

## Article

# Estimation of Soil Nutrient Content Using Hyperspectral Data

Yiping Peng<sup>1,†</sup>, Lu Wang<sup>1,2,3,4,5,†</sup>, Li Zhao<sup>1</sup>, Zhenhua Liu<sup>1,3,4,5</sup> , Chenjie Lin<sup>1</sup>, Yueming Hu<sup>1,2,3,4,5,\*</sup>  
and Luo Liu<sup>1,3,4,5</sup> 

- <sup>1</sup> College of Natural Resources and Environment, South China Agricultural University, Guangzhou 510642, China; pengyp@stu.scau.edu.cn (Y.P.); selinapple@scau.edu.cn (L.W.); zhaoli@stu.scau.edu.cn (L.Z.); zhenhua@scau.edu.cn (Z.L.); linchenjie@stu.scau.edu.cn (C.L.); liuluo@scau.edu.cn (L.L.)
- <sup>2</sup> College of Tropical Crops, Hainan University, Haikou 570228, China
- <sup>3</sup> Guangdong Provincial Key Laboratory of Land Use and Consolidation, South China Agricultural University, Guangzhou 510642, China
- <sup>4</sup> Guangdong Province Engineering Research Center for Land Information Technology, South China Agricultural University, Guangzhou 510642, China
- <sup>5</sup> Key Laboratory of Construction Land Transformation, Ministry of Land and Resources, South China Agricultural University, Guangzhou 510642, China
- \* Correspondence: ymhu@scau.edu.cn; Tel.: +86-020-852-88307
- † These authors contributed equally to this work.

**Abstract:** Soil nutrients play a vital role in plant growth and thus the rapid acquisition of soil nutrient content is of great significance for agricultural sustainable development. Hyperspectral remote-sensing techniques allow for the quick monitoring of soil nutrients. However, at present, obtaining accurate estimates proves to be difficult due to the weak spectral features of soil nutrients and the low accuracy of soil nutrient estimation models. This study proposed a new method to improve soil nutrient estimation. Firstly, for obtaining characteristic variables, we employed partial least squares regression (PLSR) fit degree to select an optimal screening algorithm from three algorithms (Pearson correlation coefficient, PCC; least absolute shrinkage and selection operator, LASSO; and gradient boosting decision tree, GBDT). Secondly, linear (multi-linear regression, MLR; ridge regression, RR) and nonlinear (support vector machine, SVM; and back propagation neural network with genetic algorithm optimization, GABP) algorithms with 10-fold cross-validation were implemented to determine the most accurate model for estimating soil total nitrogen (TN), total phosphorus (TP), and total potassium (TK) contents. Finally, the new method was used to map the soil TK content at a regional scale using the soil component spectral variables retrieved by the fully constrained least squares (FCLS) method based on an image from the HuanJing-1A Hyperspectral Imager (HJ-1A HSI) of the Conghua District of Guangzhou, China. The results identified the GBDT-GABP was observed as the most accurate estimation method of soil TN ( $R_{cv}^2$  of 0.69, the root mean square error of cross-validation (RMSECV) of 0.35 g kg<sup>-1</sup> and ratio of performance to interquartile range (RPIQ) of 2.03) and TP ( $R_{cv}^2$  of 0.73, RMSECV of 0.30 g kg<sup>-1</sup> and RPIQ = 2.10), and the LASSO-GABP proved to be optimal for soil TK estimations ( $R_{cv}^2$  of 0.82, RMSECV of 3.39 g kg<sup>-1</sup> and RPIQ = 3.57). Additionally, the highly accurate LASSO-GABP-estimated soil TK ( $R^2 = 0.79$ ) reveals the feasibility of the LASSO-GABP method to retrieve soil TK content at the regional scale.

**Keywords:** VIS-NIR spectroscopy; screening algorithm; estimation model; HJ-1A imagery



**Citation:** Peng, Y.; Wang, L.; Zhao, L.; Liu, Z.; Lin, C.; Hu, Y.; Liu, L. Estimation of Soil Nutrient Content Using Hyperspectral Data. *Agriculture* **2021**, *11*, 1129. <https://doi.org/10.3390/agriculture11111129>

Academic Editors: Gniewko Niedbała and Sebastian Kujawa

Received: 9 October 2021  
Accepted: 9 November 2021  
Published: 11 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The rapid and efficient monitoring of soil nutrients has become an important prerequisite for agricultural production management and ensuring the healthy development of crops. However, current soil nutrient estimations are often obtained using field sampling and laboratory analysis, which is time-consuming and costly. The monitoring of soil nutrients via hyperspectral remote-sensing techniques is rapid and efficient, and numerous related studies have been performed within the past 30 years [1–4].

Current research on the retrieval of soil nutrients via hyperspectral remote-sensing technology typically focuses on two factors: the determination of characteristic variables and the construction of the estimation model. The determination of suitable characteristic variables ensures high-precision estimations. Statistical analyses (e.g., Pearson correlation coefficient (PCC) and partial least squares regression (PLSR)) are frequently employed to determine these variables [5,6]. For example, Liu et al. (2007) obtained the 620–810 nm characteristic variables of soil organic matter by correlation and multiple regression analyses [5]. Vibhute et al. (2019) determined the characteristic variables of soil nitrogen at 480, 511, 653, 997, 1472, 1795, 2210, and 2296 nm based on correlation analysis [6]. However, statistical variable selection methods in high-dimensional space can fail due to a lack of significance testing and parameter estimations in the model [7]. Following the development of data mining technology, several machine learning algorithms have been introduced to determine the characteristic variables of soil nutrients. Zhang et al. (2019) proposed a method combining mutual information and ant colony optimization to select soil total nitrogen (TN) characteristic bands at 943, 1004, 1097, 1351, 1550, 1710, 2123, and 2254 nm [8]. Despite the great progress made by these studies, determining the characteristic variables remains to be difficult due to the weak spectral responses to soil nutrients. Therefore, additional screening algorithms, particularly machine learning approaches, are required in order to accurately determine the characteristic variables.

Existing relationship models between spectral variables and soil nutrient contents can be classified into two categories; linear and nonlinear models. Linear estimation methods build linear mathematical relationships between spectral variables and soil nutrient contents. Multiple linear regression (MLR) and partial least squares regression (PLSR) are the most commonly used linear estimation methods for soil nutrients [9–11]. However, correlations between spectral variables and soil nutrients are rarely linear in nature [12]. Thus, machine learning models were introduced to solve this problem. The random forest (RF), support vector machine (SVM), and back propagation neural network (BPNN) algorithms are frequently employed to estimate soil nutrients [13–15]. Compared with linear models, nonlinear methods improve on the explanatory power of the spectral changes related to soil nutrients. However, large-scale training samples for SVM approaches are difficult to obtain and implement due to their complexity, huge memory requirements, and extensive computational time in quadratic programming routines [16]. In addition, RF is prone to overfitting in regression models when learning specific details and noise in the training data [17,18]. BPNN is associated with large weights and threshold uncertainties, affecting the estimation accuracy [19,20]. Therefore, there is a great need to determine an optimal algorithm for high accuracy soil nutrient content estimations.

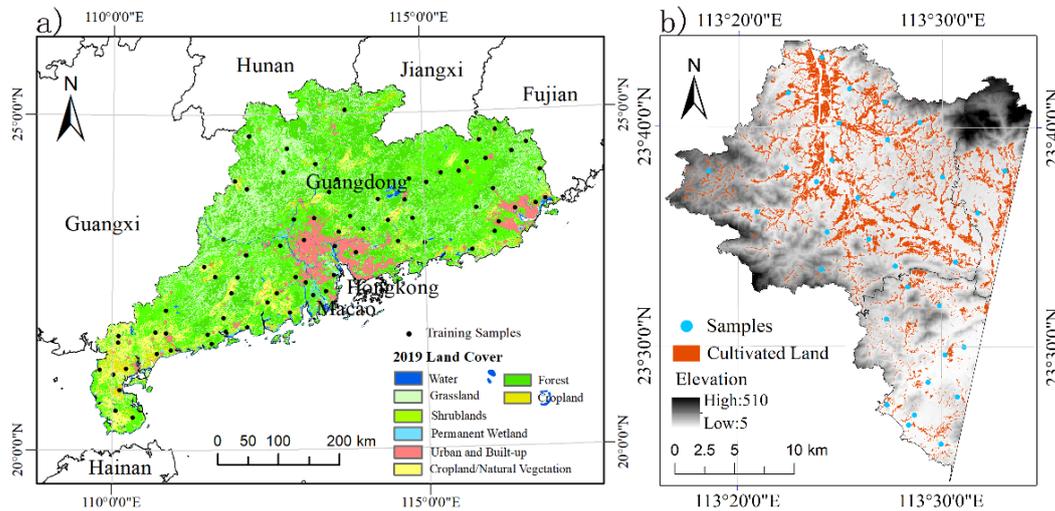
This study has the aim of developing a new method to accurately estimate soil nutrient contents. In order to achieve this aim, we set the following objectives: (1) to determine the optimal screening algorithm from three algorithms (Pearson correlation coefficient, PCC; least absolute shrinkage and selection operator, LASSO; and gradient boosting decision tree, GBDT) for the accurate selection of soil nutrient characteristic variables; (2) to implement the MLR, ridge regression (RR), back propagation neural network with genetic algorithm optimization (GABP), and SVM to determine a high-accuracy model for the estimation of soil nutrient contents; and (3) to apply a high-accuracy method to map the soil nutrient contents at the regional scale using HuanJing-1A Hyperspectral Imager (HJ-1A HSI) imagery. Both the hyperspectral data and HJ-1A HSI images were collected in the Guangdong province and Conghua District of Guangzhou, China.

## 2. Materials and Methods

### 2.1. Study Area

Guangdong province, China was selected as the study area in order to build the optimal hyperspectral estimation model of soil nutrients (Figure 1a), while Conghua District within Guangdong was selected to map the soil nutrient contents (Figure 1b). The East-West and North-South spans of Guangdong province are approximately 800 and

600 km, respectively. The province belongs to the East Asian monsoon region, with middle subtropical, south subtropical, and north tropical zone climate types from the north to south. Mean annual temperature and precipitation of the area are 21.8 °C and 1789.3 mm, respectively. Guangdong is an important grain production region, with a crop planting area of  $4.28 \times 10^4 \text{ km}^2$  in 2019 and total grain yield of  $1.19 \times 10^{10} \text{ kg}$ .



**Figure 1.** (a) MODIS land cover map from the MCD12 product of the study area with a spatial distribution of 75 soil samples; (b) the test study area determined from the 2016 Cultivated Land Map planted with rice of the Conghua National Land Department, with a spatial distribution of 33 soil samples used to assess the accuracy of the estimated soil nutrient content.

## 2.2. Data and Pre-Processing

### 2.2.1. Soil Sampling and Chemical Analysis

A total of 75 soil samples were gathered for constructing hyperspectral estimate models of soil nutrients contents based on a  $50 \times 50 \text{ km}$  sampling grid within Guangdong province and field actual conditions to ensure uniform distribution of the soil samples [21,22]. Surface soil samples (0–20 cm) were collected at five sampling locations at each site. To remove stones and other large debris, the samples were air-dried and sieved through a 2 mm polyethylene sieve. After that, the samples were pulverized into fine powder. The soil nutrient content and soil spectral reflectance were then determined by dividing each sample into two parts. Soil TN was measured using the semi-micro Kjeldahl method described by Walkley and Black [23]. Soil TP and TK were determined via an ultraviolet spectrophotometer (UV-2600, Shimadzu CO, LTD., Kyoto, Japan) and a flame photometer (FP640, INESA Analytical Instrument CO, LTD., Shanghai, China), respectively. The soil nutrient content statistics from the 75 soil samples are presented in Table 1.

**Table 1.** Statistic information for soil nutrient contents in the study area.

Soil Nutrients	Min	Q1	Median	Q3	Max	Mean	SD	Skewness	Kurtosis	CV
TN	0.21	0.99	1.33	1.70	2.79	1.36	0.57	0.43	0.21	41.91
TP	0.13	0.37	0.59	1.00	3.15	0.75	0.55	1.90	5.21	73.33
TK	0.62	4.75	9.66	16.84	30.39	10.55	7.61	0.61	−0.23	72.13

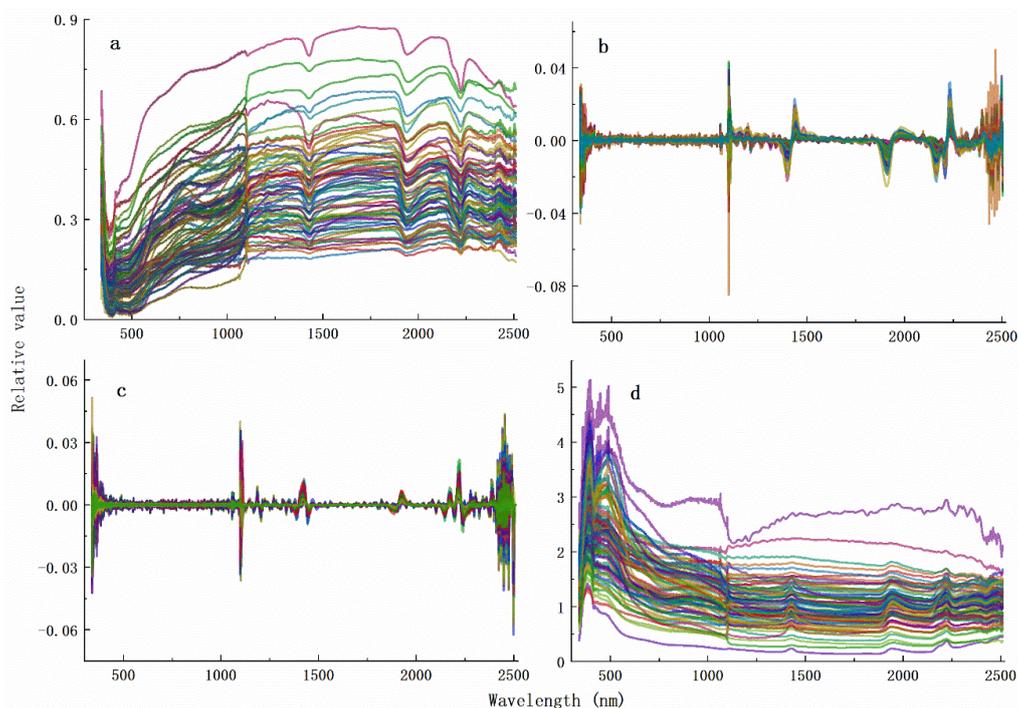
Note: soil total nitrogen, TN; total phosphorus, TP; and total potassium, TK; unit:  $\text{g kg}^{-1}$ . Q1, first quartile; Q3, third quartile; SD, standard deviation; CV, coefficient of variation (%).

Moreover, the mapping of Guangdong Province needs multiple HJ-1A images with 100 m spatial resolution. In addition, it is very difficult to obtain multiple high-quality satellite images of the whole province on the same day. Therefore, in this study, Conghua district was selected for conducting the soil nutrient mapping experiment. A total of

33 soil samples were collected in Conghua District (Figure 1b) to verify the feasibility of mapping soil nutrient content at the regional scale. The acquisition time of the HJ-1A HSI image coincided with the collection of the samples, which are evenly distributed in the whole image. The soil sample collection principle and pretreatment are consistent with the Guangdong province samples.

### 2.2.2. Spectral Measurements and Pre-Processing of Soil Samples

Soil spectral measurements were performed on 75 soil samples collected across the province. An AvaField portable spectrometer (Avantes, Inc., Apeldoorn, Holland) was used to measure soil spectral reflectance, which has a spectral range and resolution of 340–2511 and 0.6 nm, respectively. The spectral measurements were carried out in a dark room to regulate the lighting environment and minimize the influence of stray light. The soil spectral reflectance values were measured using a 50 W halogen lamp with a  $10^\circ$  field of view in vertical contact with the soil sample. Each sample was uniformly tiled on a black cloth and measured five times. The average spectrum was calculated and used in further processing. Prior to the collection of the reflectance readings, the spectrometer was calibrated every three samples with a white Spectralon. To decrease signal noise, we used Savitzky–Golay smoothing with a window size of 10. In addition, the smoothed spectral data (raw spectral, R) were processed with the first derivative (FD), second derivative (SD), and reciprocal logarithmic (RL) to eliminate or reduce the effect of background noise and account for signal intensity fluctuations induced by soil surface spectral scattering and absorption. The outcomes of the processing are shown in Figure 2.

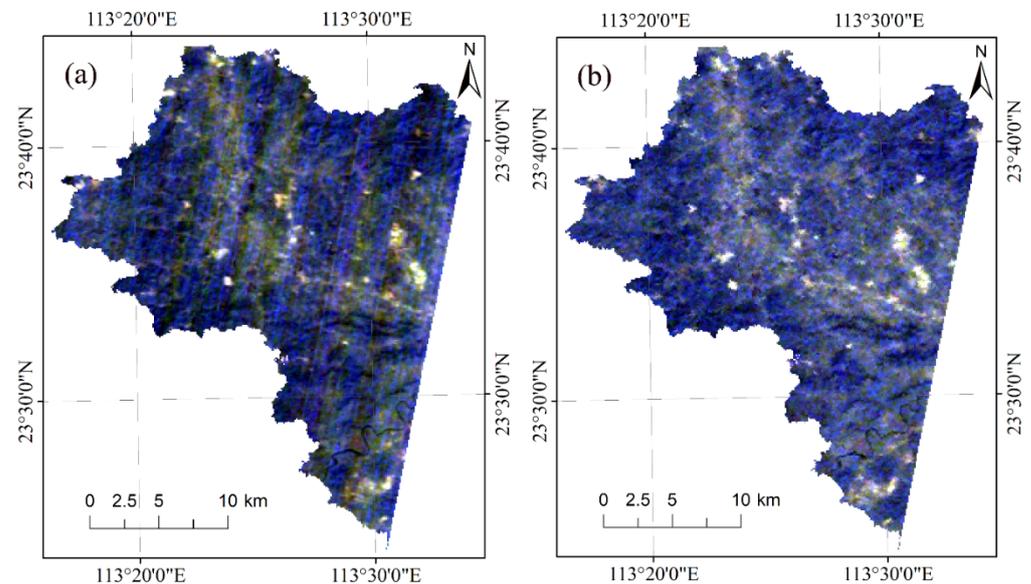


**Figure 2.** Transformed spectral indices of soil samples: (a) raw spectral curves; (b) first derivative spectral curves; (c) second derivative spectral curves; and (d) reciprocal logarithmic spectral curves.

### 2.2.3. Image Acquisition and Pre-Processing

In order to extend the application of the established model at the regional scale, a HJ-1A image acquired on 30 October 2017 with a 100 m spatial resolution and 115 bands (459–956 nm) was used to map the soil nutrient contents. The image was subjected to radiometric correction, atmospheric correction, geometric precision correction, and stripe noise reduction (Figure 3) using ENVI 5.3 (Exelis Visual Information Solutions, Inc., Boulder,

CO, USA). The image's spectral resolution was 5 nm, which was substantially coarser than the AvaField portable spectrometer's measured spectral interval of 0.6 nm. ENVI's spectral resampling technique was utilized to spectrally resample the measured soil spectral data gathered with the AvaField portable spectrometer in order to match the spectral resolution of the HJ-1A HSI data.



**Figure 3.** The HJ-1A image: (a) untreated and (b) stripe noise reduction.

### 2.3. Methods

This section is organized into four parts. In Section 2.3.1, we describe how the optimal algorithm of screening the soil nutrient characteristic variables can be determined using the PLSR fit degrees. The second section explains how the optimal prediction algorithm for soil nutrients can be screened from four algorithms by their prediction accuracy. In Section 2.3.3, we detail the mapping of soil nutrient base on HJ-1A image data using the above the optimal screening and predicting algorithms. Section 2.3.4 describes accuracy validation methods for the predicting models and mapping.

#### 2.3.1. Determining the Optimal Screening Algorithm of the Characteristic Variables

One of the most important steps in the development of the optimal hyperspectral estimation method of the soil nutrient contents was the determination of the characteristic variables [8,24,25]. Additionally, the determination of the accurate screening algorithms is key for characteristic variables of the soil nutrient content. In order to determine the optimal screening algorithm of the characteristic variables, we compared traditional linear screening algorithm (PCC) and nonlinear screening algorithms (LASSO and GBDT) based on two evaluation steps. First, the characteristic variables were screened using PCC, GBDT and LASSO might be correlated with each other. That is, there are collinearities among the variables. Therefore, the variance inflation factor (VIF) of a stepwise regression was applied to eliminate the collinearity of the selected characteristic variables. The set of variables having a VIF lower than 10 [26] was retained. The three screening algorithms are described in detail as follows:

**LASSO:** The least absolute shrinkage and selection operator, proposed by Tibshirani (1996), minimizes the sum of squares of residuals under the constraint that the sum of the absolute values of the regression coefficients (penalty coefficient) is less than a pre-defined constant. This produces regression coefficients (RC) strictly equal to 0 and removes low-

weight variables and can therefore effectively deal with complex high-dimensional data problems [27]. LASSO can be defined as follows:

$$\begin{aligned} \arg \min_{\mathbf{B}} \left\{ \sum_{j=1}^n y_j - \sum_{j=1}^p x_{ij} B_j \right\}, \\ \text{subject to } \sum_{j=1}^p |B_j| \leq t, \end{aligned} \quad (1)$$

where  $y_i$  represents the measured spectral data in the  $i$ th band;  $n$  is the spectral dimensionality;  $B_j$  denotes the input weight in the  $j$ th spectral sample;  $x_{ij}$  is the covariate vector of the  $i$ th measured spectral data and  $j$  spectral sample; and  $p$  is the spectral sample number.

GBDT: The gradient boosting decision tree is a boosting algorithm that calculates the information gain during the branching of the decision tree to determine the spectral variable to be split and the corresponding split value. Once all decision trees are constructed, the feature importance (FI) is obtained by calculating the information gain of the decision tree feature and dividing by the total frequency of the feature in all trees of the GBDT strong learner [28]:

$$\text{FI} = \frac{\sum I(a, D)}{N_a}, \quad (2)$$

where  $I(a, D)$  denotes the feature (spectral variable) information gain;  $a$  is the feature;  $D$  is the soil sample; and  $N_a$  is the total frequency of feature  $a$  in all trees.

PCC: The Pearson correlation coefficient is commonly employed to screen characteristic variables. Here, the PCC was implemented between the spectral variables and soil nutrient content to determine characteristic variables with the largest correlation coefficient ( $p \leq 0.05$  significance level). The Pearson correlation coefficient can be expressed as:

$$r_i = \frac{\sum_{n=1}^N (R_{ni} - \bar{R}_i)(y_n - \bar{y})}{\sqrt{\sum_{n=1}^N (R_{ni} - \bar{R}_i)^2 \sum_{n=1}^N (y_n - \bar{y})^2}}, \quad (3)$$

where  $R_{ni}$  is the spectral value of the  $i$ th spectral variable of the  $n$ th soil sample point;  $\bar{R}_i$  is the average spectral value of the  $i$ th spectral variable;  $y_n$  is the soil nutrient content of the  $n$ th soil sample point; and  $\bar{y}$  is the average value of the soil nutrient content.

Once the characteristic variables were selected by each algorithm, PLSR fit degrees ( $R^2$ ) [29–31] between the measured soil nutrient contents and characteristic variables were compared. The screening algorithm with the maximum fit degree was determined as the optimal.

### 2.3.2. Determining the Accurate Model for Estimating Soil Nutrients

In this study, the screened characteristic variables were used as independent variables and each of the soil nutrient (TN, TP and TK) contents were used as the dependent variable. Additionally, four different algorithms were applied to build the relationship models between characteristic variables and soil nutrients: MLR, RR, SVM, and GABP. The four algorithms are described in detail as follows:

#### (1) Multi-Linear Regression

Multi-linear regression is a type of regression analysis for multiple independent variables. The optimal combination of these independent variables is taken to estimate the dependent variables. This model can describe the influence of each variable on the soil properties and is widely used in soil property estimations [32–34]. We adopted MLR to estimate soil nutrient content using the following formula:

$$Z_{MLR} = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n, \quad (4)$$

where  $Z_{MLR}$  is the dependent variable (soil nutrient content);  $x_i$  ( $i = 1, 2, \dots, n$ ) is the independent variable (spectral variables);  $a_i$  ( $i = 1, 2, \dots, n$ ) represents the regression fitting coefficient; and  $a_0$  is the intercept.

(2) Ridge Regression

Ridge regression (RR) is a least square estimation method that improves on its predecessors. In particular, it abandons the unbiasedness of the least square method, thus losing part of the information and reducing the accuracy and making the regression coefficient more realistic and reliable [35]. The existence of multiple collinear relations between independent variables magnifies the mean square error. This error is reduced by using RR estimation rather than the standard least square estimation [36,37]. The RR is expressed as:

$$\hat{\beta}(k) = (X'X + kI)^{-1}X'Y, \tag{5}$$

where  $\hat{\beta}(k)$  is the ridge regression estimate of  $\beta$ ; and  $k$  is the ridge parameter. When  $k = 0$ , the least square estimate of  $\beta$  is equal to  $\hat{\beta}(0)$ .

(3) Support Vector Machine

SVM, proposed by Cortes and Vapnik (1995), is a robust supervised learning model with a capacity for solving practical problems (e.g., nonlinearity and high dimensionality). SVM greatly simplifies the traditional regression process through efficient “transduction inference” from training samples to predictions [38]. The SVM model can be expressed as:

$$f(x) = w_i \cdot \varnothing_i(x) + b, \tag{6}$$

where  $f(x)$  is the soil nutrient estimate;  $x$  is the characteristic variable;  $w_i$  is the weight coefficient;  $b$  is the error term;  $\varnothing_i$  denotes a nonlinear transfer function; and  $w$  and  $b$  are calculated by the following convex optimization problem with an e-insensitivity loss function [39]:

$$\min : \frac{1}{2} \|w\|^2 + C \sum_i^N (\xi_i + \xi_i^*), \tag{7}$$

$$\text{s.t.} \begin{cases} y_i - w\varnothing(x) - b \leq \varepsilon + \xi_i^* \\ w\varnothing(x) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}, (i = 1, \dots, n), \tag{8}$$

where  $\|w\|^2$  represents the flatness of the m-dimensional space;  $\varepsilon$  is a parameter that indicates the maximum allowed error between the measured and estimated values;  $\xi_i$  and  $\xi_i^*$  are slack variables and  $C$  is the penalty factor. Equations (7) and (8) belong to the convex quadratic programming problem with inequality constraints. In order to obtain the Lagrangian multipliers, the equations are converted into a dual problem via the Lagrange multiplier method. The constrained original objective function (Equation (8)) is then transformed into the unconstrained Lagrangian objective function:

$$\min : \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) (\varnothing(x_i)\varnothing(x_j)) + \varepsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i) - \sum_{i=1}^n y_i (\alpha_i^* - \alpha_i), \tag{9}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i^* \leq C, i = 1, \dots, n \end{cases} \tag{10}$$

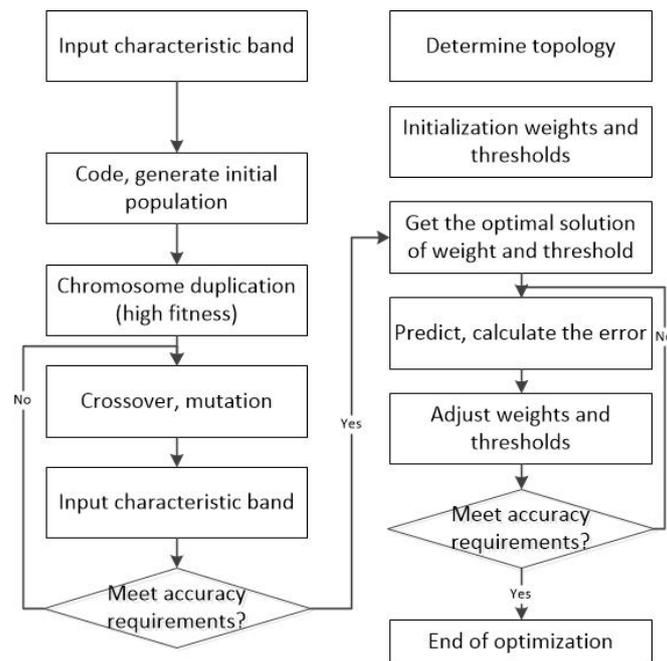
where  $\alpha_i - \alpha_i^*$  is the transformation of  $w$ . The SVM function is expressed as:

$$f(x) = w_i \cdot \varnothing_i(x) + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b, \tag{11}$$

where  $K(x_i, x) = \varnothing(x_i)\varnothing(x_j)$  is the kernel function. The radial basis function was selected as the kernel function.

(4) Genetic Algorithm-Back Propagation Neural Network

The GABP algorithm optimizes the structure and connection weight of the back propagation neural network using the parallel random search ability of the genetic algorithm, effectively avoiding a local optimal solution [40]. We adopted the population search method to optimize the weights and thresholds of the neural network (Figure 4).

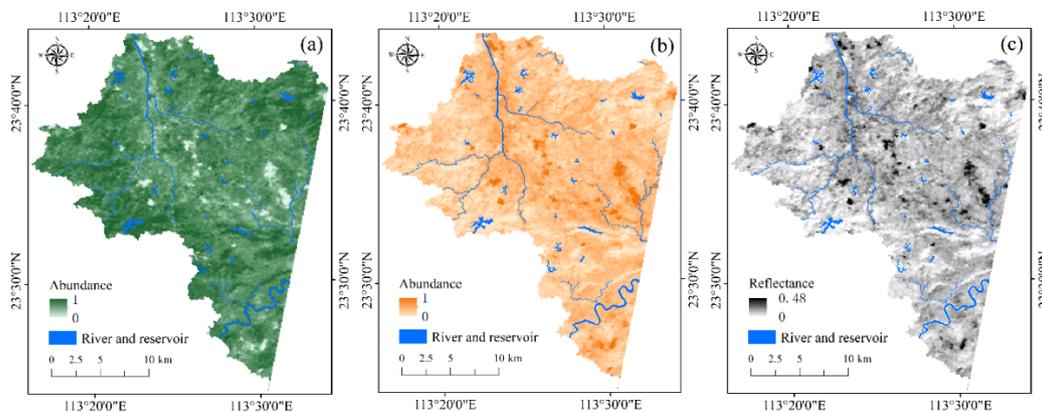


**Figure 4.** Flow chart for the genetic algorithm-back propagation neural network.

### 2.3.3. Estimating Regional-Scale Soil Nutrient Contents Using HJ-1A Hyperspectral Data

Once the optimal variable screening and predictor models were selected, the method was applied to mapping the contents of the soil nutrient using HJ-1A image with 115 bands (459–956 nm) and 5 nm spectral resolution, which will not provide the above characteristic variables with beyond 956 nm wavelength. Thus, the characteristic variables should be re-screened from the resampling measured soil spectral data with 5 nm spectral resolution using the above optimal screening algorithm and to develop the corresponding estimation models. Then, the model was applied to mapping the contents of the soil TK using the HJ-1A HSI image for the Conghua district at the regional scale.

Moreover, in order to apply the methods to Conghua district, the HJ-1A image was considered to contain pure pixels. However, the coarse image spatial resolution of  $100 \times 100$  m generally prevents the existence of pure pixels, with mixed pixels (including crop and soil) typically dominating the study area. Thus, the fully constrained least squares (FCLS) method [41] was used to obtain pure pixels and spectral reflectance of soil (Figure 5).



**Figure 5.** Component decomposition maps of the mixed pixels using FCLS: (a) vegetation abundance maps, (b) soil abundance maps, and (c) spectral reflectance of soil at the 900 nm band of the HJ-1A image.

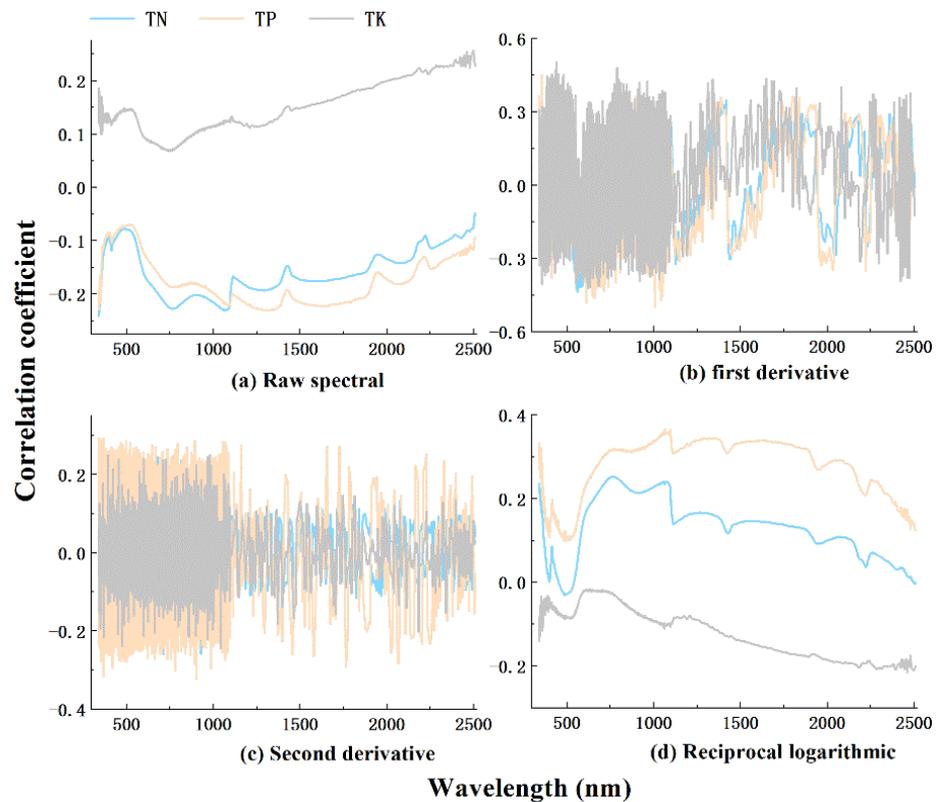
### 2.3.4. Accuracy Validation

The coefficient of determination ( $R^2$ ), concordance correlation coefficient (CCC), ratio of performance to interquartile range (RPIQ), the root mean square error of calibration (RMSEC), and cross-validation (RMSECV) were used as statistical measures to assess the performance of estimation models. The RPIQ is defined as the ratio of IQ to RMSECV [42]. IQ is the interquartile range ( $IQ = Q3 - Q1$ ) of the observed values. Q1 and Q3 denote the first and third quartile, respectively.

## 3. Results

### 3.1. Optimal Algorithm for the Screening of the Characteristic Variables

In order to determine the characteristic variables, the three choosing algorithms (PCC, LASSO, and GBDT) were implemented on 6272 spectral data of the R, FD, SD, and RL (Figure 2) and soil nutrient contents in the 75 sample points collected across the province. Figure 6 illustrates the correlation coefficients of the spectral variables. Stepwise regression with VIF analysis was further used to eliminate the collinearity among the spectral variables screened by the PCC algorithms (Table 2).



**Figure 6.** Correlation coefficients between the soil total nitrogen (TN), total phosphorus (TP), and total potassium (TK) concentrations and the various spectral variables.

**Table 2.** PCC-determined characteristic variables of the three soil nutrients.

Soil Nutrient	Spectral Variables	Correlation Coefficients	VIF
TN	FD <sub>562</sub> , SD <sub>714</sub>	−0.44, −0.26	1.70, 1.51
TP	FD <sub>1009</sub> , FD <sub>356</sub> , SD <sub>905</sub>	−0.50, 0.45, −0.32	2.65, 1.32, 1.42
TK	R <sub>2498</sub> , FD <sub>442</sub>	0.20, 0.50	1.08, 4.14

Considering the possible existence of a nonlinear relationship between the spectral variables and soil nutrient contents, we introduced the GBDT and LASSO algorithms for

the screening task. Numerous experiments were performed, identifying the prediction error of the GBDT algorithm to tend towards stability for screening criteria of soil TN, TP, and TK in the GBDT algorithm equal to  $FI > 0.015$ ,  $FI > 0.015$ , and  $FI > 0.01$ , respectively. Figure 7 depicts the feature importance and regression coefficient of the screened spectral variables. For the LASSO algorithm we employed  $RC \neq 0$  as the screening criteria. Stepwise regression with VIF analysis was further employed to eliminate the collinearity among the spectral variables screened by the GBDT and LASSO algorithms. Table 3 reports the final results of the screening characteristic variables for the three soil nutrients (TN, TP, and TK).

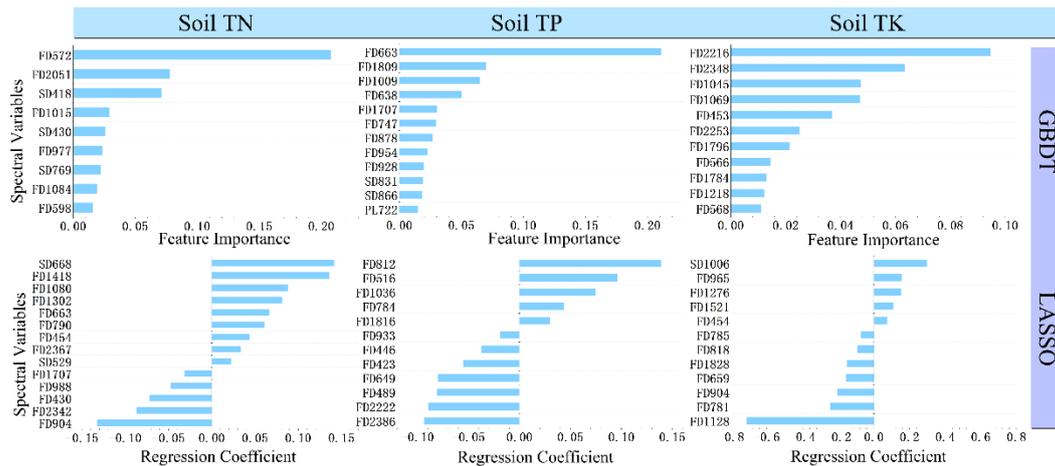


Figure 7. GBDT feature importance and LASSO regression coefficient.

Table 3. GBDT- and LASSO-determined characteristic variables of the three soil nutrients.

Models	Soil Nutrient	Spectral Variables	VIF
LASSO	TN	FD <sub>454</sub> , FD <sub>904</sub> , FD <sub>1302</sub> , FD <sub>1418</sub> , FD <sub>1707</sub> , FD <sub>2342</sub> , FD <sub>2367</sub> , SD <sub>529</sub> , SD <sub>668</sub>	3.14, 2.85, 4.42, 6.35, 3.48, 3.68, 6.44, 3.99, 3.78
	TP	FD <sub>423</sub> , FD <sub>489</sub> , FD <sub>516</sub> , FD <sub>649</sub> , FD <sub>1816</sub> , FD <sub>2222</sub> , FD <sub>2386</sub>	2.34, 2.48, 2.28, 2.13, 3.01, 1.66, 6.07
	TK	FD <sub>659</sub> , FD <sub>904</sub> , FD <sub>965</sub> , FD <sub>1128</sub> , FD <sub>1521</sub> , SD <sub>1006</sub>	2.89, 4.21, 2.78, 4.40, 3.10, 1.48
GBDT	TN	FD <sub>572</sub> , FD <sub>977</sub> , FD <sub>1084</sub> , FD <sub>1015</sub> , FD <sub>2051</sub> , SD <sub>418</sub>	8.97, 4.56, 1.45, 3.87, 1.37, 3.42
	TP	FD <sub>663</sub> , FD <sub>747</sub> , FD <sub>1009</sub> , SD <sub>831</sub>	2.68, 3.00, 3.42, 7.45
	TK	FD <sub>1045</sub> , FD <sub>1069</sub> , FD <sub>1784</sub> , FD <sub>1796</sub> , FD <sub>2348</sub>	4.42, 2.52, 5.57, 8.53, 6.17

In order determine the most accurate screening algorithm, the PLSR approach was selected to construct the model between soil nutrients and the characteristic variables from the three algorithms based on 75 soil samples from the province. The PLSR relationship models are described as follows:

$$\text{PCC - PLSP} \begin{cases} Y_N = 1.505 - 224 \times FD_{562} - 9717 \times SD_{714} & (R^2 = 0.17) \\ Y_P = 1.698 - 1091 \times FD_{1009} + 46 \times FD_{356} + 224 \times SD_{905} & (R^2 = 0.35) \\ Y_K = 12.91 + 9 \times R_{2498} + 6173 \times FD_{442} & (R^2 = 0.40) \end{cases}$$

$$\begin{cases}
 \text{LASSO - PLSP} \left\{ \begin{aligned}
 Y_N &= 1.530 + 4928 \times SD_{668} + 38 \times FD_{1418} + 53 \times FD_{1302} + 786 \times FD_{454} + 26 \\
 &\quad \times FD_{2367} - 4079 \times SD_{529} - 37 \times FD_{1707} + 32 \times FD_{2342} - 157 \times FD_{904} & (R^2 = 0.15) \\
 Y_P &= 0.646 + 1362 \times FD_{516} + 159 \times FD_{1816} - 414 \times FD_{423} - 57 \times FD_{649} \\
 &\quad - 121 \times FD_{489} - 50 \times FD_{2222} + 67 \times FD_{2386} & (R^2 = 0.15) \\
 Y_K &= 8.871 + 290,321 \times SD_{1006} + 15,335 \times FD_{965} + 2839 \times FD_{1521} - 11,052 \\
 &\quad \times FD_{659} - 15,362 \times FD_{904} - 1372 \times FD_{1128} & (R^2 = 0.47)
 \end{aligned} \right. \\
 \\
 \text{GBDT - PLSP} \left\{ \begin{aligned}
 Y_N &= 1.908 - 366 \times FD_{572} - 265 \times FD_{2051} + 237 \times FD_{1084} - 1086 \times SD_{418} \\
 &\quad - 61 \times FD_{977} + 364 \times FD_{1015} & (R^2 = 0.26) \\
 Y_P &= 1.657 - 504 \times FD_{663} - 1232 \times FD_{1009} + 915 \times FD_{747} - 4946 \times SD_{831} & (R^2 = 0.37) \\
 Y_K &= 19.188 - 2601 \times FD_{2348} - 7380 \times FD_{1045} - 14,742 \times FD_{1069} \\
 &\quad + 1605 \times FD_{1796} + 1207 \times FD_{1784} & (R^2 = 0.24)
 \end{aligned} \right.
 \end{cases}$$

The results identify the optimal algorithms of soil TN, TP, and TK as GBDT, GBDT, and LASSO, with  $R^2$  values of 0.26, 0.37, and 0.47, respectively. Among three nutrients, the LASSO-PLSR showed the best estimation of soil TK.

### 3.2. Determining the Optimal Model for Soil Nutrient Content Estimations

The MLR, RR, SVM, and GABP models were adopted to determine the relationship between the characteristic variables and soil nutrients (Figure 8). The GABP model exhibited the highest predicative capability for the three soil nutrients, with scatter plots closer to the 1:1 line compared to MLR, RR, and SVM. Additionally, it offered the most accurate estimates in cross-validation with  $R_{cv}^2$  of 0.69, RMSECV of 0.35, and RPIQ = 2.03 for TN;  $R_{cv}^2$  of 0.73, RMSECV of 0.30 and RPIQ = 2.10 for TP,  $R_{cv}^2$  of 0.82, RMSECV of 3.39, and RPIQ = 3.57 for TK, respectively (Table 4). The prediction effect of soil TK is obviously better than that of TN and TP, which may be due to potassium being a metal element, with a spectral response sensitivity that exceeds other non-metal elements (e.g., nitrogen and phosphorus).

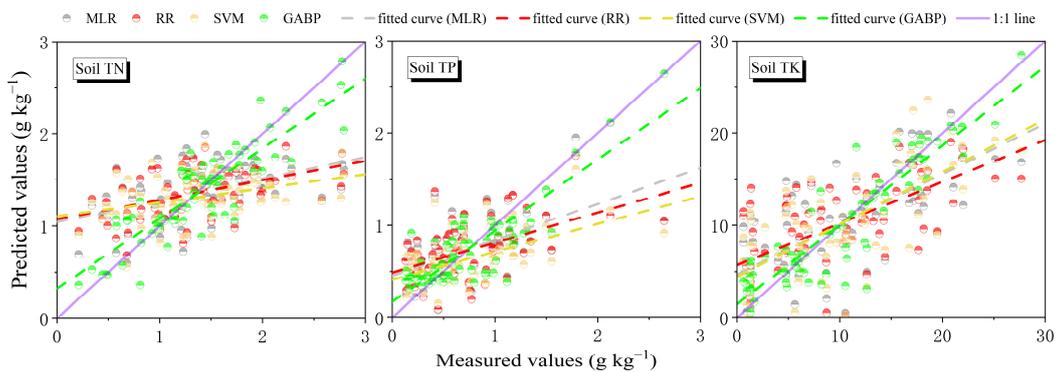


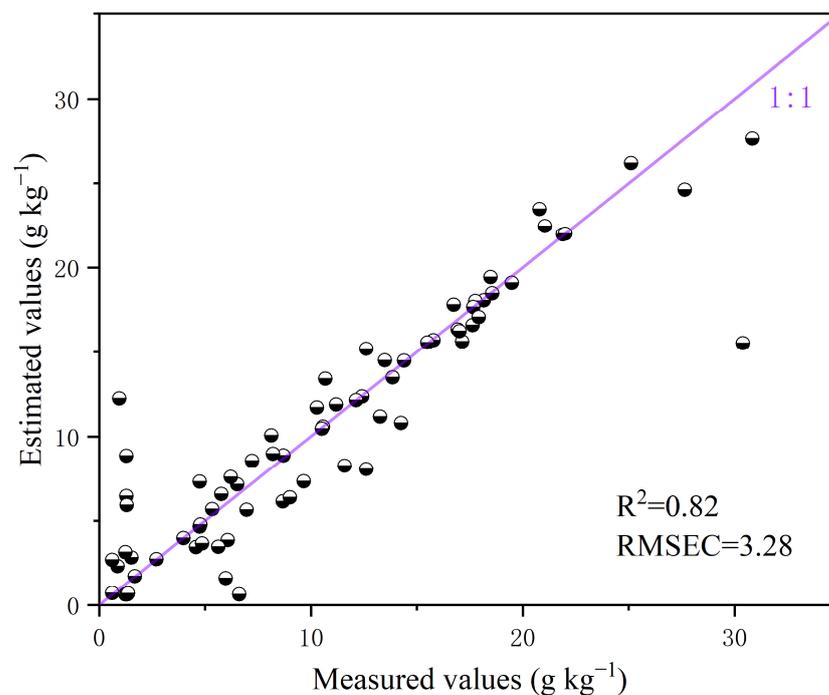
Figure 8. Scatter plots of measured and estimated values.

**Table 4.** Accuracy assessment of estimated soil nutrient contents (unit:  $\text{g kg}^{-1}$ ).

Soil Nutrients	Model	$R^2$ (C)	CCC	RMSEC	$R_{cv}^2$	RMSECV	RPIQ
TN	MLR	0.22	0.37	0.50	0.17	0.51	1.39
	RR	0.21	0.35	0.50	0.18	0.51	1.39
	SVM	0.13	0.26	0.53	0.11	0.57	1.25
	GABP	0.76	0.86	0.28	0.69	0.35	2.03
TP	MLR	0.36	0.55	0.40	0.32	0.47	1.34
	RR	0.34	0.47	0.43	0.33	0.44	1.43
	SVM	0.36	0.49	0.41	0.35	0.41	1.54
	GABP	0.77	0.87	0.26	0.73	0.30	2.10
TK	MLR	0.48	0.67	5.30	0.42	5.52	2.19
	RR	0.44	0.61	5.32	0.43	5.33	2.27
	SVM	0.54	0.72	5.17	0.52	5.31	2.28
	GABP	0.86	0.92	2.88	0.82	3.39	3.57

### 3.3. Mapping Soil Nutrient Contents Using the Proposed Method

Table 4 demonstrates the soil TK estimation accuracy exceeding that of TN and TP. Therefore, we applied the proposed method to map soil TK contents in Conghua District at the regional scale using HJ-1A imagery because the spectral wavelength of HJ-1A data ranged from 459 to 956 nm, which had different range and spectral bands of wavelengths from those of the spectral variables involved in the above estimation models. The model based on 75 sample points collected across the province could not be utilized for the HJ-1A images. We employed the LASSO-GABP method to re-screen the optimal spectral variables from the resample measured soil spectral data with 5 nm spectral resolution and to develop the corresponding estimation models. The screened spectral variables were determined as band462, band464, band466, band470, band477, band484, band574, and band652. The soil TK was estimated with reliable accuracy ( $R^2$  of 0.82, RMSEC of  $3.28 \text{ g kg}^{-1}$ ; Figure 9).



**Figure 9.** Scatter plots of the measured and estimated values based on resample measured soil spectral data.

Figure 10 demonstrates the spatial distribution of the soil TK contents obtained using the estimation model. The soil TK content is generally concentrated within 10–20 g kg<sup>-1</sup>, with flat areas exhibiting a higher content and areas with high slopes and close proximity to rivers associated with lower content. This may be linked to soil erosion, which is consistent with the actual situation.

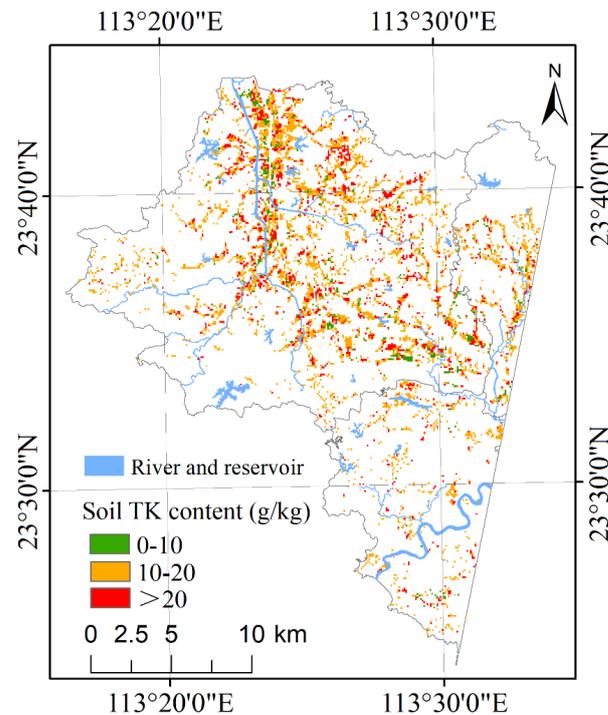


Figure 10. Spatial distribution of the soil total potassium content for the study area.

The 33 sample plots (Figure 1b) were used to verify the feasibility of mapping soil nutrient content by calculating the  $R^2$ , RMSE, and RPIQ values (Table 5). The estimation accuracy of soil TK content was relatively high, with an  $R^2$  of 0.79 and RMSE of 4.01 g kg<sup>-1</sup>. This indicates that the GABP model is capable of mapping the soil TK content. However, the estimation accuracy of the regional-scale retrievals is lower than that of the point-scale. This may be due to the limitation of the narrow spectral region of the HJ-1A HSI data (450–960 nm).

Table 5. Estimation accuracy of soil total potassium content using the GABP model based on the 33 validation sample plots (unit: g kg<sup>-1</sup>).

	Dataset	Mean	Max	Min	SD	$R^2$	RMSE	RPIQ
Soil TK	Measured Value	18.35	30.57	2.64	6.67	0.79	4.01	1.86
	Estimated Value	20.01	36.42	1.36	8.86			

#### 4. Discussion

In the current paper we compared three algorithms (PCC, LASSO, and GBDT) and four models (MLR, RR, SVM, and GABP) in terms of soil nutrient estimations in order to determine a method for the prediction of high-accuracy soil nutrients.

In this method, to the best of our knowledge, this is the first attempt to use the LASSO and GBDT algorithms to determine the characteristic variables for soil nutrient estimations. LASSO with PLSR fit degree ( $R^2$ ) of 0.47 was determined as optimal for the accurate selection of soil TK characteristic variables, and GBDT for TN and TP with  $R^2$  of 0.26 and 0.37. This indicates the significant nonlinear spectral response mechanism of the soil nutrients. The result found that 16 characteristic variables obtained using the

optimal screening algorithms are sensitive to soil nutrients: FD<sub>572</sub>, FD<sub>977</sub>, FD<sub>1084</sub>, FD<sub>1015</sub>, FD<sub>2051</sub>, SD<sub>418</sub> for TN, FD<sub>663</sub>, FD<sub>747</sub>, FD<sub>1009</sub>, SD<sub>831</sub> for TP and FD<sub>659</sub>, FD<sub>904</sub>, FD<sub>965</sub>, FD<sub>1128</sub>, FD<sub>1521</sub>, SD<sub>1006</sub> for TK. Some selected wavelengths are in general agreement with previous research [43–45].

Previous studies generally employ linear models to estimate soil nutrients [2,46–48]. In order to improve the estimation accuracy of soil nutrients, we adopted linear (MLR and RR) and nonlinear (SVM and GABP) algorithms to construct the soil nutrient estimation models based on the determined spectral characteristic variables (Table 3). The validation results (Table 4) revealed the GBDT-GABP algorithm to perform the best in soil TN ( $R_{cv}^2$  of 0.69, RMSECV of 0.35, and RPIQ = 2.03) and TP ( $R_{cv}^2$  of 0.73, RMSECV of 0.30, and RPIQ = 2.10) estimations, while LASSO-GABP was optimal for soil TK ( $R_{cv}^2$  of 0.82, RMSECV of 3.39, and RPIQ = 3.57), which are in general agreement with previous research results with  $R^2$  from 0.56 to 0.84 (TN), 0.65 to 0.81 (TP), and 0.67 to 0.82 (TK) [43,49–54]. The proposed model constructed using machine learning algorithms outperformed the linear models. This indicates the existence of a significant nonlinear relationship between the soil nutrients and spectral characteristic variables.

In order to validate the regional-scale applicability of the new method, HJ-1A image data obtained from pure pixels using the fully constrained least squares (FCLS) method was used to map soil TK with the best estimation accuracy ( $R^2 = 0.86$ ) on point scale. Results using the 33 validation sample plots demonstrate the screened spectral characteristic variables to explain 79% of the variance in the TK content, with an RMSE of 4.01 g kg<sup>-1</sup> for the mapping of TK content. This indicates the great potential of GABP to map the soil TK content at a large scale. However, the point-scale estimation accuracy is higher than that of the regional-scale due to the narrow spectral range of the HJ-1A HSI data. Future research will map the TK contents using satellite hyperspectral images covering a wider spectral region (350–2500 nm).

The prediction effect of soil TK is obviously better than that of TN and TP (Table 3). This may be because potassium is a metal element, with a spectral response sensitivity that exceeds other non-metal elements (e.g., nitrogen and phosphorus). The introduction of additional soil elements (including metals and nonmetals) to explore this phenomenon will be the focus of further work.

We employed 75 soil samples to develop the models and validate the method for the whole Guangdong province, while 33 sample plots were used to verify the feasibility of mapping soil nutrient content in Conghua District. Although the sampling design was conducted based on different soil characteristics and soil types, the sample sizes were relatively small. Future studies will employ larger sample sizes to further develop and validate the proposed method.

## 5. Conclusions

The determination of characteristic variables is key for accurate hyperspectral estimation models of the soil nutrient content. This paper introduced the LASSO and GBDT algorithms to screen the optimal relevant characteristic variables of soil TN, TP, and TK. The estimation models of soil nutrient content were subsequently developed using the selected characteristic variables and field observations of soil nutrient content. The most accurate estimation model was then adopted to explore the possibility of spatially mapping the soil nutrient content using HJ-1A data. The results demonstrated that compared with the statistical analysis method, the machine learning method effectively screened the characteristic variables. In addition, based on the RMSECV values, the GABP models of the soil nutrient contents determined the most accurate estimates at the soil sample point level. The new method provides the potential for soil nutrient mapping at the regional scale with a reasonable accuracy using hyperspectral imagery. Results indicate the ability of the LASSO and GBDT algorithms to improve the estimation accuracy of soil TN, TP, and TK, which are crucial for agricultural management. The proposed machine learning method

has the potential to effectively select the spectral characteristic indices of soil nutrients, increasing the accuracy of the results.

**Author Contributions:** Conceptualization, Y.P. and L.W.; methodology, Y.P.; software, C.L.; validation, Y.P., L.Z. and L.L.; investigation, Y.P.; resources, L.W. and Y.H.; data curation, Y.P., L.W. and L.Z.; writing—original draft preparation, Y.P.; writing—review and editing, Y.P., L.W. and Z.L.; funding acquisition, Y.H. and L.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China (No. 2020YFD1100203), National Natural Science Foundation of China (No. U1901601), and Guangdong Province Agricultural Science and Technology Innovation and Promotion Project (No. 2021KJ102).

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We gratefully acknowledge the paper writing assistance of Mingbang Zhu as well as the experimental assistance of Ziqing Xia.

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

## References

- Jacquemoud, S.; Baret, F.; Hanocq, J.F. Modeling spectral and bidirectional soil reflectance. *Remote Sens. Environ.* **1992**, *41*, 123–132. [[CrossRef](#)]
- Yu, X.; Liu, Q.; Wang, Y.B.; Liu, X.Y.; Liu, X. Evaluation of MLSR and PLSR for estimating soil element contents using visible/near-infrared spectroscopy in apple orchards on the Jiaodong peninsula. *Catena* **2016**, *137*, 340–349. [[CrossRef](#)]
- Rossel, R.A.V.; Walvoort, D.J.J.; Mcbratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [[CrossRef](#)]
- An, X.F.; Li, M.Z.; Zheng, L.H.; Liu, Y.M.; Sun, H. A portable soil nitrogen detector based on NIRS. *Precis. Agric.* **2014**, *15*, 3–16. [[CrossRef](#)]
- Liu, H.J.; Zhang, B.; Zhao, J.; Zhang, X.Y.; Song, K.S.; Wang, Z.M.; Duan, H.T. Spectral models for prediction of organic matter in black soil. *Acta Pedol. Sin.* **2007**, *44*, 27–32.
- Vibhute, A.D.; Kale, K.V.; Gaikwad, S.V.; Dhupal, R.K. Estimation of soil nitrogen in agricultural regions by VNIR reflectance spectroscopy. *SN Appl. Sci.* **2020**, *2*, 1523. [[CrossRef](#)]
- Liu, Q.; Qin, Z.G.; Luo, X.C.; Cheng, H.R. *Summary of Feature Selection Methods in Statistical Machine Learning*; China National Computer Congress: Tianjin, China, 2009.
- Zhang, Y.; Li, M.Z.; Zheng, L.H.; Qin, Q.M.; Lee, W.S. Spectral features extraction for estimation of soil total nitrogen content based on modified ant colony optimization algorithm. *Geoderma* **2019**, *333*, 23–34. [[CrossRef](#)]
- Casa, R.; Castaldi, F.; Pascucci, S.; Basso, B.; Pignatti, S. Geophysical and Hyperspectral Data Fusion Techniques for In-Field Estimation of Soil Properties. *Vadose Zone J.* **2013**, *12*, vzj2012.0201. [[CrossRef](#)]
- Cao, F.X.; Yang, Z.J.; Ren, J.C.; Jiang, M.Y.; Ling, W.K. Linear vs Nonlinear Extreme Learning Machine for Spectral-Spatial Classification of Hyperspectral Image. *Sensors* **2017**, *17*, 2603. [[CrossRef](#)]
- Leone, A.P.; Viscarra-Rossel, R.A.; Amenta, P.; Buondonno, A. Prediction of Soil Properties with PLSR and vis-NIR Spectroscopy: Application to Mediterranean Soils from Southern Italy. *Curr. Anal. Chem.* **2012**, *8*, 283–299. [[CrossRef](#)]
- Song, Y.Q.; Xin, Z.; Su, H.Y.; Li, B.; Hu, Y.M.; Cui, X.S. Predicting Spatial Variations in Soil Nutrients with Hyperspectral Remote Sensing at Regional Scale. *Sensors* **2018**, *18*, 3086. [[CrossRef](#)]
- Mouazen, A.M.; Kuang, B.; De Baerdemaeker, J.; Ramon, H. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* **2010**, *158*, 23–31. [[CrossRef](#)]
- Peng, X.T.; Shi, T.Z.; Song, A.H.; Chen, Y.Y.; Gao, W.X. Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods. *Remote Sens.* **2014**, *6*, 2699–2717. [[CrossRef](#)]
- Moura-Bueno, J.M.; Dalmolin, R.S.D.; Caten, A.T.; Dotto, A.C.; Demattê, J.A.M. Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. *Geoderma* **2019**, *337*, 565–581. [[CrossRef](#)]
- Tang, F.; Chen, M.; Wang, Z. New approach to training support vector machine. *J. Syst. Eng. Electron* **2006**, *17*, 200–219.
- Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2793. [[CrossRef](#)] [[PubMed](#)]
- Ma, L.; Chen, C.; Shen, Y.; Wu, L.H.; Huang, Z.L.; Gao, H.L. Determinants of tree survival at local scale in a sub-tropical forest. *Ecol. Res.* **2014**, *29*, 69–80. [[CrossRef](#)]
- Wen, J. *Research of Neural Network Theory and Application*; Southwest Jiaotong University Press: Chengdu, China, 1996.
- Wang, F.; Gao, J.; Zha, Y. Hyperspectral sensing of heavy metals in soil and vegetation: Feasibility and challenges. *ISPRS J. Photogramm. Remote Sens.* **2018**, *136*, 73–84. [[CrossRef](#)]

21. Zhang, S.M.; Xu, M.X.; Zhang, Z.X.; Li, B.B. Methods of sampling soil organic carbon in farmlands with different landform types on the Loess Plateau. *J. Nat. Resour.* **2018**, *33*, 634–643.
22. Yang, J.Y.; Tang, S.; Yun, W.J.; Zhang, C.; Zhu, D.H.; Chen, Y.Q. Sampling method for monitoring classification of cultivated land in county area based on Kriging estimation error. *Trans. CSAE* **2013**, *29*, 223–230.
23. Walkley, A.J.; Black, C.A. An estimation of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Sci.* **1934**, *37*, 29–38. [[CrossRef](#)]
24. Petropoulos, G.P.; Arvanitis, K.; Sigrimis, N. Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping. *Expert Syst. Appl.* **2012**, *39*, 3800–3809. [[CrossRef](#)]
25. Sorol, N.; Arancibia, E.; Bortolato, S.A.; Olivieri, A.C. Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice: A test field for variable selection methods. *Chemom. Intell. Lab. Syst.* **2010**, *102*, 100–109. [[CrossRef](#)]
26. Allouis, T.; Durrieu, S.; Véga, V.; Coueron, P. Stem Volume and Above-Ground Biomass Estimation of Individual Pine Trees from LiDAR Data: Contribution of Full-Waveform Signals. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 924–934. [[CrossRef](#)]
27. Tibshirani, R.J. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [[CrossRef](#)]
28. Xu, W.Q.; Ning, L.K.; Luo, Y. Wind Speed Forecast Based on Post-Processing of Numerical Weather Predictions Using a Gradient Boosting Decision Tree Algorithm. *Atmosphere* **2020**, *11*, 738. [[CrossRef](#)]
29. Wold, H. Nonlinear Estimation by Iterative Least Squares Procedure. *Res. Pap. Stat.* **1966**, 441–444.
30. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemometr. Intell. Lab.* **2001**, *58*, 109–130. [[CrossRef](#)]
31. Ni, W.C. Discussion on the definition of  $R^2$  equivalence. *Stat. Decis.* **2009**, *09*, 141–142.
32. Zhang, H.M.; Liu, W.; Han, W.T.; Liu, Q.Z.; Song, R.J.; Hou, G.H. Inversion of Summer Maize Leaf Area Index Based on Gradient Boosting Decision Tree Algorithm. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 251–259.
33. Wang, F.; Shi, Z.; Biswas, A.; Yang, S.T.; Ding, J.L. Multi-algorithm comparison for predicting soil salinity. *Geoderma* **2020**, *365*, 114211. [[CrossRef](#)]
34. Fatholouloumi, S.; Vaezi, A.R.; Alavipanah, S.K.; Ghorbani, A.; Biswas, A. Comparison of spectral and spatial-based approaches for mapping the local variation of soil moisture in a semi-arid mountainous area. *Sci. Total Environ.* **2020**, *724*, 138319. [[CrossRef](#)]
35. Zhang, Z.T.; Wang, H.F.; Karnieli, A.; Chen, J.Y.; Han, W.T. Inversion of Soil Moisture Content from Hyperspectra Based on Ridge Regression. *Trans. Chin. Soc. Agric. Mach.* **2018**, *49*, 240–248.
36. Kennard, H.R.W. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* **1970**, *12*, 69–82.
37. Hernandez, J.; Lobos, G.A.; Matus, I.; Del Pozo, A.; Silva, P.; Galleguillos, M. Using Ridge Regression Models to Estimate Grain Yield from Field Spectral Data in Bread Wheat (*Triticum Aestivum* L.) Grown under Three Water Regimes. *Remote Sens.* **2015**, *7*, 2109–2126. [[CrossRef](#)]
38. Wang, X.; Wang, Z.Q.; Jin, G.; Yang, J. Land reserve prediction using different kernel based support vector regression. *Trans. Chin. Soc. Agric. Eng.* **2014**, *30*, 204–211.
39. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
40. Saleh, S.M.; Ibrahim, K.H.; Magdi Eiteba, M.B. Study of genetic algorithm performance through design of multi-step LC compensator for time-varying nonlinear loads. *Appl. Soft Comput.* **2016**, *48*, 535–545. [[CrossRef](#)]
41. Xie, H.; Luo, X.; Xu, X.; Pan, H.; Tong, X. Automated Subpixel Surface Water Mapping from Heterogeneous Urban Environments Using Landsat 8 OLI Imagery. *Remote Sens.* **2016**, *8*, 584. [[CrossRef](#)]
42. Hermansen, C.; Norgaard, T.; Jonge, L.; Moldrup, P.; Müller, K.; Knadel, M. Predicting glyphosate sorption across New Zealand pastoral soils using basic soil properties or Vis–NIR spectroscopy. *Geoderma* **2020**, *360*, 114009. [[CrossRef](#)]
43. Jia, S.; Li, H.; Wang, Y.; Tong, R.; Li, Q. Hyperspectral Imaging Analysis for the Classification of Soil Types and the Determination of Soil Total Nitrogen. *Sensors* **2017**, *17*, 2252. [[CrossRef](#)]
44. Kawamura, K.; Tsujimoto, Y.; Rabenarivo, M.; Asai, H.; Andriamananjara, A.; Rakotoson, T. Vis-NIR Spectroscopy and PLS Regression with Waveband Selection for Estimating the Total C and N of Paddy Soils in Madagascar. *Remote Sens.* **2017**, *9*, 1081. [[CrossRef](#)]
45. Lin, C.; Ma, R.H.; Zhu, Q.; Li, J.T. Using hyper-spectral indices to detect soil phosphorus concentration for various land use patterns. *Environ. Monit. Assess.* **2015**, *187*, 4130. [[CrossRef](#)]
46. Yang, M.H.; Mouazen, A.; Zhao, X.M.; Guo, X. Assessment of a soil fertility index using visible and near-infrared spectroscopy in the rice paddy region of southern China. *Eur. J. Soil Sci.* **2020**, *71*, 615–626. [[CrossRef](#)]
47. Baldock, J.A.; Beare, M.H.; Curtin, D.; Hawke, B. Stocks, composition and vulnerability to loss of soil organic carbon predicted using mid-infrared spectroscopy. *Soil Res.* **2018**, *56*, 468–480. [[CrossRef](#)]
48. Munnaf, M.A.; Guerrero, A.; Nawar, S.; Haesaert, G.; Meirvenne, M.V.; Mouazen, A.M. A combined data mining approach for on-line prediction of key soil quality indicators by Vis-NIR spectroscopy. *Soil Tillage Res.* **2021**, *205*, 104808. [[CrossRef](#)]
49. An, X.F.; Zheng, L.H.; Li, M.Z. Real-Time Analysis of Soil Total Nitrogen and Soil Total Phosphorus with NIR Spectroscopy. *Sens. Lett.* **2010**, *8*, 163–166. [[CrossRef](#)]
50. Udelhoven, T.; Emmerling, C.; Jarmer, T. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. *Plant Soil* **2003**, *251*, 319–329. [[CrossRef](#)]
51. Shi, T.Z.; Cui, L.J.; Wang, J.J.; Fei, T.; Chen, Y.Y.; Wu, G.F. Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy. *Plant Soil* **2013**, *366*, 363–375. [[CrossRef](#)]

- 
52. Xue, Y.H.; Vasques, G.M.; Grunwald, S. Application of Visible/Near-Infrared Spectra in Modeling of Soil Total Phosphorus. *Pedosphere* **2013**, *23*, 417–421.
  53. Hu, G.T.; He, D.J.; Kenneth, A.S. Soil Phosphorus and Potassium Estimation Using Visible-near Infrared Reflectance Spectroscopy with Direct Orthogonal Signal Correction. *Trans. Chin. Soc. Agric. Mach.* **2015**, *46*, 139–145.
  54. Li, M.; Qin, K.; Zhao, N.B.; Tian, F.; Zhao, Y.J. Study on the Relationship Between Black Soil Emissivity Spectrum and Total Potassium Content Based on TASI Thermal Infrared Data. *Spectrosc. Spectr. Anal.* **2020**, *40*, 2862–2868.