Mapping Cropland Soil Nutrients Contents Based on Multi-Spectral Remote Sensing and Machine Learning

5 authors, including:

Liang Zhu
Chinese Academy of Sciences
4 PUBLICATIONS   **9** CITATIONS

SEE PROFILE

Qifeng Zhuang
Nanjing Tech University
14 PUBLICATIONS   **196** CITATIONS

SEE PROFILE

Article

Mapping Cropland Soil Nutrients Contents Based on Multi-Spectral Remote Sensing and Machine Learning

Wenjie Zhang [1], Liang Zhu [2,*], Qifeng Zhuang [1,*] , Dong Chen [1] and Tao Sun [1]

[1] College of Geomatics Science and Technology, Nanjing Tech University, Nanjing 211816, China; 202161223017@njtech.edu.cn (W.Z.); 202061223025@njtech.edu.cn (D.C.); 202261223038@njtech.edu.cn (T.S.)

[2] State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China

[*] Correspondence: zhuliang@radi.ac.cn (L.Z.); zhuangqf@njtech.edu.cn (Q.Z.)

Abstract: Nitrogen (N) and phosphorus (P) are primary indicators of soil nutrients in agriculture. Accurate management of these nutrients is essential for ensuring food security. High-resolution, multi-spectral remote sensing images can provide crucial information for mapping soil nutrients at the field scale. This study compares the capabilities of ZH-1 and Sentinel-2 satellite data, along with different spectral indices, in mapping soil nutrients (total N and Olsen-P) using two machine learning algorithms, random forest (RF) and XGBoost (XGB). Two agricultural fields in Suihua City were selected as the study areas for this investigation. The results showed that Sentinel-2 data performed best in computing the total N content in soil using the RF model (R^2 = 0.74, *RMSE* = 0.10 g/kg). However, for the soil Olsen-P content, the XGBoost model performed better with ZH-1 data (R^2 = 0.75, *RMSE* = 9.79 mg/kg) than the RF model. This study demonstrates that both ZH-1 and Sentinel-2 satellite data perform well in terms of accurately mapping soil total N and Olsen-P contents using machine learning. Due to its higher spectral and spatial resolution, ZH-1 remote sensing data provides more detailed information on soil nutrient content during Olsen-P inversion and exhibits comparable accuracy.

Keywords: Sentinel-2; ZH-1; random forest; extreme gradient boosting; soil fertility; digital mapping

1. Introduction

In modern agricultural practices, guiding precision agriculture development and ensuring food security are of paramount importance [1–3]. Digital soil nutrient mapping plays a crucial role in achieving these objectives as it provides essential information about the soil properties that directly influence crop growth and development [4,5]. Among various soil nutrients, total nitrogen and available phosphorus are key indicators of soil fertility and plant nutrition [6,7]. Accurate mapping of the spatial distribution of total nitrogen and available phosphorus through precise mapping techniques is critical for optimizing agricultural productivity and resource management [8].

Traditional digital soil nutrient mapping methods typically rely on interpolation of ground survey data, resulting in coarse spatial resolution and limited guidance for precision agriculture [9–11]. Alternatively, using ground spectrometers combined with spectral information on nutrients for estimation may face challenges in broad-scale applications [12,13]. In large-scale farmlands or extensive regions, traditional approaches may encounter issues such as high data acquisition costs, time-consuming processes, and reliance on ground field surveys. These limiting factors hinder the application of nutrient mapping techniques in guiding precision agriculture and achieving sustainable agricultural development [14,15]. Sentinel-2, with a revisit period of 5–10 days and a spatial resolution of 10 m, captures 13 image bands, including visible light, near-infrared, and short-wave infrared, providing valuable spectral information for inferring the soil nutrient content [16,17]. Additionally, the ZH-1 hyperspectral satellite has a revisit period of six days for a single satellite and

an extended revisit period of approximately one day for eight hyperspectral satellites. It possesses a spatial resolution of 10 m, a spectral resolution of 2.5 nanometers, and a wavelength range of 400–1000 nanometers, enabling detailed high-spectral data to be gathered for a more accurate characterization of soil properties [18,19].

In recent years, with the development of emerging technologies, digital soil mapping has been applied using various techniques. Commonly used models include multiple linear regression [20], principal component analysis regression [21], the generalized additive model [22], and kriging interpolation [23]. Moreover, machine learning algorithms (e.g., support vector machines, decision trees, random forests, artificial neural networks) have been widely employed in remote sensing studies [24–27]. These algorithms offer advantages by learning from limited data and reducing errors through adaptive learning processes [24]. However, research on soil total nitrogen and available phosphorus mapping at higher spatial resolutions is still lacking [28]. Machine learning algorithms may not be universally applicable in different environments. Therefore, it is necessary to evaluate the applicability of different machine learning algorithms in our own context to understand the distribution of soil total nitrogen and available phosphorus content.

Hence, this study adopts the random forest (RF) and extreme gradient boosting (XGB) regression methods and introduces Zhuhai-1 (ZH-1) hyperspectral data for the first time on a field scale to explore their potential and effectiveness in mapping total nitrogen (total N) and available phosphorus (Olsen-P), providing valuable insights for soil nutrient estimation.

2. Materials and Methods

The technical workflow of this study is illustrated in Figure 1. It is primarily divided into three parts: data preprocessing, model training and validation, and model application. In the data preprocessing stage, both Sentinel-2 and ZH-1 data underwent radiometric calibration, followed by atmospheric correction using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) method to ensure the accuracy of surface reflectance [29,30]. Subsequently, the coarse spatial resolution bands of Sentinel-2 were resampled to 10 m using the nearest neighbor interpolation method, aligning the spatial resolution of both Sentinel-2 and ZH-1. Finally, specific bands from Sentinel-2 and ZH-1 were selected to calculate the vegetation and soil indices.

For the model training and validation stage, the surface reflectance of the two remote sensing datasets, along with vegetation and soil index values at their respective sampling points, were utilized as feature values, with total nitrogen (total N) and Olsen-P serving as label values. The pixel values for the mentioned sampling points were extracted using the rasterio library in Python. Due to different feature combinations, eight different datasets were formed while training two nutrient prediction models. Considering prior relevant studies on soil nutrient inversion, RF and XGB were selected as prediction models for machine learning regression [27,31].

To be specific, 90% of each dataset was used as training data, while the remaining 10% was reserved as a validation set to assess the model's accuracy. In the model application portion, the best-performing model was saved. Data consistent with model features is used as input for soil nutrient inversion at the field scale, resulting in a spatial distribution map of the soil nutrient content.

2.1. Study Area

The experimental site is located in Suihua City, which is a significant core area of the black soil zone and an important grain production region in central Heilongjiang Province, China (Figure 2). It is situated to the east of the Songnen Plain, at the junction of the Xiaoxing'an Mountains and the Songnen Plain in the middle reaches of the Hulan River (latitude: 46°19′ N to 47°09′ N, longitude: 126°25′ E to 127°23′ E) [32]. The forested area in the north belongs to the semi-humid and semi-arid monsoon climate zone in the northern temperate zone. Spring is relatively dry, with little rainfall, while summer is humid and

hot, with more rainfall. Autumn is cool, but the temperature drops quickly. Winter is cold, with a freezing period of up to six months, as the area belongs to a distinct continental climate [33]. The average annual temperature is about 2.9 °C, and the annual average precipitation is 552.5 mm, of which 70% is concentrated from June to August [34]. The annual average sunshine hours are 2395 h, the annual effective accumulated temperature is 2852.6 °C, and the area belongs to the second accumulated temperature zone in the province [35].
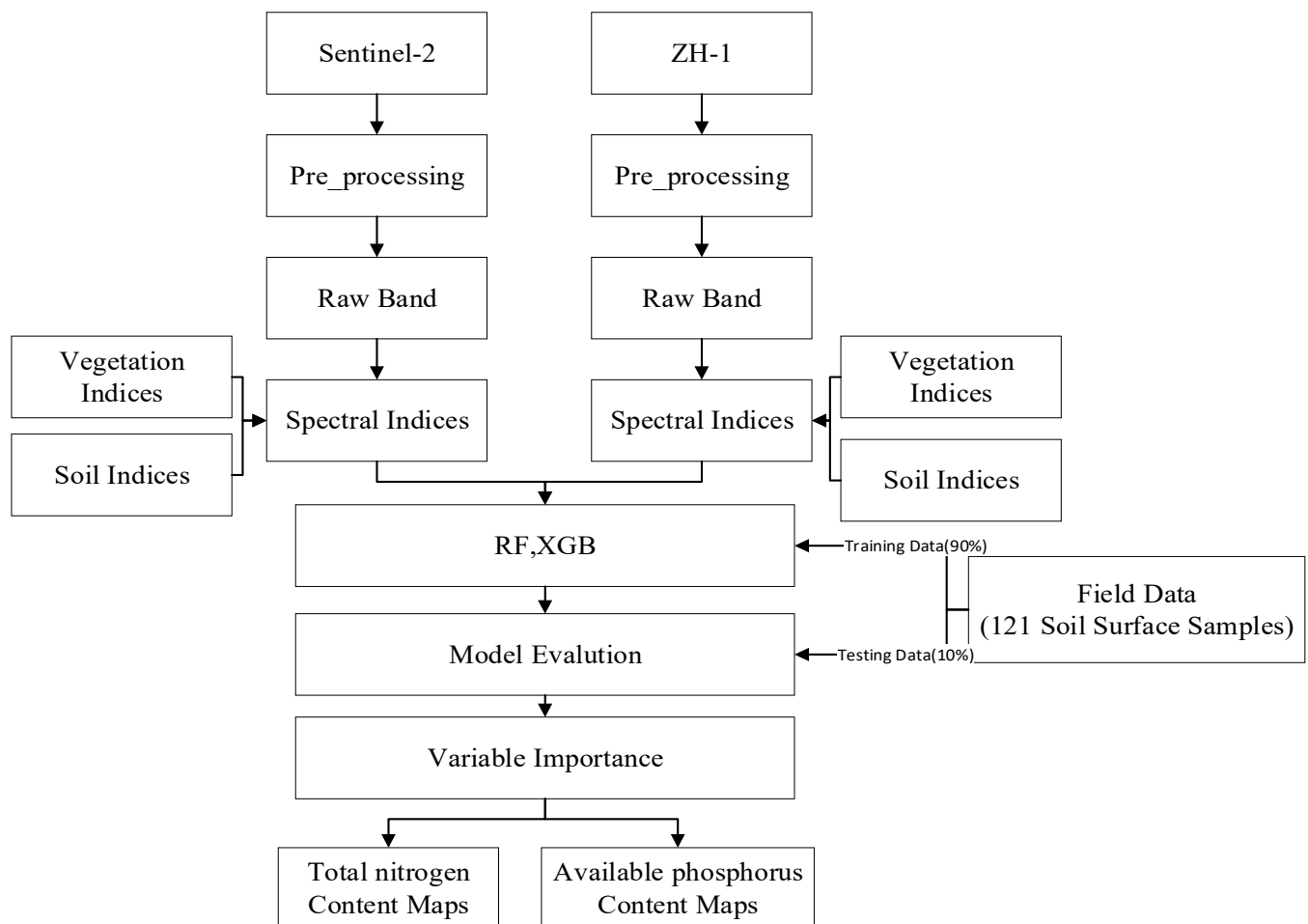

Figure 1. The proposed methodological framework for mapping soil nutrient content.

2.2. Field Data Collection and Laboratory Analysis

This study used soil samples collected from two experimental fields on 17 September 2020 and 22 October 2020. The samples were collected during the non-planting stage of the field, and the nutrient content in the samples reflects the soil nutrient status of the sampling site. Soil samples were collected using the systematic sampling method from the top 20 cm of the field surface, with approximately 500 g of samples being introduced into the sample bag. The sample bags were marked with sample numbers, and the latitude and longitude coordinates of the sampling points were recorded using GPS. A total of 72 and 49 soil samples were collected from these two experimental fields, respectively.
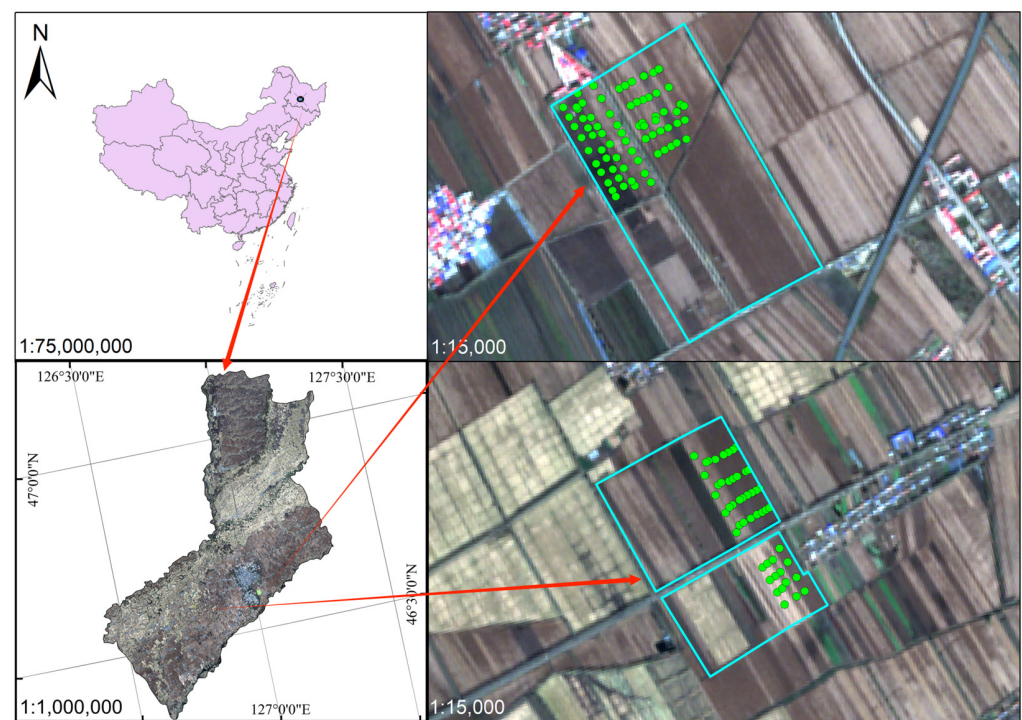
Figure 2. Illustration of the research area.

The soil samples were collected and sent to the Laboratory of the Institute of Soil Science, Chinese Academy of Sciences, for physicochemical analysis and processing. The samples were further dried, and stones, residues, and impurities were removed. The samples were then ground into a powder for subsequent physical and chemical analysis, including the determination of total nitrogen and Olsen-P in the soil using the Kelvin digestion method and the colorimetric method. The Kelvin digestion method is a widely used approach for determining soil total nitrogen content, involving high-temperature digestion with strong oxidizing agents to convert organic and inorganic nitrogen compounds into analyzable forms. The determination of Olsen-P using the colorimetric method involved the conversion of soil samples into phosphate ions within a solution, followed by their reaction with reagents containing chromogenic agents to generate colored compounds. Subsequently, the absorbance of these compounds was measured in order to rapidly and accurately quantify the available phosphorus content in the soil [36]. The statistical results obtained after testing of the nitrogen and phosphorus content in the soil at each sampling point are presented in Table 1.

Table 1. Statistical table of physical and chemical data for 121 soil samples.

	Total Nitrogen (g/kg)	Available Phosphorus (mg/kg)
Mean	1.44	40.14
Maximum	1.91	89.74
Minimum	0.95	11.01
Variance	0.04	264.64
Kurtosis	-0.20	0.05
Skewness	-0.13	0.61

2.3. Remote Sensing Data Acquisition and Preprocessing

Sentinel-2, as a part of the Copernicus program of the European Space Agency (ESA), is composed of multiple satellites that acquire medium-resolution images for various applications, such as forest monitoring, water quality assessment, land cover change detection, and disaster management [37]. This mission includes two satellites, namely,

Sentinel-2A and Sentinel-2B, which share similar designs and orbits. Each satellite is equipped with a multi-spectral instrument (MSI) and utilizes a three-mirror astigmatic telescope with a 150 mm aperture and 600 mm focal length [16]. It captures 13 image bands, including different spectral ranges such as visible light, near-infrared, and short-wave infrared [17]. Using a push-broom method, MSI can achieve high-resolution images spanning 290 km [38]. Sentinel-2 offers temporal data continuity with a revisiting period of 5 days, which is accessible to all users [39]. However, their spatial resolution varies as shown in Table 2: bands 2, 3, 4, and 8 are 10 m, bands 5, 6, 7, 8A, 11, and 12 are 20 m, and bands 9 and 10 are 60 m [40].

Table 2. Sentinel-2 band properties.

Payload Band	Central Wavelength (nm)	Spectrum Width (nm)	Payload Band	Central Wavelength (nm)	Spectrum Width (nm)
Band1	442.7	21	Band8	832.8	106
Band2	492.4	66	Band8A	864.7	21
Band3	559.8	36	Band9	945.1	20
Band4	664.6	31	Band10	1373.5	31
Band5	704.1	15	Band11	1613.7	91
Band6	740.5	15	Band12	2202.4	175
Band7	782.8	20			

The ZH-1 hyperspectral satellite (OHS) adopts a push-broom imaging technique with a spatial resolution of 10 m, a spectral resolution of 2.5 nanometers, and a wavelength range of 400–1000 nanometers as shown in Table 3. Due to storage limitations and compression design, it transmits 32 spectral bands and has a weight of 71 kg [41]. Each hyperspectral satellite can orbit the Earth approximately 15–16 times a day, with a maximum data acquisition time of about 8 min per orbit. Currently, a single hyperspectral satellite has a revisit period of six days, while the extended revisit period of eight hyperspectral satellites is about one day [18].

Table 3. ZH-1 band properties.

Payload Band	Central Wavelength (nm)	Spectrum Width (nm)	Payload Band	Central Wavelength (nm)	Spectrum Width (nm)
Band1	466	5	Band17	716	7
Band2	480	5	Band18	730	6
Band3	500	4	Band19	746	6
Band4	520	6	Band20	760	6
Band5	536	5	Band21	776	6
Band6	550	5	Band22	790	5
Band7	566	4	Band23	806	6
Band8	580	5	Band24	820	6
Band9	596	5	Band25	836	4
Band10	610	5	Band26	850	4
Band11	626	5	Band27	866	6
Band12	640	6	Band28	880	6
Band13	656	5	Band29	896	3
Band14	670	5	Band30	910	9
Band15	686	6	Band31	926	4
Band16	700	6	Band32	940	4

The experiment aimed to ensure that the surface reflectance of the research area obtained could faithfully represent the situation regarding soil sample collection. We acquired the Sentinel-2 image from 15 October 2020 through the Google Earth Engine

(GEE), which covered the date of field sampling. Considering that the image was free of clouds, it was highly suitable for our study. Additionally, the ZH-1 image from 19 October 2020 was obtained from the ZH-1 Remote Sensing Data Service Platform (https://www.obtdata.com, accessed on 24 April 2022) and was downloaded through an application specifically designed for educational and research purposes.

Prior to the experiment, both images underwent preprocessing, which included radiometric correction, atmospheric correction, and geometric correction. Furthermore, in the Sentinel-2 imagery, bands with spatial resolutions other than 10 m were resampled to achieve a uniform 10 m resolution.

2.4. Spectral Indices

In this study, we derived spectral indices from the bands of Sentinel-2 and ZH-1. This selection was motivated by various vegetation indices reported in previous studies, aiming to augment spatial information and enhance regression accuracy to some extent by incorporating additional vegetation indices. In similar studies conducted by Zinhle, vegetation indices were carefully screened, and the following indices were considered to play a significant role in soil nutrient inversion [42]. Accordingly, we continued to utilize these vegetation indices in our experiments. The vegetation indices based on vegetation reflectance include normalized difference vegetation indices (NDVIRE1n, NDVIRE2n, NDVIRE3n) in the narrow bands, as well as a modified simple ratio (MSRRE). Furthermore, these indices encompass the plant senescence reflectance index (PSRI), the enhanced vegetation index (EVI), and the green normalized difference vegetation index (GNDVI) [28,43]. The final spectral indices derived from Sentinel-2 which were used in this study are summarized in Table 4. Additionally, for this investigation, we selected corresponding bands from the Sentinel-2 and ZH-1 original data to calculate the same indices, and the results are presented in Table 5.

Table 4. Spectral indices used in this study with Sentinel-2 data: The table consists of seven vegetation indices and five soil indices. From left to right, the columns represent the names of the spectral indices, their corresponding calculation formulas, the meanings of the spectral indices, and the reference numbers for the literature sources that utilized these spectral indices.

Vegetation Index	Equation	Purpose	Source
PSRI	$\frac{(B4 - B3)}{B8A}$	Senescence-induced reflectance changes	[44]
NDVIRE1n	$\frac{(B8A - B5)}{(B8A + B5)}$	Sparse biomass	[45]
NDVIRE2n	$\frac{(B8A - B6)}{(B8A + B6)}$	Sparse biomass	[45]
NDVIRE3n	$\frac{(B8A - B7)}{(B8A + B7)}$	Sparse biomass	[45]
MSRRE	$\frac{(B8/B8A) - 1}{\sqrt{(B8/B8A) + 1}}$	Correction for leaf specular reflection	[46]
EVI	$2.5 \times \frac{(B8 - B4)}{(B8 + 6 \times B4 - 7.5 \times B2) + 1}$	Chlorophyll-sensitive	[47]
GNDVI	$\frac{(B8 - B3)}{(B8 + B3)}$	Chlorophyll-sensitive	[48]
Soil index	**Equation**	**Property**	**Source**
BI	$\left(\frac{(B4^2 + B3^2 + B2^2)}{3}\right)^{0.5}$	Average reflectance magnitude	[49]
CI	$\frac{(B4 - B3)}{(B4 + B3)}$	Soil color	[49]
HI	$\frac{(2 \times B4 - B3 - B2)}{(B3 - B2)}$	Primary colors	[49]
RI	$\frac{B4^2}{(B2 \times B3^3)}$	Hematite content	[50]
SI	$\frac{(B4 - B2)}{(B4 + B2)}$	Spectral slope	[49]

Table 5. Spectral indices used in this study with ZH-1 data: The table consists of seven vegetation indices and five soil indices. From left to right, the columns represent the names of the spectral indices, their corresponding calculation formulas, the meanings of the spectral indices, and the reference numbers for the literature sources that utilized these spectral indices.

Vegetation Index	Equation	Property	Source
PSRI	$\frac{(B13 - B8)}{B27}$	Senescence-induced reflectance changes	[44]
NDVIRE1n	$\frac{(B27 - B16)}{(B27 + B16)}$	Sparse biomass	[45]
NDVIRE2n	$\frac{(B27 - B19)}{(B27 + B19)}$	Sparse biomass	[45]
NDVIRE3n	$\frac{(B27 - B22)}{(B27 + B22)}$	Sparse biomass	[45]
MSRRE	$\frac{(B25/B27) - 1}{\sqrt{(B26/B27) + 1}}$	Correction for leaf specular reflection	[46]
EVI	$2.5 \times \frac{(B25 - B13)}{(B25 + 6 \times B13 - 7.5 \times B3) + 1}$	Chlorophyll-sensitive	[47]
GNDVI	$\frac{(B25 - B8)}{(B25 + B8)}$	Chlorophyll-sensitive	[48]
Soil index	**Equation**	**Property**	**Source**
BI	$\left(\frac{(B13^2 + B8^2 + B3^2)}{3}\right)^{0.5}$	Average reflectance magnitude	[49]
CI	$\frac{(B13 - B8)}{(B13 + B8)}$	Soil color	[49]
HI	$\frac{(2 \times B13 - B8 - B3)}{(B8 - B3)}$	Primary colors	[49]
RI	$\frac{B13^2}{(B3 \times B8^3)}$	Hematite content	[50]
SI	$\frac{(B13 - B3)}{(B13 + B3)}$	Spectral slope	[49]

2.5. Machine Learning Regression Models

2.5.1. Random Forest Regression

Random forest is a supervised ensemble learning method that operates based on the principles of decision trees as shown in Figure 3. This versatile algorithm is capable of effectively handling both classification and regression problems [51,52]. The fundamental concept underlying random forest involves creating a forest comprising multiple decision trees, where each tree serves as a base learner and the entire ensemble embodies the concept of ensemble learning [53,54]. The final model is generated by aggregating the average output of each tree in the forest. Additionally, the algorithm utilizes out-of-bag samples, representing unused data points that can be leveraged for model evaluation and assessing variable importance [55,56]. Notably, random forest exhibits the ability to handle high-dimensional feature data without necessitating feature selection. Scholar John employed a set of machine learning algorithms, including an artificial neural network (ANN), a support vector machine (SVM), cubist regression, random forest (RF), and multiple linear regression (MLR), to predict SOC levels. Among these models, RF demonstrated the best performance, with an R-squared value of 0.68 [27]. Furthermore, the algorithm boasts a concise set of parameters, including the number of decision trees (n_estimators), the maximum depth of each decision tree (max_depth), and the minimum number of samples required for a node to split (min_samples_split) [57,58]. For this investigation, we conducted our analysis utilizing the Scikit-Learn module within the Python environment. Employing a grid search approach, we ascertained the optimal parameters and established a robust model to accurately predict soil nutrient content.
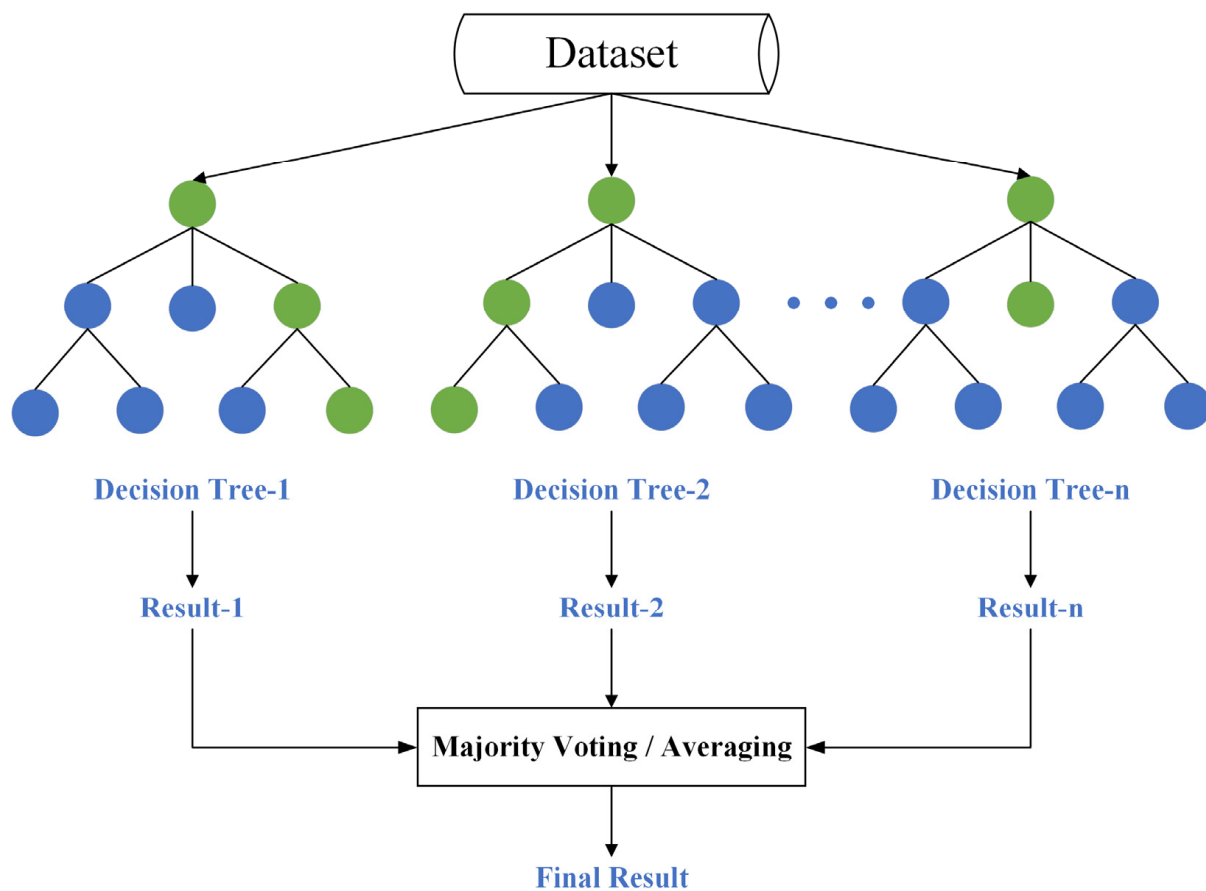
Figure 3. A simplified schematic diagram of the random forest regression model.

2.5.2. Extreme Gradient Boosting Regression

The extreme gradient boosting (XGB) algorithm, introduced by Chen and Guestrin in 2016, is a novel machine learning approach as shown in Figure 4. It has demonstrated remarkable performance in numerous international data mining competitions, surpassing even deep learning algorithms. XGB falls under the category of gradient boosting algorithms for classification and regression ensembles, making it applicable to both classification and regression tasks [59,60]. The XGB training process involves two stages: fitting the input training dataset and fitting the residuals. The main hyperparameters of XGBoost include the number of decision trees, the learning rate, the maximum depth of trees, the minimum sample weight, the subsample ratio used in each iteration, and the weight of the L1 regularization term [57,61]. This training method significantly enhances the performance of weakly supervised learning. Scholar Miao employed three machine learning models, i.e., XGBoost, RF, and LightGBM, based on Sentinel-2 images for the purpose of estimating leaf nutrient levels. The results demonstrated that XGBoost outperformed the other models in terms of estimating leaf C, (with R^2 values of 0.655, 0.799, and 0.829 for spring, summer, and winter, respectively), N (with R^2 values of 0.668, 0.743, and 0.704), and P (with R^2 values of 0.539, 0.622, and 0.596) [62]. The fitting process underwent multiple iterations until it met the convergence criterion. In this study, the XGB algorithm was adopted due to its ability to mitigate overfitting issues and its superior performance [63]. The Xgboost library in the Python environment was utilized for modeling purposes in this research.
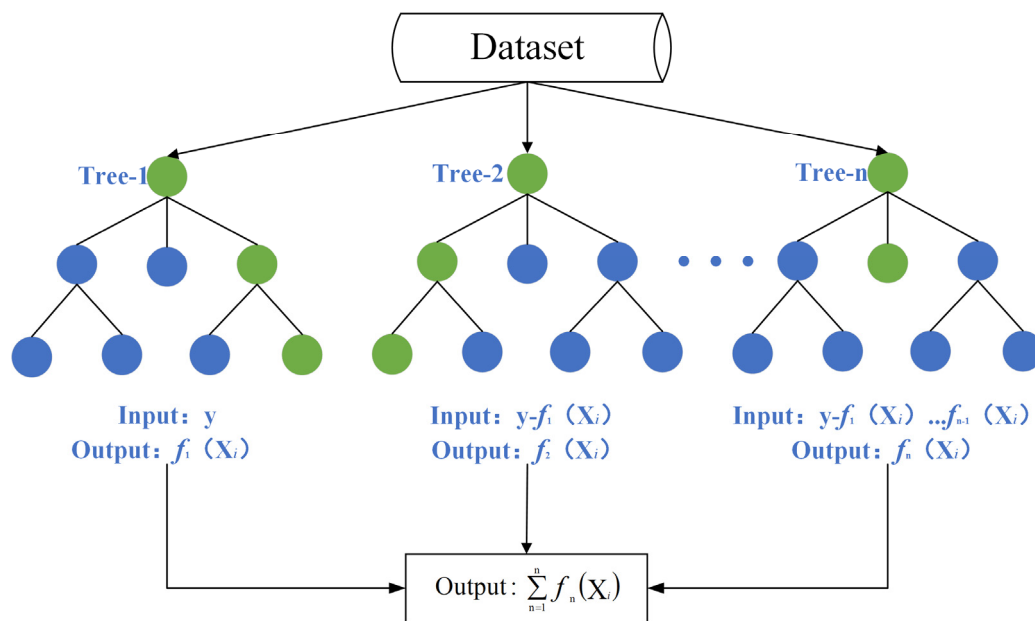
Figure 4. A concise schematic diagram illustrating the XGBoost regression model. In the figure, '*x*' represents features, '*y*' represents labels, '*i*' indicates sequence values, and '*n*' denotes the number of trees.

2.5.3. Experiments

In this study, we conducted research on characteristic variable images to simulate different soil nutrient contents (total nitrogen and Olsen-P). In order to enhance the accuracy and generalization capability of the model and to achieve a more stable performance, we employed a larger set of samples for training, allowing the model to better learn the features and patterns within the data. The dataset was divided into 90% training and 10% testing subsets. Drawing inspiration from previous research, our aim was to compare the effectiveness of different feature combinations to capture spectral information related to vegetation and soil, thereby enhancing feature representation for the purpose of obtaining more accurate and effective soil parameter estimation models. To achieve this, we employed two models, random forest (RF) and XGBoost (XGB), along with two types of remote sensing data with various combinations of variables, which are summarized in Table 6.

Table 6. The different data configurations for the machine learning regression experiments.

Experiment	Number of Variables	Data Configuration
1	12	Sentinel-2 raw bands
2	19	Sentinel-2 raw bands + vegetation indices
3	17	Sentinel-2 raw bands + soil indices
4	24	Sentinel-2 raw bands + vegetation indices + soil indices
5	32	ZH-1 raw bands
6	39	ZH-1 raw bands + vegetation indices
7	37	ZH-1 raw bands + soil indices
8	44	ZH-1 raw bands + vegetation indices + soil indices

For the two soil nutrients (total nitrogen and Olsen-P), the experimental setups included the following combinations: (1) Sentinel-2 raw bands, (2) Sentinel-2 raw bands + vegetation indices, (3) Sentinel-2 raw bands + soil indices, (4) Sentinel-2 raw bands + vegetation indices + soil indices, and (5) Sentinel-2 raw bands + soil indices. Additionally, the experimental setups comprised: (5) ZH-1 raw bands, (6) ZH-1 raw bands + vegetation indices, (7) ZH-1 raw bands + soil indices, and (8) ZH-1 raw bands + vegetation indices + soil indices.

In this research, we employed a grid search as the method for model tuning. Grid search is a widely used parameter optimization technique aimed at determining the optimal hyperparameter combinations for machine learning models. It involves traversing a predefined grid of parameter values, exploring different combinations, and evaluating the performance of each combination to identify the best parameter settings [64].

2.6. Model Evaluation

This study used common machine learning verification indices to evaluate the prediction performances of the RF and XGB models. These included mean absolute error (*MAE*), root mean square error (*RMSE*), percent bias (*PBIAS*), and r-squared(R^2), as shown in Equations (1)–(4):

$$R^2 = 1 - \frac{\sum\limits_{i=1}^{n}(O_i - P_i)^2}{\sum\limits_{i=1}^{n}(\overline{O}_i - P_i)^2} \tag{1}$$

$$MAE = \frac{1}{n}\sum\limits_{i=1}^{n}|P_i - O_i| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}(P_i - O_i)^2} \tag{3}$$

$$PBIAS = \left[\frac{\sum\limits_{i=1}^{n}(O_i - P_i)*100}{\sum\limits_{i=1}^{n}O_i}\right] \tag{4}$$

where *n* represents the number of sample points, P_i is the predicted soil content, and O_i is the observed soil content at site *i*.

3. Results

3.1. Model Evaluation

In this study, we conducted model performance statistics as shown in Tables 7 and 8 on the testing data (*n* = 13 samples) and obtained the following results. Regarding the estimation of the total nitrogen content in soil, the random forest model showed remarkable performance, especially the RF1 variant (represented by experiment number 1 in Table 6), which demonstrated outstanding results. This model exhibited the lowest root mean square error (*RMSE*) and mean absolute error (*MAE*), indicating the highest accuracy in estimating soil nitrogen content (*RMSE* = 0.10 g/kg, *MAE* = 0.07 g/kg), and it also achieved the highest R-squared value (R^2 = 0.74). It is noteworthy that, based on the prediction bias (*PBIAS* = −2.66), the predicted values of total nitrogen were slightly higher than the observed values.

Overall, among all models, the XGBoost (XGB) model in Experiment 8 performed the most poorly. This model incorporated the raw bands, soil indices, and vegetation indices of ZH-1. It had a higher error rate, as reflected by the higher *RMSE* and *MAE* values (*RMSE* = 0.16 g/kg, *MAE* = 0.11 g/kg), and it achieved the lowest R-squared value (R^2 = 0.31). Moreover, this model overestimated the total nitrogen content, as indicated by the prediction bias (*PBIAS* = −5.01).

Table 7. Model evaluation statistics for the total nitrogen in different experiments.

Model	R^2	MAE	RMSE	PBIAS
RF1	0.74	0.07	0.10	−2.66
RF2	0.70	0.07	0.11	−2.74
RF3	0.70	0.07	0.11	−3.33
RF4	0.71	0.07	0.10	−2.99
RF5	0.55	0.09	0.13	−4.74
RF6	0.56	0.09	0.13	−4.12
RF7	0.56	0.09	0.13	−4.60
RF8	0.50	0.09	0.14	−4.73
XGB1	0.72	0.07	0.10	−2.03
XGB2	0.71	0.08	0.10	−2.32
XGB3	0.64	0.08	0.12	−2.64
XGB4	0.67	0.09	0.11	−3.32
XGB5	0.48	0.10	0.14	−4.76
XGB6	0.43	0.10	0.15	−3.92
XGB7	0.41	0.12	0.19	−5.35
XGB8	0.31	0.11	0.16	−5.01

Table 8. Model evaluation statistics for the Olsen-P in different experiments.

Model	R^2	MAE	RMSE	PBIAS
RF1	0.69	7.87	10.76	3.05
RF2	0.65	8.27	11.40	3.46
RF3	0.70	7.57	10.68	2.12
RF4	0.63	8.26	11.85	3.21
RF5	0.74	7.26	9.95	6.08
RF6	0.70	8.21	10.57	7.81
RF7	0.75	7.29	9.74	7.73
RF8	0.40	11.48	14.98	5.50
XGB1	0.56	9.32	12.95	10.80
XGB2	0.67	9.29	11.23	3.88
XGB3	0.63	8.78	11.82	7.50
XGB4	0.40	11.48	14.98	5.50
XGB5	0.75	6.41	9.79	2.31
XGB6	0.63	8.30	11.82	5.71
XGB7	0.74	7.72	9.83	0.99
XGB8	0.69	8.35	10.76	0.55

In the inversion model for total soil nitrogen, the overestimation of predicted values is attributed to the presence of features with low correlations in the dataset, which negatively impact the model. In future research, we will address this issue and work towards mitigating its influence. It is evident that *PBIAS* increases (decreases) with the addition of model features. Considering the characteristics of the model, this trend may be attributed to the introduction of noise from features with lower correlations in the dataset.

The XGB model from Experiment 5 emerged as the top-performing model for Olsen-P estimation among all experiments. It demonstrated the highest accuracy in terms of predicting Olsen-P content, with the lowest root mean square error (*RMSE* = 9.79 mg/kg) and mean absolute error (*MAE* = 6.41 mg/kg), along with the highest R-squared value (R^2 = 0.75). The predicted values were slightly lower than the observed values, as indicated by the prediction bias (*PBIAS* = 2.31).

On the other hand, the XGB model from Experiment 4 exhibited the poorest performance. This model involved the original bands of Sentinel-2, soil indices, and vegetation indices. It had a higher error rate, which was evident from the elevated *RMSE* (14.97 mg/kg) and *MAE* (11.48 mg/kg), and the lowest R-squared value (R^2 = 0.40). Furthermore, this model underestimated the Olsen-P content, as indicated by the percent bias (*PBIAS* = 5.50).

The effectiveness of the inversion models for total nitrogen and Olsen-P was evaluated using the Taylor diagram shown in Figure 5 (generated using the plotrix package in R). The Olsen-P inversion model exhibited an outstanding performance, with correlation coefficients ranging from 0.8 to 0.95, demonstrating a close agreement with the actual values, as illustrated in the Taylor diagram. For the total nitrogen inversion model, the correlation coefficients were primarily distributed between 0.8 and 0.9. The inversion results were consistent with observations from multiple models and demonstrated excellent performance. According to the Taylor diagram, the best-performing total nitrogen inversion model was RF1, which corresponds with the model evaluation results presented in Table 7. RF1 displayed the highest accuracy in terms of estimating the total nitrogen content in the soil. RF7 emerged as the optimal Olsen-P inversion model. This model exhibited strong correlation and minimal errors when compared to the actual measured values.

(a)   (b)
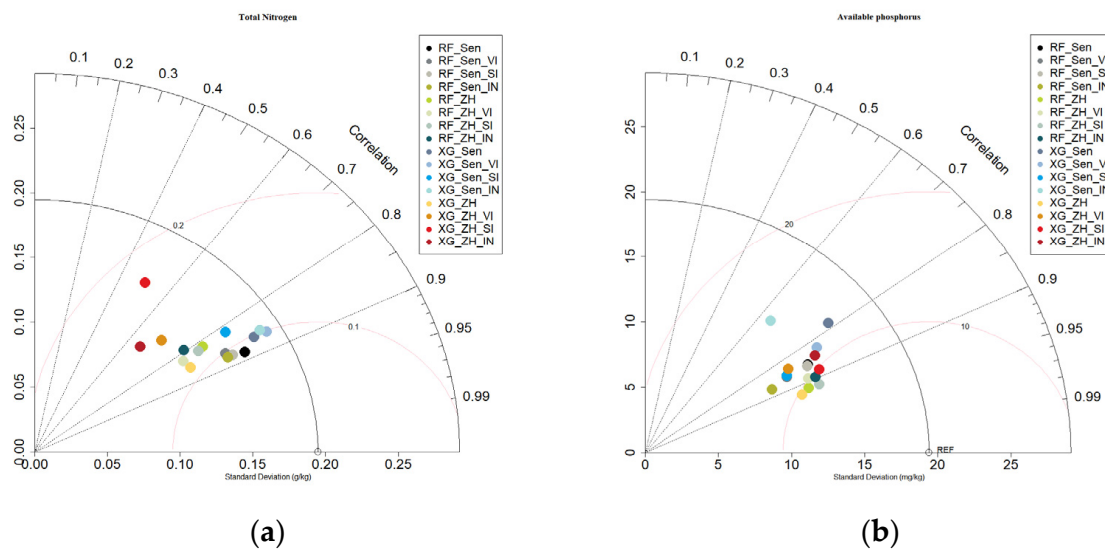
Figure 5. Taylor diagram for the 16 experiments for the two nutrients: (**a**) Taylor diagram for total nitrogen; (**b**) Taylor diagram for Olsen-P.

3.2. Variable Importance

In this study, we selected the best-performing RF and XGB models for the inversion of each soil nutrient content, then compared and analyzed the significance of each feature. The feature contributions of the soil total nitrogen and Olsen-P content inversion models, as shown in Figures 6 and 7, varied. In the soil total nitrogen content inversion model, RF1 demonstrated the best performance, with Sentinel-2′s B3 band contributing over 20%. Among the four XGB models, XGB1 performed the best, with its highest contributing feature being the same as RF1 nearly 15%. In the Olsen-P content inversion model, the top-performing models were RF7 and XGB5. In the RF7 model, the feature with the highest contribution was ZH-1′s B2 band, which contributed nearly 8%. In the XGB5 model, the feature with the largest contribution was ZH-1′s B27 band, also contributing nearly 8%. This indicates that Sentinel-2 data played a significant role in the inversion of soil total nitrogen, while for the Olsen-P content inversion model, ZH-1 data had a more prominent contribution. The observed changes in feature importance in the results can be attributed to the presence of highly correlated features in the input data, which overshadowed the importance of other features, resulting in variations in their contributions.
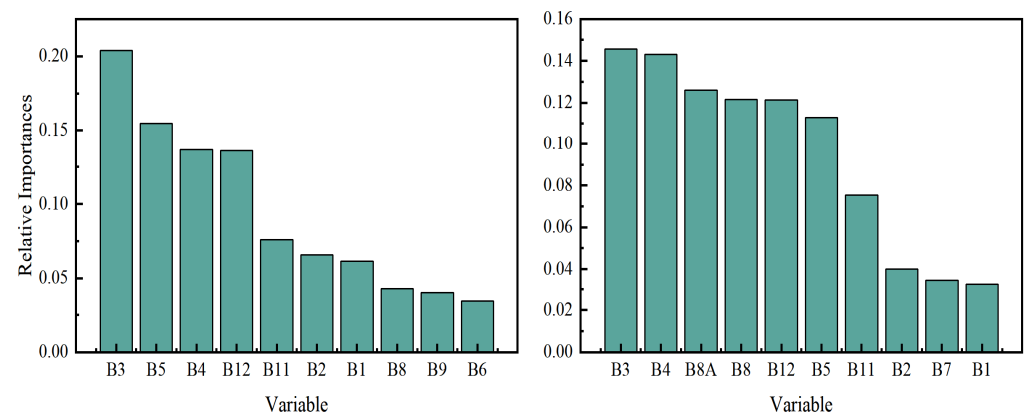
Figure 6. The contributions of key features to the prediction of total nitrogen content in soil using the RF1 and XG1 models are depicted in the graph. The y-axis represents the contribution of each feature, while the x-axis represents the different features.
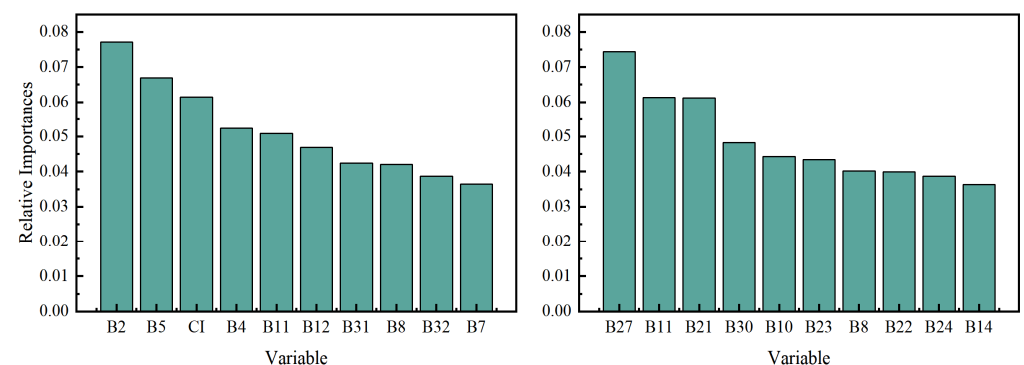

Figure 7. This graph illustrates the contributions of key features to the prediction of soil Olsen-P content using the RF7 and XG5 models. The y-axis represents the contribution of each feature, while the x-axis represents different features.

3.3. Mapping Soil Nutrients Content

Zinhle has demonstrated in study that the combination of machine learning methods with remote sensing data and derived spectral indices can accurately predict soil total nitrogen content and generate spatial distribution maps [42]. Therefore, in this study, different models were selected to predict and map the spatial distribution of soil total nitrogen and Olsen-P content. RF1 and XGB1 models were chosen to predict and map soil total nitrogen content, as shown in Figure 8a,b. From Figure 8a,b, significant spatial variations in soil total nitrogen content between the two experimental fields can be observed. In experimental field 1, the soil total nitrogen content in the southern part was notably higher than that in the northern part, while in experimental field 2, the overall soil total nitrogen content was higher in the northern part, ranging from 1.28 to 1.70 g/kg. The spatial distribution of the soil Olsen-P content shown in Figure 8c,d was generally consistent with that of soil total nitrogen content, with the Olsen-P content ranging from 16.34 to 68.76 mg/kg. In the scatter plots in Figure 9, it can be clearly observed that the data points are not highly clustered. Additionally, the spatial distribution of soil nutrient content predicted by the two models aligns, indicating that the experimental design in this study is reliable and capable of producing trustworthy results.
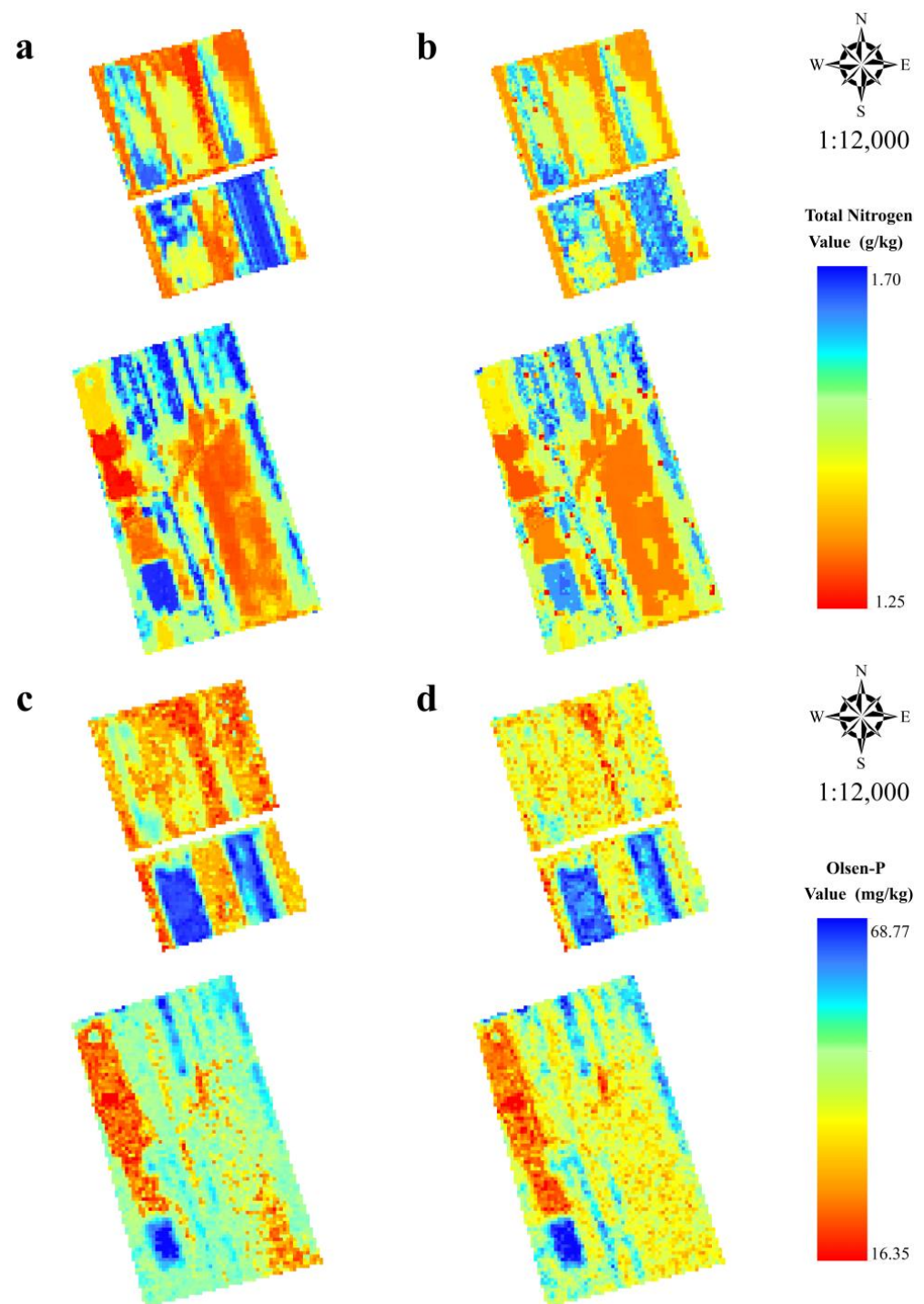
Figure 8. Spatial distribution map of soil nutrient content: (**a**) The spatial distribution of total nitrogen was mapped with the random forest model for experiment 1. (**b**) The spatial distribution of total nitrogen was mapped with the extreme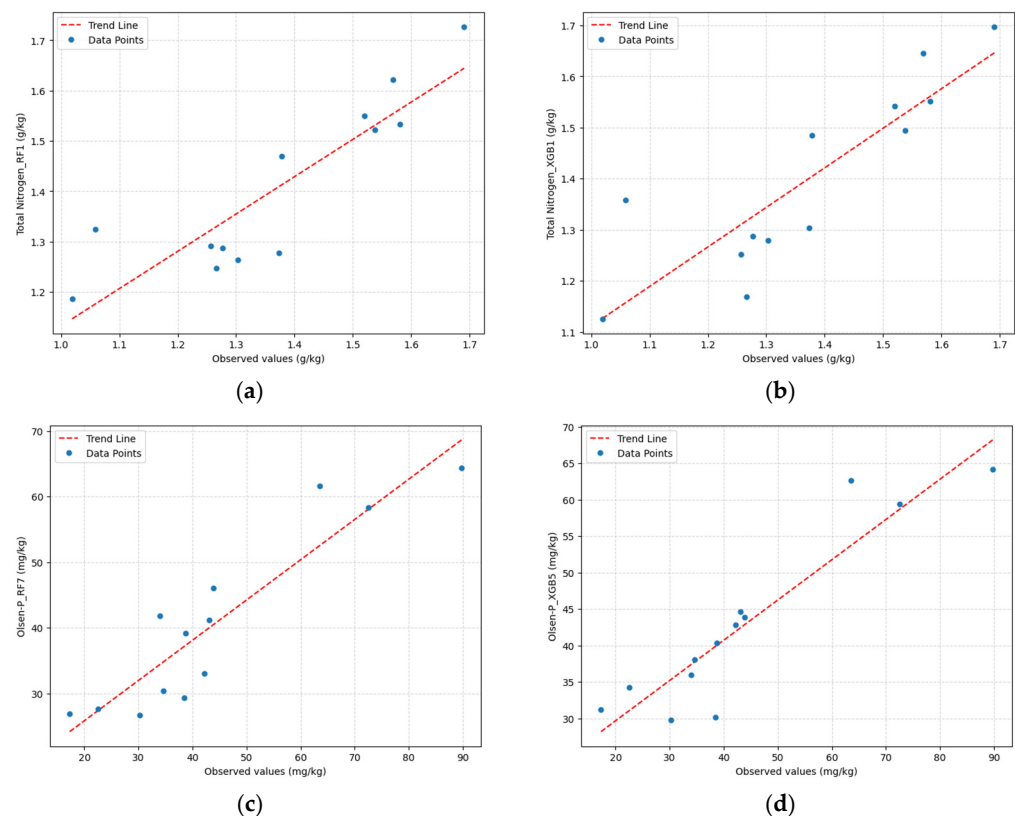 gradient boosting model for experiment 1. (**c**) The spatial distribution of Olsen-P was mapped with the random forest model for experiment 7. (**d**) The spatial distribution of Olsen-P was mapped with the extreme gradient boosting model for experiment 5.

Figure 9. Scatter plots of soil nutrient content, where the horizontal axis represents observed values and the vertical axis represents predicted values. The red lines indicate the trend lines. (**a**,**b**) represent RF1 and XGB1 in the soil total nitrogen inversion model, respectively. (**c**,**d**) represent RF7 and XGB5 in the Olsen-P inversion model, respectively.

4. Discussion

This study aims to evaluate the applicability of ZH-1 and Sentinel-2 satellite data for mapping soil nutrients (total nitrogen and Olsen-P) in farmland soil in Suihua City, China. Two machine learning algorithms, random forest (RF) and XGBoost (XGB), were employed to assess the predictive capabilities of these data for soil nutrient content.

Regarding the soil total nitrogen content, our results demonstrate that the RF model performed optimally when using Sentinel-2 data, with an R^2 of 0.74 and *RMSE* of 0.10 g/kg. Conversely, for the soil Olsen-P content, the XGB model outperformed using ZH-1 data, showing an R^2 of 0.75 and *RMSE* of 9.79 mg/kg, surpassing the RF model. The superior performance of the Sentinel-2 data model in predicting soil total nitrogen can be attributed to its sensitivity in detecting nitrogen compounds in the short-wave infrared range, as indicated by the prominent contributions of Sentinel-2's B11 and B12 bands in Figure 6 [65]. In contrast, ZH-1's spectral range is 400–1000 nm; hence, for soil total nitrogen inversion, the combination of Sentinel-2 data with machine learning algorithms yields better results. The inversion of soil total nitrogen has shown a common overestimation phenomenon, mainly due to the redundancy of features in the dataset. It can be observed that when ZH-1 data appears as a feature in the dataset, the overestimation of soil total nitrogen significantly increases, likely due to the rapid increase in the number of features. In future research, this issue will be addressed by optimizing the selection of model features to improve model accuracy.

This study incorporated vegetation indices and soil indices to construct diverse datasets for model training, with the aim of enhancing model accuracy and comparing the performance disparities among models with different input features. Although, overall, the model performance did not exhibit a significant improvement when using vegetation indices as partial features, it is noteworthy that the combination of Sentinel-2 vegetation in-

dices slightly outperformed the model utilizing soil indices for soil total nitrogen modeling, as indicated in Tables 7 and 8. Furthermore, in the models utilizing vegetation indices, their contribution was generally greater than that of other spectral bands. These findings are consistent with the observations made by Zhang in their research, wherein conventional spectral indices also played a role in nitrogen estimation [25]. Thus, judiciously selecting vegetation indices and soil indices as model input features can indeed enhance spatial information and improve regression accuracy.

It is worth noting that both the ZH-1 and Sentinel-2 satellite data performed well in accurately mapping soil total nitrogen and Olsen-P contents using machine learning regression models. Due to its higher spectral and spatial resolution, ZH-1 remote sensing data provided more detailed information on soil nutrient content during Olsen-P inversion, displaying considerable accuracy. This finding highlights the potential of ZH-1 data to provide valuable soil nutrient variation information at a finer scale. This aligns with the viewpoint presented in the studies by Sebastian [66] and Kawamura [67], indicating that hyperspectral remote sensing images exhibit a certain advantage over multispectral remote sensing images in terms of capturing key soil parameters. In the context of precise soil nutrient mapping, the spatial resolution of digital soil mapping products for total nitrogen and Olsen-P has increased from 250 m to 30 m [31]. During the production of these products, the inversion accuracy and spatial resolution of nutrient distribution largely rely on the spatial and spectral resolution of input remote sensing data. This study introduced ZH-1 hyperspectral remote sensing data and demonstrated its excellent potential in soil Olsen-P content retrieval experiments. The Olsen-P inversion mapping method proposed in this study contributes to the potential of obtaining the spatial distribution of soil Olsen-P content in farmland more rapidly and accurately through remote sensing data, thus promoting the development and implementation of precision agriculture.

Undeniably, our research has certain limitations. The sample collection in the study was relatively limited, focusing solely on a relatively small geographic area. However, for larger-scale precision agriculture projects, a more extensive and comprehensive collection of soil samples becomes crucial. With broader coverage, future research will incorporate ground-based environmental factors and other data (such as soil type, DEM, slope, and aspect) as supplementary features into the model, aiming to significantly enhance the predictive accuracy. This important enhancement will make a substantial contribution to the refined inference and prediction of soil nutrient contents, thereby providing more reliable support for decision-making regarding agricultural production. Additionally, greater attention should be given to feature selection and interpretability of the models. These considerations are of paramount importance for the advancement of precision agriculture.

5. Conclusions

Based on machine learning regression methods, combined with Sentinel-2 multispectral data and ZH-1 hyperspectral data, this study inversely estimated the total nitrogen content and Olsen-P content in farmland soils. Through experiments on different combinations of predictive factors, this study found that different factors have different effects on the prediction of soil parameters, and the best-performing model varies depending on the different soil parameters. Among them, the RF1 model performed best in the inverse estimation of the total nitrogen content, reaching $R^2 = 0.74$ and $RMSE = 0.10$ g/kg, while the XGB5 model performed best in the inverse estimation of Olsen-P content, reaching $R^2 = 0.75$ and $RMSE = 9.79$ mg/kg. In addition, this study also found, through comparative analysis, that when predicting soil total nitrogen content, the original bands of Sentinel-2 contribute more to the prediction results, proving that Sentinel-2 plays an important role in predicting soil total nitrogen content. When predicting soil Olsen-P content, the original bands of ZH-1 hyperspectral data contribute more to and have a positive impact on the prediction results. These high-contribution features are the basis for establishing soil parameter prediction models. Finally, this study has generated spatial distribution maps of soil total nitrogen and Olsen-P, which can serve as valuable tools to guide agricultural

production decision-making and aid in formulating field-scale soil nutrient management plans to increase crop yields and enhance food security. The application of these maps holds great potential, not only for promoting the development, but also for facilitating the implementation of precision agriculture.

References

1. Sishodia, R.P.; Ray, R.L.; Singh, S.K. Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sens.* **2020**, *12*, 3136. [CrossRef]
2. Khanna, A.; Kaur, S. Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture. *Comput. Electron. Agric.* **2019**, *157*, 218–231. [CrossRef]
3. Prosekov, A.Y.; Ivanova, S.A. Food security: The challenge of the present. *Geoforum* **2018**, *91*, 73–77. [CrossRef]
4. Dong, W.; Wu, T.; Luo, J.; Sun, Y.; Xia, L. Land parcel-based digital soil mapping of soil nutrient properties in an alluvial-diluvia plain agricultural area in China. *Geoderma* **2019**, *340*, 234–248. [CrossRef]
5. Iticha, B.; Takele, C. Digital soil mapping for site-specific management of soils. *Geoderma* **2019**, *351*, 85–91. [CrossRef]
6. Potdar, R.P.; Shirolkar, M.M.; Verma, A.J.; More, P.S.; Kulkarni, A. Determination of soil nutrients (NPK) using optical methods: A mini review. *J. Plant Nutr.* **2021**, *44*, 1826–1839. [CrossRef]
7. Masrie, M.; Rosman, M.S.A.; Sam, R.; Janin, Z. Detection of nitrogen, phosphorus, and potassium (NPK) nutrients of soil using optical transducer. In Proceedings of the 2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA), Putrajaya, Malaysia, 28–30 November 2017; pp. 1–4.
8. Dharumarajan, S.; Hegde, R.; Janani, N.; Singh, S.K. The need for digital soil mapping in India. *Geoderma Reg.* **2019**, *16*, e204. [CrossRef]
9. Zandi, S.; Ghobakhlou, A.; Sallis, P. Evaluation of spatial interpolation techniques for mapping soil pH. In Proceedings of the International Congress on Modelling and Simulation (MODSIM 2011), Perth, Australia, 12–16 December 2011.
10. Robinson, T.P.; Metternicht, G. Testing the performance of spatial interpolation techniques for mapping soil properties. *Comput. Electron. Agric.* **2006**, *50*, 97–108. [CrossRef]
11. Bogunovic, I.; Mesic, M.; Zgorelec, Z.; Jurisic, A.; Bilandzija, D. Spatial variation of soil nutrients on sandy-loam soil. *Soil Tillage Res.* **2014**, *144*, 174–183. [CrossRef]
12. Lu, P.; Wang, L.; Niu, Z.; Li, L.; Zhang, W. Prediction of soil properties using laboratory VIS-NIR spectroscopy and Hyperion imagery. *J. Geochem. Explor.* **2013**, *132*, 26–33. [CrossRef]
13. McCarty, G.W.; Reeves, J.B. Comparison of near infrared and mid infrared diffuse reflectance spectroscopy for field-scale measurement of soil fertility parameters. *Soil Sci.* **2006**, *171*, 94–102. [CrossRef]
14. Zhang, Q.; Yang, Z.; Li, Y.; Chen, D.; Zhang, J.; Chen, M. Spatial variability of soil nutrients and GIS-based nutrient management in Yongji County, China. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 965–981. [CrossRef]
15. Yang, Y.; Zhang, S. Approach of developing spatial distribution maps of soil nutrients. In *Computer and Computing Technologies in Agriculture, Volume I: First IFIP TC 12 International Conference on Computer and Computing Technologies in Agriculture (CCTA 2007), Wuyishan, China, 18–20 August 2007*; Springer: Boston, MA, USA, 2008; pp. 565–571.
16. Maselli, F.; Battista, P.; Chiesi, M.; Rapi, B.; Gozzini, B. Use of Sentinel-2 MSI data to monitor crop irrigation in Mediterranean areas. *Int. J. Appl. Earth Obs.* **2020**, *93*, 102216. [CrossRef]
17. Pellegrini, P.; Cossani, C.M.; Bella, C.; Pieiro, G.; Oesterheld, M. Simple regression models to estimate light interception in wheat crops with Sentinel-2 and a handheld sensor. *Crop Sci.* **2020**, *60*, 1607–1616. [CrossRef]

18. Zhang, Y.; Yang, J.; Du, L. Analyzing the effects of hyperspectral ZhuHai-1 band combinations on LAI estimation based on the PROSAIL model. *Sensors* **2021**, *21*, 1869. [CrossRef]

19. Du, S.; Huang, H.; He, F.; Luo, H.; Yin, Y.; Li, X.; Xie, L.; Guo, R.; Tang, S. Unsupervised stepwise extraction of offshore aquaculture ponds using super-resolution hyperspectral images. *Int. J. Appl. Earth Obs.* **2023**, *119*, 103326. [CrossRef]

20. Shi, T.; Cui, L.; Wang, J.; Fei, T.; Chen, Y.; Wu, G. Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy. *Plant Soil* **2013**, *366*, 363–375. [CrossRef]

21. Munawar, A.A.; Yunus, Y.; Devianti; Satriyo, P. Calibration models database of near infrared spectroscopy to predict agricultural soil fertility properties. *Data Brief* **2020**, *30*, 105469. [CrossRef]

22. Song, Y.; Shen, Z.; Wu, P.; Viscarra Rossel, R.A. Wavelet geographically weighted regression for spectroscopic modelling of soil properties. *Sci. Rep.* **2021**, *11*, 17503. [CrossRef]

23. Panday, D.; Maharjan, B.; Chalise, D.; Shrestha, R.K.; Twanabasu, B. Digital soil mapping in the Bara district of Nepal using kriging tool in ArcGIS. *PLoS ONE* **2018**, *13*, e206350. [CrossRef]

24. Guo, P.; Li, T.; Gao, H.; Chen, X.; Cui, Y.; Huang, Y. Evaluating Calibration and Spectral Variable Selection Methods for Predicting Three Soil Nutrients Using Vis-NIR Spectroscopy. *Remote Sens.* **2021**, *13*, 4000. [CrossRef]

25. Song, Y.; Zhao, X.; Su, H.; Li, B.; Hu, Y.; Cui, X. Predicting spatial variations in soil nutrients with hyperspectral remote sensing at regional scale. *Sensors* **2018**, *18*, 3086. [CrossRef] [PubMed]

26. Li, X.; Chen, W.; Cheng, X.; Wang, L. A Comparison of Machine Learning Algorithms for Mapping of Complex Surface-Mined and Agricultural Landscapes Using ZiYuan-3 Stereo Satellite Imagery. *Remote Sens.* **2016**, *8*, 514. [CrossRef]

27. John, K.; Abraham Isong, I.; Michael Kebonye, N.; Okon Ayito, E.; Chapman Agyeman, P.; Marcus Afu, S. Using Machine Learning Algorithms to Estimate Soil Organic Carbon Variability with Environmental Variables and Soil Nutrient Indicators in an Alluvial Soil. *Land* **2020**, *9*, 487. [CrossRef]

28. Yiming, X.; Scot, E.S.; Grunwald, S.; Abd-Elrahman, A.; Wani, S.; Nair, V. Estimating soil total nitrogen in smallholder farm settings using remote sensing spectral indices and regression kriging. *Catena* **2018**, *163*, 111–122. [CrossRef]

29. Perkins, T. Speed and accuracy improvements in FLAASH atmospheric correction of hyperspectral imagery. *Opt. Eng.* **2012**, *51*, 111707. [CrossRef]

30. Module, F. Atmospheric correction module: Quac and flaash user's guide. *Version* **2009**, *4*, 44.

31. Hengl, T.; Leenaars, J.G.B.; Shepherd, K.D.; Walsh, M.G.; Heuvelink, G.B.M.; Mamo, T.; Tilahun, H.; Berkhout, E.; Cooper, M.; Fegraus, E. Soil nutrient maps of Sub-Saharan Africa: Assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutr. Cycl. Agroecosyst.* **2017**, *109*, 77–102. [CrossRef]

32. Zhang, J.; Zhao, Y.; Xin, Y. Changes in and evaluation of surface soil quality in *Populus × xiaohei* shelterbelts in midwestern Heilongjiang province, China. *J. For. Res.* **2021**, *32*, 1221–1233. [CrossRef]

33. Zhang, L.; Liu, Z.; Liu, D.; Xiong, Q.; Yang, N.; Ren, T.; Zhang, C.; Zhang, X.; Li, S. Crop Mapping Based on Historical Samples and New Training Samples Generation in Heilongjiang Province, China. *Sustainability* **2019**, *11*, 5052. [CrossRef]

34. Li, L.; Wang, K.; Chen, W.; Zhao, Q.; Liu, L.; Liu, W.; Liu, Y.; Jiang, J.; Liu, J.; Zhang, M. Atmospheric pollution of agriculture-oriented cities in Northeast China: A case in Suihua. *J. Environ. Sci.* **2020**, *97*, 85–95. [CrossRef]

35. Li, X.; Shang, B.; Wang, D.; Wang, Z.; Wen, X.; Kang, Y. Mapping soil organic carbon and total nitrogen in croplands of the Corn Belt of Northeast China based on geographically weighted regression kriging model. *Comput. Geosci.* **2020**, *135*, 104392. [CrossRef]

36. Shao, H.; Huang, X.; Wang, R.; Eminniyaz, A.; Wang, J.; Wu, S. Potential allelopathic effects of Xanthium italicum Moretti on wheat. *J. Med. Plants Res.* **2013**, *7*, 587–592.

37. Isola, C.; Drusch, M.; Gascon, F.; Martimort, P.; Bello, U.D.; Spoto, F.; Sy, O.; Laberinti, P. Sentinel-2 Optical High Resolution Mission for GMES Land Operational Services. In Proceedings of the IEEE International Geoscience & Remote Sensing Symposium, Cape Town, South Africa, 12–17 July 2009.

38. Novelli, A.; Aguilar, M.A.; Nemmaoui, A.; Aguilar, F.J.; Tarantino, E. Performance evaluation of object based greenhouse detection from Sentinel-2 MSI and Landsat 8 OLI data: A case study from Almería (Spain). *Int. J. Appl. Earth Obs.* **2016**, *52*, 403–411.

39. Zhang, Q.; Zhang, P.; Hua, X. Unsupervised GRNN flood mapping approach combined with uncertainty analysis using bi-temporal Sentinel-2 MSI imageries. *Int. J. Digit. Earth* **2021**, *14*, 1561–1581. [CrossRef]

40. Chemura, A.; Mutanga, O.; Odindi, J.; Kutywayo, D. Mapping spatial variability of foliar nitrogen in coffee (*Coffea arabica* L.) plantations with multispectral Sentinel-2 MSI data. *ISPRS J. Photogramm. Remote Sens.* **2018**, *138*, 1–11. [CrossRef]

41. Qin, P.; Cai, Y.; Wang, X. Small waterbody extraction with improved U-Net using Zhuhai-1 hyperspectral remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]

42. Mashaba-Munghemezulu, Z.; Chirima, G.J.; Munghemezulu, C. Modeling the Spatial Distribution of Soil Nitrogen Content at Smallholder Maize Farms Using Machine Learning Regression and Sentinel-2 Data. *Sustainability* **2021**, *13*, 11591. [CrossRef]

43. Wang, S.; Adhikari, K.; Zhuang, Q.; Yang, Z.; Jin, X.; Wang, Q.; Bian, Z. An improved similarity-based approach to predicting and mapping soil organic carbon and soil total nitrogen in a coastal region of northeastern China. *PeerJ* **2020**, *8*, e9126. [CrossRef]

44. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plant.* **1999**, *106*, 135–141. [CrossRef]

45. Fernández-Manso, A.; Fernández-Manso, O.; Quintano, C. SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. *Int. J. Appl. Earth Obs.* **2016**, *50*, 170–175. [CrossRef]

46. Chen, J.M. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Can. J. Remote Sens.* **2014**, *22*, 229–242. [CrossRef]
47. Miura, T.; Huete, A.R.; Yoshioka, H. Evaluation of sensor calibration uncertainties on vegetation indices for MODIS. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 1399–1409. [CrossRef]
48. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]
49. Madeira, J.; Bedidi, A.; Cervelle, B.; Pouget, M.; Flay, N. Visible spectrometric indices of hematite (Hm) and goethite (Gt) content in lateritic soils: The application of a Thematic Mapper (TM) image for soil-mapping in Brasilia, Brazil. *Int. J. Remote Sens.* **2010**, *18*, 2835–2852. [CrossRef]
50. Bullard, J.E. Quantifying iron oxide coatings on dune sands using spectrometric measurements: An example from the Simpson-Strzelecki Desert, Australia. *J. Geophys. Res.* **2002**, *107*, ECV 5-1–ECV 5-11. [CrossRef]
51. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
52. Peng, J.; Manevski, K.; Krup, K.; Larsen, R.; Andersen, M.N. Random forest regression results in accurate assessment of potato nitrogen status based on multispectral data from different platforms and the critical concentration approach. *Field Crops Res.* **2021**, *268*, 108158. [CrossRef]
53. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2007**, *26*, 217–222. [CrossRef]
54. Patel, M.K.; Ryu, D.; Western, A.W.; Fitzgerald, G.; Young, I. *Mapping Canopy Nitrogen Concentration across Ryegrass and Barley Crop Using Random Forest Regression*; American Geophysical Union (AGU): Washington, DC, USA, 2021.
55. Hutengs, C.; Vohland, M. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sens. Environ.* **2016**, *178*, 127–141. [CrossRef]
56. Liang, L.; Di, L.; Huang, T.; Wang, J.; Lin, L.; Wang, L.; Yang, M. Estimation of leaf nitrogen content in wheat using new hyperspectral indices and a random forest regression algorithm. *Remote Sens.* **2018**, *10*, 1940. [CrossRef]
57. Emadi, M.; Taghizadeh-Mehrjardi, R.; Cherati, A.; Danesh, M.; Mosavi, A.; Scholten, T. Predicting and Mapping of Soil Organic Carbon Using Machine Learning Algorithms in Northern Iran. *Remote Sens.* **2020**, *12*, 2234. [CrossRef]
58. Shah, S.H.; Angel, Y.; Houborg, R.; Ali, S.; Mccabe, M.F. A Random Forest Machine Learning Approach for the Retrieval of Leaf Chlorophyll Content in Wheat. *Remote Sens.* **2019**, *11*, 920. [CrossRef]
59. Giannakas, F.; Troussas, C.; Krouska, A.; Sgouropoulou, C.; Voyiatzis, I. XGBoost and Deep Neural Network Comparison: The Case of Teams' Performance. In *Intelligent Tutoring Systems: 17th International Conference, ITS 2021, Virtual Event, 7–11 June 2021, Proceedings 17*; Cristea, A.I., Troussas, C., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 343–349.
60. Zamani Joharestani, M.; Cao, C.; Ni, X.; Bashir, B.; Talebiesfandarani, S. PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. *Atmosphere* **2019**, *10*, 373. [CrossRef]
61. Dvornikov, Y.; Slukovskaya, M.; Yaroslavtsev, A.; Meshalkina, J.; Ryazanov, A.; Sarzhanov, D.; Vasenev, V. High-resolution mapping of soil pollution by Cu and Ni at a polar industrial barren area using proximal and remote sensing. *Land Degrad. Dev.* **2022**, *33*, 1731–1744. [CrossRef]
62. Miao, J.; Zhen, J.; Wang, J.; Zhao, D.; Jiang, X.; Shen, Z.; Gao, C.; Wu, G. Mapping Seasonal Leaf Nutrients of Mangrove with Sentinel-2 Images and XGBoost Method. *Remote Sens.* **2022**, *14*, 3679. [CrossRef]
63. Georganos, S.; Grippa, T.; Vanhuysse, S.; Lennert, M.; Shimoni, M.; Wolff, E. Very High Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 607–611. [CrossRef]
64. Liashchynskyi, P.; Liashchynskyi, P. Grid search, random search, genetic algorithm: A big comparison for NAS. *arXiv* **2019**, arXiv:1912.06059.
65. Yue, Z.; Biao, S.; Hai-ou, S.; Ling, O. Mapping stocks of soil total nitrogen using remote sensing data: A comparison of random forest models with different predictors. *Comput. Electron. Agric.* **2019**, *160*, 23–30. [CrossRef]
66. Candiago, S.; Remondino, F.; De Giglio, M.; Dubbini, M.; Gattelli, M. Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote Sens.* **2015**, *7*, 4026–4047. [CrossRef]
67. Kawamura, K.; Tsujimoto, Y.; Nishigaki, T.; Andriamananjara, A.; Rabenarivo, M.; Asai, H.; Rakotoson, T.; Razafimbelo, T. Laboratory Visible and Near-Infrared Spectroscopy with Genetic Algorithm-Based Partial Least Squares Regression for Assessing the Soil Phosphorus Content of Upland and Lowland Rice Fields in Madagascar. *Remote Sens.* **2019**, *11*, 506. [CrossRef]