



Article

Selection of Spectral Parameters and Optimization of Estimation Models for Soil Total Nitrogen Content during Fertilization Period in Apple Orchards

Zhilin Gao ^{1,†}, Wenqian Wang ^{2,*,†}, Hongjia Wang ^{1,*} and Ruiyan Li ³

¹ College of Resources and Environment, Shandong Agricultural University, Tai'an 271018, China; katherinegzl@163.com

² Shandong Yuanhong Survey Planning and Design Co., Ltd., Jinan 250014, China

³ Jinan Survey and Mapping Research Institute, Jinan 250101, China; jnlry0418@126.com

* Correspondence: wenqian9506@163.com (W.W.); hongjiawang0101@163.com (H.W.)

† These authors contributed equally to this work.

Abstract: The rapid and accurate diagnosis of nitrogen content in apple orchard soil is of great significance for the rational application of nitrogen fertilizer in orchards to improve apple yield and quality. An apple orchard in Shuangquan Town, Changqing District, Jinan City, Shandong Province, was taken as the experimental area. The optimal method for extracting spectral characteristic bands and screening spectral characteristic indices (SCIs) of soil total nitrogen (TN) for independent and comprehensive fertilization periods was explored. Independent and comprehensive soil TN content estimation models were constructed and optimized for each and the entire fertilization period, respectively. The results show that compared with the correlation coefficient method, stepwise multiple linear regression (SMLR) performs better in extracting hyperspectral characteristic bands of soil TN content. It helps to achieve a higher modeling accuracy, smaller root mean square error (RMSE), and is more conducive to avoiding the influence of multicollinearity of model variables. The sensitive areas of soil TN content in the SCI do not undergo significant changes due to different fertilization periods. Among them, the ratio spectral indices (RSIs) are in the range of 800–900 nm, 1900–1950 nm, and 2200–2300 nm, while the sensitive areas of the difference spectral index (DI) and Normalized difference spectral index (NDSI) are in the range of 1900–1950 nm and 2200–2300 nm. The combination of SCI and characteristic bands significantly improves the prediction accuracy of soil TN estimation models. The independent and comprehensive estimation models for each fertilization period based on the BP (back propagation) neural network optimized by the Mind Evolution Algorithm (MEA-BPNN) can achieve a more stable and accurate estimation of soil TN. Finally, using soil spectral characteristic bands selected through continuum removal (CR) transformation and SMLR, combined with SCI, the model based on the MEA-BPNN (CR-SCI-MEA-BPNN) has the best prediction performance. The modeling determination coefficients R^2 for each fertilization period reached 0.94, 0.95, 0.92, and 0.94, respectively, with RMSE of 0.0032, 0.0024, 0.0035, and 0.0027. The R^2 and RMSE of the modeling and validation set of the entire fertilization period comprehensive model are 0.899, 0.0038, and 0.89, 0.0041, respectively. The results of this article provide technical support for promoting the timely monitoring of soil TN content and guiding rational fertilization in apple orchards.

Keywords: spectral parameters; estimation model; soil total nitrogen; apple orchards



Citation: Gao, Z.; Wang, W.; Wang, H.; Li, R. Selection of Spectral Parameters and Optimization of Estimation Models for Soil Total Nitrogen Content during Fertilization Period in Apple Orchards. *Horticulturae* **2024**, *10*, 358. <https://doi.org/10.3390/horticulturae10040358>

Academic Editor: Moreno Toselli

Received: 11 March 2024

Revised: 31 March 2024

Accepted: 2 April 2024

Published: 4 April 2024



Copyright: © 2024 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

Apple is one of the four major fruits in the world, cultivated in over 80 countries worldwide. China is the world's largest producer and consumer of apples. According to the 2022 China Apple Industry Report released by the China Apple Industry Association, the national planting area in 2022 was 2 million hectares, with a total production of 45.9734 million tons, which accounted for 58.3% of the global total production [1]. Nitrogen

is an essential nutrient for maintaining the growth of apple trees [2], and plays an irreplaceable role in organ development, material metabolism, fruit quality, and yield [3–7]. Soil nitrogen is the main source of nutrients for fruit trees, and its content is the basis for orchard fertilization and an important indicator for evaluating and managing orchard quality [6]. Due to the challenge of estimating the nitrogen content of orchard soil in advance, fruit farmers often apply too much or too little nitrogen in their application of nitrogen fertilizer in orchards. This leads to the abnormal growth of fruit trees, susceptibility to diseases, and soil and groundwater pollution [8]. Therefore, knowing how to diagnose the soil nitrogen in apple orchards quickly and accurately is of great significance for the timely monitoring and scientific management of reasonable nitrogen application to improve yield and quality [2–4].

At present, in the monitoring of soil nitrogen in apple orchards, the laboratory Kjeldahl nitrogen determination method is commonly used to determine the soil total nitrogen (TN) content. Although it can be highly accurate, the experimental steps are cumbersome, especially the drawbacks of sampling causing significant damage to soil and roots [9], which greatly limits its application in practice. The emergence of hyperspectral remote sensing technology provides a new way for non-destructive estimation of soil nutrients, which has advantages such as being fast, accurate, non-destructive, and pollution-free. It is widely used in the estimation and inversion of soil physicochemical properties [10,11]. Hyperspectral inversion has achieved good research results in soil iron oxides [12], organic carbon [13,14], organic matter [10,15,16], and soil moisture content [17], providing a reference for the rapid monitoring of soil nitrogen in orchards.

In recent years, scholars at home and abroad have conducted extensive research on spectral estimation techniques for soil TN content and found that the spectral corresponding characteristics of soil TN content are mainly concentrated in the visible near-infrared region [18–20]. Multiple modeling methods have been attempted to prove the feasibility of quantitative inversion of soil TN spectra [21–23]. Niu et al. constructed a hyperspectral estimation model for TN content in Shajiang black soil and found that the accuracy of the SVM model was slightly higher than that of the indices model, but both models could facilitate the rapid estimation of TN content in Shajiang black soil [24]. The spectral estimation technology of soil nutrients has achieved some preliminary results in the research in apple orchards [25,26], navel orange orchards [27], and rubber orchards [28]. Liu et al. constructed an estimation model for soil organic matter content in apple orchards using the random forest method, and the R^2 of the modeling sample set reached 0.88 [29]. Tang et al. extracted hyperspectral features of rubber plantation soil using the successive projection algorithm (SPA) and established a partial least squares regression (PLSR) model for soil TN content [30]. Liu et al. used the competitive adaptive reweighting algorithm (CARS), SPA, and CARS-SPA algorithm to select the characteristic bands of soil TN spectrum. The SVM model established using the characteristic wavelengths selected by the CARS method had the best performance [31]. Overall, most spectral inversion methods for nitrogen content in orchard soil currently used conventional spectral transformation and correlation analysis to screen characteristic bands, and based on this, commonly used regression models were constructed. There is a lack of in-depth exploration of soil spectral information, which affects the predictive performance of the models. And there is a lack of systematic analysis and screening of data processing methods. In addition, existing research often randomly selected a certain point in the year for sample collection, and the model lacked a duration test, resulting in low practicality of the model.

The purpose of this study is as follows: (i) Screen the optimal soil spectra characteristic band extraction method for soil TN during each fertilization period; (ii) select the most suitable soil TN spectral characteristic index (SCI); (iii) construct and optimize independent and comprehensive soil TN estimation models for each fertilization period and the entire fertilization period, respectively. The layout of this article is as follows: firstly, collect and transform the soil spectral reflectance for each fertilization stage; then, conduct in-depth research step by step according to the order of the three objectives mentioned above.

2. Materials and Methods

2.1. Data Acquisition

2.1.1. Overview of Experimental Orchard

The experimental orchard is located at the foot of a hill in Shuangquan Town, Changqing District, Jinan City, Shandong Province, China, with an area of approximately 33 hm² (116.7426–116.7507° E, 36.3139–36.3216° N). It is located at the northwest foot of Mountain Taishan, and the soil type is brown soil. The apple tree spacing is 1.5 m, the row spacing is 4.5 m, the number of plants per 667 m² is about 100, the crown width is 1–1.5 m, and the overall canopy closure of the orchard is about 0.6. Apple varieties include Gala and Fuji, with an age of about 5 years. During the 4 stages of apple growth: young fruit (in early May, young fruit formation with slower growth), swelling (in mid-June, fruit volume rapidly increases, the flesh thickens, and the skin color lightens), quality (in mid-July, fruit volume is basically fixed, and the color and taste are formed), and postpartum (in early October, after fruit picking, the roots grow at its peak before winter) period, nitrogen fertilizer is applied using drip irrigation, respectively.

2.1.2. Soil Sample Collection and Analysis

The experimental orchard was divided into modeling and validation areas, with 100 and 40 sampling points evenly distributed, respectively, and a fruit tree was selected as the center of each sampling point. Soil samples were collected 10 days before each of the 4 fertilization dates. To ensure the collection of soil as close as possible to the capillary roots and reduce damage to fruit trees during sampling, a soil drill was used to collect mixed soil samples from 0 to 30 cm below the soil surface at a distance of 30 cm horizontally from the apple tree roots [31]. Sand, gravel, and animal and plant residues were removed from the collected soil samples. Soil samples were placed in a dry and ventilated place for natural air drying. During the air-drying process, the samples were kept clean and not contaminated with each other. Then, the samples were ground through a 60-mesh sieve and mixed evenly before being divided into two parts using the quartering method. One sample was used to determine the soil TN content using the micro Kjeldahl nitrogen determination method (Table 1). Another one was placed in a completely black container (diameter 10 cm, depth 2 cm, default infinite depth) for obtaining soil spectral data.

Table 1. Descriptive statistical characteristics of soil TN content in each fertilization periods.

Study Area	Sample Size	Index	Soil TN Content (%)			
			Young Fruit	Swelling	Quality	Postpartum
Modeling area	100	Max	0.0840	0.0726	0.0882	0.0812
		Min	0.0353	0.0254	0.0406	0.0307
		Mean	0.0512	0.0427	0.0523	0.0482
Validation area	40	Max	0.0813	0.0764	0.0796	0.0779
		Min	0.0339	0.0233	0.0328	0.0290
		Mean	0.0526	0.0428	0.0528	0.0491

2.1.3. Soil Spectral Collection

The collection of soil spectral data was carried out outdoors in an open and unobstructed area with artificial control of the surrounding environment, using natural sunlight as the light source. The ASD Fieldspec3 ground object spectrometer was adopted, with a spectral range of 350–2500 nm and a resampling interval of 1 nm for data output. A total of 2151 bands of reflectance were collected. To ensure the solar altitude angle, measurements were taken from 10:00 to 14:00. When measuring, the dish filled with the soil sample to be tested was placed on black rubber with a field of view angle of 25° and a distance of 15 cm from the soil sample. To reduce the influence of soil spectral anisotropy [32], each soil sample was rotated at 90° three times during measurement, and a total of 20 spectral curves

were collected in each direction. After arithmetic averaging, the reflectance spectrum of the sample was obtained. During the measurement process, a white board with a reflectivity of 1 was used for calibration in a timely manner [33].

2.1.4. Spectral Data Preprocessing

On the basis of breakpoint correction, edge noise removal, and SG smoothing, the original soil spectral reflectance (R) was transformed into eight forms: reciprocal ($1/R$), logarithmic ($\text{Log}R$), square root (\sqrt{R}), first-order derivative $[(R)']$, reciprocal first-order derivative $[(1/R)']$, logarithmic first-order derivative $[(\text{Log}R)']$, square root first-order derivative $[(\sqrt{R})']$, and continuum removal (CR). The transformation result data were used together with the original spectrum as modeling and analysis data.

2.2. Modeling Parameter Screening

2.2.1. Characteristic Band Screening

1. Correlation coefficient method

Correlation analysis band by band was performed between the TN content of the sample soil and the original soil reflectance, as well as various transformations. The bands at the extreme or sudden change point of the correlation coefficient were selected as the characteristic bands [34]. The commonly used Pearson coefficient was selected as the correlation coefficient, and the calculation formula is as follows:

$$r_i = \frac{\sum_{n=1}^N (R_{ni} - \bar{R}_i)(TN_n - \bar{TN})}{\sqrt{\sum_{n=1}^N (R_{ni} - \bar{R}_i)^2 \sum_{n=1}^N (TN_n - \bar{TN})^2}} \quad (1)$$

where r_i represents the correlation coefficient between soil TN content and input spectrum and i is the band number; R is the input spectrum; R_{ni} represents the input value of the n th sample at the i -th band; \bar{R}_i represents the mean of the input spectral values of all soil samples in the i -th band; TN_n is the soil TN content of the n th soil sample; \bar{TN} is the average TN content of all soil samples.

2. Stepwise multiple linear regression (SMLR)

The spectral data of the original soil reflectance and its various transformation forms in the range of 400–2450 nm were taken as the input spectrum, and the SMLR method was used to screen the characteristic bands of soil TN content. The basic operation steps were divided into two steps: first, variables that have not been tested as significant based on a certain level of significance were eliminated; second, new variables were introduced into the regression model [35]. The representation of the filtering model is as follows:

$$TN = a_0 + \sum_{n=1}^N a_n R_n \quad (2)$$

where TN represents the soil TN content, a_0 and a_n are the constant terms and coefficients of the regression model, N is the number of filtered bands, and R_n is the input spectral value of the wavelength at the n -th modeling band.

2.2.2. Spectral Characteristic Index (SCI) Screening

In the range of 400–2450 nm, SCIs were calculated for the original soil spectra of each fertilization period band by band [36], and different spectral index matrices were obtained for the entire band range; SCI included the ratio spectral index (RSI), difference spectral index (DI), and Normalized difference spectral index (NDSI). A correlation coefficient matrix was obtained through performing correlation analysis with soil TN content and analyzed by using contour maps [37]. Then, the SCIs for the sensitive areas of soil TN content was analyzed, the combination of SCI with high correlation with soil TN were screened and determined.

The formulas for *RSI*, *DI*, and *NDSI* are as follows:

$$RSI(R_1, R_2) = R_1 / R_2, \tag{3}$$

$$DI(R_1, R_2) = R_1 - R_2, \tag{4}$$

$$NDSI(R_1, R_2) = (R_1 + R_2) / (R_1 - R_2), \tag{5}$$

where R_1 and R_2 represent the spectral reflectance of any two bands in the sample spectrum.

2.3. Modeling Methods

Firstly, R and its 8 transformations during each fertilization period were used as input spectra, respectively, and the soil TN characteristic bands were extracted based on correlation coefficient method and SMLR method, respectively. Afterwards, the modeling effects of univariate regression and MLR, as well as the MEA-BPNN method (Back propagation neural network optimized by Mind Evolution Algorithm) were compared. Subsequently, the optimal characteristic bands extraction method was selected based on the model's coefficient of determination (R^2) and root mean square error (RMSE). Then, using band by band analysis techniques and correlation analysis methods, the sensitive areas of soil TN spectral index were determined, and SCIs with high correlation with soil TN were screened out. Finally, based on SCIs and characteristic bands combinations, MLR and MEA-BPNN methods were used to construct and optimize independent and comprehensive soil TN content estimation models for each fertilization period, and the optimal models were validated. The above steps were implemented using MATLAB (version: 2016b) software.

2.3.1. Regression Analysis

Univariate regression is used for regression analysis that includes only one independent variable and one dependent variable [38]. MLR is the most common method for the quantitative relationship analysis between multiple variables [39]. The calculation method is as follows:

$$TN = \begin{bmatrix} TN_1 \\ TN_2 \\ TN_3 \\ \vdots \\ TN_n \end{bmatrix}, R = \begin{bmatrix} R_{11} & R_{12} & R_{13} & \cdots & R_{1m} \\ R_{21} & R_{22} & R_{23} & \cdots & R_{2m} \\ R_{31} & R_{32} & R_{33} & \cdots & R_{3m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{n1} & R_{n2} & R_{n3} & \cdots & R_{nm} \end{bmatrix}, \tag{6}$$

where TN is the TN content value of the soil sample and R is the spectral characteristic band matrix of the soil sample; TN_i refers to the TN content of the i -th soil sample; R_{ij} ($i = 1, 2, 3 \dots n; j = 1, 2, 3 \dots m$) represents the reflectance value of the i -th sample at the j -th characteristic band; n is the soil sample capacity, i.e., the number of samples; and m is the number of characteristic bands.

Assuming that the soil TN value can be expressed linearly by R_i ($i = 1, 2, 3 \dots, m$), the linear form is as follows:

$$TN = a_0 + a_1R_1 + \dots + a_mR_m + \ell, \text{ converted into a matrix representation as: } TN = TA + E,$$

$$T = \begin{bmatrix} 1 & R_{11} & R_{12} & \cdots & R_{1m} \\ 1 & R_{31} & R_{22} & \cdots & R_{2m} \\ 1 & R_{31} & R_{32} & \cdots & R_{3m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & R_{n1} & R_{n2} & \cdots & R_{nm} \end{bmatrix}, A = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}, E = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \vdots \\ \ell_n \end{bmatrix}, \tag{7}$$

where matrix A is the regression coefficient of the model, which is generally solved by partial least squares method, that is, $A = (R^T R)^{-1} R^T TN$, and the predicted value of the model is $TN^* = T(R^T R)^{-1} R^T TN$.

2.3.2. Optimization of BP (Back Propagation) Neural Network Based on Mind Evolution Algorithm (MEA-BPNN)

BP (back propagation) neural network was proposed by Rnnelhart and Hinton in the mid-1980s as a multi-layer feedforward neural network information processing system that simulates human brain function. The two key points in the BP neural network algorithm are the initial weights and thresholds of the network, which play a decisive role in the computational speed and prediction accuracy of the model. Knowing how to quickly and accurately determine the appropriate weights and thresholds is the most critical issue in BP neural network inversion [40,41]. Evolutionary Computation (EC) can be used to solve complex combinatorial optimization problems, but early evolutionary algorithms have problems such as premature convergence and slow convergence when optimizing BP neural networks. The MEA overcomes the drawback of EC [42] and has significant advantages in finding global optimal solutions. Therefore, this study selected the MEA algorithm to optimize the BP neural network estimation model for soil TN content in order to achieve better prediction results.

The MEA is a new type of algorithm optimized on the basis of genetic algorithm. It simulates the evolutionary process of human thinking, and through operations such as “convergence” and “dissimilation” operators, continuously iterates to minimize the error between predicted values and expected values. Figure 1 shows the design process of a BP neural network based on MEA optimization. The basic idea is to first map the solution space to the encoding space based on the topology of the BP neural network after obtaining soil sample data, that is, each encoding corresponds to an individual; then, select the scoring functions of individuals and populations, and continuously iterate and analyze them using the MEA algorithm; finally output the optimal initial weights and thresholds, and use this to train the BP neural network.

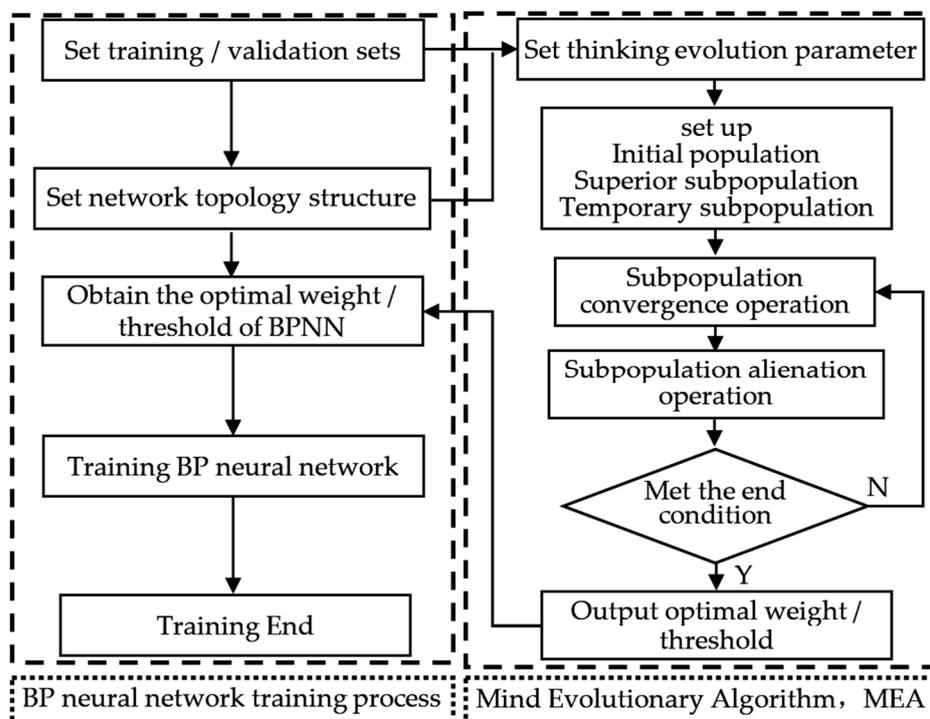


Figure 1. BP (back propagation) neural network (BPNN) flowchart optimized based on Mind Evolutionary Algorithm (MEA).

3. Results

3.1. Selected Soil TN Characteristic Bands and Their Modeling Effects

The screening results based on correlation coefficients show that in each fertilization period, the characteristic bands of the same transformation at different stages are different

but their positions are similar, and the distance between bands does not exceed 10 nm. The characteristic bands of different transformation forms during the same fertilization period are not entirely the same, with the characteristic band positions of $1/R$, $\text{Log}R$, and \sqrt{R} being the same as that of R ; the characteristic bands of all first-order derivative transformations are located similarly, while the characteristic bands of CR are mainly concentrated around 600–900 nm and 2220 nm. For the entire fertilization period, the characteristic band positions of R and $1/R$, $\text{Log}R$, and \sqrt{R} are basically consistent, which are all around 666 nm. The characteristic band positions of each first-order derivative and CR have changed, but most of them are concentrated around 562, 808, 1429, 1924, and 2223 nm.

The characteristic band screening results based on SMLR show that there is only one sensitive band for R , $1/R$, $\text{Log}R$, and \sqrt{R} in each fertilization period, which is consistent with the correlation analysis and is close to the results of the entire fertilization period. The bands are 663 nm, 666 nm, 666 nm, and 664 nm, respectively. The number of characteristic bands selected by the first derivative and CR transformation during each fertilization period ranges from 8 to 12, while the number of characteristic bands selected during the entire fertilization period increases from 1 to a maximum of 15, with a few cross bands between different transformations. This is consistent with the results of the correlation analysis method. At the same time, most of the characteristic bands selected based on SMLR for the entire fertilization period include bands selected by the correlation coefficient method or their nearby position bands, but all first-order derivatives and CR transformations have not been selected into the 666 nm band. Based on the screening results of various transformation forms, the bands 562 nm, 706 nm, 808 nm, 1933 nm, and 2345 nm are significantly affected by soil TN content.

3.1.1. Modeling Analysis Based on Univariate Regression

Using the selected characteristic bands as independent variables, univariate regression analysis on soil TN for each fertilization period, including univariate linear, polynomial, and exponential function regression was performed. The results are shown in Table 2. It can be seen that although the univariate regression model based on a single characteristic band is simple and convenient, the overall accuracy is relatively low. Except for the postpartum period (with a maximum R^2 of 0.479), the accuracy R^2 of the models for other fertilization periods is lower than 0.3, indicating poor modeling performance.

In the same fertilization period, from the perspective of fitting equations, regardless of the transformation form, for univariate regression analysis, the fitting effect of the third-order polynomial model is better than other univariate linear models and exponential models, and the modeling accuracy is slightly higher. From the perspective of spectral transformation forms, the model established based on \sqrt{R} has slightly higher accuracy, followed by $\text{Log}R$ and $1/R$. Among all the univariate regression models, the third-order polynomial model with \sqrt{R} is the best for all four fertilization periods. The young fruit stage corresponds to the 663 nm wavelength band, and the modeling determination coefficient $R^2 = 0.19$; the swelling period corresponds to the 666 nm wavelength band, and $R^2 = 0.264$; the quality period corresponds to the 666 nm wavelength band, and $R^2 = 0.246$; and the postpartum period corresponds to the 664 nm wavelength band, with $R^2 = 0.479$. However, overall, compared with R , the accuracy difference between univariate regression models based on $1/R$, $\text{Log}R$, and \sqrt{R} is very small, and all of them are relatively low. The results indicate that all three transformations have little significance in extracting effective information from soil spectra. Therefore, in the subsequent research on spectral estimation of soil TN content, univariate linear regression analysis may not be necessary.

Table 2. Univariate regression analysis of soil original reflectance and its transformation form with soil TN content in each fertilization period.

Sampling Period	Input Spectrum	Regression Equation	R ²	RMSE
Young fruiting	$x = R_{663}$	$y = -0.2146x + 0.0954$	0.101	0.0087
		$y = -349.8x^3 + 228.8x^2 - 49.818x + 3.6604$	0.183	0.0083
		$y = 0.1095e^{3.755x}$	0.094	0.0091
	$x = 1/R_{663}$	$y = 0.0098x + 0.0033$	0.137	0.0086
		$y = 0.0251x^3 - 0.3515x^2 + 1.6407x - 2.5031$	0.186	0.0083
		$y = 0.022e^{0.171x}$	0.104	0.0090
	$x = \text{Log}R_{663}$	$y = -0.0461x - 0.0217$	0.107	0.0087
		$y = -3.0554x^3 - 14.048x^2 - 21.523x - 10.938$	0.189	0.0083
		$y = 0.0142e^{0.804x}$	0.099	0.0089
	$x = \sqrt{R_{663}}$	$y = -0.1991x + 0.1416$	0.104	0.0087
		$y = -263.37x^3 + 368.04x^2 - 171.39x + 26.646$	0.190	0.0083
		$y = 0.2445e^{3.477x}$	0.097	0.0092
Swelling fruit	$x = R_{666}$	$y = -0.2376x + 0.0982$	0.238	0.0081
		$y = -34.815x^3 + 28.138x^2 - 7.6531x + 0.7374$	0.264	0.0080
		$y = 0.1394e^{5.166x}$	0.227	0.0082
	$x = 1/R_{666}$	$y = 0.013x - 0.0134$	0.251	0.0081
		$y = -0.0004x^3 + 0.013x^2 - 0.0782x + 0.1682$	0.264	0.0080
		$y = 0.0124e^{0.282x}$	0.2376	0.0084
	$x = \text{Log}R_{666}$	$y = -0.0558x - 0.0387$	0.245	0.0081
		$y = -0.1746x^3 - 0.5915x^2 - 0.6622x - 0.2077$	0.264	0.0082
		$y = 0.0071e^{1.212x}$	0.233	0.0083
	$x = \sqrt{R_{666}}$	$y = -0.2306x + 0.154$	0.241	0.0081
		$y = -21.463x^3 + 34.296x^2 - 18.313x + 3.304$	0.264	0.0082
		$y = 0.4688e^{5.01x}$	0.230	0.0084
Quality period	$x = R_{666}$	$y = -0.3209x + 0.1203$	0.215	0.0081
		$y = 24.822x^3 - 9.6462x^2 + 0.4149x + 0.1602$	0.246	0.0080
		$y = 0.1721e^{5.679x}$	0.220	0.0081
	$x = 1/R_{666}$	$y = 0.0145x - 0.0166$	0.228	0.0081
		$y = -0.0054x^3 + 0.0873x^2 - 0.4458x + 0.7828$	0.245	0.0081
		$y = 0.0153e^{0.2568x}$	0.233	0.0081
	$x = \text{Log}R_{666}$	$y = -0.0686x - 0.0543$	0.222	0.0082
		$y = 0.438x^3 + 2.2879x^2 + 3.8672x + 2.1798$	0.246	0.0080
		$y = 0.0078e^{1.213x}$	0.227	0.0081
	$x = \sqrt{R_{666}}$	$y = -0.2971x + 0.189$	0.219	0.0081
		$y = 28.267x^3 - 34.101x^2 + 13.124x - 1.5215$	0.246	0.0080
		$y = 0.5798e^{5.255x}$	0.224	0.0082
Postpartum period	$x = R_{664}$	$y = -0.4319x + 0.1418$	0.371	0.0074
		$y = -621.75x^3 + 418.04x^2 - 93.809x + 7.0712$	0.478	0.0071
		$y = 0.3988e^{9.803x}$	0.4287	0.0072
	$x = 1/R_{664}$	$y = 0.0206x - 0.0471$	0.393	0.0074
		$y = 0.0473x^3 - 0.635x^2 + 2.8474x - 4.2225$	0.477	0.0071
		$y = 0.0056e^{0.4627x}$	0.4432	0.0072
	$x = \text{Log}R_{664}$	$y = -0.0945x - 0.0964$	0.382	0.0074
		$y = -5.5455x^3 - 24.909x^2 - 37.326x - 18.615$	0.478	0.0071
		$y = 0.0018e^{2.135x}$	0.436	0.0073
	$x = \sqrt{R_{664}}$	$y = -0.4043x + 0.2364$	0.377	0.0075
		$y = -417.18x^3 + 670.28x^2 - 317.29x + 50.13$	0.479	0.0071
		$y = 3.377e^{9.156x}$	0.433	0.0072

3.1.2. Modeling Analysis Based on Multiple Regression

1. Modeling effect of characteristic bands based on correlation analysis screening

Based on the correlation coefficient method, the characteristic bands of R and its eight transformations were screened, and multiple sensitive bands could be selected. Therefore, the MLR method was used for modeling and fitting analysis. The results are shown in Table 3. The coefficient of determination R^2 of each model is above 0.7. Among them, $(\sqrt{R})'$ has the best fitting effect in the young and swollen fruit stages; the model determination coefficient R^2 and root mean square error RMSE are 0.74, 0.75, and 0.0047, respectively, in which R^2 are 0.55 and 0.49 higher than the univariate regression. CR has the best fitting effect during the quality period, where $R^2 = 0.78$, RMSE = 0.0038, and R^2 has a maximum increase of 0.53; $(\text{Log}R)'$ has the best fitting effect during the postpartum period, with $R^2 = 0.77$, RMSE = 0.0039, and R^2 increasing by a maximum of 0.37.

Table 3. Multiple linear regression (MLR) results of soil TN based on characteristic bands screened with correlation analysis.

Sampling Period	Input Spectrum	Regression Equation	R^2	RMSE
Young fruiting	$x = (R)'$	$y = 0.1049 - 46.70x_{571} + 116.09x_{849} - 40.17x_{1425} - 6.29x_{1925}$	0.71	0.0049
	$x = (1/R)'$	$y = 0.0409 - 9.17x_{809} - 3.46x_{849} + 5.57x_{1427} + 0.71x_{1914}$	0.73	0.0047
	$x = (\text{Log}R)'$	$y = 0.0663 - 5.49x_{559} + 47.46x_{809} - 10.96x_{1426} + 0.27x_{1927}$	0.70	0.0050
	$x = (\sqrt{R})'$	$y = 0.0915 - 38.39x_{559} + 150.86x_{809} - 45.39x_{1425} - 2.49x_{1926}$	0.74	0.0047
	$x = CR$	$y = -0.0424 - 0.49x_{664} - 2.39x_{694} + 2.77x_{878} + 0.22x_{2222}$	0.73	0.0048
Swelling fruit	$x = (R)'$	$y = 0.0758 - 22.29x_{567} + 110.10x_{836} - 41.68x_{1430} - 8.42x_{1923}$	0.71	0.0043
	$x = (1/R)'$	$y = 0.0274 - 6.50x_{803} - 8.53x_{837} + 5.07x_{1430} + 0.87x_{1916}$	0.71	0.0051
	$x = (\text{Log}R)'$	$y = 0.0986 - 9.46x_{561} + 37.68x_{836} - 4.75x_{1430} + 0.33x_{1923}$	0.72	0.0049
	$x = (\sqrt{R})'$	$y = 0.0979 - 32.91x_{548} + 103.57x_{836} - 52.49x_{1429} - 8.73x_{1922}$	0.75	0.0047
	$x = CR$	$y = 0.9725 + 0.07x_{499} - 5.77x_{692} + 4.57x_{840} + 0.29x_{2215}$	0.74	0.0047
Quality period	$x = (R)'$	$y = 0.0847 - 28.52x_{563} + 101.52x_{836} - 28.97x_{1428} - 15.3x_{1915}$	0.73	0.0042
	$x = (1/R)'$	$y = 0.0451 - 4.82x_{800} - 2.84x_{827} + 4.87x_{1415} + 1.15x_{1915}$	0.72	0.0043
	$x = (\text{Log}R)'$	$y = 0.0751 - 3.94x_{561} + 31.18x_{827} - 5.83x_{1428} - 5.42x_{1915}$	0.72	0.0043
	$x = (\sqrt{R})'$	$y = 0.1007 - 40.73x_{533} + 93.54x_{827} - 38.18x_{1428} - 17.50x_{1915}$	0.76	0.0040
	$x = CR$	$y = 1.5867 - 1.2x_{640} - 3.63x_{699} + 2.77x_{849} + 0.57x_{2219}$	0.78	0.0038
Postpartum period	$x = (R)'$	$y = 0.0799 - 36.67x_{585} + 151.09x_{836} - 33.74x_{1428} - 13.57x_{1924}$	0.73	0.0044
	$x = (1/R)'$	$y = 0.0345 - 6.06x_{817} - 7.12x_{837} + 6.76x_{1415} + 2.09x_{1914}$	0.74	0.0043
	$x = (\text{Log}R)'$	$y = 0.0951 - 9.69x_{562} + 39.75x_{817} - 3.77x_{1428} - 7.73x_{1914}$	0.77	0.0039
	$x = (\sqrt{R})'$	$y = 0.1074 - 47.69x_{548} + 137.48x_{837} - 37.11x_{1428} - 13.51x_{1915}$	0.76	0.0040
	$x = CR$	$y = 0.9601 + 2.73x_{662} - 0.81x_{678} + 1.89x_{863} + 0.79x_{2223}$	0.75	0.0041

Using data from the 40 sampling points in the validation area for each fertilization period, the four selected models with acceptable accuracy were tested. The validation accuracy was evaluated using two indicators: R^2 and RMSE, and a 1:1 histogram was created (Figure 2).

It can be seen that, except for the fruit swelling period ($R^2 = 0.803$) and the postpartum period ($R^2 = 0.779$), the validation accuracy of the other period models is less than 0.75.

2. Modeling effect of characteristic bands based on SMLR screening

Using the characteristic bands screened based on SMLR, MLR fitting was performed on the soil TN content for each fertilization period. The results are shown in Table 4.

From Table 4, it can be seen that based on the principle of maximum R^2 , models with better soil TN estimation for each fertilization period were selected. Among them, the optimal models for each fertilization period are the MLR models based on CR transformation, with R^2 reaching 0.82, 0.84, 0.85, and 0.82, and RMSE being 0.0039, 0.0037, 0.0031, and 0.0034, respectively. Compared with the model established based on correlation analysis screening method, the modeling accuracy increased by 0.095, 0.099, 0.067, and 0.068, respectively, and the RMSE decreased by about 0.001. Secondly, the MLR model based on $(\text{Log}R)'$ performed well, with R^2 of 0.81, 0.81, 0.80, and 0.80, and RMSE of 0.0040, 0.0041,

0.0036, and 0.0036, respectively. The maximum variance inflation factor (Max VIF) of all the selected model variables above is less than 10, avoiding the influence of multicollinearity of model variables. To further test the stability and reliability of the model prediction, the four selected models were tested using data from 40 sampling points in the validation area during each fertilization period, and a 1:1 histogram was created (Figure 3). As shown in the figure, except for the quality period ($R^2 = 0.702$), the modeling accuracy of the other three models is higher than 0.8, indicating good validation results.

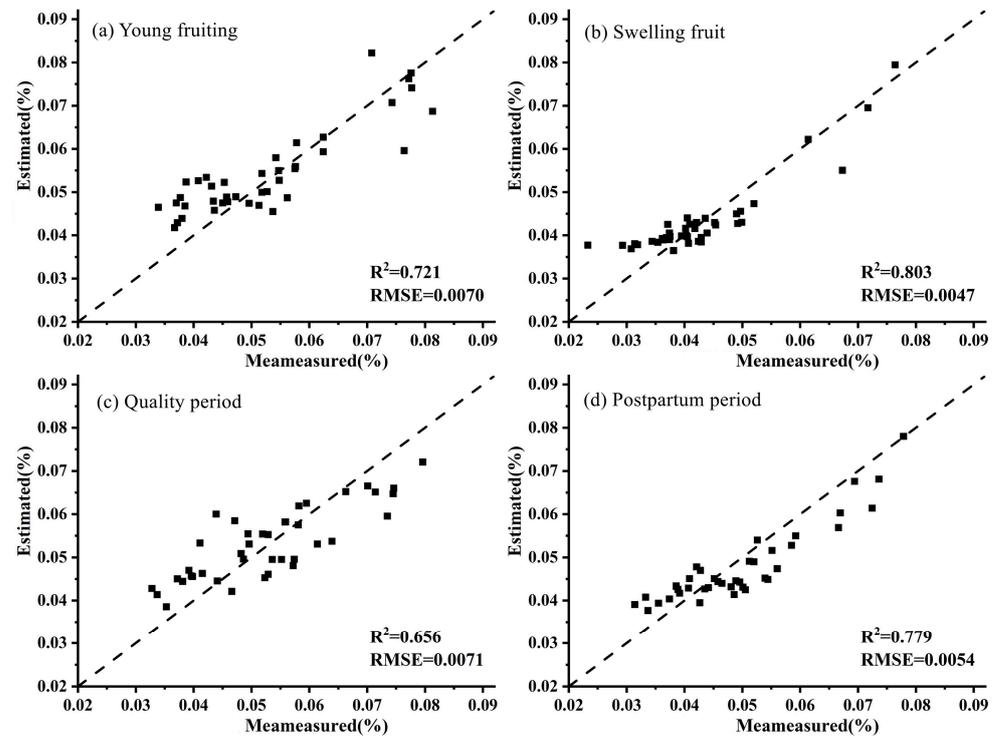


Figure 2. Comparison of measured and estimated values of soil TN content during different fertilization periods using model based on characteristic bands screened with correlation analysis. (a) young fruiting, (b) swelling fruit, (c) quality period, (d) postpartum period.

Table 4. Multiple linear regression (MLR) fitting results of soil TN content based on stepwise multiple linear regression (SMLR) screening characteristic bands.

Sampling Period	Input Spectrum	Variables Number	Regression Equation	R ²	RMSE	Max VIF
Young fruiting	$x = (R)'$	8	$y = 0.0848 - 66.35x_{570} + 82.79x_{679} + 79.45x_{808} + 1.66x_{845} + 10.11x_{1297} - 41.49x_{1423} - 14.43x_{1967} + 1.21x_{2375}$	0.77	0.0043	4.69
	$x = (1/R)'$	10	$y = 0.0396 + 0.83x_{541} + 0.72x_{574} - 2.75x_{678} - 3.90x_{706} - 6.14x_{808} + 3.34x_{1462} + 2.28x_{1589} + 3.37x_{1661} + 0.19x_{1780} - 0.30x_{2339}$	0.81	0.004	8.29
	$x = (\text{Log}R)'$	11	$y = 0.0817 - 9.73x_{541} + 26.18x_{808} + 31.94x_{848} - 3.99x_{1462} - 8.08x_{1561} - 11.64x_{1696} - 1.36x_{1804} - 4.08x_{2046} - 0.46x_{2138} - 0.70x_{2366} - 0.34x_{2423}$	0.81	0.004	7.55
	$x = (\sqrt{R})'$	12	$y = 0.0874 - 36.52x_{541} + 164.14x_{808} - 13.20x_{975} + 21.72x_{1001} + 28.58x_{1281} - 58.75x_{1422} + 14.54x_{1588} + 1.73x_{2138} + 9.52x_{2236} + 3.41x_{2311} + 3.76x_{2342} - 1.04x_{2423}$	0.76	0.0044	6.85
	$x = CR$	10	$y = 1.5527 - 5.71x_{693} + 4.95x_{712} - 9.21x_{806} + 1.90x_{876} - 0.08x_{972} + 2.10x_{1092} + 3.72x_{1695} + 1.16x_{1830} - 0.52x_{2287} + 0.20x_{2351}$	0.82	0.0039	8.57

Table 4. Cont.

Sampling Period	Input Spectrum	Variables Number	Regression Equation	R ²	RMSE	Max VIF
Swelling fruit	$x = (R)'$	8	$y = 0.0761 - 46.65x_{580} + 55.23x_{691} + 70.7x_{813} - 9.69x_{1009} + 23.2x_{1141} + 3.28x_{1374} - 65.84x_{1428} + 3.66x_{2153}$	0.73	0.0049	5.01
	$x = (1/R)'$	10	$y = 0.0373 + 0.75x_{554} + 2.18x_{583} - 3.45x_{680} - 2.6x_{691} - 4.5x_{701} - 4.78x_{812} - 1.88x_{1099} + 7.73x_{1185} - 1.59x_{2097} - 0.08x_{2301}$	0.79	0.0043	9.58
	$x = (\text{Log}R)'$	11	$y = 0.0852 - 5.08x_{554} - 9.97x_{560} + 14.49x_{703} + 27.24x_{777} + 3.26x_{1062} + 24.78x_{1192} - 2.8x_{1367} + 0.61x_{1627} - 4.98x_{1931} + 0.39x_{2301} - 0.19x_{2438}$	0.81	0.0040	7.17
	$x = (\sqrt{R})'$	12	$y = 0.0837 - 27.62x_{565} - 37.51x_{580} + 105.86x_{719} + 27.28x_{775} + 26.34x_{901} + 7.42x_{1130} + 16.83x_{1292} - 18x_{1375} - 53.69x_{1418} + 17.26x_{1740} - 15.78x_{1931} + 0.06x_{2296}$	0.78	0.0044	6.57
	$x = CR$	10	$y = 3.3908 - 2.85x_{675} + 4.72x_{837} - 1.9x_{983} - 1.81x_{1008} + 2.58x_{1087} - 3.77x_{1352} + 0.5x_{1841} - 1.36x_{2092} + 1.17x_{2188} - 0.57x_{2293}$	0.84	0.0037	7.11
Quality period	$x = (R)'$	8	$y = 0.0992 + 7.34x_{402} - 27.67x_{556} - 42.26x_{562} + 126.85x_{893} + 16.7x_{1000} + 23.34x_{1020} + 9.85x_{1330} + 4.28x_{2147}$	0.66	0.0047	5.75
	$x = (1/R)'$	10	$y = 0.0519 + 0.05x_{428} - 0.05x_{435} + 0.53x_{511} - 7.06x_{816} - 6.96x_{826} + 0.37x_{864} + 0.37x_{2041} + 0.33x_{2184} - 0.47x_{2313} + 0.06x_{2450}$	0.74	0.0041	6.28
	$x = (\text{Log}R)'$	11	$y = 0.0836 - 7.04x_{555} - 6.22x_{590} + 15.12x_{677} + 20.32x_{826} + 10.25x_{864} + 15.25x_{1328} + 0.27x_{1752} - 1.6x_{2041} + 3.15x_{2179} - 1.08x_{2292} - 0.21x_{2450}$	0.80	0.0036	5.59
	$x = (\sqrt{R})'$	12	$y = 0.1187 - 48.73x_{543} - 33.76x_{549} + 109.16x_{802} + 4.64x_{865} - 21.4x_{1547} + 61.57x_{1572} + 41.88x_{1579} - 8.63x_{1862} + 17.39x_{2130} - 3.69x_{2192} + 0.32x_{2240} - 1.43x_{2213}$	0.79	0.0037	8.31
	$x = CR$	10	$y = -20.7013 + 0.11x_{490} - 4.88x_{698} - 6.61x_{752} - 5.02x_{777} + 75.52x_{779} - 37.54x_{780} + 4.01x_{844} - 5.66x_{1322} + 1.39x_{1332} - 0.54x_{2066}$	0.85	0.0031	9.21
Postpartum period	$x = (R)'$	8	$y = 0.0765 - 24.13x_{562} - 86.39x_{660} + 33.61x_{679} + 91.15x_{809} + 83.75x_{815} - 22.64x_{1922} - 3.92x_{2255} + 4.53x_{2347}$	0.79	0.0038	7.04
	$x = (1/R)'$	10	$y = 0.0351 + 0.15x_{476} + 1.41x_{551} - 6.48x_{679} - 7.35x_{837} - 7.37x_{1541} - 9.32x_{1575} + 3.78x_{1666} + 0.85x_{1676} + 1.24x_{1771} - 0.53x_{2269}$	0.81	0.0034	8.81
	$x = (\text{Log}R)'$	11	$y = 0.0994 - 11.48x_{541} + 28.76x_{829} + 23.56x_{836} + 6.79x_{1542} + 5.84x_{1772} - 0.71x_{2135} + 2.55x_{2198} + 1.42x_{2269} - 0.42x_{2316} + 0.84x_{2391} + 0.4x_{2401}$	0.80	0.0036	5.42
	$x = (\sqrt{R})'$	12	$y = 0.1133 - 65.15x_{562} + 76.01x_{809} + 12.11x_{858} + 26.67x_{1044} + 30.36x_{1227} + 1.04x_{1550} - 14.9x_{1602} + 15.07x_{1676} - 27.51x_{1912} - 15.56x_{2070} + 7.35x_{2199} + 8.81x_{2269}$	0.79	0.0037	7.57
	$x = CR$	10	$y = 39.6808 - 3.47x_{676} - 41.67x_{773} + 2.92x_{858} - 1.78x_{974} + 4.26x_{1120} - 2.69x_{1732} + 2.71x_{1746} - 0.66x_{1841} + 0.85x_{1874} - 0.08x_{2437}$	0.82	0.0034	6.91

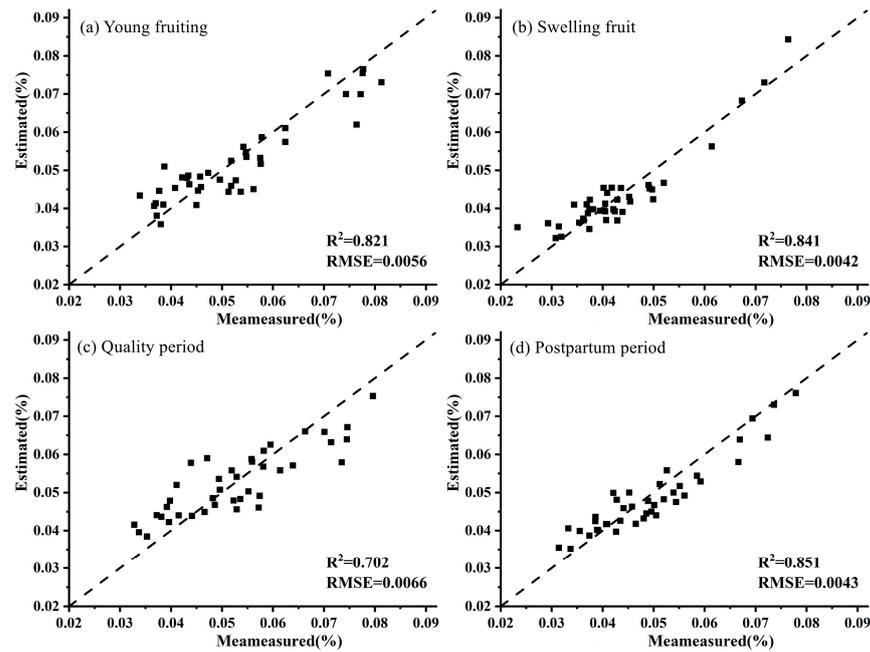


Figure 3. Comparison of measured and estimated values of soil TN content during different fertilization periods using model based on characteristic bands screened with stepwise multiple linear regression. (a) young fruiting, (b) swelling fruit, (c) quality period, (d) postpartum period.

3.1.3. Modeling and Analysis Based on MEA-BPNN

The characteristic bands selected through the correlation coefficient method and SMLR analysis in R and its various transformation forms were used as input spectra, and the MEA-BPNN was used for modeling. The results are shown in Table 5.

Table 5. Modeling and validation results of soil TN content estimation based on BP (back propagation) neural network optimized by Mind Evolution Algorithm (MEA-BPNN).

Sensitive Band Screening			Correlation Analysis				Stepwise Multiple Linear Regression				
Sampling Period	Input Spectrum	Variables Number	Modeling		Validation		Variables Number	Modeling		Validation	
			R ²	RMSE	R ²	RMSE		R ²	RMSE	R ²	RMSE
Young fruiting	(R)'	5	0.75	0.0045	0.73	0.0068	8	0.78	0.0048	0.76	0.0065
	(1/R)'	5	0.75	0.0046	0.72	0.0070	10	0.83	0.0045	0.82	0.0056
	(LogR)'	5	0.73	0.0048	0.75	0.0066	11	0.86	0.0043	0.85	0.0052
	(√R)'	5	0.77	0.0044	0.76	0.0065	12	0.83	0.0045	0.81	0.0058
	CR	5	0.74	0.0048	0.71	0.0072	10	0.87	0.0043	0.87	0.0047
Swelling fruit	(R)'	5	0.74	0.0042	0.68	0.0060	8	0.85	0.0042	0.84	0.0043
	(1/R)'	5	0.72	0.0045	0.66	0.0063	10	0.85	0.0041	0.84	0.0042
	(LogR)'	5	0.73	0.0042	0.71	0.0058	11	0.85	0.0037	0.87	0.0038
	(√R)'	5	0.75	0.0041	0.76	0.0053	12	0.84	0.0037	0.87	0.0038
	CR	5	0.81	0.0040	0.79	0.0041	10	0.86	0.0033	0.89	0.0035
Quality period	(R)'	5	0.75	0.0041	0.68	0.0069	8	0.82	0.0048	0.69	0.0068
	(1/R)'	5	0.73	0.0042	0.66	0.0070	10	0.84	0.0046	0.81	0.0053
	(LogR)'	5	0.72	0.0043	0.60	0.0077	11	0.85	0.0046	0.82	0.0052
	(√R)'	5	0.76	0.0040	0.65	0.0072	12	0.78	0.0051	0.74	0.0062
	CR	5	0.78	0.0038	0.66	0.0071	10	0.86	0.0043	0.84	0.0049
Postpartum period	(R)'	5	0.75	0.0042	0.70	0.0063	8	0.87	0.0043	0.85	0.0045
	(1/R)'	5	0.77	0.0055	0.69	0.0064	10	0.87	0.0043	0.86	0.0044
	(LogR)'	5	0.77	0.0056	0.70	0.0063	11	0.84	0.0046	0.83	0.0047
	(√R)'	5	0.77	0.0055	0.71	0.0062	12	0.85	0.0050	0.80	0.0051
	CR	5	0.79	0.0053	0.75	0.0057	10	0.88	0.0038	0.87	0.0042

Similar to the MLR method, the model based on SMLR method for screening variables has better performance, with modeling accuracy of over 0.8 except for the young fruit stage (R' ($R^2 = 0.78$)). Based on the performance of the models in the comprehensive modeling and validation sets, the selected models for each period are all MEA-BPNN models based on CR, with R^2 of 0.87, 0.86, 0.86, and 0.88, and RMSE of 0.0043, 0.0033, 0.0043, and 0.0038, respectively. Compared with the model established by MLR method, the accuracy increased by 0.054, 0.021, 0.012, and 0.058, respectively, and the validation accuracy is 0.87, 0.89, 0.84, and 0.87, respectively. In summary, it can be seen that the SMLR method is more effective in screening sensitive bands, and this method will be used for characteristic band screening in future research.

3.2. SCI Screening Results

3.2.1. Independent Soil SCIs for Each Fertilization Period

Figure 4 shows the Isoline map of the correlation coefficient between soil SCI and TN content during each fertilization period. It can be seen that the form of contour maps for each fertilization period is not significantly different, with only slight differences in correlation size. Among them, RSI is the most stable, and the sensitive areas of soil TN are the same in the four fertilization periods, all located in the range of 800–900 nm, 1900–1950 nm, and 2200–2300 nm bands. Under DI and NDSI, the sensitive regions of soil TN during the four fertilization periods are all located in the range of 1900–1950 nm and 2200–2300 nm bands. However, the correlation coefficient of soil TN varies over time in the wavelength range of 800–900 nm, and the quality period performance is poor. The absolute value of the correlation coefficient R is less than 0.6. Therefore, this region will not be considered in the subsequent screening of sensitive differences and normalized spectral indices.

To avoid the multicollinearity effect between variables caused by close band distance, an SCI band combination with the highest correlation was selected for each fertilization period in different sensitive areas of the spectral index (Table 6). It can be seen that the combination bands of each SCI are not within the screening range of the spectral characteristic bands mentioned above, so calculating spectral indices based solely on characteristic bands cannot fully explore spectral information.

Table 6. Spectral characteristic index of each fertilization period.

Spectral Characteristic Index	Fertilization Period							
	Young Fruiting		Swelling Fruit		Quality Period		Postpartum Period	
	Band Combination	R						
RSI	(R_{860}, R_{870})	0.84	(R_{835}, R_{844})	0.80	(R_{829}, R_{814})	0.78	(R_{826}, R_{842})	0.89
	(R_{1907}, R_{1941})	−0.80	(R_{1905}, R_{1936})	−0.84	(R_{1902}, R_{1949})	−0.80	(R_{1909}, R_{1926})	−0.91
	(R_{2203}, R_{2283})	−0.77	(R_{2210}, R_{2292})	−0.82	(R_{2203}, R_{2216})	−0.78	(R_{2213}, R_{2300})	−0.89
DI	(R_{1907}, R_{1940})	−0.80	(R_{1906}, R_{1935})	−0.86	(R_{1903}, R_{1949})	−0.81	(R_{1909}, R_{1926})	−0.91
	(R_{2208}, R_{2285})	−0.77	(R_{2210}, R_{2291})	−0.82	(R_{2215}, R_{2303})	−0.79	(R_{2230}, R_{2267})	−0.91
NDSI	(R_{1907}, R_{1943})	0.83	(R_{1907}, R_{1937})	0.82	(R_{1909}, R_{1948})	0.77	(R_{1910}, R_{1934})	0.92
	(R_{2202}, R_{2283})	0.80	(R_{2209}, R_{2286})	0.83	(R_{2163}, R_{2218})	0.78	(R_{2211}, R_{2285})	0.89

The combination bands of various soil spectral characteristic indices during each fertilization period have little interval, and the sensitive areas are mainly concentrated in the range of 800–900 nm, 1900–1950 nm, and 2200–2300 nm bands. Among them, the correlation between characteristic parameters is higher in the range of 1900–1950 nm. To reduce data redundancy, only these three regions are used as screening areas for spectral characteristic parameters in conducting large-scale data analysis during the entire fertilization period.

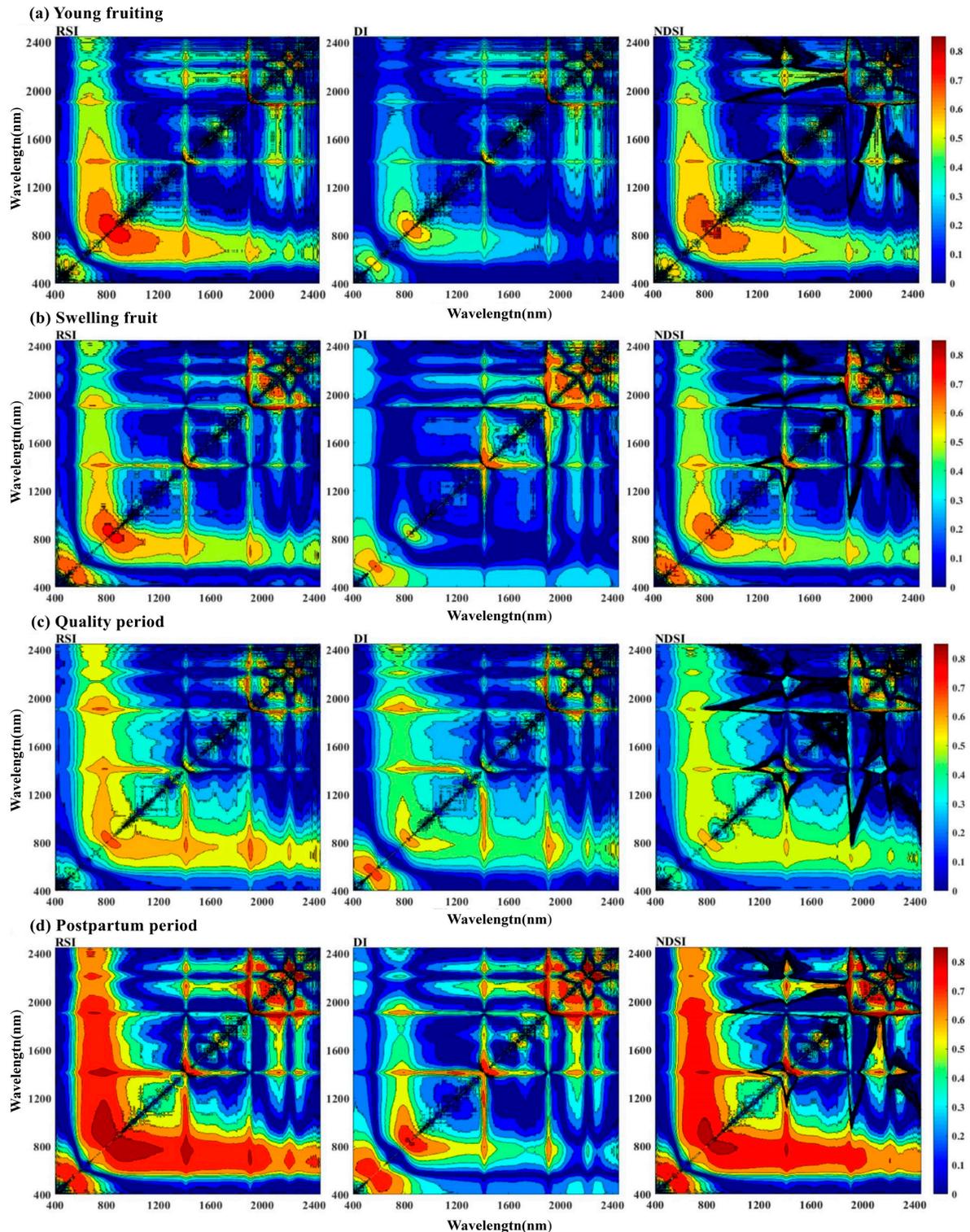


Figure 4. Isoline maps of correlation coefficient between soil spectral characteristic index (SCI) and TN content for each fertilization period. (a) young fruiting, (b) swelling fruit, (c) quality period, (d) postpartum period. Note: RSI: ratio spectral index; DI: difference spectral index; NDSI: Normalized difference spectral index. The same below.

3.2.2. Comprehensive Soil SCI during the Entire Fertilization Period

Based on the sensitive areas of SCI during each fertilization period, the screening of comprehensive soil SCI for the entire fertilization period of the orchard was conducted. A

band combination with the highest correlation was selected in different sensitive areas of the spectral indices. The results are shown in Table 7. Correlation coefficient contour analysis between SCI and TN during the entire fertilization period was performed (Figure 5). It can be seen that 1904 nm and 1949 nm perform well in both RSI and DI, consistent with the results of each period, indicating that this band range has advantages in SCI screening.

Table 7. Spectral characteristic index of the entire fertilization period.

Spectral Characteristic Index	Band Combination	R
RSI	(R_{808} , R_{810})	0.62
	(R_{1904} , R_{1949})	−0.83
	(R_{2221} , R_{2300})	−0.72
DI	(R_{1904} , R_{1949})	−0.85
	(R_{2210} , R_{2286})	0.79
NDSI	(R_{1908} , R_{1954})	0.79
	(R_{2210} , R_{2286})	0.68

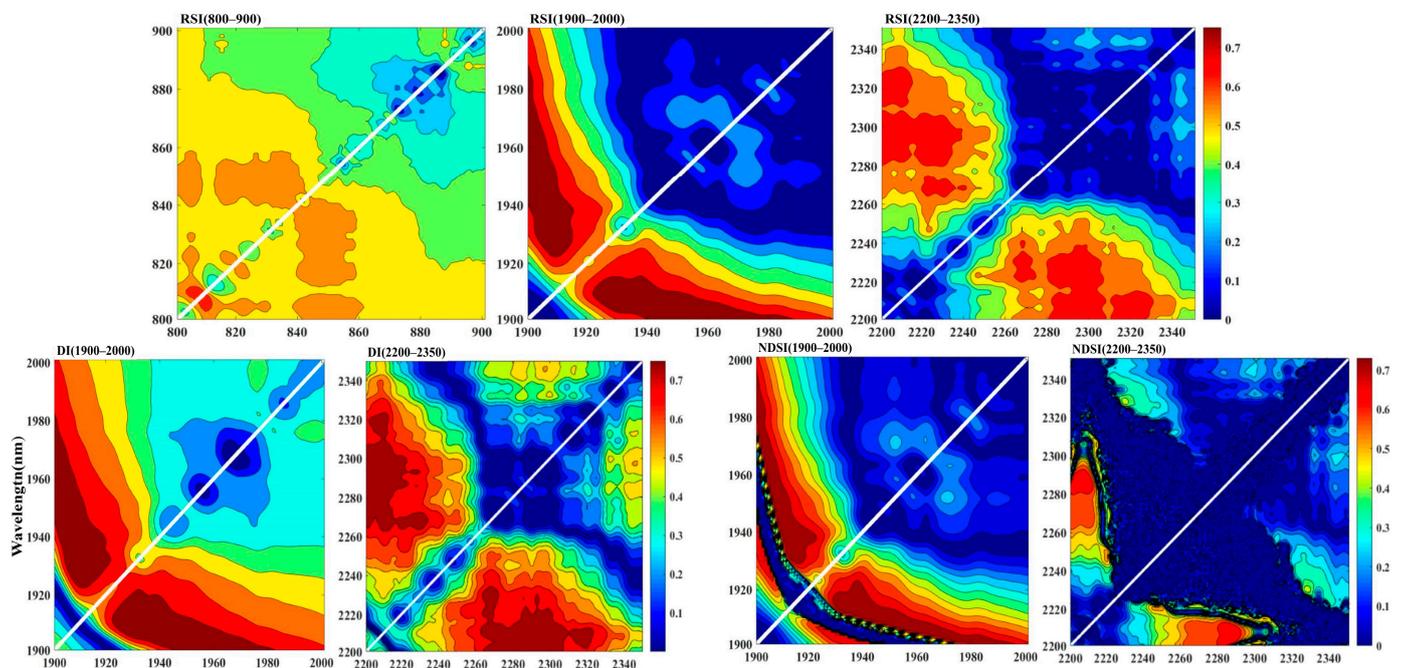


Figure 5. Isoline maps of the correlation coefficient between soil spectral characteristic index (SCI) and soil TN content during the entire fertilization period.

3.3. Estimation Model Based on SCI and Characteristic Band Combination

3.3.1. Independent Estimation Models for Each Fertilization Period

The seven SCIs (three RSIs, three DIs, and two NDSIs) were used together with the SMLR selected characteristic bands as independent variables, and the soil TN content in each period was used as the training target. MLR and MEA-BPNN were used for modeling, respectively. The results are shown in Table 8.

Table 8. Estimation model for each fertilization period based on spectral characteristic index (SCI) and characteristic band combination.

Model		Multiple Linear Regression					Mind Evolutionary Algorithm-BPNN				
Sampling Period	Input Spectrum	Variables Number	Modeling		Validation		Variables Number	Modeling		Validation	
			R ²	RMSE	R ²	RMSE		R ²	RMSE	R ²	RMSE
Young fruiting	(R)'	15	0.83	0.0037	0.77	0.0064	15	0.89	0.0044	0.88	0.0047
	(1/R)'	17	0.85	0.0036	0.80	0.0058	17	0.88	0.0045	0.85	0.0045
	(LogR)'	18	0.86	0.0035	0.78	0.0062	18	0.91	0.0040	0.88	0.0047
	(√R)'	19	0.83	0.0038	0.74	0.0068	19	0.92	0.0038	0.90	0.0040
	CR	17	0.87	0.0032	0.81	0.0058	17	0.94	0.0032	0.92	0.0033
Swelling fruit	(R)'	15	0.81	0.0041	0.73	0.0056	15	0.93	0.0290	0.92	0.0031
	(1/R)'	17	0.86	0.0035	0.79	0.0050	17	0.92	0.0030	0.91	0.0032
	(LogR)'	18	0.83	0.0038	0.84	0.0043	18	0.93	0.0027	0.91	0.0031
	(√R)'	19	0.82	0.0039	0.78	0.0050	19	0.92	0.0031	0.91	0.0031
	CR	17	0.86	0.0035	0.85	0.0042	17	0.95	0.0024	0.93	0.0029
Quality period	(R)'	15	0.85	0.0031	0.68	0.0069	15	0.87	0.0043	0.85	0.0045
	(1/R)'	17	0.85	0.0031	0.66	0.0071	17	0.90	0.0039	0.88	0.0041
	(LogR)'	18	0.85	0.0030	0.68	0.0068	18	0.88	0.0042	0.85	0.0045
	(√R)'	19	0.86	0.0030	0.69	0.0068	19	0.89	0.0040	0.88	0.0040
	CR	17	0.88	0.0028	0.71	0.0065	17	0.92	0.0035	0.91	0.0037
Postpartum period	(R)'	15	0.82	0.0033	0.82	0.0048	15	0.92	0.0033	0.89	0.0035
	(1/R)'	17	0.83	0.0032	0.81	0.0051	17	0.92	0.0032	0.90	0.0034
	(LogR)'	18	0.83	0.0032	0.82	0.0049	18	0.91	0.0035	0.87	0.0040
	(√R)'	19	0.82	0.0034	0.79	0.0052	19	0.90	0.0036	0.87	0.0041
	CR	17	0.87	0.0032	0.86	0.0043	17	0.94	0.0027	0.93	0.0033

Overall, regardless of the modeling method, the addition of SCI has improved the modeling accuracy. Among them, the two models constructed based on CR transformation have slightly higher modeling and validation accuracy than the other transformation methods. Therefore, it can be proven that this transformation form has advantages in spectral processing. Compared with the two modeling methods, the MEA-BPNN model still maintains its modeling advantages, with better performance in R² and RMSE than the MLR model, and some models have modeling accuracy greater than 0.9.

Based on the comprehensive performance of each fertilization period model, the optimal models were selected based on the CR-MEA-BP method, with R² of 0.94, 0.95, 0.92, and 0.94, and RMSE of 0.0032, 0.0024, 0.0035, and 0.0027, respectively. Compared with MLR, the modeling accuracy increased by 0.068, 0.089, 0.054, and 0.066, respectively, and the validation accuracy was 0.92, 0.93, 0.91, and 0.93, respectively.

In each fertilization period, the optimized model with SCI correction and its original model were used to predict soil TN, and the prediction curves of SCI-MEA-BPNN and MEA-BPNN were obtained for each fertilization period (Figure 6). It can be seen that the model prediction curve with the addition of SCI is more consistent with the measured value curve, and the prediction effect is better. The overall trend of the MEA-BP predicted value curve without SCI input is consistent with the measured value curve, but the degree of conformity is slightly poor. It has a good prediction effect on samples with gentle changes, and a large prediction error at extreme points. This demonstrates the positive effect of adding SCI to the construction of soil TN content prediction models.

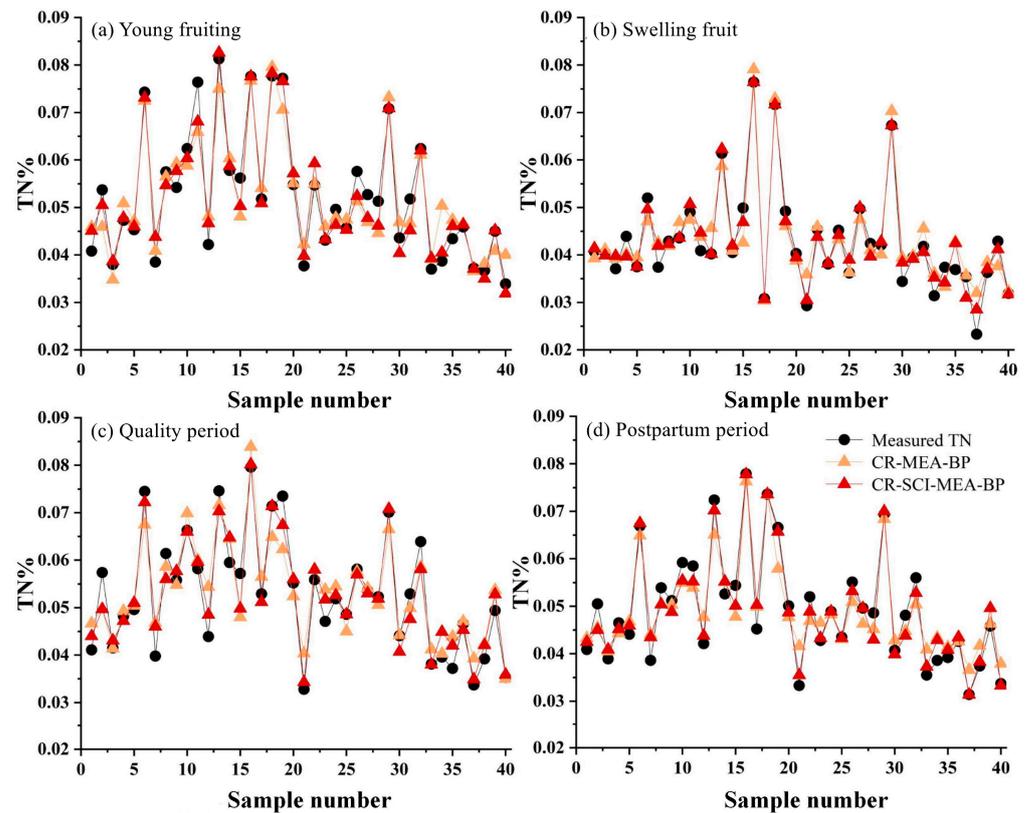


Figure 6. Comparison of measured and estimated values of soil TN content in different models during different fertilization periods. (a) young fruiting, (b) swelling fruit, (c) quality period, (d) postpartum period.

3.3.2. Comprehensive Estimation Model for the Entire Fertilization Period

Based on the soil TN content, original spectral reflectance, and various transformation forms of 400 sample data from 100 sampling points during four fertilization periods in the modeling area, the SMLR method was used to extract characteristic bands, and together with the seven extracted SCIs, a unified soil TN estimation model for the orchard’s entire fertilization period was constructed. And a total of 160 sample data from 40 sampling points during four fertilization periods in the validation area were used for validation.

The MLR and MEA-BPNN modeling results of the soil TN content based on the original soil spectrum and its first derivative transformation during the entire fertilization period of the orchard are shown in Table 9.

Table 9. Estimating and modeling results of soil TN content during the entire fertilization period.

Model	Input Spectrum	Variables Number	Modeling		Validation		Input Spectrum	Variables Number	Modeling		Validation	
			R ²	RMSE	R ²	RMSE			R ²	RMSE	R ²	RMSE
MLR	R	1	0.35	0.0081	0.23	0.012	R + SCI	8	0.788	0.0045	0.611	0.0079
	(R)′	15	0.814	0.0042	0.793	0.0057	(R)′ + SCI	22	0.829	0.0040	0.798	0.0057
	(1/R)′	7	0.838	0.0039	0.797	0.0057	(1/R)′ + SCI	14	0.853	0.0038	0.804	0.0056
	(LogR)′	12	0.835	0.004	0.796	0.0057	(LogR)′ + SCI	19	0.854	0.0037	0.813	0.0055
	(√R)′	14	0.841	0.0039	0.826	0.0053	(√R)′ + SCI	21	0.859	0.0035	0.831	0.0049
	CR	10	0.823	0.0041	0.812	0.0055	CR + SCI	17	0.838	0.0039	0.828	0.0051
MEA-BP	(R)′	15	0.826	0.0052	0.814	0.0054	(R)′ + SCI	22	0.865	0.0044	0.857	0.0045
	(1/R)′	7	0.844	0.0051	0.826	0.0052	(1/R)′ + SCI	14	0.869	0.0042	0.861	0.0045
	(LogR)′	12	0.845	0.004	0.822	0.004	(LogR)′ + SCI	19	0.863	0.0045	0.855	0.0046
	(√R)′	14	0.855	0.0039	0.837	0.0051	(√R)′ + SCI	21	0.872	0.004	0.867	0.0046
	CR	10	0.861	0.0038	0.841	0.0041	(CR)′ + SCI	17	0.899	0.0038	0.890	0.0041

It can be seen that the addition of SCI has less advantages in the estimation model of the entire fertilization period than in the small sample size estimation model of each period. The modeling accuracy of each model based on the first derivative transformation and the CR transformation after adding the index is relatively small, and the validation accuracy is not significantly different from before. However, the prediction performance of the model based on the original spectral reflectance is significantly improved after adding an index. This indicates that the SCI has a better correction effect on linear models with fewer variables and lower modeling accuracy.

Unlike the results of soil TN estimation models for different fertilization periods, the advantage of the CR method in the study of soil TN content MLR models for the entire fertilization period is not significant, while the modeling effect of $(\sqrt{R})'$ and $(\text{Log}R)'$ is better. According to the size of R^2 , it can be seen that the addition of SCI has a more positive effect on the MEA-BPNN. For instance, The accuracy of modeling and validation increases more compared to MLR models. The performance of each spectral transformation is basically consistent with the results of soil TN content estimation modeling in each fertilization period. The CR-SCI-MEA-BP model has the best prediction performance, with R^2 and RMSE of 0.899, 0.0038, and 0.89, 0.0041, respectively, in the modeling and validation sets.

4. Discussion

In this article, the extraction method of hyperspectral parameters for TN content in apple orchard soil was explored and its estimation model was optimized. The results show that SMLR is effective and has significant advantages in extracting hyperspectral characteristic bands of soil TN content; moreover, the combination of SCI and characteristic bands, and the use of the MEA-BPNN method for independent and comprehensive estimation of soil TN content during each fertilization period and entire fertilization period, have the best modeling effect.

4.1. Collection and Preprocessing of Hyperspectral Data on TN in Orchard Soil

The soil spectrum is the result of the joint action of various components in the soil, and the relationship between soil spectral reflectance and various nutrients in the soil is extremely complex [9,11]. Most apple orchards are located in mountainous and hilly areas, which are affected by factors such as geography, climate, and soil texture. The soil nitrogen content is low, the spectral signal is weak, and it is easy to be disturbed or masked, making it difficult to obtain effective spectral information [31]. For this reason, previous studies have mostly collected spectra of ground and sieved samples under strictly controlled indoor conditions, resulting in a higher modeling accuracy of the measured spectra. This approach is far from orchards and extends the estimation time, which is not suitable for the real-time management needs of modern orchards [29,32]. However, field measurements are greatly influenced by the weather and surrounding environmental conditions, which can lead to the inability to guarantee the accuracy of the model [43]. Especially, modern orchards generally adopt ecological planting methods such as intercropping or grass planting [44], where light is often obstructed by fruit trees and other plants, and it is not suitable to directly collect soil spectra. Therefore, this article adopts the method of manually controlling the surrounding outdoor environment and using natural sunlight as a light source to collect spectral data. In this way, spectral collection has taken a step closer from indoors to orchard fields, providing a beneficial exploration for future spectral collection in complex orchard environments. At the same time, in order to assist with orchard fertilization management, soil samples were collected 10 days before the fertilization day. This is beneficial for understanding the soil nutrient situation after the previous fertilization period and providing guidance for the next fertilization period.

Performing $(1/R)$, $(\text{Log}R)$, (\sqrt{R}) , and their first derivative transformations on the original spectral reflectance R can further reduce the noise effects caused by non-target factors such as lighting conditions, soil particle size, and air moisture in spectral measurements; highlight effective information in the spectrum; improve the sensitivity of characteristic

bands [28]; and explore the optimal spectral transformation form [45]. CR, also known as envelope division, can effectively enhance the spectral characteristics of the region of interest [46–48]. The results of this study also demonstrated that CR is beneficial for extracting characteristic spectral information and effectively improving the response ability of spectral data to soil TN. Through modeling and prediction performance analysis, it was found that various first-order derivative transformations and CR transformations of reflectance can improve the prediction accuracy of the corresponding fertilization period estimation model, which has positive significance for the spectral estimation research of soil TN. However, the prediction performance of the $1/R$, LogR , and \sqrt{R} models is not ideal. In addition, the characteristic bands selected by various transformations are different, and the estimation models for soil TN content during different fertilization periods also differ.

4.2. Extraction of Characteristic Bands and Selection of SCI

Due to the varying degrees of influence of different spectral bands on the soil TN content, in order to establish the optimal model, it is necessary to filter out the bands that have little impact on the soil TN content and leave behind the most suitable characteristic bands for modeling [11,43,45]. This article compares the screening effects of the correlation coefficient method and SMLR based on R and its eight transformations.

The results indicate that the correlation coefficient method can determine the characteristic band range of soil TN to some extent, providing a basis for the selection of characteristic bands. Choosing the band at the extreme point of the correlation coefficient as the input variable to construct the model has the advantages of simplicity, clarity, and strong operability. However, the selection of bands located at the extreme points of the correlation coefficient is subjective and cannot estimate the interrelationships between bands at the extreme points, thus ultimately failing to achieve ideal model prediction results [49].

SMLR is a good variable screening method, and a large number of studies have shown that all independent variables retained after stepwise regression analysis have a significant impact on the dependent variable [26,50]. In SMLR, this article refers to the correlation between variables and the variance inflation factor (VIF), and selects variables included in the equation that pass the 0.05 significance level test and have a maximum variance inflation factor of no more than 10 as the characteristic bands of soil TN [51]. In this way, the problem of multicollinearity between bands that are closely spaced within the same regression equation is avoided, and the modeling effect is improved. The results of this article indicate that compared with the correlation coefficient method, using the SMLR method to screen characteristic bands can avoid collinearity issues caused by band spacing while considering spectral characteristic information [35]. Although the selected bands are not located at the extreme points of the correlation coefficient, the predictive performance and stability of the model are slightly better than the previous screening method. It should be noted that there is a significant difference between the characteristic bands selected based on SMLR and the corresponding correlation coefficient characteristic bands during each fertilization period. Except for the characteristic bands transformed by the first derivative of the original reflectance, the positions of the other transformed characteristic bands are mostly far apart. It is difficult to screen the same spectral band for multiple spectral transformations during different fertilization periods. Therefore, this paper established an estimation model and validated it for evaluation in order to select a more suitable extraction method for characteristic bands. Compared with the correlation coefficient method, the model established based on the characteristic bands selected by the SMLR method can achieve higher modeling accuracy and smaller RMSE under the same spectral transformation. Moreover, it is easier to select models with a maximum variance inflation factor of less than 10, thus avoiding the influence of multicollinearity of model variables.

It can be seen that regardless of which characteristic band selection method is used, the modeling effect of the MEA-BPNN is superior to the MLR method, and the modeling

accuracy has been improved, but the RMSE of some models has only slightly improved. The SMLR method performs better in screening sensitive bands and exhibits greater advantages. It has good practicality in constructing and predicting both linear and nonlinear models. Therefore, in the study of spectral estimation of soil TN content during the entire fertilization period, this article did not further construct a sensitive band model based on correlation analysis. Instead, the significant bands selected by SMLR and the SCI selected were mainly used as input spectra for the study.

In terms of SCI, research on models in different periods has found that the addition of SCI has a positive effect on the prediction accuracy of various types of models [36]. Overall, the addition of SCI can improve the prediction accuracy of the model, reduce RMSE, and make the estimation model more stable [37]. Therefore, this article determines the strategy for modeling using the combination of SCI and characteristic bands.

4.3. About Modeling Methods

Regression analysis is a common statistical analysis method used to determine the quantitative relationship between two or more variables [52] and is widely used in the quantitative estimation of soil nutrients [53].

The results of this article show that the modeling effect of univariate regression is not satisfactory. This indicates that even if a certain band has a high correlation with soil TN, it cannot cover all the effective information. Relying solely on a single band for the spectral estimation of soil TN in apple orchards is not ideal. However, simple univariate regression models can still be used to evaluate the significance and value of various transformations in extracting effective spectral information.

The use of MLR for the quantitative estimation of soil TN content is simple. Compared to univariate regression models, the MLR models have higher modeling accuracy [50]. When there is a strong linear relationship between soil TN and characteristic bands, the modeling effect is better, but the selection of characteristic bands requires higher requirements. There should be no strong collinearity between bands, and the number of bands should not exceed the number of modeling samples. This can be solved by the SMLR filtering of characteristic bands [33]. Overall, the linear regression model is relatively simple, computationally small, and has strong operability. The predictive effect of the model is acceptable, and it can achieve a rough prediction of soil TN content in the orchard.

The BP neural network can handle nonlinear relationships between variables well and has great advantages in the spectral prediction of soil physical and chemical properties [54–56]. The MEA overcomes the shortcomings of EC such as premature convergence and slow convergence speed [42], solves complex combinatorial optimization problems, and facilitates the search for global optimal solutions, ultimately minimizing the error between the predicted and expected values. The modeling results of this article indicate that the model based on MEA-BPNN has higher accuracy and stability and can more accurately estimate soil TN content. The disadvantage is that the model is relatively complex and computationally intensive. This can utilize computers to achieve automated calculations and improve the practicality of the model. Among the estimation models for each fertilization period and the entire fertilization period, the model based on CR-SCI-MEA-BPNN has the best prediction performance. It should be pointed out that there are currently not many studies on the hyperspectral inversion of total nitrogen in apple orchard soil, especially for different fertilization periods, which have not been reported yet [25,26,30]. Our research was conducted in a relatively small apple orchard in Shandong Province; due to differences in natural conditions and management practices among orchards in different regions, this may limit the generalizability of research results to other regions or orchards with different soil and climate conditions. Therefore, in the future, research on the application of orchards in different regions, climates, and soil conditions should be carried out.

5. Conclusions

This article collected the soil spectral reflectance of apple orchards during different fertilization periods. The original spectral reflectance (R) and its eight transformations, including $(1/R)$, $(\log R)$, (\sqrt{R}) and their first derivative transformation, as well as CR, were used to extract the characteristic bands of soil TN using the correlation coefficient method and SMLR. The best characteristic band extraction method was selected by comparing and analyzing the modeling effects of linear regression and the MEA-BPNN method. Band by band analysis techniques and correlation analysis methods were used to select the SCI suitable for soil TN. Based on the selected SCI and characteristic band combinations, MLR and MEA-BPNN were used to construct and optimize independent and comprehensive soil TN content estimation models for each fertilization period and the entire fertilization period, respectively.

The results indicate that SMLR is effective and has significant advantages in extracting hyperspectral characteristic bands of soil TN content. Compared with the correlation coefficient method, under the same spectral transformation, the model based on SMLR filtering for characteristic bands can achieve a higher modeling accuracy and smaller RMSE. Moreover, it is easier to select models with a maximum VIF of less than 10 to avoid the influence of multicollinearity of model variables.

The sensitive areas of soil TN content in the SCI do not undergo significant changes due to different fertilization periods. Among them, the RSIs are in the range of 800–900 nm, 1900–1950 nm, and 2200–2300 nm, while the sensitive areas of the DI and NDSI are in the range of 1900–1950 nm and 2200–2300 nm.

The addition of SCI and characteristic band combination significantly improves the prediction accuracy of soil TN model. The independent and comprehensive soil TN content estimation models for each fertilization period based on MEA-BPNN can achieve stable and accurate soil TN estimation. Among the estimation models for each fertilization period and the entire fertilization period, the model based on CR-SCI-MEA-BPNN has the best prediction performance.

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

Funding: This research was funded by the Key R&D Program Project in Shandong Province (2018CXGC0209), and Natural Science Foundation of Shandong Province (ZR2021MD018).

Data Availability Statement: Data are contained within the article.

Acknowledgments: We thank the anonymous reviewers for their valuable comments and suggestions.

Conflicts of Interest: Wenqian Wang was employed by the company Shandong Yuanhong Survey Planning and Design Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Wang, Y. China's apple production steadily ranks first in the world. *Farmers Daily*, 18 November 2023.
2. Peng, F.T.; Jiang, Y.M.; Gu, M.R.; Shu, H.R. Advances in research on nitrogen nutrition of deciduous fruit crops. *J. Fruit Sci.* **2003**, *20*, 54–58.
3. Gao, Y.M.; Tong, Y.A.; Lu, Y.L.; Wang, X.Y. Effects of soil available nutrients and long-term fertilization on yield of Fuji apple orchard of Weibei area in Shaanxi, China. *Acta Hort. Sin.* **2013**, *40*, 613–622. [[CrossRef](#)]
4. Zhang, Y.Z.; Cheng, C.G.; Zhao, D.Y.; Zhou, J.T.; Chen, Y.H.; Zhang, H.T.; Xie, B. Effects of nitrogen application levels on fruit Ca forms and quality of 'Fuji' apples. *J. Plant Nutr. Fert.* **2021**, *27*, 87–96.

5. Li, Y.; Xiao, J.; Yan, Y.; Liu, W.; Cui, P.; Xu, C.; Nan, L.; Liu, X. Multivariate analysis and optimization of the relationship between soil nutrients and berry quality of vitis vinifera cv. cabernet franc vineyards in the eastern foothills of the Helan mountains, China. *Horticulturae* **2024**, *10*, 61. [[CrossRef](#)]
6. Raese, J.T.; Drake, S.R.; Curry, E.A. Nitrogen fertilizer influences fruit quality, soil nutrients and cover crops, leaf color and nitrogen content, biennial bearing and cold hardiness of ‘Golden Delicious’. *J. Plant Nutr.* **2007**, *30*, 1585–1604. [[CrossRef](#)]
7. Deng, J.; Feng, X.Q.; Wang, D.Y.; Lu, J.; Chong, H.T.; Shang, C.; Liu, K.; Huang, L.Y.; Tian, X.H.; Zhang, Y.B. Root morphological traits and distribution in direct-seeded rice under dense planting with reduced nitrogen. *PLoS ONE* **2020**, *15*, e0238362. [[CrossRef](#)]
8. Zhang, J.H. Nitrogen imbalance in apple trees and its application in science and technology. *Hortic. Seed.* **2016**, *35*, 56–59. [[CrossRef](#)]
9. Bai, Z.J.; Peng, J.; Luo, D.F.; Cai, H.H.; Ji, W.J.; Shi, Z.; Liu, W.Y.; Yin, C.Y. A mid-infrared spectral inversion model for total nitrogen content of farmland soil in southern Xinjiang. *Spectrosc. Spect. Anal.* **2022**, *42*, 2768–2773.
10. Zhang, G.B.; Li, X.C.; Qi, F.Y.; Wu, B.; Cheng, S.H. Grey cluster estimating model of soil organic matter content based on hyperspectral data. *J. Grey Syst. UK* **2014**, *26*, 28–37.
11. Peng, J.; Zhang, Y.Z.; Zhou, Q.; Pang, X.A.; Wu, W.M. The progress on the relationship physics-chemistry properties with spectrum characteristic of the soil. *Chin. J. Soil. Sci.* **2009**, *40*, 1204–1208. [[CrossRef](#)]
12. Guo, Y.; Guo, Z.X.; Liu, J.; Yuan, Y.Z.; Sun, H.; Chai, M.; Bi, R.T. Hyperspectral inversion of paddy soil iron oxide in typical subtropical area with Pearl River Delta, China as illustration. *J. Appl. Ecol.* **2017**, *28*, 3675–3683.
13. Vasques, G.M.; Grunwald, S.; Harris, W.G. Spectroscopic models of soil organic carbon in Florida. USA. *J. Environ. Qual.* **2010**, *9*, 23–934. [[CrossRef](#)]
14. Liu, J.B.; Han, J.C.; Zhang, Y.; Wang, H.Y.; Kong, H.; Shi, L. Prediction of soil organic carbon with different parent materials development using visible-near infrared spectroscopy. *Spectrochim. Acta A* **2018**, *204*, 33–39. [[CrossRef](#)]
15. Liao, Q.H.; Gu, X.H.; Li, C.J.; Chen, L.P.; Huang, W.J.; Du, S.Z.; Fu, Y.Y.; Wang, J.H. Estimation of fluvo-aquic soil organic matter content from hyperspectral reflectance based on continuous wavelet transformation. *Trans. CSAE* **2012**, *28*, 132–139.
16. Lin, L.X.; Wang, Y.J.; Teng, J.Y.; Wang, X.C. Hyperspectral analysis of soil organic matter in coal mining regions using wavelets, correlations, and partial least squares regression. *Environ. Monit. Assess* **2016**, *188*, 1–11. [[CrossRef](#)]
17. Divya, Y.; Gopinathan, P. Soil water content measurement using hyper-spectral remote sensing techniques—A case study from north-western part of Tamil Nadu, India. *Remote Sens. Appl. Soc. Environ.* **2019**, *14*, 1–7. [[CrossRef](#)]
18. Yang, H.; Kuang, B.; Mouazen, A.M. Quantitative analysis of soil nitrogen and carbon at a farm scale using visible and Near Infrared spectroscopy coupled with wavelength reduction. *Eur. J. Soil. Sci.* **2012**, *63*, 410–420. [[CrossRef](#)]
19. Xie, H.T.; Yang, X.M.; Drury, C.F.; Yang, J.Y.; Zhang, X.D. Predicting soil organic carbon and total nitrogen using Mid-and Near-infrared spectra for brookston clay loam soil in south-western Ontario, Canada. *Can. J. Soil. Sci.* **2011**, *91*, 53–63. [[CrossRef](#)]
20. Zhang, J.J.; Tian, Y.C.; Yao, X.; Cao, W.X.; Ma, X.M.; Zhu, Y. Estimating model of soil total nitrogen content based on near-infrared spectroscopy analysis. *Trans. CSAE* **2012**, *28*, 183–188.
21. Liu, X.Y.; Wang, L.; Chang, Q.R.; Wang, X.X.; Shang, Y. Prediction of total nitrogen and alkali hydrolysable nitrogen content in loess using hyperspectral data based on correlation analysis and partial least squares regression. *Chin. J. Appl. Ecol.* **2015**, *26*, 2107–2114.
22. Li, Y.; Wang, R.H.; Guan, Y.L.; Jiang, H.L.; Wu, X.Q.; Peng, Q. Prediction analysis of soil total nitrogen content based on hyperspectral. *Remote Sens. Technol. Appl.* **2017**, *32*, 173–179.
23. Jia, S.Y.; Li, H.Y.; Wang, Y.J.; Tong, R.Y.; Li, Q. Hyperspectral Imaging Analysis for the classification of soil types and the determination of soil total nitrogen. *Sensors* **2017**, *17*, 2252. [[CrossRef](#)]
24. Niu, Z.; Shi, L.; Qiao, H.; Xu, X.; Wang, W.; Ma, X.; Zhang, J. Construction of a hyperspectral estimation model for total nitrogen content in Shajiang black soil. *J. Plant Nutr. Soil. Sci.* **2023**, *186*, 196–208. [[CrossRef](#)]
25. Li, Z.; Zhang, G.C.; An, X.H.; Li, M.; Li, E.M.; Cheng, G.G. Estimation for available nitrogen content of apple orchard soil by near infrared spectroscopy and partial least squares regression. *North. Hortic.* **2018**, *3*, 113–119.
26. Yu, X.; Liu, Q.; Wang, Y.B.; Liu, X.Y.; Liu, X. Evaluation of MLR 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](#)]
27. Liu, Y.D.; Xiong, S.S.; Chen, D.B. Estimation of the TN and SOM contents in soil from GAN NAN navel orange plant area by NIR diffuse spectroscopy. *Spectrosc. Spect. Anal.* **2013**, *33*, 2679–2682.
28. Tang, R.N.; Li, X.W.; Li, C.; Jiang, K.X.; Hu, W.F.; Wu, J.J. Estimation of total nitrogen content in rubber plantation soil based on hyperspectral and fractional order derivative. *Electronics* **2022**, *11*, 1956. [[CrossRef](#)]
29. Liu, T.L.; Zhu, X.C.; Bai, X.Y.; Peng, Y.F.; Li, M.X.; Tian, Z.Y.; Jiang, Y.M.; Yang, G.J. Hyperspectral estimation model construction and accuracy comparison of soil organic matter content. *Smart Agric.* **2020**, *2*, 129–138.
30. Liu, Z.M.; Li, X.Q.; Chen, G.L.; Ding, H.P.; Xu, M.G.; Yang, C.X. Inversion of soil total nitrogen content in Yunnan Rubber plantation based on hyperspectral. *Trop. Agric. Sci. Technol.* **2023**, *46*, 35–41.
31. Wang, W.Q.; Gao, M.X.; Lang, K.; Wang, J.F. Hyperspectral estimation of total nitrogen in apple orchard soil based on wavelet transform. *J. Shandong Agric. Univ.* **2021**, *52*, 845–852. [[CrossRef](#)]
32. Chen, H.Y.; Zhao, G.X.; Li, X.C.; Zhu, X.C.; Sui, L.; Wang, Y.J. Hyper-spectral estimation of soil organic matter content based on wavelet transformation. *Chin. J. Appl. Ecol.* **2011**, *22*, 2935–2942.
33. Wang, L.W.; Wei, Y.X. Estimating the total nitrogen and total phosphorus content of wetland soils using hyperspectral models. *Acta Ecol. Sin.* **2016**, *36*, 5116–5125.

34. Qiu, H.; Chen, H.Y.; Xing, S.H.; Zhang, L.M.; Dong, X.Y. Soil organic matter estimation models based on Hyperion data. *J. Fujian Agric. For. Univ.* **2017**, *46*, 460–467.
35. Li, X.J.; Wang, M.L.; Zhao, M.; Shi, Y.H.; Gao, Q. Vineyard drought monitoring model based on multiple stepwise regression analysis. *Agric. Res. Arid. Areas* **2022**, *40*, 249–254.
36. Nie, X.J.; Zhao, X.H.; Ammara, G.; Hong, W.W.; Zhang, H.B. Inversion of coal-derived carbon content in mine soils based on hyperspectral index and machine learning. *J. Chin. Coal Soc.* **2023**, *48*, 2869–2880.
37. Li, C.; Zhang, G.W.; Zhou, Z.G.; Zhao, W.Q.; Meng, Y.L.; Chen, B.L.; Wang, Y.H. Hyperspectral parameters and prediction model of soil moisture in coastal saline. *Chin. J. Appl. Ecol.* **2016**, *27*, 525–531.
38. Liu, J.F.; Zhao, Z.Y.; Li, C.Y.; Gao, Y.; Zhao, X.; Jiang, W.G.; Gong, Z. Estimation of apple tree canopy SPAD based on UAV multispectral remote sensing. *J. Drain. Irrig. Mach. Eng.* 2024. Available online: <https://link.cnki.net/urlid/32.1814.TH.20240115.1641.004> (accessed on 16 February 2024).
39. Leng, J.F.; Gao, X.; Zhu, J.P. The application of multiple linear regression statistical prediction models Statistics and decision-making. *Stat. Decis.* **2016**, *7*, 82–85.
40. Wen, X. *Applying MATLAB to Implement Neural Networks*; National Defense Industry Press: Beijing, China, 2015; p. 95.
41. Liu, B.Y.; Zhang, J.H.; Geng, S.C.; Tao, J.J.; Feng, Y. Estimation of N₂O emission flux from heterotrophic nitrification process in global forest soils based on BP neural network model. *Chin. J. Ecol.* **2022**, *41*, 209–217. [[CrossRef](#)]
42. Sun, C.Y.; Xie, K.M.; Cheng, M.Q. Mind-evolution-based machine learning framework and new development. *J. Taiyuan Univ. Technol.* **1999**, *43*, 3–7.
43. Liu, H.J.; Zhang, X.L.; Zheng, S.F.; Tang, N.; Hu, Y.L. Black soil organic matter prediction model based on field hyperspectral reflectance. *Spectrosc. Spect. Anal.* **2010**, *30*, 3355–3358.
44. Bai, G.S.; Zou, C.Y.; Du, S.N. Effects of self-sown grass on soil moisture and tree growth in apple orchard on Weibei dry plateau. *Trans. CSAE* **2018**, *34*, 151–158. [[CrossRef](#)]
45. Shen, Y.; Zhang, X.P.; Yang, X.M.; Liang, A.Z.; Shi, X.H.; Fan, R.Q. Effects of data pretreatment and spectrum bands on near infrared spectroscopy model of soil organic carbon in black soil. *ACTA Pedol. Sin.* **2010**, *47*, 1005–1012.
46. Philippe, L.; Baret, F.; Feret, J.B.; José, M.N.; Jean, M. R-M. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sens. Environ.* **2007**, *112*, 825–835.
47. Peng, J.; Chi, C.M.; Xiang, H.Y.; Teng, H.F.; Shi, Z. Inversion of soil salt content based on continuum-removal method. *ACTA Pedol. Sin.* **2014**, *51*, 459–469.
48. Xiang, H.Y.; Liu, W.Y.; Peng, J.; Wang, J.Q.; Chi, C.M.; Niu, J.L. Predicting organic matter content in paddy soil using method of continuum removal in southern Xinjiang, China. *Soils* **2016**, *48*, 389–394.
49. Yang, D.H.; Li, Y.Q.; Dao, J.; Wang, J.X. Estimation of soil main nutrient content based on unmanned aerial vehicle multispectral and ground hyperspectral remote sensing. *Jiangsu Agri. Sci.* **2022**, *50*, 178–186. [[CrossRef](#)]
50. Wei, L.S. *Mathematical Statistics*; Science Press: Beijing, China, 2024; p. 1.
51. Xu, Y.; Shao, G.C.; Ding, M.M.; Zhang, E.Z.; He, J. Inversion model of soil total nitrogen content based on ridge regression. *J. Drain. Irrig. Mach. Eng.* **2022**, *40*, 1159–1166.
52. Ren, H.; Guo, C.L.; Zhao, Z.P.; He, Q.M.; Chen, D.M. Current status of vision and its influencing factors among primary students in Meilan district of Haikou city. *Pract. Prev. Med.* **2020**, *27*, 54–56.
53. Xu, Y.M.; Lin, Q.Z.; Huang, X.H.; Shen, Y.; Wang, L. Experimental study on total nitrogen concentration in soil by VNIR reflectance spectrum. *Geogr. Geo. Inform. Sci.* **2005**, *21*, 19–22.
54. Lu, Z.H.; Liu, X.Y.; Chang, S.J.; Yang, S.L.; Zhao, W.W.; Yang, Y.; Liu, A.J. Hyperspectral inversion of the surface soil N/P ratio in a grassland mining area based on the BP neural network. *Pratac. Sci.* **2018**, *35*, 2127–2136.
55. Zhao, J.H.; Zhang, C.Y.; Min, L.; Li, N.; Wang, Y. Retrieval for soil moisture in farmland using multi-source remote sensing data and feature selection with GA-BP neural network. *Trans. CSAE* **2021**, *37*, 112–120.
56. Wang, S.D.; Shi, P.J.; Zhang, H.B.; Wang, X.C. Retrieval of soil total nitrogen content in reclaimed farmland of mining area based on hyperspectral imaging. *Chin. J. Ecol.* **2019**, *38*, 294–301.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.