

Estimating the Texture of Purple Soils Using Vis-NIR Spectroscopy and Optimized Conversion Models

Baina Chen^{1,2}, Jie Wei^{1,2*}, Qiang Tang³, Yu Gou¹, Chunhong Liu^{1,2}

¹Chongqing Key Laboratory of Surface Process and Environment Remote Sensing in the Three Gorges Reservoir Area, School of Geography and Tourism, Chongqing Normal University, Chongqing, China

²Chongqing Field Observation and Research Station of Surface Ecological Process in the Three Gorges Reservoir Area, School of Geography and Tourism, Chongqing Normal University, Chongqing, China

³Chongqing Jinfo Mountain Karst Ecosystem National Observation and Research Station, School of Geographical Sciences, Southwest University, Chongqing, China

Email: *wei_jie@mails.ucas.ac.cn

How to cite this paper: Chen, B.N., Wei, J., Tang, Q., Gou, Y. and Liu, C.H. (2023) Estimating the Texture of Purple Soils Using Vis-NIR Spectroscopy and Optimized Conversion Models. *Agricultural Sciences*, **14**, 202-218.

https://doi.org/10.4236/as.2023.142014

Received: December 31, 2022 Accepted: February 19, 2023 Published: February 22, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

 Open Access

Abstract

Soil texture is an indicator of soil physical structure which delivers many ecological functions of soils such as thermal regime, plant growth, and soil quality. However, traditional methods for soil texture measurement are time-consuming and labor-intensive. This study attempts to explore an indirect method for rapid estimating the texture of three subgroups of purple soils (i.e. calcareous, neutral, and acidic). 190 topsoil (0 - 10 cm) samples were collected from sloping croplands in Tongnan and Beibei Districts of Chongqing Municipality in China. Vis-NIR spectrum was measured and processed, and stepwise multiple linear regression (SMLR), partial least squares regression (PLSR), and back propagation neural network (BPNN) models were constructed to inform the soil texture. The clay fractions ranged from 4.40% to 27.12% while sand fractions ranged from 0.34% to 36.57%, hereby soil samples encompass three textural classes (i.e. silt, silt loam, and silty clay loam). For the original spectrum, the texture of calcareous and neutral purple soils was not significantly correlated with spectral reflectance and linear models (SMLR and PLSR) exhibited low prediction accuracy. The correlation coefficients and the goodness-of-fits between soil texture and the transformed spectra of all soil groups increased by continuum-removal (CR), first-order differential (R'), and second-order differential (R'') transformations. Among them, the R'' had the best performance in terms of improving the correlation coefficients and the goodness-of-fits. For the calcareous purple soil, the SMLR exceeds PLSR and BPNN with a higher coefficient of determination (R^2) and the ratio of performance to inter-quartile distance (RPIQ) values and lower root mean square error of validation (RMSEV), but for the neutral and acidic purple soils, the PLSR model has a better prediction accuracy. In summary, the linear methods (SMLR and PLSR) are more reliable in estimating the texture of the three purple soil groups when using Vis-NIR spectroscopy inversion.

Keywords

Soil Texture, Vis-NIR Spectra, Stepwise Multiple Linear Regression, Partial Least Squares Regression, Backpropagation Neural Network

1. Introduction

Soil texture depicts soil particle composition expressed as the mass percentage of soil particles at all levels [1]. It is a relatively stable natural attribute of the soil and affects the fertilization response, water-holding capacity, aeration, and the difficulty of tillage [2]. More specifically, soil texture regulates the growth and development of crops by determining runoff infiltration and soil water content, air, and heat, as well as the transformation rate and existing state of soil nutrients, thus affecting the growth and development of crop roots and the yield and quality of crops [3]. Different soil textures always have distinct agricultural production traits; therefore, it is necessary to understand the soil texture when using, managing, and improving soil [4]. Moreover, soil texture also modulates land surface and atmospheric processes, such as soil-plant-atmosphere interaction, soil erosion, and soil solute transport [5]. Therefore, a rapid acquisition of soil texture information is of great significance for land management to improve agricultural productivity and ecosystem services.

Soil texture is always determined by laboratory methods and *in-situ* field measurements. Laboratory measurements generally adopt specific gravity and suction tube methods, and field measurements employ dry test and wet test methods. These methods largely require large-scale field sampling and pretreatment efforts [6]. Thus, it is time-consuming and labor-intensive and causes irrecoverable damage to samples [7]. The laser diffraction method offers a relatively alternative option [8]. In recent years, soil Vis-NIR reflectance spectroscopy has been increasingly used to solve the trade-offs between large-scale soil information and high cost [9]. Its attraction lies in that one single measurement can be used to assess a wide range of soil properties, thus facilitating the analysis of many samples in a short time [10]. It can predict the chemical and physical properties of soils such as soil organic matter, texture, and clay minerals [11].

It is difficult to directly predict soil texture from the spectrum. Spectral preprocessing is an important initial step and improves the accuracy of the prediction results [12]. Approaches used for spectral preprocessing include Savitzky-Golay (SG) smoothing [13] and multiplicative scattering correction (MSC) [14]. Given soil components interact with each other in a complex way to produce their spectra; there is a strong correlation between the adjacent bands of the spectral data. However, not all bands have an equal impact on the subsequent spectral processing [15]. Therefore, the importance of characteristic bands selection in spectral data processing and analysis has been widely recognized and has become a key step in hyperspectral data modeling and analysis [16]. It is also necessary to select the appropriate calibration model for quantitative measurement. Using appropriate calibration methods is key in determining the success of simulated reflection spectrum calibration. Common calibration methods include linear and nonlinear approaches. The linear regression model establishes simple linear and multivariate linear regression equations based on the variables and auxiliary information of known points, to realize the prediction of unknown points [17] [18]. The commonly used linear methods include stepwise multiple linear regression (SMLR) [19] and partial least squares regression (PLSR) [20]. Machine learning in nonlinear methods has good applicability in dealing with multi-dimensional, nonlinear massive data and improving the generalization ability of the model [21] [22]. Common models include random forest (RF) [23], support vector machine (SVM) [24], neural networks (NNs) [25], etc.

Purple soil, one of the soil types defined within the genetic soil classification of China, is the most important and widely distributed type of arid land soil in southern China. Purple soil is formed by weathering of the Mesozoic and Cenozoic purple sandstone (including Triassic, Jurassic, Cretaceous, and Tertiary) [26]. It is fast weathered and it has high soil mineral content, but also has the defects of the shallow soil layer and low aggregate content, coupled with the influence of human farming activities, thus purple soil is susceptible to environmental disturbances such as reservoir fluctuations and heavy rainfall [27], which seriously threatens the safety of village buildings and roads and hinders the development of agricultural production [28]. However, there are few studies on the estimation of purple soil texture, especially for the three types of purple soil. Therefore, it is necessary to study the Vis-NIR reflectance spectra of the calcareous (pH > 7.5), neutral ($6.5 \le pH \le 7.5$), and acidic (pH < 6.5) purple soil and their relationship with the texture of the soil and assess the suitability of Vis-NIR spectroscopy coupled with SMLR, PLSR, and BPNN methods for quickly obtaining textural information of soil in Chongqing.

2. Materials and Methods

2.1. Site Description

The sampling site is located in Tongnan and Beibei Districts of Chongqing. Tongnan District is characterized by a shallow hill landform of the basin. It lies between $105^{\circ}31' - 106^{\circ}00'E$ and $29^{\circ}47' - 30^{\circ}26'N$ and rises to altitudes of 300 to 400 m. The purple soil in this area accounts for 46.30% of the total cultivated land, and the main soil type is calcareous purple soil. Beibei District belongs to the parallel ridgeline area of Chongqing. It lies between $106^{\circ}18' - 106^{\circ}40'E$ and $29^{\circ}27' - 30^{\circ}05'N$ and rises to altitudes of 500 to 900 m with steeper terrain than that in Tongnan. Paddy soil and purple soil accounted for 81.7% of the soil area in the Beibei District, and the main soil types are neutral and acidic purple soils. Both two districts have a humid subtropical monsoon climate, with annual average temperatures ranging from $16^{\circ}C$ to $18^{\circ}C$, annual rainfall ranging from 1000 to 1350 mm, and total annual sunshine of 1000 to 1200 hours.

2.2. Soil Sampling and Analysis

Soil sampling was carried out using a five-point mixing method. 62 calcareous purple soil samples were collected from Gaohe catchment in Tongnan, 63 neutral purple soil samples and 65 acidic purple soil samples were collected from Jigongshan and Baihelin catchments in Beibei. All samples were air-dried, disaggregated, and sieved using a 2-mm mesh after removal of stones, plant roots, and litter. Soil texture was determined using laser particle size analysis (Malvern MS2000) and grouped into three subcategories in terms of clay (<0.002 mm), silt (0.002 - 0.05 mm), and sand (>0.05 - 2 mm). The wide difference in the percentage contents of clay and sand in soil samples leads to varied spectral properties [29]; therefore, this paper only studied the contents of clay and sand. Two-thirds of the datasets of the clay and sand contents were used for calibration and one-third for validation according to a stratified sampling approach.

2.3. Spectra Measurements

The spectral reflectance of the soil samples was measured using an ASD FieldSpec 4 spectroradiometer (Analytical Spectral Devices Inc., USA). We used a wavelength range of 350 - 2500 nm, with a sampling interval of 1.4 nm between 350 and 1000 nm, and 2 nm between 1000 and 2500 nm, and measured a total of 2150 nm. A halogen light source (50 W) that could provide parallel light was placed at a distance of 30 cm from the soil (which was held in a black sample dish with a diameter of 6.5 cm and a depth of 2 cm) and at a zenith angle of 45°. To reduce the diffuse reflection of light, the surface was scraped flat and calibrated before each measurement. The spectral reflectance of each sample was measured at four angles (rotating the petri-dish 90° clockwise) with five replications, and the average of these 20 measurements was deemed as the standard spectral reflectance of the sample.

2.4. Spectrum Preprocessing

The SG convolution smoothing method [30] was used to pretreat spectral reflectance data, and then the smoothed spectral data can be transformed to increase the correlation between spectral reflectance and the soil's physical and chemical properties [31]. Spectral transformation mainly uses three methods including continuum-removal (CR), first-order differential reflectivity (R), and second-order differential reflectivity (R'). The three methods can highlight effectively the reflection and absorption characteristics of the spectral curve. Among them, the R' and R'' can also be used to quickly determine the inflection point and the position of the maximum and minimum reflectance in the wavelength range. The calculation method is as follows:

$$R'(\lambda_i) = R(\lambda) / R_C(\lambda)$$
⁽¹⁾

$$R'(\lambda_i) = \left[R(\lambda_i + 1) - R(\lambda_i) \right] / \Delta \lambda$$
(2)

$$R''(\lambda_i) = \left[R'(\lambda_i + 1) - R'(\lambda_i) \right] / \Delta \lambda$$
(3)

Here, λ_i is the wavelength band at *i* nm, $\lambda_i + 1$ is the wavelength band at *i* + 1 nm, and $R(\lambda)$ is the spectral reflectance of band λ_i . $R'(\lambda_i)$ and $R''(\lambda_i)$ are the first- and second-order differential reflectivity of band λ_i , where $\Delta \lambda = \lambda_i + 1 - \lambda_i = 10$ nm.

2.5. Model Description

2.5.1. Linear Modeling

In this study, we used the SMLR and PLSR methods for the modeling. SMLR is used to predict the dependent variable according to the best combination of several independent variables. The key is that the selection of the dependent variable is based on the principle of keeping the most significant band of the dependent variable and keeping the number of the dependent variable as small as possible. PLSR is a fully independent linear regression model, which combines the methods of multivariate regression analysis and principal component analysis. It can solve the problems of independent variable multicollinearity and sample numbers less than the number of variables [32], and it can also effectively identify spectral information and noise, thereby reducing the spectral dimension and data redundancy in spectral modeling [33].

2.5.2. Non-Linearity Modeling

A BPNN is an information processing system based on the structure and function of the brain's neural network [34]. It can fit complex nonlinear functions by learning from a large number of sample data and has a strong nonlinear mapping and generalization ability [35]. In this study, the BPNN was operated in MATLAB software, and the network consisted of an input layer, a hidden layer, and an output layer. The soil samples were divided into a model set, test set, and verification set according to the proportion of 4:2:3, and they were normalized in the learning process.

2.5.3. Prediction Accuracy

The indicators that we used to test the three models were the coefficient of determination (R^2), root mean square error (RMSE), and the ratio of performance to inter-quartile distance (RPIQ). A larger RPIQ indicates a better-fitting effect of the model. The values of R^2 , RMSE, and RPIQ were calculated as follows:

$$R^{2} = \frac{\sum_{i=1}^{n} (\hat{y}_{i} = \overline{y}_{i})^{2}}{\sum_{i=1}^{n} (y_{i} = \overline{y}_{i})^{2}}$$
(4)

RMSE =
$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}$$
 (5)

$$RPIQ = \frac{IQ}{RMSEV}$$
(6)

Here, y_i is the measured value and y_i is the predicted textural value, and n is the number of samples. IQ is the difference between the third and the first quartile of the sample.

3. Results

3.1. Descriptive Statistics of the Soil Texture

The descriptive statistics of clay and sand contents in the three purple soils are shown in **Figure 1** and **Figure 2**. The clay content of all soil samples ranged from 4.40% to 27.12%, and the sand content ranged from 0.34% to 36.57%. Combined with the USDA soil texture classification triangle, it could conclude that soil samples had three different textural classes (silt, silt loam, and silty clay loam). Among them, the skewness and kurtosis values of the clay content varied from -0.17 to 1.88, roughly fitting the usual normal distribution. The mean and standard deviation of clay content in neutral purple soil was 19.18% and 2.14%, respectively, and the difference in clay content was larger than that of calcareous and acidic purple soils. While the mean and standard deviation of sand content in acidic purple soil was 18.26% and 8.71%, respectively, and the difference in sand content was larger than that of the other two soils. In addition, the clay and sand contents of the three purple soils fell within the range of 10% to 100%, indicating moderate variation.



Figure 1. Descriptive statistics of soil clay content. SD: standard deviation, CV: variation coefficient.



Figure 2. Descriptive statistics of soil sand content. SD: standard deviation, CV: variation coefficient.

3.2. Correlation of Transformed Reflectance Spectra and Soil Texture

The correlation between the spectral reflectance of the three purple soils and their clay content in different bands is shown in **Figure 3**. In the original spectrum, the clay content of the calcareous and neutral purple soils had a low correlation with spectral reflectance, whereas the clay content of the acidic purple soil had a very significant correlation with the spectral reflectance, and all bands were belonging to the sensitive bands. After the CR, R', and R'' transformation, the correlation of the original reflectance increased obviously. The sensitive bands corresponding to the highest correlation coefficients after the CR and R' transformation were concentrated at 2417 - 2437 nm and 433 - 566 nm, respectively, and the absolute value of the correlation coefficients were 0.72 and 0.83.

The correlation between the spectral reflectance of the three purple soils and their various transformations with the sand content over a wavelength range of 350 - 2500 nm is shown in **Figure 4**. The original reflectance of calcareous and acidic purple soils with their sand content showed a very significant correlation, its sensitive bands were located at 879 - 1364 nm and 350 - 2500 nm, respectively. However, the correlation between the original reflectance of neutral purple soil and the sand content was not significant. After the CR transformation, the spectral reflectance with the sand content of calcareous purple soil had the highest correlation coefficient of 0.79, and the sensitive bands changed to 2417 - 2437 nm. But the correlation coefficient between the spectral reflectance of the R' transformation and sand content was not as good as that of the CR transformation, and the correlation coefficient the R' transformation had the highest correlation coefficient with sand content, with a value up to 0.85. Its sensitive bands were concentrated between 448 and 467 nm.



Figure 3. Correlation between the spectral reflectance and clay content under different spectral transformations for calcareous, neutral, and acidic purple soils.



Figure 4. Correlation between sand content and the spectral reflectance under different spectral transformations for the calcareous, neutral, and acidic purple soils.

3.3. Model Calibration and Validation

3.3.1. SMLR and PLSR Modeling

Table 1 summarizes the SMLR estimation results of the clay and sand contents

	Coil Trmo	Calibration		Validation			
	son Type		R_c^2	RMSEC	R_v^2	RMSEV	RPIQ
Clay	Calcareous	R	0.123	1.775	0.011	1.707	0.369
		CR	0.519	1.315	0.473	1.156	0.543
		R'	0.417	1.448	0.195	2.114	0.611
		R''	0.528	1.329	0.561	1.099	1.191
	Neutral	R	0.038	5.307	0.228	5.415	0.121
		CR	0.310	1.609	0.361	1.299	0.611
		R'	0.510	1.305	0.475	1.168	1.360
		<i>R</i> ″	0.542	1.311	0.503	1.182	1.600
	Acidic	R	0.541	1.437	0.515	1.260	1.052
		CR	0.351	1.390	0.228	1.629	0.020
		R'	0.707	1.407	0.710	1.075	1.560
		R''	0.828	0.754	0.822	0.868	2.135
	Calcareous	R	0.194	1.519	0.071	1.515	0.518
		CR	0.605	1.063	0.650	0.939	1.185
Sand		R'	0.314	1.402	0.208	1.356	0.511
		R''	0.660	0.962	0.682	0.932	1.526
	Neutral	R	0.325	1.403	0.327	1.995	0.397
		CR	0.414	1.340	0.610	1.683	0.842
		R'	0.422	1.331	0.495	1.940	0.694
		R''	0.555	1.167	0.649	1.721	1.113
	Acidic	R	0.442	6.579	0.207	7.760	1.038
		CR	0.353	7.081	0.283	7.170	1.164
		R'	0.492	6.274	0.330	6.943	1.451
		R''	0.729	4.583	0.675	4.772	1.871

Table 1. SMLR model results for clay and sand contents in the three purple soils after applying the four transformations.

Notes: *R*: reflectivity; CR: continuum-removal; *R*': first-order differential reflectivity; *R*'': second-order differential reflectivity; RMSE: root mean square error; RPIQ: the ratio of performance to inter-quartile distance.

in the three purple soils under the 10 bands with the highest correlations. Based on the four spectral indices, except that the R^2 of clay content in acidic purple soil after the CR transformation was lower than that of the original reflectance, the R^2 of clay and sand contents in other soil types after the CR, R' and R'' transformation was higher than that of the original reflectance. And the modeling effect of clay and sand contents in the three purple soils after the R'' transformation was better than that of the other two transformations. Among them, the acidic purple soil after the R'' transformation had the best prediction effect with the R^2 of 0.822, 0.675, the RMSEV of 0.868, 4.772, and the RPIQ of 2.135, 1.871, respectively. By contrast, the modeling effect of clay content was better than that of sand content. In short, the clay content in acidic purple soil after the R'' transformation had the best modeling effect.

The Calibration and validation sets of the PLSR model were obtained by extracting the first 10 characteristic bands with the highest correlation coefficient for the clay and sand contents for the three purple soils (**Table 2**). The modeling effect of the original reflectance after mathematical transformations was better

Table 2. PLSR model results for clay and sand contents in the three purple soils after four transformations.

			Calibration		Validation		
	Soil Type	Transformation -	R_c^2	RMSEC	R_{ν}^2	RMSEV	RPIQ
Clay	Calcareous	R	0.123	1.775	0.011	1.708	0.366
		CR	0.489	1.270	0.476	1.050	0.763
		R'	0.502	1.287	0.165	2.868	0.707
		R''	0.516	1.235	0.500	1.278	0.959
	Neutral	R	0.034	1.900	0.014	1.531	0.257
		CR	0.275	1.609	0.356	1.298	0.491
		R'	0.521	1.257	0.414	1.233	1.349
		R''	0.543	1.227	0.510	1.072	1.929
	Acidic	R	0.530	1.437	0.513	1.262	1.049
		CR	0.445	1.523	0.451	1.415	0.373
		R'	0.719	1.097	0.728	1.036	1.735
		R''	0.829	0.823	0.832	0.851	2.056
	Calcareous	R	0.162	1.511	0.072	1.742	0.415
Sand		CR	0.579	1.029	0.584	0.905	1.666
		R'	0.342	1.322	0.060	1.545	0.394
		<i>R</i> ″	0.670	0.911	0.630	0.966	1.714
	Neutral	R	0.317	1.411	0.328	2.001	0.402
		CR	0.380	1.342	0.609	1.686	0.233
		R'	0.432	1.272	0.439	1.971	0.652
		<i>R</i> ″	0.558	1.120	0.658	1.874	1.030
	Acidic	R	0.424	6.604	0.198	7.823	1.020
		CR	0.338	7.082	0.283	7.170	1.162
		R'	0.514	5.771	0.371	6.701	1.313
		R''	0.736	4.251	0.744	4.290	2.440

Notes: *R*: reflectivity; CR: continuum-removal; *R*! first-order differential reflectivity; *R*". second-order differential reflectivity; RMSE: root mean square error; RPIQ: the ratio of performance to inter-quartile distance.

than that of the original reflectance, especially the model prediction effect after the R''transformation was better than other transformations. The modeling results of clay and sand contents in acidic purple soil were better than those of neutral and calcareous purple soils, with the modeling set of $R_c^2 > 0.736$, and the verification set of $R_v^2 > 0.744$. Among them, the predicted value of clay content was higher than that of sand content (the R_v^2 was 0.832, the RMSEV was 0.851, and the RPIQ was 2.056), and the modeling and validation sets of the sand content had a larger RMSE of 4.772, indicating that there is a large deviation between the predicted value and the measured value. In conclusion, the clay content of acidic purple soil by the R'' transformation has the best model prediction ability.

3.3.2. BPNN Modeling

As mentioned above, soil texture and spectral reflectance had the highest correlation coefficient and the best modeling effect after the R'' transformation. Thus, the spectral data from the 350 - 2500 mm waveband transformed by R'' was used as the input layer, and tansig and purelin were employed as the transfer function of the hidden and output layers, respectively. In addition, the learning rate, the maximum training time, and the expected error of the model were set to 0.02, 1000, and 0.0001, respectively. The results of the modeling and verification sets are shown in **Table 3**, the clay and sand contents of the three purple soils have high prediction accuracy ($R^2 > 0.5$), indicating that the model effect after the R''transformation has a generalization ability and a high fitting degree. Among them, the modeling effect of acidic purple soil was better than that of calcareous and neutral purple soils ($R^2 > 0.7$, RPIQ > 1.9), but the sand content of acidic purple soil had a larger RMSE of 4.114. In contrast, the BPNN model of clay content in acidic purple soil had a better model prediction effect.

4. Discussion

4.1. Effect of Mathematical Transformation

There is a certain correlation between soil texture and spectral reflectance. The

	Coil True o	Calibration		Validation			
	Son Type	R_c^2	RMSEC	R_{ν}^2	RMSEV	RPIQ	
Clay	Calcareous	0.608	1.467	0.554	1.714	1.138	
	Neutral	0.547	2.612	0.501	2.801	0.603	
	Acidic	0.812	1.045	0.807	1.081	1.976	
Sand	Calcareous	0.672	1.294	0.661	1.387	1.496	
	Neutral	0.583	2.418	0.591	2.405	1.018	
	Acidic	0.751	4.150	0.763	4.114	2.509	

Table 3. BPNN modeling results of clay and sand contents for the three purple soils following the R''transformation.

Notes: RMSE: root mean square error; RPIQ: the ratio of performance to inter-quartile distance. bands respond more sensitive to the texture with the increase of correlation, which is conducive to the determination of characteristic bands. To improve the correlation between soil texture and spectral reflectance, the original spectral reflectance of mathematical transformation is necessary. In this study, the correlation was improved after the CR, R' and R'' transformation of the original reflectance, especially the correlation coefficient was increased to 0.8 by the R'' transformation. This is consistent with the highest correlation between soil moisture content and spectral reflectance of the three purple soils after the R'' transformation [36]. These results demonstrated that the second derivative extremely eliminates the baseline effect and enhances the micro-absorption characteristics [37]. Mathematical transformation can also improve the accuracy of the model [32]. In this study, the prediction model after the R'' transformation had the best effect ($R^2 > 0.5$), indicating that the model after the R'' transformation has a high fitting degree and generalization ability.

4.2. The Performance of the Three Models

The linear (SMLR and PLSR) and the nonlinear (BPNN) prediction models of the texture (clay and sand contents) in the three purple soils were established and the adaptability of the three models was evaluated by calculating R^2 , RMSE, and RPIQ. For the prediction value of calcareous purple soil, the R^2 of the SMLR model was slightly higher than that of the PLSR and BPNN models ($R^2 = 0.682$), and it had the lowest RMSE (RMSE = 0.932). While for the predicted value of neutral and acidic purple soils, the prediction effect of the PLSR model was better than that of SMLR and BPNN. This may be because the three purple soils have different physical and chemical properties, resulting in different spectral characteristics, thus there are different prediction models in the three purple soils [38].

Moreover, it is found that the linear model has a better fitting degree than the nonlinear model in this study. This may be due to the fact that the linear model used the characteristic bands for modeling, while the nonlinear model needed the training of big data, and the full-band was used for modeling. The full-band (350 - 2500 nm) exists certain redundant information (including noise and repeated variables), and these variables in the construction of soil physical and chemical properties' estimation model may reduce the prediction accuracy and reliability of the model [39]. Therefore, compared with the nonlinear model, if the linear model can overcome the multicollinearity problem directly by applying statistical rotation to simulate the relationship [40], coupled with its simple operation, and good stability. It is recommended to use the linear method to quickly estimate the soil texture of the three purple soils. However, compared with the traditional measurement methods, the linear model has the advantages of less time-consuming and low cost, but it has a low prediction accuracy in estimating soil texture.

4.3. Limitations of This Study

Soil spectral reflectance is affected by texture, moisture, and other physical and

chemical factors. The different percentage contents of clay, sand, and silt will lead to the scattering effect in the process of testing the spectrum, and the correlation information between the soil components to be tested and the spectrum is covered [41]. Therefore, in this study, soil samples were ground and screened to ensure the uniformity of soil particle size. Moreover, the influence of soil moisture on the soil surface spectral reflectance is complex, which makes it difficult to directly interpret the spectral parameters corresponding to soil characteristics from the obtained soil current spectral data [42]. Hereby, in this study, the effect of soil moisture on spectral reflectance was eliminated by using air-dried soil, and the clay content of air-dried soil can be better estimated than wet soil [43].

The deficiency lies in that the estimation accuracy of soil texture in the fields could be affected negatively by several factors such as spectral mixture, atmospheric conditions, and spatial variability of soil moisture content, whereas our soil spectrum was collected indoors [44] [45]. In particular, there may be a large variation in soil moisture in the field, which will reduce the prediction accuracy of soil texture [46]. Therefore, to apply the models of this study to the field, they need to be further optimized. In the future research, it is better to focus on integration between spectra obtained in the field or in the laboratory, and spectra from airborne or satellite imagery [47] [48], and evaluate the external parameter orthogonalisation (EPO) method that can project all the soil spectra orthogonal to the space of unwanted variation (*i.e.* moisture) to eliminate the effects of moisture on the spectra [49].

5. Conclusion

In the study, different pre-processing methods were utilized for the estimation of the soil texture of the three purple soils based on Vis-NIR spectroscopy. The soil samples had three textural classes (including silt, silt loam, and silty clay loam). In addition, the CR, R', and R'' pre-processing methods had strong positive influence on the improvement of correlation analysis and the performance of the three models, among which the R'' had the greatest performance. On this basis, the SMLR, PLSR, and BPNN models were established, and the accuracy among them was compared. The results show that the SMLR model is more suitable for estimating the texture of calcareous purple soil, and the PLSR model is the optimum approach for neutral and acidic purple soils. The SMLR and PLSR models also provide better estimations than the BPNN model in the calibration verification step. But these model theories were studied under well-controlled laboratory conditions, future research will consider extending models for field and large-scale applications.

Acknowledgements

This research was supported by the Chongqing Talent Program (CQYC201905009), the Science Fund for Distinguished Young Scholars of Chongqing (cstc2019jcyjjqX0025), Sichuan Science and Technology Program (2020YJ0202,2021YFS0288), and Technology Research Project of Chongqing Municipal Education Commission (JQN201800531).

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Daniels, W.L. (2016) The Nature and Properties of Soils, 15th Edition Ray R. Weil and Nyle C. Brady. Pearson Press, Upper Saddle River NJ, 2017. 1086 p. \$164.80. ISBN-10: 0-13-325448-8; ISBN-13: 978-0-13-325448-8. Also Available as eText for \$67.99. Soil Science Society of America Journal, 80, 1428. https://doi.org/10.2136/sssaj2016.0005br
- [2] Greve, M.H., Kheir, R.B., Greve, M.B. and Bocher, P.K. (2012) Quantifying the Ability of Environmental Parameters to Predict Soil Texture Fractions Using Regression-Tree Model with GIS and LIDAR Data: The Case Study of Denmark. *Ecological Indicators*, 18, 1-10. https://doi.org/10.1016/j.ecolind.2011.10.006
- Brians, A., Debbiel, S. and Sachaj, M. (2009) The Impact of Soil Structure on the Establishment of Winter Wheat (*Triticum aestivum*). *European Journal of Agronomy*, 30, 243-257. <u>https://doi.org/10.1016/j.eja.2008.12.002</u>
- Shahriari, Delbari, M., Afrasiab, P. and Pahlavan-Rad, M.R. (2019) Predicting Regional Spatial Distribution of Soil Texture in Floodplains Using Remote Sensing Data: A Case of Southeastern Iran. *Catena*, **182**, Article ID: 104149. https://doi.org/10.1016/j.catena.2019.104149
- [5] Lakshmi, V., Albertson, J. and Schaake, J. (2001) Land Surface Hydrology, Meteorology and Climate: Observations and Modeling. American Geophysical Union, Washington DC, 77-94. <u>https://doi.org/10.1029/WS003</u>
- [6] Zhang, N., Zhang, D.L., Li, L.X. and Qu, Z.Y. (2014) Establishment and Evaluation of Model for Predicting Soil Texture Based on Hyperspectral Data—Case Study of Jiefangzha Irrigation Area in Hetao Irrigation District. *Journal of Arid Land Resources* and Environment, 28, 67-72. (In Chinese)
- [7] X.D., Yao, X.J., Yun, H., Xu, Y.H. and Nan, Y.M. (2022) Rapid Detection of Surface Soil Texture Based on Noise Estimation. *China Resources Comprehensive Utilization*, 40, 38-40. (In Chinese)
- [8] Beuselinck, L., Govers, G., Poesen, J., Degraer, G. and Froyen, L. (1998) Grain-Size Analysis by Laser Diffractometry: Comparison with the Sieve-Pipette Method. *Catena*, 32, 193-208. <u>https://doi.org/10.1016/S0341-8162(98)00051-4</u>
- [9] Douglas, R.K., Nawar, S., Alamar, M.C., Mouazen, A.M. and Coulon, F. (2018) Rapid Prediction of Total Petroleum Hydrocarbons Concentration in Contaminated Soil Using Vis-NIR Spectroscopy and Regression Techniques. *Science of the Total Environment*, 616-617, 147-155. <u>https://doi.org/10.1016/j.scitotenv.2017.10.323</u>
- [10] Mohanty, B., Gupta, A. and Das, B.S. (2016) Estimation of Weathering Indices Using Spectral Reflectance over Visible to Mid-Infrared Regions. *Geoderma*, 26, 111-119. <u>https://doi.org/10.1016/j.geoderma.2015.11.030</u>
- [11] Stenberg, B., Mouazen, A.M., Wetterlind, J. and Rossel, R.A.V. (2010) Visible and Near-Infrared Spectroscopy in Soil Science. *Advances in Agronomy*, **107**, 163-215. <u>https://doi.org/10.1016/S0065-2113(10)07005-7</u>
- [12] Buddenbaum, H. and Steffens, M. (2012) The Effects of Spectral Pretreatments on

Chemometric Analyses of Soil Profiles Using Laboratory Imaging Spectroscopy. *Applied and Environmental Soil Science*, **2012**, Article ID: 274903. https://doi.org/10.1155/2012/274903

- [13] Ba, Y., Liu, J., Han, J. and Zhang, X. (2019) Application of Vis-NIR Spectroscopy for Determining the Content of Organic Matter in Saline-Alkali Soils. *Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy*, **229**, Article ID: 117863. <u>https://doi.org/10.1016/j.saa.2019.117863</u>
- [14] Nawar, S., Buddenbaum, H., Hill, J., Kozak, J. and Mouazen A.M. (2016) Estimating the Soil Clay Content and Organic Matter by Means of Different Calibration Methods of Vis-NIR Diffuse Reflectance Spectroscopy. *Soil and Tillage Research*, 155, 510-522. https://doi.org/10.1016/j.still.2015.07.021
- [15] Jia, X., Kuo, B. and Crawford, M.M. (2013) Feature Mining for Hyperspectral Image Classification. *Proceedings of the IEEE*, **101**, 676-697. https://doi.org/10.1109/JPROC.2012.2229082
- [16] Bin, J., Fan, W., Zhou, J.H., Li, X. and Liang, Y.Z. (2017) Application of Intelligent Optimization Algorithms to Wavelength Selection of Near-Infrared Spectroscopy. *Spectroscopy and Spectral Analysis*, 37, 95-102.
- [17] Jahan, N. and Gan, T.Y. (2011) Modeling the Vegetation-Climate Relationship in a Boreal Mixed Wood Forest of Alberta Using Normalized Difference and Enhanced Vegetation Indices. *International Journal of Remote Sensing*, **32**, 313-335. <u>https://doi.org/10.1080/01431160903464146</u>
- [18] Thompson, J.A., Pena-Yewtukhiw, E.M. and Grove, J.H. (2006) Soil-Landscape Modeling across a Physiographic Region: Topographic Patterns and Model Transportability. *Geoderma*, **133**, 57-70. <u>https://doi.org/10.1016/j.geoderma.2006.03.037</u>
- [19] Xu, C., Zeng, W.Z., Huang, J.S., Wu, J.W. and Willem, V.L. (2016) Prediction of Soil Moisture Content and Soil Salt Concentration from Hyperspectral Laboratory and Field Data. *Remote Sensing*, 8, 42-62. <u>https://doi.org/10.3390/rs8010042</u>
- [20] Recena, R., Fernández-Cabanás, V.M. and Delgado, A. (2019) Soil Fertility Assessment by Vis-NIR Spectroscopy: Predicting Soil Functioning Rather than Availability Indices. *Geoderma*, 37, 368-374. <u>https://doi.org/10.1016/j.geoderma.2018.09.049</u>
- [21] Filla, V.A., Coelho, A.P., Ferroni, A.D., Bahia, A.S.R.D.S. and José, M.J. (2021) Estimation of Clay Content by Magnetic Susceptibility in Tropical Soils Using Linear and Nonlinear Models. *Geoderma*, 403, Article ID: 115371. https://doi.org/10.1016/j.geoderma.2021.115371
- [22] Wang, Z., Xin, Y. and Ren, W. (2015) Nonlinear Structural Model Updating Based on Instantaneous Frequencies and Amplitudes of the Decomposed Dynamic Responses. *Engineering Structures*, **100**, 189-200. <u>https://doi.org/10.1016/j.engstruct.2015.06.002</u>
- Breiman, L. (2001) Statistical Modeling: The Two Cultures. *Statistical Science*, 16, 199-215. <u>https://doi.org/10.1214/ss/1009213726</u>
- [24] Lu, Y.Y., Liu, F., Zhao, Y.G., Song, X.D. and Zhang, G.L. (2019) An Integrated Method of Selecting Environmental Covariates for Predictive Soil Depth Mapping. *Journal of Integrative Agriculture*, 18, 301-315. https://doi.org/10.1016/S2095-3119(18)61936-7
- [25] Malone, B.P., Mcbratney, A.B., Minasny, B. and Laslett, G.M. (2009) Mapping Continuous Depth Functions of Soil Carbon Storage and Available Water Capacity. *Geoderma*, **154**, 138-152. <u>https://doi.org/10.1016/j.geoderma.2009.10.007</u>
- [26] Zhong, S.Q., Zhong, M., Wei, C.F., Zhang, W.H. and Hu, F.N. (2016) Shear Strength

Features of Soils Developed from Purple Clay Rock and Containing Less than Two-Millimeter Rock Fragments. *Journal of Mountain Science*, **13**, 1464-1480. https://doi.org/10.1007/s11629-015-3524-8

- [27] Qin, W., Zuo, C.Q., Yan, Q.H., Wang, Z.Y., Du, P.F. and Yan, N. (2015) Regularity of Soil Erosion during Single Rainfall on Red Soil Bare Slope Land. *Transactions of the Chinese Society of Agricultural Engineering*, **31**, 124-132. (In Chinese)
- [28] Xu, L., Zhang, D., Xiang, Y., Chen, F. and Huang, T. (2020) Runoff and Sediment Yield Characteristics of Slope Farmland in Purple Soil Region of Lower Jinsha River under Different Cultivation Measures. *Journal of Mountain Science*, 38, 851-860.
- [29] Ma, C., M., Shen, G.R., Wang, Z.J. and Wang, Z. (2015) Analysis of Spectral Characteristics for Different Soil Particle Sizes. *Chinese Journal of Soil Science*, 46, 292-298. (In Chinese)
- [30] Metzger, K., Daly, K., Ward, M. and Zhang, C. (2020) Mid-Infrared Spectroscopy as an Alternative to Laboratory Extraction for the Determination of Lime Requirement in Tillage Soils. *Geoderma*, **364**, Article ID: 114171. https://doi.org/10.1016/j.geoderma.2020.114171
- [31] Guan, H., Jia, K., Zhang, Z. and Xin, M.A. (2015) Research on Remote Sensing Monitoring Model of Soil Salinization Based on Spectrum Characteristic Analysis. *Remote Sensing*, 27, 100-104.
- [32] Tao, P.F., Wang, J.H., Li, Z.Z., Zhou, P., Yang, J.J. and Gao, F.Q. (2020) Research of Soil Nutrient Content Inversion Model Based on Hyperspectral Data. *Geology and resources*, 29, 68-75. (In Chinese)
- [33] Smith, G.S., Workman, C.I., Protas, H., Su, Y., Savonenko, A., Kuwabara, H., Gould, N.F., Kraut, M., Jin, H.J. and Nandi, A. (2021) Positron Emission Tomography Imaging of Serotonin Degeneration and Beta-Amyloid Deposition in Late-Life Depression Evaluated with Multi-Modal Partial Least Squares. *Translational Psychiatry*, 11, Article No. 473. <u>https://doi.org/10.1038/s41398-021-01539-9</u>
- [34] Yu, D. and Deng, L. (2011) Deep Learning and Its Applications to Signal and Information Processing. *IEEE Signal Processing Magazine*, 28, 145-154. https://doi.org/10.1109/MSP.2010.939038
- [35] Saikia, P., Baruah, R.D., Singh, S.K. and Chaudhuri, P.K. (2020) Artificial Neural Networks in the Domain of Reservoir Characterization: A Review from Shallow to Deep Models. *Computers & Geosciences*, **135**, Article ID: 104357. https://doi.org/10.1016/j.cageo.2019.104357
- [36] Gou, Y., Wei, J., Li, J.L., Han, C., Tu, Q.Y. and Liu, C.H. (2020) Estimating Purple-Soil Moisture Content Using Vis-NIR Spectroscopy. *Journal of Mountain Science*, 17, 2214-2223. https://doi.org/10.1007/s11629-019-5848-2
- [37] Kemper, T. and Sommer, S. (2022) Estimate of Heavy Metal Contamination in Soils after a Mining Accident Using Reflectance Spectroscopy. *Environmental Science & Technology*, 36, 2742-2747. <u>https://doi.org/10.1021/es015747j</u>
- [38] Marco, N., Antoine, S., Bas, W., Erick, B., Ronald, V. and Luca, M. (2015) Soil Spectroscopy: An Opportunity to Be Seized. *Global Change Biology*, 21, 10-11. https://doi.org/10.1111/gcb.12632
- [39] Yu, L., Hong, Y.S., Zhou, Y., Zhu, Q., Xu, N., Li, J.Y. and Nie, Y. (2016) Wavelength Variable Selection Methods for Estimation of Soil Organic Matter Content Using Hyperspectral Technique. *Transactions of the Chinese Society of Agricultural Engineering*, **32**, 95-102. (In Chinese)
- [40] Shi, T.Z., Cui, L.J., Wang, J.J., Fei, T., Chen, Y.Y. and Wu, G.F. (2013) Comparison of

Multivariate Methods for Estimating Soil Total Nitrogen with Visible/Near-Infrared Spectroscopy. *Plant Soil*, **366**, 363-375. <u>https://doi.org/10.1007/s11104-012-1436-8</u>

- [41] Wang, H.J., Liu F., Cui, J., Ma, L. and Yunger. J.A. (2019) Fitting Model of Soil Total Nitrogen Content in Different Soil Particle Sizes Using Hyperspectral Analysis. *Transactions of the Chinese Society for Agricultural Machinery*, **50**, 195-204. (In Chinese)
- [42] Mouazen, M., Baerdemaeker, J.D. and Ramon, H. (2006) Effect of Wavelength Range on the Measurement Accuracy of Some Selected Soil Properties Using Visual-Near Infrared Spectroscopy. *Journal of Near Infrared Spectroscopy*, 14, 189-199. https://doi.org/10.1255/jnirs.614
- [43] Chang, C.W., Laird, D.A. and Hurburgh, C.R. (2005) Influence of Soil Moisture on Near-Infrared Reflectance Spectroscopic Measurement of Soil Properties. *Soil Science*, 170, 244-255. <u>https://doi.org/10.1097/00010694-200504000-00003</u>
- [44] Ben-Dor, E., Chabrillat, S., Dematte, M.A.J., Taylor, G.R., Hill, J., Whiting, M.L. and Sommer, S. (2009) Using Imaging Spectroscopy to Study Soil Properties. *Remote Sensing of Environment*, **113**, 538-555. <u>https://doi.org/10.1016/j.rse.2008.09.019</u>
- [45] Gomez, C.C., Viscarra, R.A. and McBratney, A.B. (2008) Soil Organic Carbon Prediction by Hyperspectral Remote Sensing and Field VIS-NIR Spectroscopy: An Australian Case Study. *Geoderma*, 146, 403-411. https://doi.org/10.1016/j.geoderma.2008.06.011
- [46] Waiser, T.H., Morgan, C.L.S., Brown, D.J., Hallmark, C.T., Waiser, T.H., Morgan, C.L.S. and Hallmark, C.T. (2007) *In Situ* Characterization of Soil Clay Content with Visible Near-Infrared Diffuse Reflectance Spectroscopy. *Soil Science Society of America Journal*, **71**, 389-396. https://doi.org/10.2136/sssaj2006.0211
- [47] Lu, P., Wang, L., Niu, Z., Li, L. and Zhang, W. (2013) Prediction of Soil Properties Using Laboratory VIS-NIR Spectroscopy and Hyperion Imagery. *Journal of Geochemical Exploration*, **132**, 26-33. <u>https://doi.org/10.1016/j.gexpl0.2013.04.003</u>
- [48] Nawar, S., Buddenbaum, H. and Hill, J. (2015) Digital Mapping of Soil Properties Using Multivariate Statistical Analysis and ASTER Data in an Arid Region. *Remote Sensing*, 7, 1181-1205. <u>https://doi.org/10.3390/rs70201181</u>
- [49] Minasny, B., Mcbratney, A.B., Bellon-Maurel, V., Roger, J.M., Gobrecht, A., Ferrand, L. and Joalland, S. (2011) Removing the Effect of Soil Moisture from NIR Diffuse Reflectance Spectra for the Prediction of Soil Organic Carbon. *Geoderma*, 167-168, 118-124. <u>https://doi.org/10.1016/j.geoderma.2011.09.008</u>