Health* 1, no. 4: 100049.

Chen, Z., Z. Wang, X. Wang, Z. Shi, and S. Chen. 2024. "Including Soil Spatial Neighbor Information for Digital Soil Mapping." *Geoderma* 451: 117072.

Chenu, C., C. Rumpel, C. Védère, and P. Barré. 2024. *Methods for Studying Soil Organic Matter: Nature, Dynamics, Spatial Accessibility, and Interactions With Minerals*, 369–406. Elsevier.

Chowdhury, S., N. Bolan, M. Farrell, et al. 2021. "Role of Cultural and Nutrient Management Practices in Carbon Sequestration in Agricultural Soil." *Advances in Agronomy* 166: 131–196.

Dai, L., J. Ge, L. Wang, et al. 2022. "Influence of Soil Properties, Topography, and Land Cover on Soil Organic Carbon and Total Nitrogen Concentration: A Case Study in Qinghai-Tibet Plateau Based on Random Forest Regression and Structural Equation Modeling." *Science of the Total Environment* 821: 153440.

Dai, L., J. Xue, R. Lu, et al. 2024. "In-Situ Prediction of Soil Organic Carbon Contents in Wheat-Rice Rotation Fields via Visible Near-Infrared Spectroscopy." *Soil & Environmental Health* 2, no. 4: 100113.

Das, P., K. Singh, G. Nagpure, et al. 2022. "Plant-Soil-Microbes: A Tripartite Interaction for Nutrient Acquisition and Better Plant Growth for Sustainable Agricultural Practices." *Environmental Research* 214: 113821.

Das, S., P. Kim, M. Nie, and A. Chabbi. 2024. "Soil Organic Matter in the Anthropocene: Role in Climate Change Mitigation, Carbon Sequestration, and Food Security." *Agriculture, Ecosystem and Environment* 375: 109180.

Davidson, E. A., and I. A. Janssens. 2006. "Temperature Sensitivity of Soil Carbon Decomposition and Feedbacks to Climate Change." *Nature* 440, no. 7081: 165–173.

Dong, Y., J. Yang, X. Zhao, et al. 2022. "Soil Acidification and Loss of Base Cations in a Subtropical Agricultural Watershed." *Science of the Total Environment* 827: 154338.

Fageria, N., and A. Moreira. 2011. "The Role of Mineral Nutrition on Root Growth of Crop Plants." *Advances in Agronomy* 110: 251–331.

Fornasier, E., F. Fornasier, and V. Di Marco. 2018. "Spectrophotometric Methods for the Measurement of Soil pH: A Reappraisal." *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 204: 113–118.

Gao, H., J. Gong, T. Ye, M. Maier, and J. Liu. 2024. "Constructing Cropland Ecological Stability Assessment Method Based on Disturbance-Resistance-Response Processes and Classifying Cropland Ecological Types." *Science of the Total Environment* 930: 172673.

Guo, A., W. Yue, J. Yang, et al. 2023. "Cropland Abandonment in China: Patterns, Drivers, and Implications for Food Security." *Journal of Cleaner Production* 418: 138154.

Guo, B., Y. Liu, J. Fan, et al. 2023. "The Salinization Process and Its Response to the Combined Processes of Climate Change–Human Activity in the Yellow River Delta Between 1984 and 2022." *Catena* 231: 107301.

Guy, P., R. Sibly, S. M. Smart, M. Tibbett, and B. J. Pickles. 2022. "Mycorrhizal Type of Woody Plants Influences Understory Species Richness in British Broadleaved Woodlands." *New Phytologist* 235, no. 5: 2046–2053.

Hao, T., X. Liu, Q. Zhu, et al. 2022. "Quantifying Drivers of Soil Acidification in Three Chinese Cropping Systems." *Soil and Tillage Research* 215: 105230.

He, X., L. Yang, A. Li, et al. 2021. "Soil Organic Carbon Prediction Using Phenological Parameters and Remote Sensing Variables Generated From Sentinel-2 Images." *Catena* 205: 105442.

Helfenstein, A., V. L. Mulder, G. B. Heuvelink, and M. J. Hack-ten Broeke. 2024. "Three-Dimensional Space and Time Mapping Reveals Soil Organic Matter Decreases Across Anthropogenic Landscapes in The Netherlands." *Communications Earth & Environment* 5, no. 1: 130.

Hou, P., C. Chien, Y. Chiang-Hsieh, et al. 2018. "Paddy-Upland Rotation for Sustainable Agriculture With Regards to Diverse Soil Microbial Community." *Scientific Reports* 8, no. 1: 7966.

Hu, B., H. Ni, M. Xie, et al. 2023. "Mapping Soil Organic Matter and Identifying Potential Controls in the Farmland of Southern China: Integration of Multi-Source Data, Machine Learning and Geostatistics." *Land Degradation & Development* 34, no. 17: 5468–5485.

Hu, B., M. Xie, Y. Zhou, et al. 2024. "A High-Resolution Map of Soil Organic Carbon in Cropland of Southern China." *Catena* 237: 107813.

Huang, H., L. Yang, L. Zhang, et al. 2022. "A Review on Digital Mapping of Soil Carbon in Cropland: Progress, Challenge, and Prospect." *Environmental Research Letters* 17, no. 12: 123004.

Ji, W., Z. Shi, Q. Zhou, and L. Zhou. 2012. "VIS-NIR Reflectance Spectroscopy of the Organic Matter in Several Types of Soils." *Journal of Infrared and Millimeter Waves* 31, no. 3: 277–282.

Jin, Q., C. Wang, J. Sardans, et al. 2022. "Effect of Soil Degradation on the Carbon Concentration and Retention of Nitrogen and Phosphorus Across Chinese Rice Paddy Fields." *Catena* 209: 105810.

Jin, Y., Y. Yuan, Z. Liu, S. Gai, K. Cheng, and F. Yang. 2024. "Effect of Humic Substances on Nitrogen Cycling in Soil-Plant Ecosystems: Advances, Issues, and Future Perspectives." *Journal of Environmental Management* 351: 119738.

Kennedy, O. 2022. "Soil pH and Its Impact on Nutrient Availability and Crop Growth." *International Journal of Geography, Geology and Environment* 4: 236–238.

Kuhn, M., S. Weston, C. Keefer, and N. Coulter. 2012. "Cubist Models for Regression." R Package Vignette R Package Version 0.0, 18: 480.

Li, D., X. Min, J. Xu, J. Xue, and Z. Shi. 2022. "Assessment of Three Gridded Satellite-Based Precipitation Products and Their Performance Variabilities During Typhoons Over Zhejiang, Southeastern China." *Journal of Hydrology* 610: 127985.

Li, S., L. Xu, Y. Jing, H. Yin, X. Li, and X. Guan. 2021. "High-Quality Vegetation Index Product Generation: A Review of NDVI Time Series Reconstruction Techniques." *International Journal of Applied Earth Observation and Geoinformation* 105: 102640.

Li, T., J. Liang, X. Chen, et al. 2021. "The Interacting Roles and Relative Importance of Climate, Topography, Soil Properties and Mineralogical Composition on Soil Potassium Variations at a National Scale in China." *Catena* 196: 104875.

Li, X., D. Chen, V. J. Carrión, et al. 2023. "Acidification Suppresses the Natural Capacity of Soil Microbiome to Fight Pathogenic Fusarium Infections." *Nature Communications* 14, no. 1: 5090.

Li, X., Y. Li, B. Wang, Y. Sun, G. Cui, and Z. Liang. 2022. "Analysis of Spatial-Temporal Variation of the Saline-Sodic Soil in the West of Jilin Province From 1989 to 2019 and Influencing Factors." *Catena* 217: 106492.

Lie, Z., W. Huang, G. Zhou, et al. 2023. "Acidity of Soil and Water Decreases in Acid-Sensitive Forests of Tropical China." *Environmental Science & Technology* 57, no. 30: 11075–11083.

Liu, F., H. Wu, Y. Zhao, et al. 2022. "Mapping High Resolution National Soil Information Grids of China." *Science Bulletin* 67, no. 3: 328–340.

Liu, Y., S. Chen, Q. Yu, et al. 2023. "Improving Digital Mapping of Soil Organic Matter in Cropland by Incorporating Crop Rotation." *Geoderma* 438: 116620.

Liu, Y., Y. Wang, and J. Zhang. 2012. *New Machine Learning Algorithm: Random Forest*, 246–252. Springer.

Ma, T., Y. Zhan, W. Chen, et al. 2024. "Microbial Traits Drive Soil Priming Effect in Response to Nitrogen Addition Along an Alpine Forest Elevation Gradient." *Science of the Total Environment* 907: 167970.

Ma, Y., D. Woolf, M. Fan, L. Qiao, R. Li, and J. Lehmann. 2023. "Global Crop Production Increase by Soil Organic Carbon." *Nature Geoscience* 16, no. 12: 1159–1165.

Mao, M. 2019. "Temporal and Spatial Variation of Cloud Cover and Its Influencing Factors in Zhejiang Province." 2019 China Environmental Science Society Science and Technology Annual Conference Proceedings, Xian Shanxi Province. 6.

Maslov, M. N., and O. A. Maslova. 2021. "Nitrogen Limitation of Microbial Activity in Alpine Tundra Soils Along an Environmental Gradient: Intra-Seasonal Variations and Effect of Rising Temperature." *Soil Biology and Biochemistry* 156: 108234.

McBratney, A., M. Santos, and B. Minasny. 2003. "On Digital Soil Mapping." *Geoderma* 117, no. 1–2: 3–52.

Meng, X., Y. Bao, Y. Wang, X. Zhang, and H. Liu. 2022. "An Advanced Soil Organic Carbon Content Prediction Model via Fused Temporal-Spatial-Spectral (TSS) Information Based on Machine Learning and Deep Learning Algorithms." *Remote Sensing of Environment* 280: 113166.

Naz, M., Z. Dai, S. Hussain, et al. 2022. "The Soil pH and Heavy Metals Revealed Their Impact on Soil Microbial Community." *Journal of Environmental Management* 321: 115770.

Ng, J., O. Ahmed, M. Jalloh, et al. 2022. "Soil Nutrient Retention and pH Buffering Capacity Are Enhanced by Calciprill and Sodium Silicate." *Agronomy* 12, no. 1: 219.

Nussbaum, M., S. Zimmermann, L. Walthert, and A. Baltensweiler. 2023. "Benefits of Hierarchical Predictions for Digital Soil Mapping—An Approach to Map Bimodal Soil pH." *Geoderma* 437: 116579.

Osman, K. T. 2013. "Plant Nutrients and Soil Fertility Management." *Soils: Principles, Properties and Management* 10 : 129–159.

Peng, S., Y. Ding, W. Liu, and Z. Li. 2019. "1 Km Monthly Temperature and Precipitation Dataset for China From 1901 to 2017." *Earth System Science Data* 11, no. 4: 1931–1946.

Pôças, I., A. Calera, I. Campos, and M. Cunha. 2020. "Remote Sensing for Estimating and Mapping Single and Basal Crop Coefficientes: A Review on Spectral Vegetation Indices Approaches." *Agricultural Water Management* 233: 106081.

Prieto, A., B. Prieto, E. Ortigosa, et al. 2016. "Neural Networks: An Overview of Early Research, Current Frameworks and New Challenges." *Neurocomputing* 214: 242–268.

Rahman, S. U., J. C. Han, M. Ahmad, et al. 2024. "Aluminum Phytotoxicity in Acidic Environments: A Comprehensive Review of Plant Tolerance and Adaptation Strategies." *Ecotoxicology and Environmental Safety* 269: 115791.

Rasooli, N., M. Farpoor, M. Mahmoodabadi, and I. Esfandiarpour-Boroujeni. 2023. "Vis-NIR Spectroscopy as an Eco-Friendly Method for Monitoring Pedoenvironmental Variations and Pedological Assessments in Lut Watershed, Central Iran." *Soil and Tillage Research* 233: 105808.

Rengasamy, P., E. Tavakkoli, and G. McDonald. 2016. "Exchangeable Cations and Clay Dispersion: Net Dispersive Charge, a New Concept for Dispersive Soil." *European Journal of Soil Science* 67, no. 5: 659–665.

Rossel, R. V., T. Behrens, E. Ben-Dor, et al. 2016. "A Global Spectral Library to Characterize the World's Soil." *Earth-Science Reviews* 155: 198–230.

Shafizadeh-Moghadam, H., F. Minaei, H. Talebi-khiyavi, T. Xu, and M. Homaee. 2022. "Synergetic Use of Multi-Temporal Sentinel-1, Sentinel-2, NDVI, and Topographic Factors for Estimating Soil Organic Carbon." *Catena* 212: 106077.

Shi, J., and G. Song. 2016. "Soil Type Database of China: A Nationwide Soil Dataset Based on the Second National Soil Survey." *China Science Data* 1, no. 1: 1–12.

Song, P., Y. Zhang, P. Yao, and T. Zhao. 2021. "Dally All-Weather 2. Surface Soil Moisture Data Set With 1 km Resolution in China (2003-2019)." National Tibetan Plateau/Third Pole Environment Data Center.

Temme, A., J. Schoorl, and W. Van der Meij. 2022. "New Geomorphometric Variables for Non-Continuous Hillslopes–Assessing the Value for Digital Soil Mapping." *Geoderma* 418: 115848.

Teng, H., Z. Shi, Z. Ma, and Y. Li. 2014. "Estimating Spatially Downscaled Rainfall by Regression Kriging Using TRMM Precipitation and Elevation in Zhejiang Province, Southeast China." *International Journal of Remote Sensing* 35, no. 22: 7775–7794.

Tian, J., Y. Lou, Y. Gao, et al. 2017. "Response of Soil Organic Matter Fractions and Composition of Microbial Community to Long-Term

Organic and Mineral Fertilization." *Biology and Fertility of Soils* 53: 523–532.

Wan, Z., S. Hook, and G. Hulley. 2021. "MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V061[DS]." NASA EOSDIS Land Processes Distributed Active Archive Center, 2021.

Wang, J., J. Zhen, W. Hu, et al. 2023. "Remote Sensing of Soil Degradation: Progress and Perspective." *International Soil and Water Conservation Research* 11, no. 3: 429–454.

Wang, S., K. Guan, C. Zhang, et al. 2022. "Using Soil Library Hyperspectral Reflectance and Machine Learning to Predict Soil Organic Carbon: Assessing Potential of Airborne and Spaceborne Optical Soil Sensing." *Remote Sensing of Environment* 271: 112914.

Wang, X., S. Chen, C. Zhou, S. Duan, and Z. Shi. 2025. "Cross-Sensor Data Reconstruction for Optical Remote Sensing Gap-Filling With Attention-Enhanced Multi-Scale Fusion Network." *International Journal of Applied Earth Observation and Geoinformation* 140: 104571.

Wang, Z., T. Tao, H. Wang, et al. 2023. "Forms of Nitrogen Inputs Regulate the Intensity of Soil Acidification." *Global Change Biology* 29, no. 14: 4044–4055.

Weng, Z., J. Lehmann, L. Van Zwieten, et al. 2022. "Probing the Nature of Soil Organic Matter." *Critical Reviews in Environmental Science and Technology* 52, no. 22: 4072–4093.

Wu, B., M. Zhang, H. Zeng, et al. 2023. "Challenges and Opportunities in Remote Sensing-Based Crop Monitoring: A Review." *National Science Review* 10, no. 4: nwac290.

Wu, Z., X. Sun, Y. Sun, J. Yan, Y. Zhao, and J. Chen. 2022. "Soil Acidification and Factors Controlling Topsoil pH Shift of Cropland in Central China From 2008 to 2018." *Geoderma* 408: 115586.

Xie, Z., Z. Yu, Y. Li, et al. 2022. "Soil Microbial Metabolism on Carbon and Nitrogen Transformation Links the Crop-Residue Contribution to Soil Organic Carbon." *NPJ Biofilms and Microbiomes* 8, no. 1: 14.

Xu, C., X. Xu, C. Ju, et al. 2021. "Long-Term, Amplified Responses of Soil Organic Carbon to Nitrogen Addition Worldwide." *Global Change Biology* 27, no. 6: 1170–1180.

Xu, D., G. H. Ros, Q. Zhu, et al. 2024. "Major Drivers of Soil Acidification Over 30 Years Differ in Paddy and Upland Soils in China." *Science of the Total Environment* 916: 170189.

Yan, F., W. Shangguan, J. Zhang, and B. Hu. 2020. "Depth-to-Bedrock Map of China at a Spatial Resolution of 100 Meters." *Scientific Data* 7, no. 1: 1–13.

Yang, J., J. Dong, L. Liu, et al. 2023. "A Robust and Unified Land Surface Phenology Algorithm for Diverse Biomes and Growth Cycles in China by Using Harmonized Landsat and Sentinel-2 Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 202: 610–636.

Yang, L., X. He, F. Shen, et al. 2020. "Improving Prediction of Soil Organic Carbon Content in Croplands Using Phenological Parameters Extracted From NDVI Time Series Data." *Soil and Tillage Research* 196: 104465.

Yu, Q., L. Duan, and J. Hao. 2021. "Acid Deposition in China: Sources, Effects and Control." *Acta Scientiae Circumstantiae* 41, no. 3: 731–746.

Zeng, Y., D. Hao, A. Huete, et al. 2022. "Optical Vegetation Indices for Monitoring Terrestrial Ecosystems Globally." *Nature Reviews Earth and Environment* 3, no. 7: 477–493.

Zhang, L., and E. Xu. 2023. "Effects of Agricultural Land Use on Soil Nutrients and Its Variation Along Altitude Gradients in the Downstream of the Yarlung Zangbo River Basin, Tibetan Plateau." *Science of the Total Environment* 905: 167583.

Zhang, X., S. Chen, J. Xue, et al. 2023. "Improving Model Parsimony and Accuracy by Modified Greedy Feature Selection in Digital Soil Mapping." *Geoderma* 432: 116383.

Zhang, X., D. Xiang, C. Yang, W. Wu, and H. Liu. 2022. "The Spatial Variability of Temporal Changes in Soil pH Affected by Topography and Fertilization." *Catena* 218: 106586.

Zhang, Y., K. Hou, H. Qian, et al. 2022. "Characterization of Soil Salinization and Its Driving Factors in a Typical Irrigation Area of Northwest China." *Science of the Total Environment* 837: 155808.

Zheng, Y., J. Jin, X. Wang, P. M. Kopittke, J. B. O'Sullivan, and C. Tang. 2024. "Disentangling the Effect of Nitrogen Supply on the Priming of Soil Organic Matter: A Critical Review." *Critical Reviews in Environmental Science and Technology* 54, no. 8: 676–697.

Zhou, K., W. Xu, L. Zhang, M. Ma, X. Liu, and Y. Zhao. 2023. "Estimating Nitrogen and Sulfur Deposition Across China During 2005 to 2020 Based on Multiple Statistical Models." *Atmospheric Chemistry and Physics* 23, no. 15: 8531–8551.

Zhu, X., G. H. Ros, M. Xu, et al. 2024. "The Contribution of Natural and Anthropogenic Causes to Soil Acidification Rates Under Different Fertilization Practices and Site Conditions in Southern China." *Science of the Total Environment* 934: 172986.

Zhuo, Z., Q. Chen, X. Zhang, et al. 2022. "Soil Organic Carbon Storage, Distribution, and Influencing Factors at Different Depths in the Dryland Farming Regions of Northeast and North China." *Catena* 210: 105934.