

Article

Estimation of Surface Soil Nutrient Content in Mountainous Citrus Orchards Based on Hyperspectral Data

Xuchao Jiao, Hui Liu , Weimu Wang, Jiaojiao Zhu and Hao Wang

College of Agricultural Science and Engineering, Hohai University, Nanjing 210098, China; jiaoxuchao17@hhu.edu.cn (X.J.); wangwm@hhu.edu.cn (W.W.); 211310040006@hhu.edu.cn (J.Z.); 211310010014@hhu.edu.cn (H.W.)

* Correspondence: liuhui@hhu.edu.cn

Abstract: Monitoring soil conditions is of great significance for guiding fruit tree production and increasing yields. Achieving a rapid determination of soil physicochemical properties can more efficiently monitor soil conditions. Traditional sampling and survey methods suffer from slow detection speeds, low accuracy, limited coverage, and require a large amount of manpower and resources. In contrast, the use of hyperspectral technology enables the precise and rapid monitoring of soil physicochemical properties, playing an important role in advancing precision agriculture. Yuxi City, Yunnan Province, was selected as the study area; soil samples were collected and analyzed for soil organic matter (SOM), total nitrogen (TN), total phosphorus (TP), and available nitrogen (AN) contents. Additionally, soil spectral reflectance was obtained using a portable spectroradiometer. Hyperspectral characteristic bands for soil nutrients were selected from different spectral preprocessing methods, and different models were used to predict soil nutrient content, identifying the optimal modeling approach. For SOM prediction, the second-order differentiation-multiple stepwise regression (SD-MLSR) model performed exceptionally well, with an R^2 value of 0.87 and RMSE of $6.61 \text{ g} \cdot \text{kg}^{-1}$. For TN prediction, the logarithm of the reciprocal first derivative-partial least squares regression (LRD-PLSR) model had an R^2 of 0.77 and RMSE of $0.37 \text{ g} \cdot \text{kg}^{-1}$. For TP prediction, the logarithmic second-order differentiation-multiple stepwise regression (LTSD-MLSR) model had an R^2 of 0.69 and RMSE of $0.04 \text{ g} \cdot \text{kg}^{-1}$. For AN prediction, the logarithm of the reciprocal second derivative-partial least squares regression (LRSD-PLSR) model had an R^2 of 0.83 and RMSE of $24.12 \text{ mg} \cdot \text{kg}^{-1}$. The results demonstrate the high accuracy of these models in predicting soil nutrient content.

Keywords: hyperspectral; soil nutrient; partial least squares regression; multiple stepwise regression



Citation: Jiao, X.; Liu, H.; Wang, W.; Zhu, J.; Wang, H. Estimation of Surface Soil Nutrient Content in Mountainous Citrus Orchards Based on Hyperspectral Data. *Agriculture* **2024**, *14*, 873. <https://doi.org/10.3390/agriculture14060873>

Academic Editor: Amin Beiranvand Pour

Received: 13 April 2024

Revised: 19 May 2024

Accepted: 23 May 2024

Published: 30 May 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

Soil nutrient management is crucial for the development of green agriculture because soil nutrients directly impact crop growth and the sustainability of agriculture [1]. In recent years, China has undertaken some work in soil nutrient management and monitoring, including increasing the demand for soil monitoring instruments, enhancing monitoring techniques, and to some extent, emphasizing the importance of soil quality [2]. With the advancement of hyperspectral technology, the related applications for determining soil nutrients are continuously evolving, demonstrating significant potential and enhancing the efficiency of soil nutrient determination. Hyperspectral remote sensing technology represents a significant advancement following traditional remote sensing techniques. Compared to traditional remote sensing, hyperspectral remote sensing can capture more detailed spectral information [3]. This technique utilizes a large number of spectral bands to capture reflectance data ranging from visible to near-infrared and even mid-infrared ranges. Soils possess complex physical and chemical properties, with different soil components such as organic matter, moisture, various nutrients, etc., exhibiting specific reflectance and absorption characteristics across different spectral bands [4]. Hyperspectral remote

sensing technology can identify these features and be used to analyze soil types, structures, nutrient statuses, etc. Hyperspectral techniques acquire spectral information at different wavelengths for analysis, providing unique advantages in soil property analysis [5]. In agriculture and environmental monitoring, hyperspectral remote sensing technology plays a crucial role. It is utilized not only for assessing soil fertility and optimizing fertilization plans but also for effectively monitoring soil degradation and pollution situations [6,7]. Through a portable spectroradiometer, narrowband spectral data can be continuously acquired in the range of 350–2500 nm [8]. The portable spectroradiometer, due to its small size, is convenient for fieldwork. Not only is it suitable for outdoor operations, it is also efficient and has a fast measurement speed coupled with a simple operation, which significantly enhances the efficiency of data acquisition [9]. Currently, there are many applications for field spectrometers, such as identifying contamination points in soil [10] to reduce the levels of heavy metal pollutants such as cadmium and lead in urban areas, for instance [11]. There have been some studies on the determination of soil nutrients using portable field spectroradiometers, such as estimating organic matter in tropical soils in Thailand [12], predicting nitrogen content in soil [13], and determining soil nutrient texture and other elements in the Huanghui River Basin [14], which all achieved good accuracy. However, there are currently research deficiencies in soil hyperspectral remote sensing, such as issues regarding the generality and applicability of the models.

Many research cases focus on modeling the inversion of individual elements in soil, and hyperspectral analysis also imposes limitations on models and algorithms across different regions or soil types, requiring the further consideration of factors such as soil texture, based on the aforementioned potential limitations of portable field spectroradiometers in soil nutrient measurement.

In this study, soil samples are air-dried, ground, and sieved before outdoor spectral collection to avoid the influence of moisture and impurities in the soil. This approach enables the capture of the spectral characteristics of soil in its natural environment. Furthermore, the study models and inverts multiple nutrients in the soil, aiming to explore and establish the generality of hyperspectral inversion models for soil nutrient assessment applicable to the study area. Despite relatively few soil nutrient inversion cases in the mountainous areas of Yunnan, China, the region has a large proportion of red soil and suitable climatic and soil conditions, providing a solid foundation for agricultural development [15]. Strengthening research on soil nutrient inversion is of significant importance for the agricultural development of the mountainous areas in Yunnan. Through hyperspectral remote sensing technology, the rapid and comprehensive monitoring and the evaluation of soil nutrients in the region can be achieved, providing a scientific basis and precise management for agricultural production.

2. Materials and Methods

2.1. Study Area Overview

As shown in Figure 1, the study area is located in Yuxi City, Yunnan Province, situated in the central-southwestern part of Yunnan Province, along the eastern foothills of the Ailao Mountains, and the upper reaches of the Honghe River. The research area is a typical mountainous region. The terrain is primarily characterized by mountains and hills, with altitudes mainly fluctuating between 700 and 900 m, presenting a complex, multi-hill terrain. The soil types are diverse, mainly consisting of red soil and brown soil. Among them, the red soil is a typical subtropical mountain red soil, rich in organic matter and with good permeability, particularly suitable for the growth of citrus. In terms of climate, the region experiences a subtropical monsoon climate with distinct wet and dry seasons. Summers are hot and rainy, while winters are cooler and relatively dry. The average annual temperature remains stable around 20 °C, with an annual accumulated temperature over 7500 °C, and precipitation primarily concentrated in the summer, providing favorable climatic conditions for the growth of citrus. The main agricultural activity in the area is citrus orchard cultivation, with citrus becoming one of the major economic crops of the

region. In addition to citrus orchards, some land is also used for the cultivation of rice, vegetables, and other crops. Moreover, due to the complex terrain, some mountainous areas are used as forest land, and the coverage rate of mountain forests is relatively high.

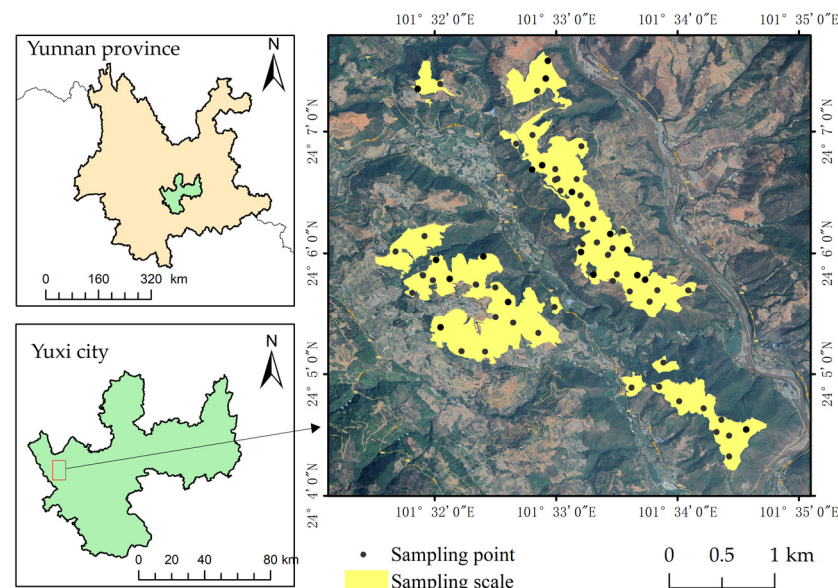


Figure 1. Location of the research area and soil sample collection points.

2.2. Soil Sample Collection

To investigate the application and effectiveness of hyperspectral technology in soil, a comprehensive soil sample collection was conducted in July 2021. A total of 66 soil samples were collected, primarily from areas with red soil types. During the collection process, particular attention was paid to the representativeness of the samples. Sampling points were chosen at the dripline of tree canopies to ensure that the collected soil samples truly reflected the soil characteristics of the area [16]. Additionally, to avoid interference from fertilization, sampling was deliberately avoided in fertilized areas.

Sampling was conducted at a depth of 0–20 cm [17]. The collected soil samples were crushed and mixed, then spread out into a square shape. Next, diagonal lines were drawn to divide it into four equal parts, and the two parts opposite each other were discarded, keeping only the other two parts. The standard quartering method was strictly followed to ensure that each sample adequately represented the soil characteristics of its location [18]. Simultaneously, detailed sampling methods, locations, times, environmental conditions, soil types, land use conditions, etc., were recorded, maintaining the accuracy and completeness of records to support data analysis and the effectiveness of research results. Then, the collected soil samples were placed in clean, sealed plastic bags and carefully labeled for subsequent analysis and processing. After air drying, the soil samples were ground. The processed soil samples were divided into two portions: one for soil analysis and the other sieved through a 2 mm mesh aperture for soil spectral measurements.

2.3. Soil Hyperspectral Data Measurement

When measuring soil spectral reflectance, the portable ASD FieldSpec4 spectrometer from the United States was utilized. This spectrometer covers a wide spectral range from 350 nm to 2500 nm, providing rich spectral information. Regarding the sampling intervals, the spectrometer sampled at intervals of 1.4 nm within the range of 350–1000 nm and at intervals of 2 nm within the range of 1000–2500 nm. To enhance the accuracy and reliability of the data, a resampling interval of 1 nm was used. During the measurement process, the processed soil samples were placed in specific containers, ensuring that the containers were filled with soil samples and the soil surface was leveled to obtain a flat measurement surface. To minimize the impact of environmental factors, spectral measurements are

conducted in a controlled outdoor environment with clear weather. During measurement, the spectrometer's fiber optic probe was positioned vertically to the soil surface at a distance of 10 cm to ensure measurement accuracy.

To ensure the accuracy of the measurement data, a whiteboard with a reflectance of 1 was promptly used for correction during the measurement process. Five spectral curves were measured for each soil sample, and then these five curves were arithmetically averaged to obtain the actual reflectance spectral data of the soil sample. To eliminate random errors caused by various factors during spectral measurement, all spectral curves were averaged using the ViewSpecPro V5.6.8 software (Malvern Panalytical Ltd., Malvern, UK) during the data processing stage. As shown in Figure 2, different colors represent different soil samples, the spectral reflectance is clearly visible. In the data processing stage, significant noise was observed in the spectral curves within the ranges of 350~399 nm, 1350~1405 nm, 1750~1950 nm, and 2320~2500 nm. To obtain more accurate data, these noisy spectral ranges were removed.

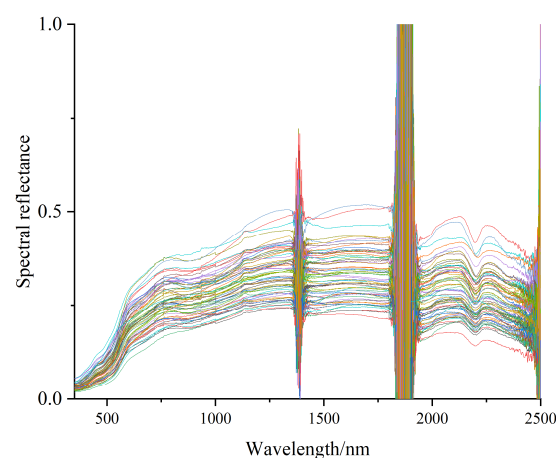


Figure 2. Spectral curves before band removal.

2.4. Soil Nutrient Determination

The determination of soil organic matter (SOM) was conducted using the external heating method [19]. The total nitrogen (TN) content in the soil was determined using the Modified Kjeldahl Method [20]. The determination of total phosphorus (TP) content in the soil was performed using the alkali fusion–Mo–Sb anti-spectrophotometric method [21]. The method used for determining available nitrogen (AN) in the soil was the alkali diffusion method [22].

2.5. Preprocessing Methods for Soil Spectral Data

The Savitzky–Golay filter was employed, which smooths spectral data while preserving important information and improving its signal-to-noise ratio, thus retaining the essential features of the data [23]. In this study, the Savitzky–Golay smoothing algorithm uses a polynomial degree of 2 and a window length of 5. Subsequently, different forms of first-order and second-order differential transformations were applied to the SG-smoothed spectral curves. For instance, the first-order differentiation (FD) preprocessing is a method for feature extraction and enhancement of spectral data. Second-order differentiation (SD) preprocessing involves calculating the second derivative of spectral intensity with respect to wavelength, reflecting changes in the rate of spectral intensity. Logarithmic first-order differentiation (LTFD) preprocessing combines logarithmic transformation with first-order differentiation to improve data resolution in spectral analysis. Logarithmic second-order differentiation (LTSD) preprocessing is an advanced spectral data preprocessing method that combines logarithmic transformation with second-order differentiation to further enhance spectral features and improve data analysis performance. The logarithm of reciprocal first derivative (LRD) preprocessing integrates logarithmic transformation, reciprocal

transformation, and first-order differentiation calculation for spectral analysis. Logarithm of reciprocal second derivative (LRSD) preprocessing combines reciprocal transformation, logarithmic transformation, and second derivative processing to further enhance spectral features, compress dynamic range, and reduce the influence of baseline drift and noise in spectral data analysis.

2.6. Technology Roadmap

In Figure 3 is the technology roadmap after data acquisition; the SG smoothing algorithm is applied followed by preprocessing to select the characteristic bands and establish the model.



Figure 3. Workflow diagram.

3. Methods for Predictive Modeling

3.1. Partial Least Squares Regression

Partial Least Squares Regression (PLSR) is a regression technique used to handle multicollinear data, particularly suitable when the number of predictor variables (independent variables) exceeds the number of observations, or when there is a high correlation among variables. PLSR attempts to find the most important relationship between predic-

tor variables and response variables to achieve effective prediction and dimensionality reduction [24].

$$X = TP^T + E \quad (1)$$

$$Y = UQ^T + F \quad (2)$$

In the equation, T and U are the score matrices extracted from X and Y , respectively. P and Q are the loading matrices, which describe how to reconstruct the score matrices from the original variables. E and F are the residual matrices.

3.2. Multiple Stepwise Regression

The principle of Multiple Stepwise Regression (MLSR) is to determine, through a series of steps, which explanatory variables should be included in the final regression model when predicting one or more response variables. This process aims to find a model that is neither overly simplified nor overly complex, thereby maximizing the model's explanatory power and predictive accuracy while maintaining simplicity [25].

4. Results

4.1. Sample Detected Abnormality

Anomaly detection is helpful in distinguishing natural variations from potential pollution or exceptional conditions. At the same time, the identification and treatment of outliers is an important part of data preprocessing, which is conducive to improving the accuracy of subsequent data analysis and modeling [26]. As shown in Figure 4, it can be observed that there are no outliers present in the box plot of the data from this study. This indicates the high reliability and stability of the data, as there are no unusual or outlier observations that could disrupt the analysis and interpretation of the overall dataset.

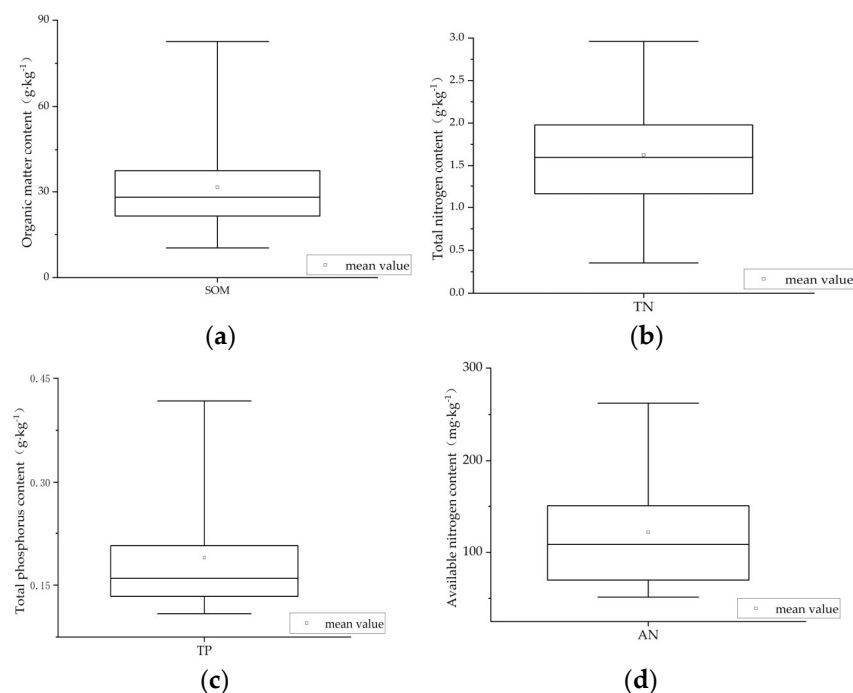


Figure 4. Anomaly testing for four types of soil nutrients. (a) SOM anomaly inspection; (b) TN anomaly inspection; (c) TP anomaly inspection; (d) AN anomaly inspection.

4.2. Descriptive Statistics of Soil Nutrients

As shown in Table 1, the heterogeneity and complexity of soil nutrient content are revealed. The data indicate that the variability of various indicators in the samples differs, with a relatively high coefficient of variation for soil nutrients.

Table 1. Descriptive statistics of soil nutrients.

Soil Nutrient	Number	Minimum Value	Maximum Value	Median	Average Value	Standard Deviation	Coefficient of Variation%
SOM	66	10.39	82.62	28.11	31.60	15.22	48.15
TN	66	0.35	3.00	1.60	1.64	0.60	36.78
TP	66	0.11	0.42	0.16	0.19	0.08	41.92
AN	66	51.45	262.15	108.50	121.59	57.41	47.22

4.3. Spectral Preprocessing

The application of SG smoothing helps to reduce noise and enhance the model's ability to identify and predict soil characteristics, while preserving important spectral features unchanged. As shown in Figure 5, different colors represent different soil samples. The spectral curve has been smoothed using a Savitzky–Golay filter. After SG smoothing, due to the multitude of hyperspectral bands and considerable data redundancy, efforts were made to identify sensitive bands, reducing noise and abrupt changes in hyperspectral data through smoothing with techniques like Savitzky–Golay (SG) filtering which makes the spectral curves smoother, aiding in a clearer identification of features and patterns within the data. This was achieved through analyzing empirical hyperspectral data and reviewing previous studies related to soil hyperspectral prediction [3,27]. So, preprocessing was conducted using transformations such as FD, SD, LTFD, LTSD, LRD, and LRSD.

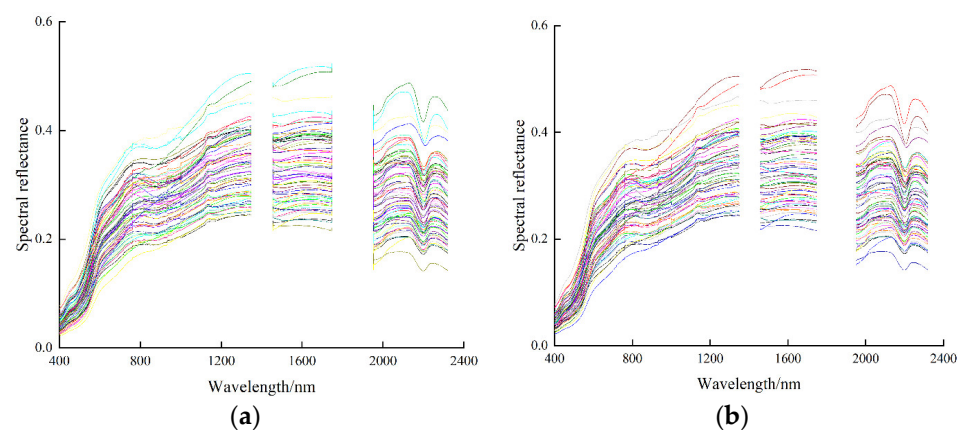


Figure 5. Comparison plot of spectral curves smoothed with Savitzky–Golay filter. (a) Spectral curves after band removal; (b) Spectral curves after SG smoothing.

As shown in Figure 6, different preprocessing methods had varying effects on the data. By comprehensively comparing the transformation results of different preprocessing methods, the most suitable preprocessing method for the current dataset and modeling task can be selected to improve the model's fitting effect and prediction accuracy. The representation of mean and standard deviation indicates the central tendency and range of variation, allowing for a clearer understanding of how different preprocessing methods affect soil hyperspectral data.

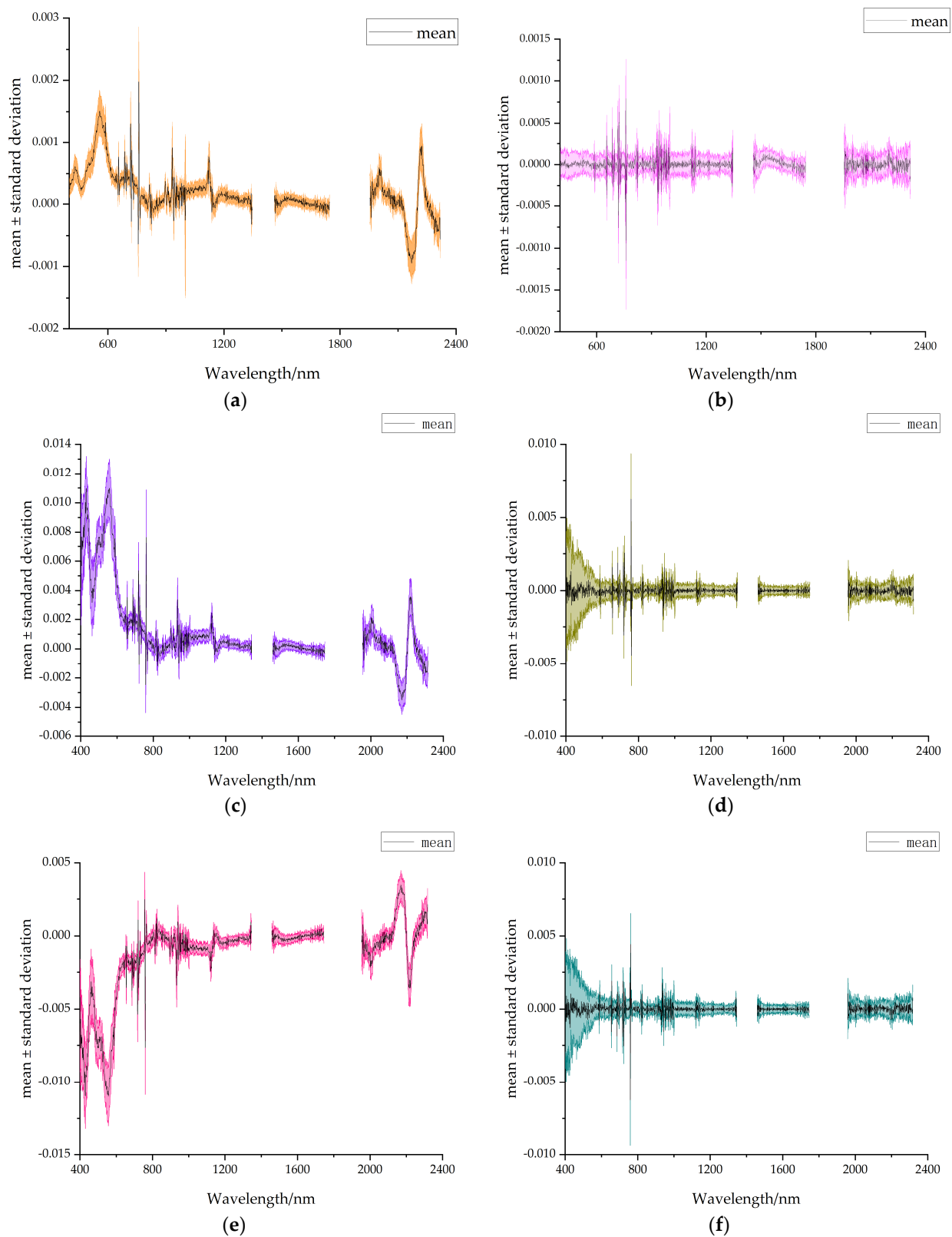


Figure 6. Mean $\pm$ standard deviation under different preprocessing conditions. (a) Mean $\pm$ standard deviation after FD preprocessing; (b) mean $\pm$ standard deviation after SD preprocessing; (c) mean $\pm$ standard deviation after LTFD preprocessing; (d) mean $\pm$ standard deviation after LTSD preprocessing; (e) mean $\pm$ standard deviation after LRD preprocessing; (f) mean $\pm$ standard deviation after LRSD preprocessing.

4.4. Spectral Transformation and Correlation Analysis

Figure 7 shows the correlation coefficients between soil nutrient content and spectral reflectance, as well as their transformations. It presents the Pearson correlation analysis between soil nutrient content and spectral reflectance after six different preprocessing methods: FD, SD, LTFD, LTSD, LRD, and LRSD. The correlation coefficients of soil nutrient content range from -0.6 to $+0.6$.

As Figure 8 shows the absolute values of Pearson correlation coefficients, and spectral bands are selected based on these values. It can be seen that the preprocessing method FD shows a higher correlation coefficient with soil nutrients, while SD and LRSD show lower correlation coefficients with soil nutrients. Additionally, to determine the optimal prediction pathway, it is necessary to establish a model for prediction.

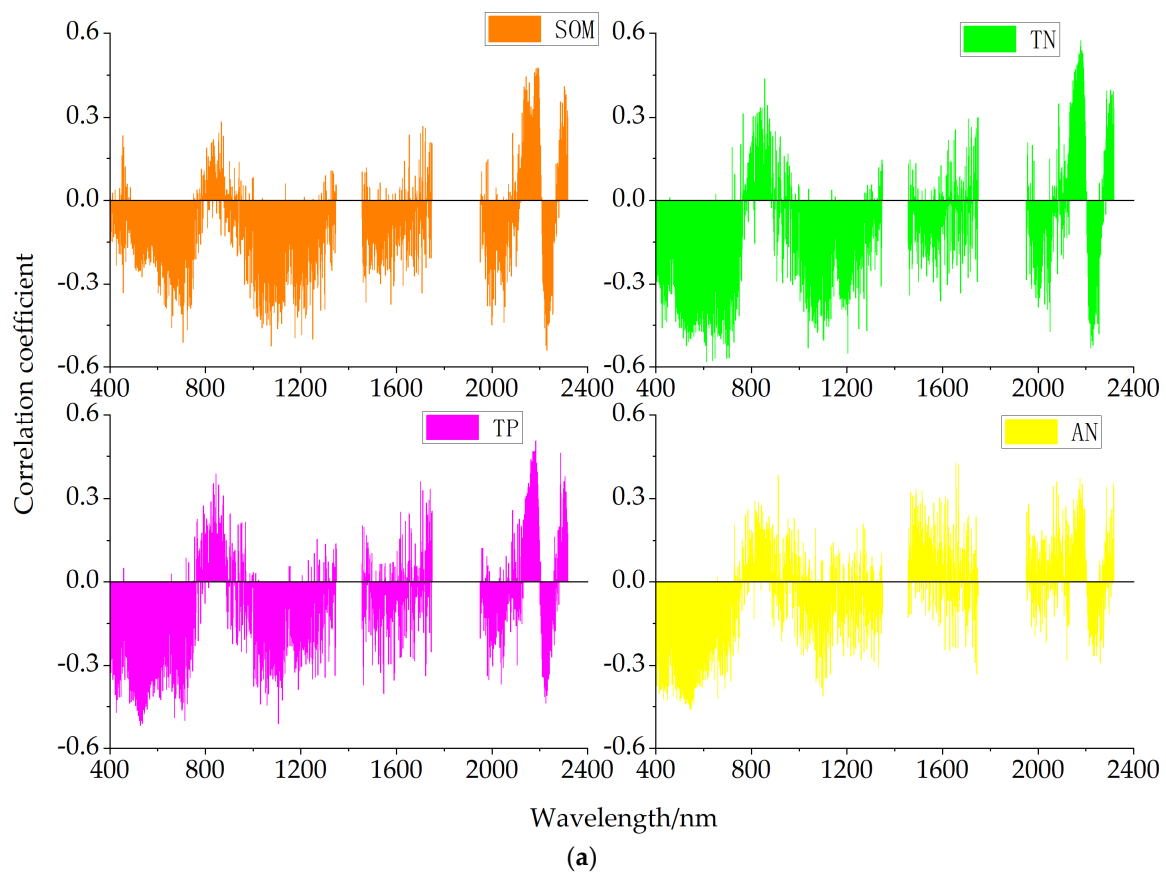
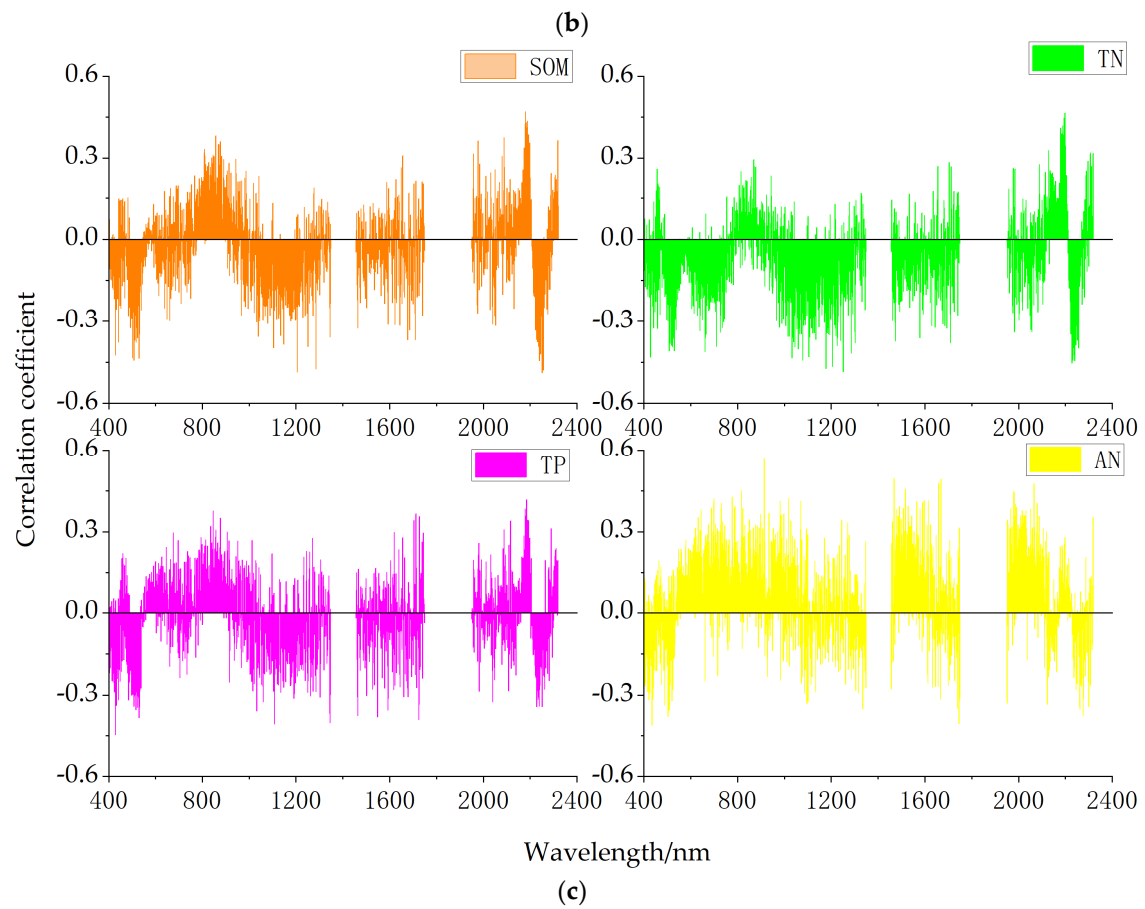
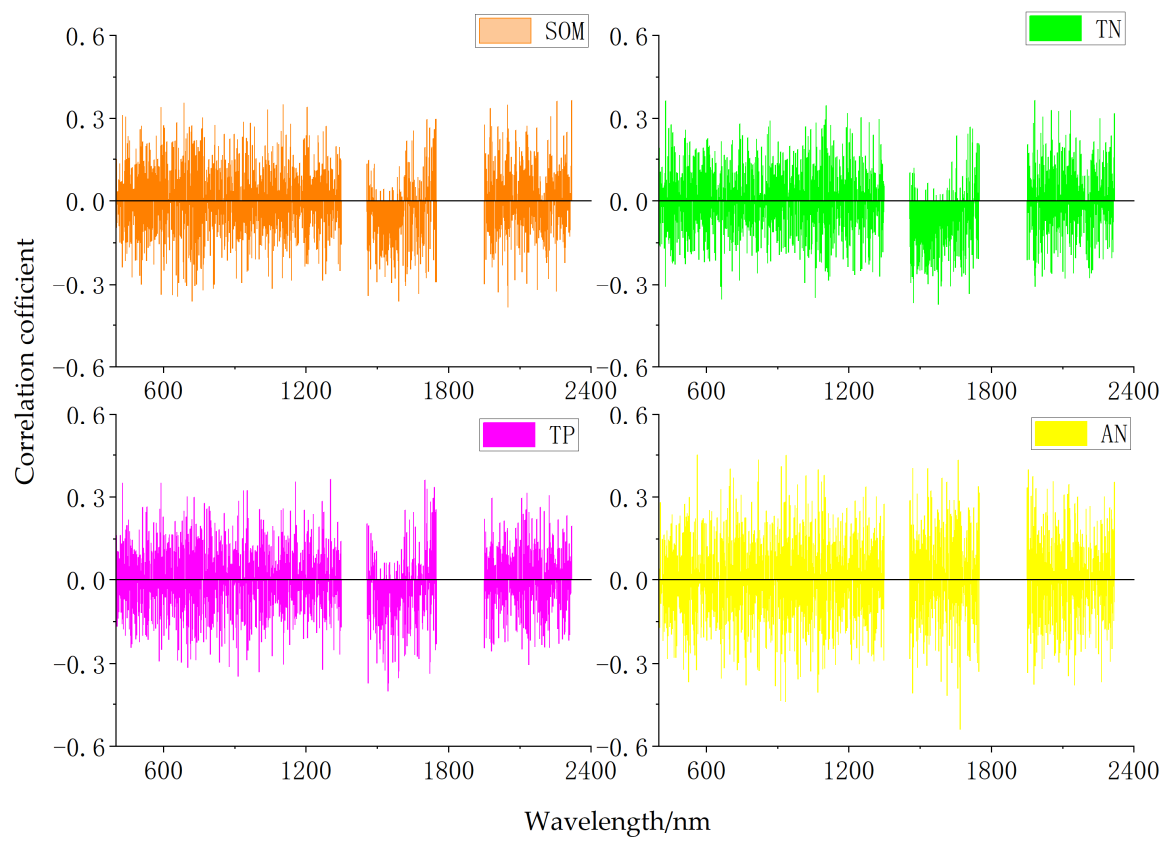


Figure 7. Cont.

**Figure 7.** *Cont.*

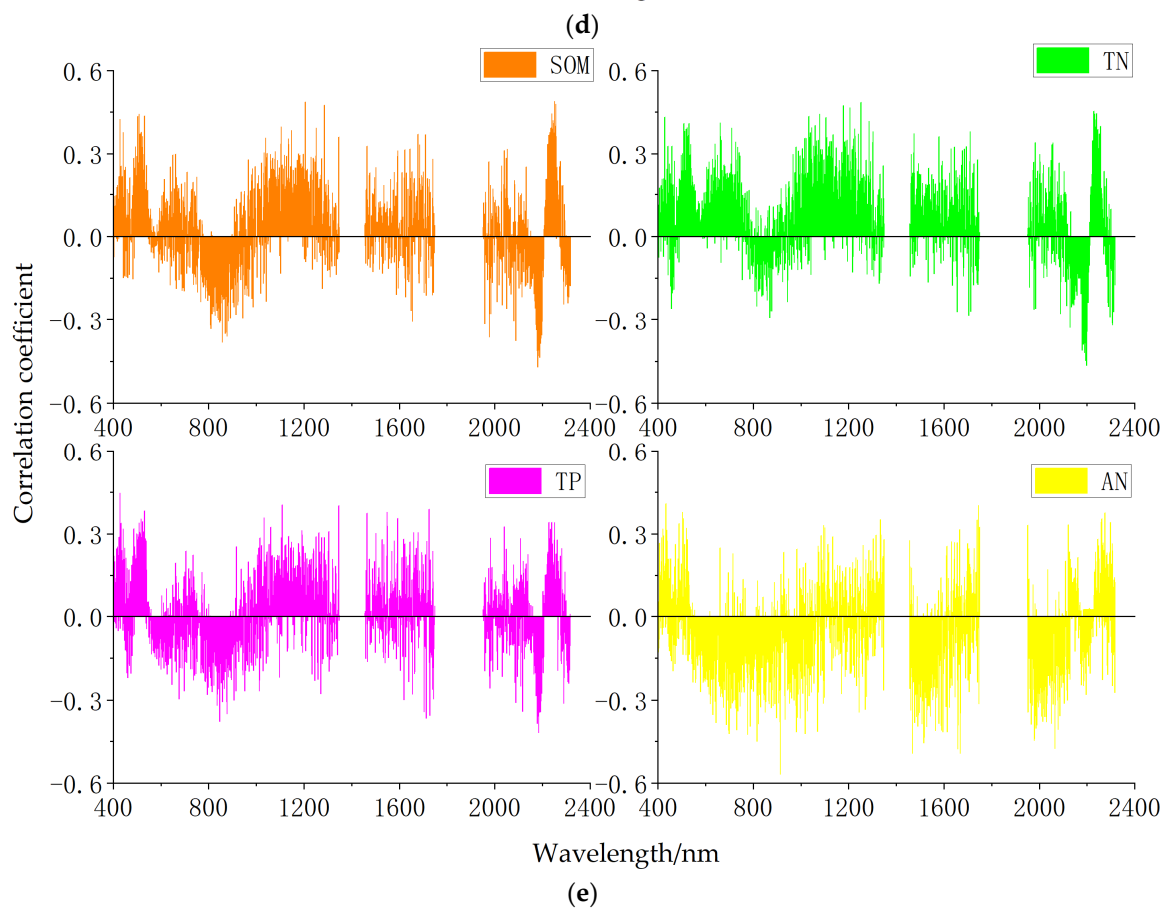
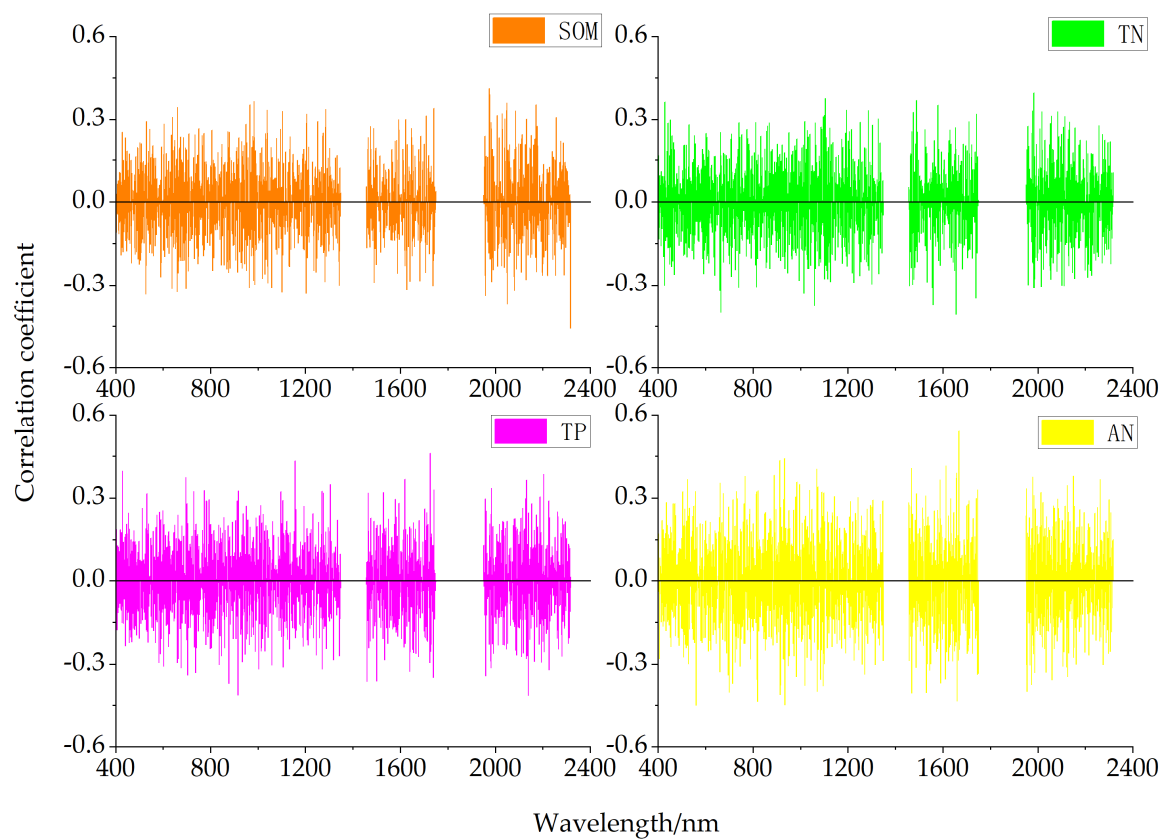


Figure 7. Cont.

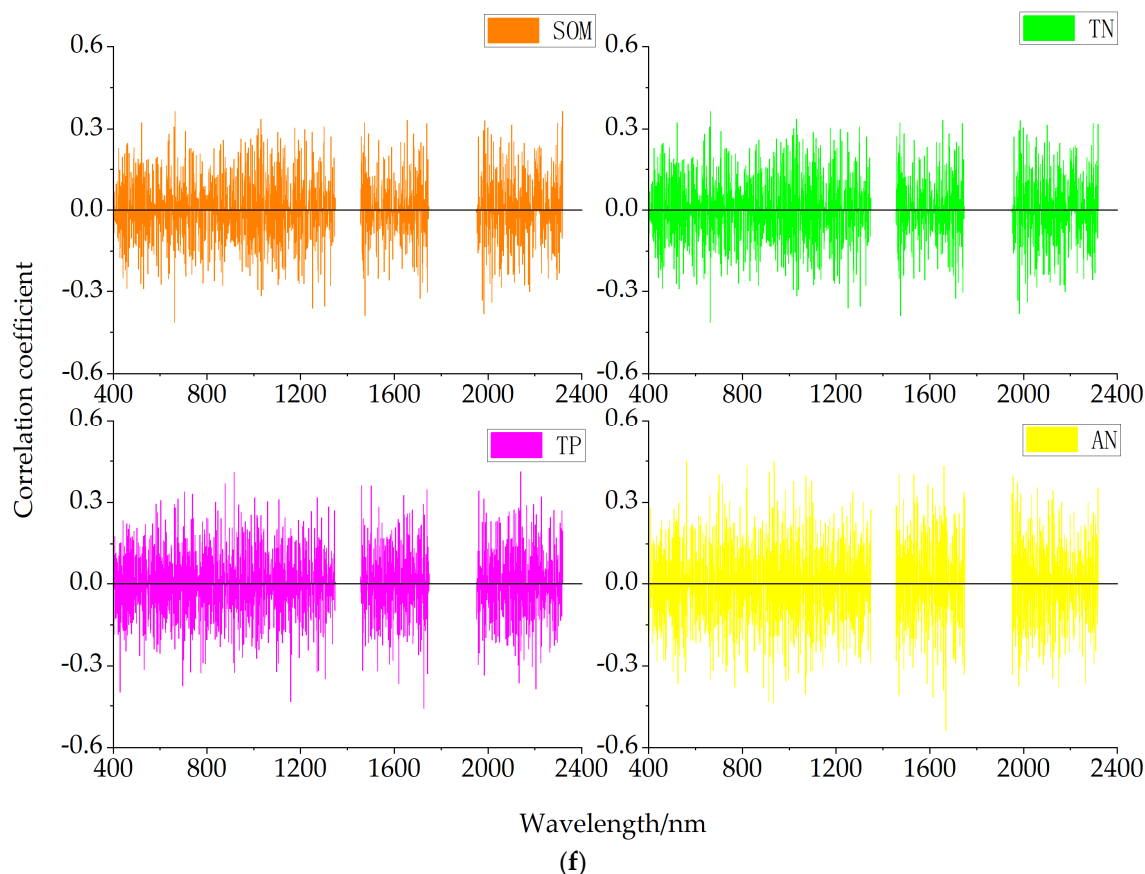


Figure 7. The correlation coefficient between soil nutrient content and spectral reflectance and its transformation. (a) Nutrient correlation coefficients under FD processing; (b) nutrient correlation coefficients under SD processing; (c) nutrient correlation coefficients under LTFD processing; (d) nutrient correlation coefficients under LTSD processing; (e) nutrient correlation coefficients under LRD processing; (f) nutrient correlation coefficients under LRSD processing.

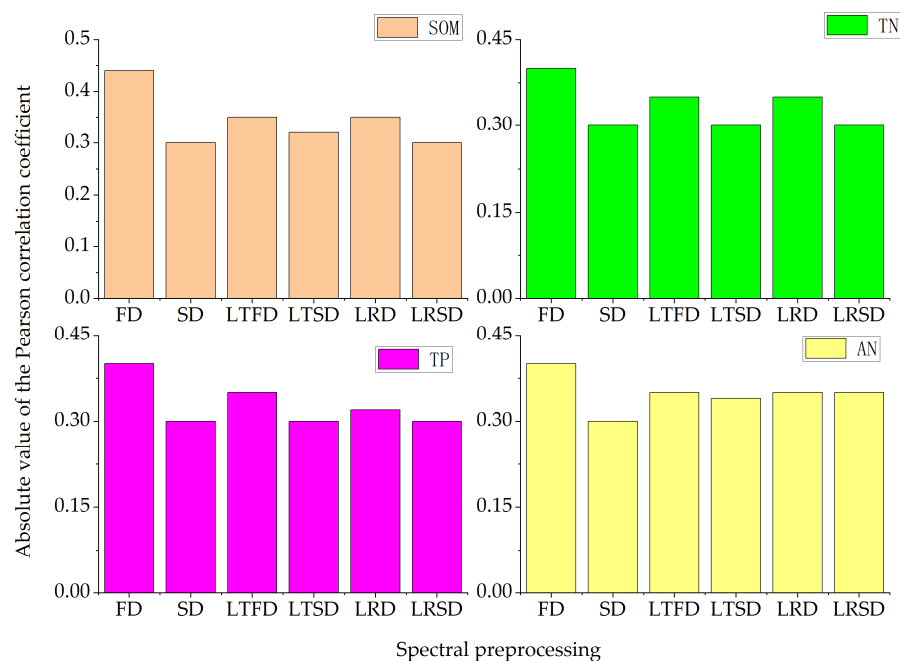


Figure 8. Absolute value of the Pearson correlation coefficient.

4.5. Feature Band Selection

In Figure 9, FD, SD, LTSD, LTSD, LRD, and LRSD are based on SG smoothing and further involve first-order differentiation, second-order differentiation, logarithmic first-order differentiation, logarithmic second-order differentiation, square root first-order differentiation, square root second-order differentiation, the logarithm of reciprocal first derivative, and the logarithm of reciprocal second derivative. Different spectral transformation methods (FD, SD, LTSD, LTSD, LRD, LRSD) have varying impacts on the selection of bands related to soil organic matter (SOM), total nitrogen (TN), total phosphorus (TP), and available nitrogen (AN). For SOM, the selected bands are mainly concentrated in the visible light region and specific infrared bands. The selected bands for TN cover multiple regions from visible light to infrared. TP has a wider range of selected bands, but specific regions such as 400–950 nm and 2131–2224 nm are prominent. The selected bands for AN are primarily concentrated in the visible light to near-infrared region.

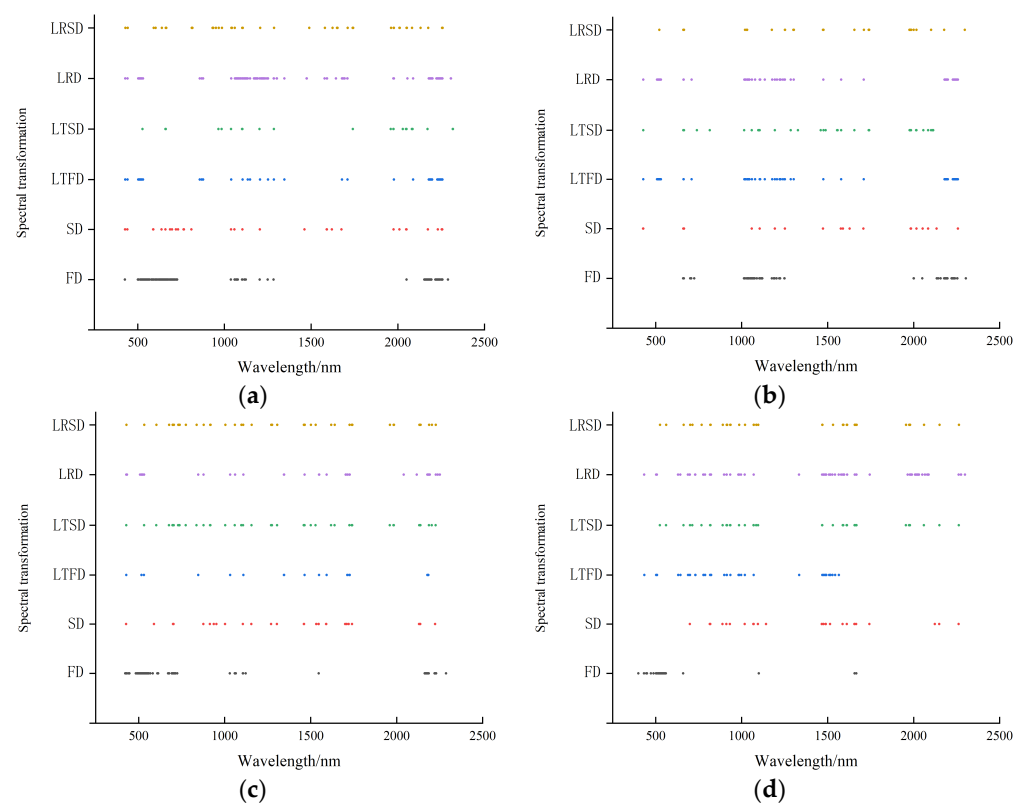


Figure 9. Spectral feature plot of selected soil nutrients. (a) Characteristic bands selected for SOM; (b) characteristic bands selected for TN; (c) characteristic bands selected for TP; (d) characteristic bands selected for AN.

In the spectral reflectance screening of various soil nutrients, the visible light region from 400 to 700 nm has demonstrated a high correlation. This region may contain spectral information directly related to soil organic matter and other nutrients. Besides the visible light region, the near-infrared region from 1000 to 1500 nm and the short-wave infrared region from 2000 to 2300 nm also exhibited high correlation in the screening of multiple nutrients. Although different spectral transformation methods, such as FD, SD, and LTSD, vary in the selection of wavelength bands, they all aim to extract information more relevant to soil nutrients using transforming spectral data. These transformation methods may have different applicability for different types of soil or different nutritional components. At the same time, the selected bands for each soil nutrient are specific, indicating that different soil nutrients exhibit distinct spectral signatures.

4.6. Building Hyperspectral Models for Soil Nutrient Content Estimation

4.6.1. Partitioning of Training and Validation Sets

For soil nutrient data, a 7:3 ratio split was adopted to ensure that the training set contains 46 samples while the validation set contains 20 samples. The key to this split is to ensure that both the training and validation sets adequately cover the entire range of variation in soil nutrients. The effectiveness of the model was measured using the correlation coefficient (R) and the root mean square error (RMSE). A higher correlation coefficient and a smaller RMSE indicate a better model performance [28].

4.6.2. Model Construction and Validation for Soil Nutrients

As shown in Table 2, the performance evaluation metrics including R^2 , RMSE, and the Ratio of Performance to Deviation (RPD) of different spectral transformation methods and models in the prediction of SOM are presented for both the training and validation sets. This table demonstrates the comparative analysis of different prediction models in terms of prediction accuracy. Two prediction models are listed: PLSR and MLSR. In the PLSR model, higher simulation accuracy was achieved using different forms of second-order differentiation transformation. For the LRD and LRSD preprocessing models, there was a significant decrease in performance, showing lower R^2 values and higher RMSE. Regarding the RPD data, the RPD values of PLSR were relatively low. The PLSR model utilizes cross-validation for hyperparameter optimization. The range for the number of components is set from 2 to 10. The best hyperparameter values are chosen through training the model on the training set and evaluating its performance on the validation set, suggesting potentially lower predictive capabilities compared to MLSR. In the MLSR model, overall performance was better compared to PLSR, especially in the FD, SD, LTFD, and LTSD transformations, with higher R^2 values and lower RMSE. Moreover, MLSR consistently exhibited higher RPD values, indicating strong predictive capabilities across different transformations. While there was a decrease in performance for the LRD and LRSD preprocessing, the decline was relatively minor. In various forms of first-order and second-order differentiation transformations, MLSR demonstrated better overall performance than PLSR in predicting SOM.

Table 2. Statistics of accuracy parameters for SOM inversion model.

Model	Transformation	Training Set		Validation Set		
		R^2	RMSE/ $\text{g} \cdot \text{kg}^{-1}$	R^2	RMSE/ $\text{g} \cdot \text{kg}^{-1}$	RPD
PLSR	FD	0.76	10.19	0.79	11.14	1.41
	SD	0.83	13.13	0.74	12.01	1.52
	LTFD	0.75	10.22	0.80	11.34	1.49
	LTSD	0.81	10.36	0.86	11.12	1.46
	LRD	0.68	11.10	0.60	11.34	1.26
	LRSD	0.52	12.16	0.51	11.16	1.25
MLSR	FD	0.84	6.40	0.86	6.45	1.56
	SD	0.87	6.61	0.87	6.54	1.55
	LTFD	0.85	6.94	0.87	6.48	1.58
	LTSD	0.86	6.76	0.88	6.76	1.59
	LRD	0.59	9.46	0.58	9.19	1.57
	LRSD	0.51	9.76	0.60	9.66	1.59

The prediction accuracy of the models established in the LRD and LRSD preprocessing was not high for both models. However, MLSR maintained better prediction accuracy and consistency across different spectral transformations. Therefore, in the accuracy parameter statistics of the SOM inversion model, adopting the MLSR model may result in higher prediction accuracy, with the highest R^2 reaching 0.87, RMSE of $6.61 \text{ g} \cdot \text{kg}^{-1}$, and an RPD of up to 1.55.

In Table 3, for the preprocessing methods LTFD, LTSD, LRD, and LRSD, their R^2 values are very close, indicating higher model stability. The RPD values range from 1.51 to 1.57, reflecting higher model prediction accuracy, especially when employing the LRD transformation, where in the PLSR model, R^2 reaches 0.77 with an RMSE of $0.37 \text{ g} \cdot \text{kg}^{-1}$. In the MLSR model, the overall accuracy is not as high as in the PLSR model, and the RPD values are relatively lower compared to the PLSR model.

Table 3. Statistics of accuracy parameters for TN inversion model.

Model	Transformation	Training Set		Validation Set		
		R^2	RMSE/ $\text{g} \cdot \text{kg}^{-1}$	R^2	RMSE/ $\text{g} \cdot \text{kg}^{-1}$	RPD
PLSR	FD	0.61	0.45	0.61	0.46	1.45
	SD	0.65	0.43	0.64	0.43	1.56
	LTFD	0.73	0.40	0.79	0.40	1.56
	LTSD	0.75	0.40	0.75	0.38	1.57
	LRD	0.77	0.37	0.76	0.37	1.55
	LRSD	0.74	0.37	0.78	0.39	1.51
MLSR	FD	0.52	0.45	0.50	0.46	1.43
	SD	0.72	0.40	0.54	0.55	1.54
	LTFD	0.65	0.36	0.62	0.40	1.45
	LTSD	0.68	0.37	0.65	0.38	1.46
	LRD	0.71	0.33	0.70	0.34	1.47
	LRSD	0.68	0.32	0.70	0.31	1.44

Table 4 provides statistical accuracy parameters for the TP inversion models, evaluated using different models: PLSR and MLSR. For the PLSR model, using the FD transformation, the lowest R^2 value is 0.42, indicating relatively poor performance in data interpretation. In contrast, with the LTSD model, the highest R^2 value reaches 0.64, with an RMSE of $0.05 \text{ g} \cdot \text{kg}^{-1}$. In the MLSR model, a higher R^2 value of 0.69 is achieved, but the RMSE is not significantly lower than other methods. Overall, there is little difference in accuracy between the two models for TP inversion. Both models achieve good inversion accuracy with transformations using SD, LTFD, and LTSD preprocessing methods.

Table 4. Statistics of accuracy parameters for TP inversion model.

Model	Transformation	Training Set		Validation Set		
		R^2	RMSE/ $\text{g} \cdot \text{kg}^{-1}$	R^2	RMSE/ $\text{g} \cdot \text{kg}^{-1}$	RPD
PLSR	FD	0.42	0.07	0.43	0.07	1.16
	SD	0.63	0.05	0.64	0.05	1.14
	LTFD	0.63	0.05	0.66	0.05	1.13
	LTSD	0.64	0.05	0.62	0.05	1.27
	LRD	0.56	0.05	0.57	0.06	1.14
	LRSD	0.42	0.05	0.43	0.06	1.17
MLSR	FD	0.47	0.05	0.48	0.06	1.16
	SD	0.63	0.06	0.64	0.06	1.17
	LTFD	0.64	0.06	0.66	0.05	1.17
	LTSD	0.69	0.04	0.68	0.06	1.11
	LRD	0.63	0.06	0.69	0.06	1.13
	LRSD	0.47	0.06	0.49	0.05	1.14

In Table 5, the accuracy parameters for the AN inversion models are provided, comparing the performance of PLSR and MLSR models. In the PLSR model, high inversion accuracy is achieved with the R^2 values for SD, LTFD, LTSD, LRD, and LRSD transformations. Particularly, the PLSR model established with the LRSD transformation achieves the

highest inversion accuracy, with an R^2 of 0.83 and an RMSE of $24.12 \text{ mg} \cdot \text{kg}^{-1}$. This indicates the strongest data interpretation ability of this method in the PLSR model, with the smallest prediction error. In the MLSR model, the LTSD transformation yields the highest R^2 value of 0.82, indicating good data interpretation ability, with a relatively low RMSE of $26.05 \text{ mg} \cdot \text{kg}^{-1}$, indicating a relatively small error. For soil AN inversion, both models demonstrate high accuracy except for the FD model.

Table 5. Statistics of accuracy parameters for AN inversion model.

Model	Transformation	Training Set		Validation Set		
		R^2	RMSE/ $\text{mg} \cdot \text{kg}^{-1}$	R^2	RMSE $\text{mg} \cdot \text{kg}^{-1}$	RPD
PLSR	FD	0.57	38.64	0.55	39.55	1.09
	SD	0.70	35.14	0.67	39.56	1.12
	LTFD	0.82	27.17	0.63	40.84	1.13
	LTSD	0.81	30.21	0.66	40.45	1.24
	LRD	0.75	25.21	0.70	38.56	1.20
	LRSD	0.83	24.12	0.68	37.54	1.27
MLSR	FD	0.57	30.12	0.66	35.41	1.46
	SD	0.68	27.45	0.68	34.11	1.45
	LTFD	0.72	27.64	0.75	35.14	1.53
	LTSD	0.82	26.05	0.80	26.45	1.52
	LRD	0.72	26.48	0.71	32.00	1.51
	LRSD	0.76	28.19	0.61	29.42	1.50

4.6.3. Optimal Estimation Model of Soil Nutrients

Figure 10a shows the prediction of SOM with an R^2 value of 0.87, indicating a strong correlation between the model and the actual data. The RMSE is $6.61 \text{ g} \cdot \text{kg}^{-1}$, measuring the deviation between estimated and actual values. A smaller RMSE suggests more accurate predictions, and most points appear close to the fitted line, indicating good agreement between estimated and actual values. SD-MLSR may demonstrate applicability to the data in testing, providing good prediction accuracy and stability. Through SD transformation, the model can effectively filter out noise in the spectral data, highlighting important features [29]; PLSR is capable of handling datasets with highly correlated variables.

In Figure 10b, for TN prediction, an R^2 value of 0.77 indicates good correlation, while an RMSE of $0.37 \text{ g} \cdot \text{kg}^{-1}$ suggests better predictive accuracy with smaller values. The LRD-PLSR model is likely chosen as the best option because its preprocessing steps enhance the signal, and PLSR's multivariate analysis method is suitable for handling the complexity of spectral data. These preprocessing steps and modeling methods enhance the model's ability to capture variations in soil TN [30].

The prediction for TP depicted in Figure 10c of the figure, an R^2 of 0.69 and an RMSE of $0.04 \text{ g} \cdot \text{kg}^{-1}$, indicate a relatively small average error between the predicted soil total phosphorus content and the actual measured values. The higher R^2 and lower RMSE suggest that LTSD-MLSR may perform relatively well on the dataset, providing more accurate predictions compared to other models, particularly in handling complex spectral data and soil TP content.

In Figure 10d, the prediction for AN demonstrates a high predictive accuracy with an R^2 of 0.83 and an RMSE of $24.12 \text{ mg} \cdot \text{kg}^{-1}$. The utilization of LRSD preprocessing might have enabled this model to better handle noise and signal variability in the spectral data, thereby enhancing the predictive accuracy.

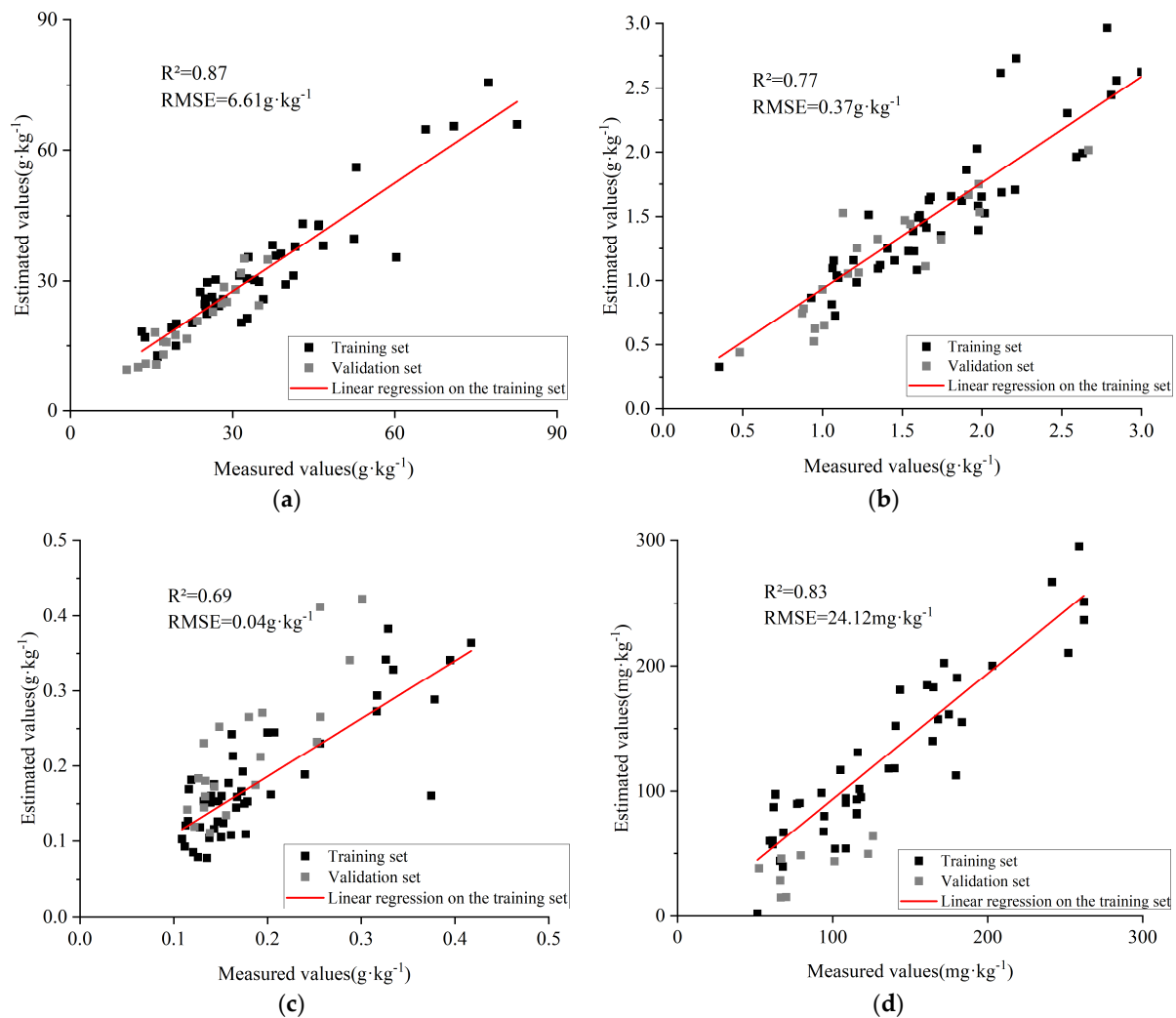


Figure 10. Optimal model for selecting soil nutrients. (a) Best model for SOM:SD—MLSR; (b) best model for TN: LRD—PLSR; (c) best model for TP: LTSD—MLSR; (d) best model for AN: LRSD—PLSR.

5. Discussion

The high correlation between the visible light region 400–700 nm and soil nutrients indicates that it may contain important information about soil organic matter and other key nutrients. The visible reflectance spectra can be used to accurately determine important soil components [31]. Visible spectra are easier to obtain using common spectral sensors and can provide information about the surface properties of the soil. Within the visible light bands, different colors of wavelengths interact differently with specific soil components [32]. Soils rich in organic matter tend to absorb more blue light and reflect more red light, making these soils appear darker. A major advantage of using visible spectra for soil analysis is that most spectral instruments are capable of capturing this range of spectral data. This data is then processed through algorithms and models to assess soil nutrients [33]. The near-infrared region 1000–1500 nm is renowned for its sensitivity to organic matter. Near-infrared spectra penetrate soil deeper than visible light, allowing for a more comprehensive analysis of soil composition, including both organic and inorganic substances. One notable advantage of near-infrared spectra over visible spectra is their ability to penetrate the soil surface more deeply, providing information about the chemical and physical state of the soil beyond just the surface layer. The short-wave infrared region 2000–2300 nm is particularly useful for identifying the specific mineral and chemical properties of the soil. It can detect soil variations that are not visible in the near-infrared and visible spectra, making it a powerful tool for monitoring soil health and texture [34]. Notably, the specificity of the

selected bands for each soil nutrient reflects the unique spectral signatures exhibited by different soil nutrients. Based on these findings, for this study, the spectral analysis of soil nutrients particularly in the visible light range of 400–700 nm, the near-infrared region of 1000–1500 nm, and the short-wave infrared region of 2000–2300 nm, is expected to assist us in gaining a deeper understanding of the spectral characteristics of soil nutrients.

The hyperspectral data of soil in this study area was collected in the field, and the use of different forms of differential transformation has been proven to be helpful in extracting soil nutrient information [27,35]. The existence of interfering factors often leads to distortions or deviations in spectral data. Differential transformation can effectively eliminate these interfering factors, thereby improving the accuracy of inversion and helping us more accurately identify and analyze soil components. Alternatively, differential transformation can capture effective information from spectral data for identifying soil nutrients. Based on this, further exploring the optimized combination of different differential transformation methods and parameter adjustment strategies can be used to better adapt to different soil types and inversion needs.

After preprocessing, PLSR and MLSR models achieved high accuracy in different soil nutrients. This may be because the two models can fully utilize the information in hyperspectral data and find the best relationship model between spectral variables and soil component content through regression analysis, thereby achieving accurate prediction of soil components [36,37]. In practical applications, the parameters of the model can be further adjusted based on different soil types and component characteristics to obtain better prediction results. Alternatively, other technical means, such as adding principal component analysis, can be combined to further improve the accuracy of soil nutrient content prediction.

For future work in hyperspectral soil sensing, to improve accuracy, we can refine the application of our models. On one hand, we can attempt to use more complex models that may be able to capture the subtle differences in soil properties more accurately. On the other hand, we can optimize the performance of models by adding more data processing steps. In terms of instrument selection, using more precise spectrometers that can capture richer spectral information is crucial for improving the accuracy of soil property measurements. In this study, to eliminate the influence of moisture on spectral data, we chose to collect samples from dry soil. However, to further enhance the convenience and efficiency of operations, we are exploring in situ spectral collection, which allows for direct spectral measurements in the soil's natural state, without the need for collecting and preprocessing soil samples. With spectral data collected in situ, we can conduct subsequent inversion analyses of soil nutrients, thus quickly obtaining soil nutrient information.

In the current application of soil nutrient inversion models, although the accuracy has reached a reasonable level, there is still room for improvement. To optimize the model, we can consider introducing more targeted models or adjusting parameters to achieve higher simulation accuracy. Currently, the research is limited to the inversion of four types of soil nutrients, but in the future, we could try to expand the model to predict other nutrients in the soil, including heavy metals. However, the current study also has some limitations. Firstly, the high cost of ASD spectral instruments may limit their widespread use in agriculture. Secondly, the large volume of soil hyperspectral data requires complex data processing and analysis techniques. This not only demands the appropriate hardware support but also requires users to have the ability to handle large datasets and a high level of technical knowledge.

6. Conclusions

(1) During the spectral reflectance screening of soil nutrients, the visible light region of 400–700 nm consistently exhibits a high correlation. Apart from the visible light region, the near-infrared region of 1000–1500 nm and the short-wave infrared region of 2000–2300 nm also show a high correlation in the screening of multiple nutrients. These spectral regions are of reference value for assessing soil nutrients, and future research directions include the

further exploration of these specific wavelengths to enhance the accuracy and efficiency of soil nutrient detection.

(2) Through various forms of first-order and second-order differentiation transformations, the prediction accuracy of the model can be improved, significantly highlighting the spectral data's information on soil nutrients such as SOM, TN, TP, and AN. Future research could focus on improving these preprocessing techniques.

(3) After applying SG smoothing to the original bands and utilizing six preprocessing methods selected based on the Pearson correlation analysis of characteristic bands, PLSR and MLSR regression models were established. These models were used to predict SOM, TN, TP, and AN, thereby identifying the optimal prediction pathways. Specifically, the prediction pathway for soil SOM content was SD-MLSR; for soil TN content, it was LRD-PLSR; for soil TP content, it was LTSD-MLSR; and for soil AN content, it was LRSD-PLSR. Utilizing different preprocessing methods in conjunction with PLSR and MLSR regression models highlights the importance of optimizing prediction pathways for accurately estimating soil nutrient content. Future research directions include the further exploration of alternative preprocessing techniques and regression models to enhance the robustness and applicability of soil nutrient prediction models. Additionally, integrating advanced machine learning algorithms and hyperspectral data analysis techniques offers promising avenues for improving the prediction accuracy and efficiency of soil nutrient assessment.

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

Funding: Supported by the major science and technology special program of Yunnan Province (202002AE090010). Supported by the China Scholarship Council (202206715002).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Yang, X.; Xiong, J.; Du, T.; Ju, X.; Gan, Y.; Li, S.; Xia, L.; Shen, Y.; Pacenka, S.; Steenhuis, T.S.; et al. Diversifying crop rotation increases food production, reduces net greenhouse gas emissions and improves soil health. *Nat. Commun.* **2024**, *15*, 1–14. [\[CrossRef\]](#)
2. Li, F.; Li, D.; Voors, M.; Feng, S.; Zhang, W.; Heerink, N. Improving smallholder farmer's soil nutrient management: The effect of science and technology backyards in the North China plain. *China Agric. Econ. Rev.* **2023**, *15*, 134–158. [\[CrossRef\]](#)
3. Ding, J.-L.; Wu, M.-C.; Liu, H.-X.; Li, Z.-G. Study on the soil salinization monitoring based on synthetical hyperspectral index. *Spectrosc. Spectr. Anal.* **2012**, *32*, 1918–1922.
4. Yu, H.; Kong, B.; Wang, Q.; Liu, X.; Liu, X. Hyperspectral remote sensing applications in soil: A review. In *Hyperspectral Remote Sensing*; Pandey, P.C., Srivastava, P.K., Balzter, H., Bhattacharya, B., Petropoulos, G.P., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 269–291.
5. Hively, W.D.; McCarty, G.W.; Reeves, J.B.; Lang, M.W.; Oesterling, R.A.; Delwiche, S.R. Use of airborne hyperspectral imagery to map soil properties in tilled agricultural fields. *Appl. Environ. Soil Sci.* **2011**, *2011*, 358193. [\[CrossRef\]](#)
6. Vibhute, A.D.; Kale, K.V. Mapping several soil types using hyperspectral datasets and advanced machine learning methods. *Results Opt.* **2023**, *12*, 100503. [\[CrossRef\]](#)
7. Yu, H.; Kong, B.; Hou, Y.; Xu, X.; Chen, T.; Liu, X. A critical review on applications of hyperspectral remote sensing in crop monitoring. *Exp. Agric.* **2022**, *58*, e26. [\[CrossRef\]](#)
8. Zhao, H. Spectral characteristics analysis of common precious jade material reflectance based on portable ground object spectrometer. *Mineral. Petrol.* **2021**, *41*, 1–12. [\[CrossRef\]](#)
9. Tan, K.L.; Wan, Y.Q.; Yang, Y.D.; Duan, Q.B. Study of hyperspectral remote sensing for archaeology. *J. Infrared Millim. Waves* **2005**, *24*, 437–440.
10. Sut, M.; Fischer, T.; Repmann, F.; Raab, T.; Dimitrova, T. Feasibility of field portable near infrared (nir) spectroscopy to determine cyanide concentrations in soil. *Water Air Soil Pollut.* **2012**, *223*, 5495–5504. [\[CrossRef\]](#)

11. Arif, M.; Qi, Y.; Dong, Z.; Wei, H. Rapid retrieval of cadmium and lead content from urban greenbelt zones using hyperspectral characteristic bands. *J. Clean. Prod.* **2022**, *374*, 133922. [\[CrossRef\]](#)
12. Daniel, K.W.; Tripathi, N.K.; Honda, K.; Apisit, E. Analysis of vnir (400–1100 nm) spectral signatures for estimation of soil organic matter in tropical soils of Thailand. *Int. J. Remote Sens.* **2004**, *25*, 643–652. [\[CrossRef\]](#)
13. Wan, M.; Jin, X.; Han, Y.; Wang, L.; Li, S.; Rao, Y.; Zhang, X.; Gao, Q. A stacking-based ensemble learning method for available nitrogen soil prediction with a handheld micronear-infrared spectrometer. *J. Appl. Spectrosc.* **2023**, *89*, 1241–1253. [\[CrossRef\]](#)
14. Hu, Y.; Gao, X.; Shen, Z.; Xiao, Y. Estimating fertility index by using field-measured vis-nir spectroscopy in the huanghui river basin. *Chin. J. Soil Sci.* **2021**, *52*, 575–584.
15. Dongmei, X.I.; Weidong, D.; Hongguang, G.A.O.; Huaming, M.A.O. Physico-chemical properties of rocks and soils and abundances of minerals in main geological background areas in yunnan province. *Soils* **2008**, *40*, 114–120.
16. Kirchhoff, M.; Romes, T.; Marzolf, I.; Seeger, M.; Hssaine, A.A.; Ries, J.B. Spatial distribution of argan tree influence on soil properties in southern morocco. *Soil* **2021**, *7*, 511–524. [\[CrossRef\]](#)
17. Mendes, W.d.S.; Demattê, J.A.; de Resende, M.E.B.; Ruiz, L.F.C.; de Mello, D.C.; Rosas, J.T.F.; Silvero, N.E.Q.; Alleoni, L.R.F.; Colzato, M.; Rosin, N.A.; et al. A remote sensing framework to map potential toxic elements in agricultural soils in the humid tropics. *Environ. Pollut.* **2022**, *292*, 118397. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Wang, Z.; Cao, M.; Cai, W.; Huang, H.; Zeng, H. Effects of land use on humus and soil enzymes of yellow soil under soil erosion. *Bull. Soil Water Conserv.* **2017**, *37*, 27–33.
19. Lesu, Y. An improvement on the heating condition of soil organic matter determination. *Ecol. Sci.* **2006**, *25*, 459–461.
20. HJ 717-2014; Soil Quality—Determination of Total Nitrogen—Modified Kjeldahl Method. Ministry of Ecology and Environment of the People's Republic of China: Beijing, China, 2015.
21. HJ 632-2011; Soil-Determination of Total Phosphorus by Alkali Fusion—Mo-Sb Anti Spectrophotometric Method. Ministry of Ecology and Environment of the People's Republic of China: Beijing, China, 2012.
22. Lu, R.K. *Analysis Method of Soil Agricultural Chemistry*; China Agricultural Science and Technology Press: Beijing, China, 2000.
23. Chen, H.; Pan, T.; Chen, J. Combination optimization of multiple scatter correction and savitzky-golay smoothing modes applied to the near infrared spectroscopy analysis of soil organic matter. *Comput. Appl. Chem.* **2011**, *28*, 518–522.
24. Guo, P.; Li, T.; Gao, H.; Chen, X.; Cui, Y.; Huang, Y. Evaluating calibration and spectral variable selection methods for predicting three soil nutrients using vis-nir spectroscopy. *Remote Sens.* **2021**, *13*, 4000. [\[CrossRef\]](#)
25. Tang, M.; Yang, Q.; Tang, H. Prediction of soil organic carbon content based on hyperspectral data in peak-cluster depressions, northeastern Guangxi. *Carsol. Sin.* **2021**, *40*, 876–883.
26. Chen, L.; Wang, H.; Sun, C. The application of anomaly contrast to extracting geochemical anomaly information: A study of duobaoshan area in heilongjiang province. *Geophys. Geochem. Explor.* **2018**, *42*, 1150–1155.
27. Yansenjiang, K.; Yang, S.; Nigara, T.; Zhang, F. Hyperspectral estimation of soil electrical conductivity based on fractional order differentially optimised spectral indices. *Acta Ecol. Sin.* **2019**, *39*, 7237–7248.
28. Li, Z.Y.; Deng, F.; He, J.L.; Wei, W. Hyperspectral estimation model of heavy metal arsenic in soil. *Spectrosc. Spectr. Anal.* **2021**, *41*, 2872–2878.
29. Nie, L.; Dou, Z.; Cui, L.; Tang, X.; Zhai, X.; Zhao, X.; Lei, Y.; Li, J.; Wang, J.; Li, W. Hyperspectral inversion of soil carbon and nutrient contents in the yellow river delta wetland. *Diversity* **2022**, *14*, 862. [\[CrossRef\]](#)
30. Wang, C.; Pan, X.; Zhou, R.; Liu, Y.; Li, Y.; Xie, X. Prediction of soil properties using pls-r-based soil-environment models. *Acta Pedol. Sin.* **2012**, *49*, 237–245.
31. Piccini, C.; Metzger, K.; Debaene, G.; Stenberg, B.; Götzinger, S.; Borůvka, L.; Sandén, T.; Bragazza, L.; Liebisch, F. In-field soil spectroscopy in vis-nir range for fast and reliable soil analysis: A review. *Eur. J. Soil Sci.* **2024**, *75*, e13481. [\[CrossRef\]](#)
32. Horta, A.; Malone, B.; Stockmann, U.; Minasny, B.; Bishop, T.; McBratney, A.; Pallasser, R.; Pozza, L. Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. *Geoderma* **2015**, *241–242*, 180–209. [\[CrossRef\]](#)
33. Liu, H.; Zhang, X.; Yu, W.; Zhang, B.; Song, K.; Blackwell, J. Simulating models for phaeozem hyperspectral reflectance. *Int. J. Remote Sens.* **2011**, *32*, 3819–3834. [\[CrossRef\]](#)
34. Zhang, S.; Lu, X.; Nie, G.G.; Li, Y.R.; Shao, Y.T.; Tian, Y.Q.; Fan, L.Q.; Zhang, Y.J. Estimation of soil organic matter in coastal wetlands by svm and bp based on hyperspectral remote sensing. *Spectrosc. Spectr. Anal.* **2020**, *40*, 556–561.
35. Li, X.; Ding, J.; Hou, Y.; Deng, K. Estimating the soil salt content and electrical conductivity in semi-arid and arid areas by using hyperspectral data. *J. Glaciol. Geocryol.* **2015**, *37*, 1050–1058.
36. Zhang, X.; Yao, Y.; Yan, X. Effects of spectral transformation and spectral resolution on the estimation accuracy of soil organic matter content. *Soil Fertil. Sci. China* **2023**, *3*, 184–193.
37. Mu, Q.; Yang, G.; Chen, H.; Zhang, T. Hyperspectral characteristics of simulated soils with different salinity in laboratory. *J. North-East For. Univ.* **2021**, *49*, 68–75.

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.