

Hyperspectral Estimation of Total Nitrogen Content in Orchard Soil Based on Successive Projections Algorithm

Zizhen Li ¹, Xicun Zhu ^{1, 2†}, Zhongyu Tian ¹, Meixuan Li ¹, Yufeng Peng ¹, Xinyang Yu ¹, Xueyuan Bai ¹, Ling Wang ¹, Yuanmao Jiang ³

1. College of Resources and Environment, Shandong Agricultural University, Tai'an 271018, China.

2. National Engineering Laboratory for Efficient Utilization of Soil and Fertilizer Resources, College of Resources and Environment, Shandong Agricultural University, Tai'an 271018, China

3. College of Horticulture Science and Engineering, Shandong Agricultural University, National Apple Engineering and Technology Research Center, Tai'an 271018, China.

†Email: zxc@sdau.edu.cn

Abstract

Soil total nitrogen content is an important factor to affect the growth of fruit trees. The objective of the paper is to propose a successive projections algorithm for hyperspectral estimation of soil total nitrogen content in orchard. A total of 60 soil samples were collected from apple orchard, Guanli Town, Qixia City, Shandong Province. The spectral reflectivity of the samples was measured by a ASD Field-Spec4.0 spectrometer in the laboratory. The total nitrogen content of the samples was measured by a Kjeldahl apparatus. The spectral data were preprocessed by Savitzky-Golay smoothing and first-order derivative transformation. The sensitive bands were screened by successive projections algorithm. Partial least squares regression (PLSR) model, random forest (RF) model and support vector machine regression (SVR) model were established respectively, and the accuracy of the models was tested and compared. The sensitive wavelengths selected by successive projections algorithm were 849, 2100, 1305 and 2427nm, the determination coefficients (R^2) of PLSR, SVR and RF models were 0.822, 0.724 and 0.683, and the relative percent deviation (RPD) were 2.329, 1.940 and 1.702, respectively. PLSR model has the highest accuracy and can play a good role in the estimation of total nitrogen content in orchard soil; the accuracy of SVR and RF models is lower than that of PLSR models, but the estimated results reach the usable level.

Keywords: Orchard Soil; Total Nitrogen Content; Successive Projections Algorithm; Hyperspectral

1 INTRODUCTION

Soil total nitrogen content is an important information to diagnose the level of soil fertility. How to obtain soil total nitrogen content data rapidly and nondestructive to carry out real-time and effective monitoring of soil total nitrogen content is an urgent demand of agricultural production. Traditional measurement methods are mostly chemical measurement in the laboratory. Although more accurate soil total nitrogen content data can be obtained, there are problems such as labor cost, high cost and poor timeliness¹. In recent years, hyperspectral technology, as an efficient and convenient analytical method, has shown a good application prospect in the estimation of nutrient elements. Hyperspectral remote sensing has the characteristics of high spectral resolution, strong band continuity and large spectral information, etc., and has achieved good results in the laboratory inversion of soil nutrients. Its rapid development provides a new technical means for real-time and efficient acquisition of soil total nitrogen content information²⁻³. Dalal et al.⁴ found that there was a close correlation between soil spectrum and nitrogen at 1700 to 2100nm, and the reflectance at 1744, 1870 and 2052nm was used to effectively estimate the total nitrogen content of various soil types. Sun et al.⁵ took the grey-brown alluvial soil in northern China as the research object, discussed the

feasibility and possibility of real-time analysis of soil parameters by NIR spectroscopy, the 1888 and 2262nm wavelengths were selected to establish multiple linear regression model to estimate soil total nitrogen content, and the results proved that the method was feasible. When Reeves et al.⁶ estimated the total nitrogen content of soil samples in the field, the reflectance of the spectrum was treated by first-order differentiation, the partial least squares regression model was established to estimate the total nitrogen content with 1100 to 2300nm as the variable, and a good effect was obtained. Ben-Dor et al.⁷ estimated six soil indexes, including clay content, surface area, cation exchange capacity, moisture, total nitrogen and carbonate content by analyzing the near-infrared band spectra of soil. Songchao Chen et al.⁸ took samples of different soil types as research objects and verified the applicability of local weighted regression method combined with VIS-NIR spectroscopy in the estimation of total nitrogen content in large samples and large areas. Took black soil as the research object, Yanli Lu et al.⁹ enhanced the relationship between spectral reflectance and soil total nitrogen content through different transformations and derivative processing of spectral reflectance, combined with principal component regression analysis, they predicted the soil total nitrogen content in black soil with the first-order derivative of the reflectance logarithm of 556, 1642 and 2491nm wavelengths. Xiuying Liu et al.¹⁰ used correlation analysis combined with partial least squares method to establish hyperspectral estimation model for soil total nitrogen content in loess soil, and achieved good results. Juanjuan Zhang et al.¹¹ studied the quantitative relationship between total nitrogen content and hyperspectral reflectance of five major types of soil in central and eastern China, based on the first-order derivative spectra of spectral reflectance smoothing by Norris filter in the region of 500 to 900nm and 1350 to 1490nm, a highly accurate estimation model of total nitrogen in soil was established. At present, there have been a large number of studies on the estimation of nitrogen content of various soil types. However, there are few reports on the estimation of total nitrogen content of orchard soil by successive projections algorithm. In this study, based on the spectral reflectance and total nitrogen content of orchard soil samples collected in Qixia City, Shandong Province, the successive projections algorithm was used to screen out the sensitive bands, then the partial least squares regression model, random forest model and support vector machine regression model for estimating soil total nitrogen content were established, and the accuracy of the models was tested and compared, in order to construct an effective estimation model of soil total nitrogen content in apple orchard, and to provide certain theoretical basis and technical support for rapidly obtaining soil total nitrogen content in apple orchard.

2 MATERIALS AND METHODS

2.1 Overview of the Study Area

Qixia City in Shandong Province (120°33' -121 ° 15'E, 37°05' -37 ° 32'N) is located in the center of Jiaodong Peninsula. The territory is dominated by mountains and hills, with an average altitude of 178m. It belongs to temperate monsoon sub-humid climate, with an annual average temperature of 11.3°C and an average ground temperature of 13.6°C. The average annual precipitation is 640 to 846mm, and the average frost-free period is 207 days. The soil type in the high mountain area is generally brown soil, the middle and lower mountain area is typical brown soil, the lowest part is moisture brown soil and moisture soil. Qixia City has a long history of apple planting and is known as "China's Apple Town".

2.2 Sample Collection and Determination

Samples were collected on March 9, 2019 at Apple Orchard, Guanli Town, Qixia City, Shandong Province. In 12 orchards, 60 soil samples were collected with soil drills at a sampling depth of 0-40cm according to the five-point sampling method. The samples were dried in the laboratory for the determination of spectral reflectivity and total nitrogen content. The spectrum of soil samples was measured by ASD Field-Spce4.0 spectrometer. The band range of the instrument was 350-2500nm, in the band range of 350-1000 nm, the spectral sampling interval is 1.4nm and the resolution is 3nm; in the band range of 1001-2500 nm, the spectral sampling interval is 2nm and the resolution is 8nm. The soil samples were placed in a vessel with a diameter of 10cm and a depth of 1.5cm, and the surface of the soil sample was smoothed down. The sensor probe was placed at a distance of 10cm perpendicular to the surface of the soil sample. Before each collection of spectral data, the instrument was calibrated. The observation of each sample was repeated for 10 times, and the average value was taken as the soil spectrum of the sample. The total

nitrogen content of the samples was measured by a Kjeldahl apparatus.

2.3 Data Preprocessing

1) Savitzky-Golay Smoothing

The spectral data of soil were derived under the software ViewSpecPro6.0. Due to the influence of illumination and other external factors, the soil spectral curve was mixed with noise signals, in order to reduce the influence of background noise and improve the accuracy of spectral data acquisition, it is necessary to denoise soil spectral data. Savitzky-Golay smoothing (SG smoothing) algorithm was used to denoise the soil spectral data in the band of 350-2500nm. SG smoothing can not only remove noise signals but also retain effective information in spectral data, which can play a good role in spectral processing, its basic principle is to achieve the least square fitting of the internal elements of the window through a sliding window based on polynomial to obtain the smoothing value of the elements at the center of the sliding window and make it a smooth curve¹², in essence, it is a kind of weighted average method, emphasizing the central role of the central point. SG smoothing of spectral data was carried out under the software Matlab2014a.

2) First-order Derivative Transformation

The first-order derivative of soil spectral data after denoising was calculated, the first-order derivative transformation can determine the inflection point of the spectral curve and the wavelength position of the maximum and minimum reflectivity, reduce the baseline drift and background interference, significantly improve the sensitivity between the soil total nitrogen content and the spectrum, and reduce the sensitivity to noise at the same time, which is convenient for the screening of sensitive bands, and shows a good effect in practical application¹³⁻¹⁵.

3) Screening of Sensitive Wavelength

Sensitive wavelengths were selected by successive projections algorithm (SPA) for the construction of a model for predicting total nitrogen content in soil. SPA is a forward variable selection method that minimizes the collinearity of vector space, it was first proposed by Bregman in 1965 when solving the convex feasible problem. SPA projects specific wavelengths onto other wavelengths, uses projection analysis to compare the size of projection vectors, takes the maximum wavelength of projection vector as the wavelength to be selected, and finally selects the characteristic wavelength from the wavelength to be selected based on the correction model¹⁶. SPA can eliminate the redundant information in the original spectral matrix to the greatest extent, and the effect is obvious when the characteristic wavelength is screened in the spectral analysis¹⁷.

4) Division of Training Set and Verification Set

Equidistant sampling was used to divide the samples into modeling training set and validation set. Equidistant sampling is a method that the total samples are arranged in a certain order and then selected at a fixed interval¹⁸, which can make the sample uniform distribution and reasonable extraction effect. All the samples were arranged in descending order according to their total nitrogen content, starting with the sample with the highest total nitrogen content, the sample was selected in a ratio of 2:1, and one sample from every two samples was selected into the verification set. The training set contained a total of 40 samples, and the verification set contained 20 samples.

Standard deviation was used as the evaluation standard for the partition results of training set and validation set, and the smaller the standard deviation was, the smaller the dispersion degree of sample extraction was. The standard deviations of the training set and the verification set were 0.028 and 0.024, respectively, indicating that the distribution of the samples extracted was relatively uniform, and the estimation model has good universality. The partition results are shown in Table 1.

TABLE 1 STATISTICS OF MEASURED TOTAL NITROGEN CONTENT (%) OF TRAINING SET AND VERIFICATION SET

	Number of samples	Maximum (%)	Minimum (%)	Average (%)	Standard deviation
Training set	40	0.153	0.020	0.059	0.029
Verification set	20	0.110	0.019	0.056	0.025

2.4 Model Construction and Precision Analysis

1) Partial Least Squares Regression

Partial Least Squares Regression (PSLR) is a statistical method which integrates the basic abilities of multivariate linear Regression analysis, canonical correlation analysis and principal component analysis, its basic principle is to find the best function match of a set of data by minimizing the sum of squares of errors, which can be used for regression modeling under the condition of serious multiple correlation of independent variables. Partial least squares regression can maximize the use of spectral information in the process of model construction and achieve better estimated results, the regression equation is shown as follows:

$$y=a_0+a_1x_1+a_2x_2+a_3x_3+\dots+a_nx_n \quad (1)$$

In the regression equation, a_0 refers to the intercept of the regression coefficient; a_i refers to the regression coefficient; x_i refers to an independent variable 1-n¹⁹.

2) Support Vector Machine Regression

Support Vector Machine (SVM) is a data mining method based on VC dimension theory, structural risk minimization principle and statistical learning theory, it can deal with many problems such as small sample size, dimensional disaster, local minimum point, over-learning and various kinds of nonlinear classification and regression. Mechanism of SVM is to find a optimal separating hyperplane that could satisfy the requirement of classification, in ensuring the classification accuracy at the same time, the blank space on either side of the hyperplane to maximize, implementation of linear can score according to the optimal classification, when carries on the regression algorithms according to the needs, through a nonlinear mapping, the data is mapped to a high-dimensional feature space, Linear regression is carried out in this space²⁰⁻²¹.

3) Random Forest

Random Forest (RF) is a kind of machine learning algorithm based on classification tree proposed by Breiman et al in 2001. RF extracts multiple samples from the sample data set through bootstrap resampling method, and constructs a decision tree for each bootstrap sample, then combined with the prediction of multiple decision trees, the predicted value with higher repetition times is selected as the final result through voting. RF has a good tolerance to outliers and noise, and is not prone to over-fitting, so it can play a good training effect on both linear and nonlinear data²².

4) Test of Model Accuracy

In this study, the determination coefficient (R^2), root mean square error (RMSE) and relative percent deviation (RPD) were selected as the evaluation criteria for the accuracy test of the estimation model. According to these indicators, the closer R^2 is to 1, the higher estimation accuracy of the model is; the smaller RMSE is, the better the stability of the model is; and RPD represents the estimation ability of the model, it means the estimated results of the model are not of reference value when $RPD < 1.0$; $1.0 \leq RPD < 1.4$ means the estimation ability of the model is poor; $1.4 \leq RPD < 1.8$ means the estimation ability of the model could reach the usable level; $1.8 \leq RPD < 2.0$ means the model could make a relatively accurate estimation, $2.0 \leq RPD < 2.5$ indicates that the model has good predictive ability, while RPD higher than 2.5 indicates that the model has excellent predictive ability²³.

3 RESULTS AND ANALYSIS

3.1 Hyperspectral Analysis of Soil

The soil spectral curve is shown in Figure 1. The soil spectral data were divided into 3 intervals according to the total nitrogen content, and the average value of the spectral data of each interval was taken as the spectral curve of the total nitrogen content range within this region. From the perspective of spectral reflectance in the full band range of 350 to 2500nm, the spectral curves of the total soil samples showed an upward trend on the whole, which was a typical spectral curve of soil characteristics. As the total nitrogen content increased, the reflectance decreased in the whole band range, the figure roughly showed that the higher the total nitrogen content was, the lower the curve position was. The reason is that soil total nitrogen and organic matter are closely related, and the higher the content

of soil organic matter, the lower the reflectance was²⁴. In the visible band, the spectral reflectance of soil was low, and the overall rising trend was fast, in the red-edge band range, the spectral reflectance of soil increased rapidly. After entering the near-infrared band, the spectral reflectance of the soil was higher than that of the visible band, but the reflectance rose at a slower rate than before, in addition, there were absorption valleys near 1420, 1930 and 2230nm, and the absorption valleys at 1930nm were more obvious. Soil clay particles have the water-holding property that can still absorb water even after drying at room temperature, and the water absorption valleys are mainly caused by water absorption in the air and residual water in the soil. The spectral reflectance of soil decreased gradually from 2250 nm to 2500nm.

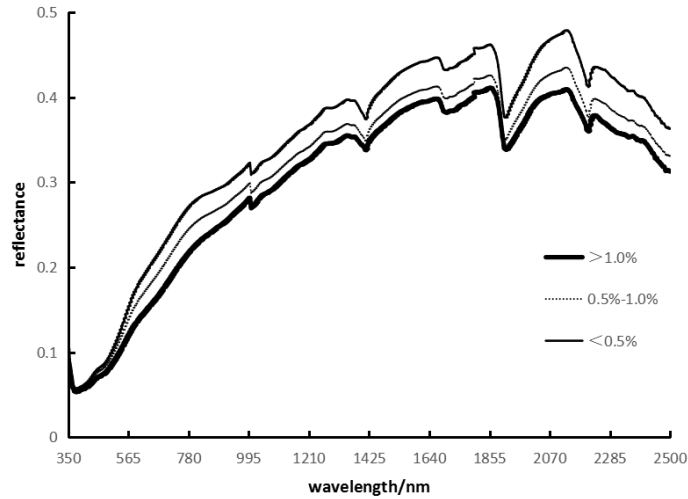


FIG. 1 SPECTRAL REFLECTANCE OF SOIL TOTAL NITROGEN CONTENT

3.2 Screening Results of Successive Projections Algorithm

The spectral band range of soil was 350-2500nm, SPA calculation was carried out under the software MATLAB2014a, and the SPA calculation process was as follows:

Step 1: $x_{k(0)}$ and N were respectively defined as the initial iteration vector and the number of variables to be extracted, and the number of column variables of the spectral matrix was defined as J .

Step 2: before the start of the first iteration ($n=1$), selected column j of the spectral matrix and assigned the j th column of the correction spectral array to x_j , denoted as $x_{k(0)}$, $k(0)=j$;

Step 3: denoted the set of unselected column vector positions as s , $s=\{j, 1 \leq j \leq J, j \notin \{k(0), \dots, k(N-1)\}\}$, calculated the projection vectors of x_j onto the vectors in s , $Px_j = x_j - (x_j^T - x_{k(n-1)})x_{k(n-1)}(x_{k(n-1)}^T - x_{k(n-1)})^{-1}$, $j \in s$;

Step 4: defined the sequence number of the maximum projection $k(n)=\arg(\max \|Px_j\|, j \in s)$;

Step 5: defined $x_j = Px_j$, $j \in s$;

Step 6: $n=n+1$, if $n < N$, go back to step 3 to continue the projection.

In the above steps, the selection of $k(0)$ and N was a key step before the extraction of variables, the range of $k(0)$ was $1-j$, due to the existence of collinearity between variables, the selection of N was generally not greater than the number of samples in the correction set. For $k(0)$ and N selected each time, a multiple linear regression model was established to obtain the root mean square error of prediction, and the minimum error indicates that the selected $k(0)$ and N were optimal^{17, 25-26}. The sensitive wavelengths screened by SPA were 849, 2100, 1305 and 2427nm, and the preprocessed spectral data of soil samples on these wavelengths were used for the establishment of the model and accuracy test.

3.3 Construction of Soil Total Nitrogen Content Estimation Model

During the establishment of the estimation model, the data of denoising and first-order derivative transformation on

the sensitive bands of 849, 2100, 1305 and 2427nm were taken as the independent variable, and the soil total nitrogen content was taken as the dependent variable.

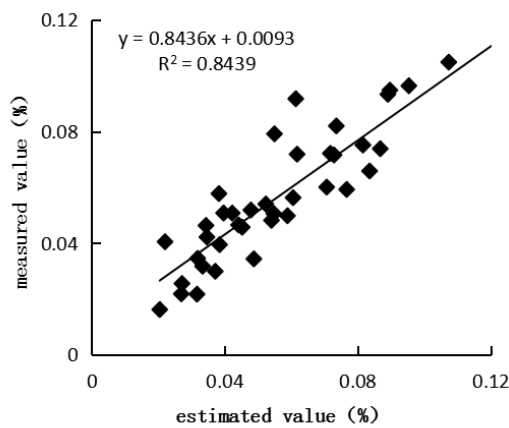
The data were imported into partial least squares regression model for regression operation, and the PLSR regression equation was obtained. PLSR regression equation is shown as:

$$y=0.06+455.31x_1-273.08x_2-201.81x_3+16.41x_4 \quad (2)$$

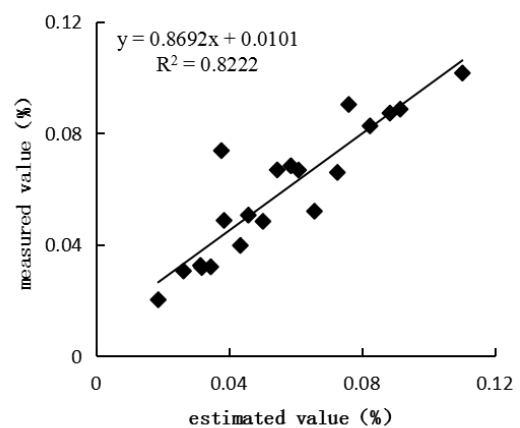
In the formula, x_1 , x_2 , x_3 and x_4 were spectral data at 849, 2100, 1305 and 2427nm after denoising and first-order derivative transformation, respectively. The estimated results of PLSR model are shown in Table 2 and Figure 2. The R^2 of PLSR model was 0.822, the RPD was 2.329, and the RMSE was 0.011, indicating that PLSR model could make a good estimation of soil total nitrogen content.

TABLE 2 ESTIMATED RESULTS OF PLSR MODEL

	Maximum (%)	Minimum (%)	Average (%)	RMSE
Fitted value of training set	0.131	0.016	0.059	0.011
Estimated value of verification set	0.102	0.020	0.059	0.011



a: fitted value of training set



b: estimated value of verification set

FIG. 2 ESTIMATED RESULTS OF PLSR MODEL

When establishing the Support Vector Regression (SVR) model, the type of kernel Function was Radial Basis Function (RBF), and other specific parameters were set as shown in the table:

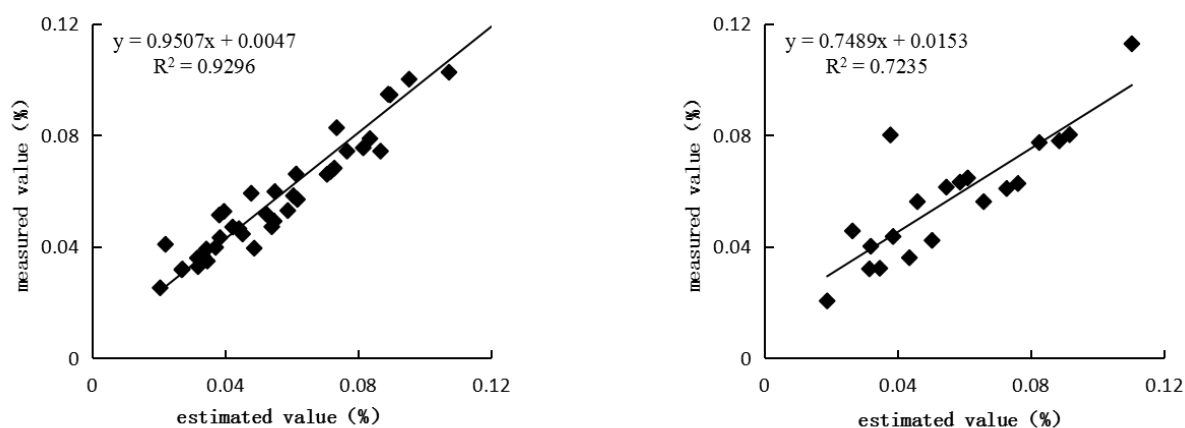
TABLE 3 PARAMETER SETTINGS OF SVR MODEL

Degree	Gamma	Coef0	Nu	Epsilon	Cashesize	Cost	Shrinking	Prob	P
3	0.5	0.001	0.5	0.001	100	1	1	1	0.01

The sample data of training set and verification set were substituted into the SVR model for regression operation, the estimated results of the SVR model are shown in Table 4 and Figure 3. The R^2 of the SVR model was 0.724, the RPD was 1.940, and the RMSE was 0.013, indicating that the SVR model could make a relatively accurate estimation of the soil total nitrogen content.

TABLE 4 ESTIMATED RESULTS OF SVR MODEL

	Maximum (%)	Minimum (%)	Average (%)	RMSE
Fitted value of training set	0.149	0.025	0.061	0.007
Estimated value of verification set	0.113	0.021	0.057	0.013



a: fitted value of training set

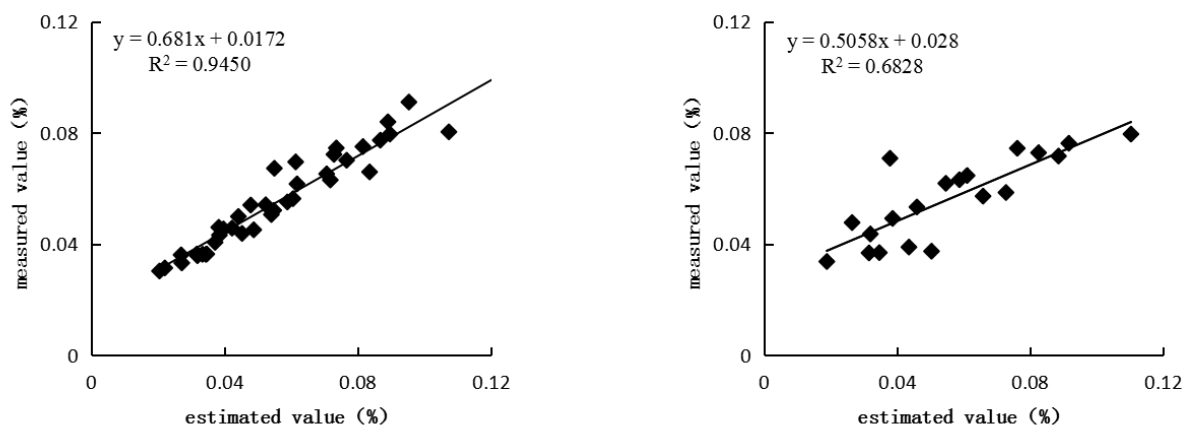
b: estimated value of verification set

FIG. 3 ESTIMATED RESULTS OF SVR MODEL

When the sample data of training set and verification set were substituted into the RF model for regression operation, the proportion of training sample was set as 50%, the number of trees of random forest was set as 300, and the number of variables randomly selected at nodes was set as 2, the estimated results of the RF model are shown in Table 5 and Figure 4. The R^2 of RF model was 0.683, RPD was 1.702, RMSE was 0.015. The estimated results of soil total nitrogen content were lower than PLSR and SVR models, but the estimation effect also reached the usable level.

TABLE 5 ESTIMATED RESULTS OF RF MODEL

	Maximum (%)	Minimum (%)	Average (%)	RMSE
Fitted value of training set	0.118	0.030	0.058	0.010
Estimated value of verification set	0.080	0.034	0.056	0.015



a: fitted value of training set

b: estimated value of verification set

FIG. 4 ESTIMATED RESULTS OF RF MODEL

4 DISCUSSION

In this study, the soil of apple orchard was taken as the research object, and a total of 60 soil samples were collected, all of which were topsoil. About 90% of total nitrogen in the surface soil is organic nitrogen, which is highly closely related to the content of organic matter in the soil, the higher the content of organic matter in the soil, the lower the spectral reflectance²⁴. Hyperspectral data can reflect the spectral characteristics of ground objects in a very detailed way, in the analysis of hyperspectral characteristics of total nitrogen content of this study, the spectral reflectance decreased with the increase of soil total nitrogen content, which reflected this.

The spectral data were preprocessed by SG smoothing and first-order derivative transformation, the first-order derivative transformation can clearly reflect the subtle changes of the spectrum, significantly improved the sensitivity between the total nitrogen content and hyperspectral data, and enhanced the relationship between spectral reflectance and soil total nitrogen content. Three commonly used machine learning algorithms were used to construct the estimation model of soil total nitrogen content, and the best estimation model of soil total nitrogen content was selected through precision test and comparison. The sensitivity of different soil types to total nitrogen content is significantly different, which is caused by different soil composition and soil texture, in this study, sensitive wavelengths were screened by successive projections algorithm. Xiuying Liu et al.¹⁰ concluded that the range of general sensitive wavelengths for predicting total nitrogen in different types of soil was 500-900, 1400-1600 and 2050-2250nm based on many researches, the sensitive wavelengths screened out of this study were 849, 2100, 1305 and 2427nm, which were relatively consistent with the range of induction mentioned above. By eliminating a large number of bands which are not related to total nitrogen content, the estimation model can be simplified and the accuracy can be improved. Equidistant sampling was used to divide the training set and the verification set, which can make the samples evenly distributed and ensure that the chemical values of the verification set samples were within the range of the chemical values of the training set samples, which was convenient to improve the effect of machine learning, however, the equidistant sampling method requires samplers to have a certain understanding of the overall structure, and invalid samples should be eliminated before modeling to ensure the availability of samples¹⁸. PLSR algorithm fully considers the comprehensive effect of single factors on the predicted value, and can provide a "many-to-many" linear regression modeling method, when the number of two groups of variables is large and there are multiple correlations, and the number of observed data is small, PLSR model has obvious advantages. In terms of regression rate, PLSR algorithm is faster than the traditional multiple regression algorithm, with more relaxed requirements for small samples and more obvious advantages when dealing with a small number of samples, which is reflected in the establishment of model and accuracy test. SVR model has good generalization and generalization ability and is less affected by the number of observed samples. In this study, it also showed relatively accurate estimation effect on soil total nitrogen content. RF algorithm has good tolerance to outliers and noise and can avoid overfitting, however, it is more suitable to handle large data samples with a larger number and more information, when the number of observation samples is medium or small, the advantage of sample noise resistance will be reduced²⁰. The performance of RF model in this study was relatively inferior to that of PLSR and SVR model.

Based on the hyperspectral data of orchard soil in Qixia City, Shandong Province, the PLSR, SVR and RF estimation models of soil total nitrogen content were established based on the successive projections algorithm, which provided a reference for hyperspectral monitoring of soil total nitrogen content in orchard. This study only discussed the effect of this estimation method on a small number of samples, and the differences of soil composition and texture will lead to great differences in the sensitivity of total nitrogen content, whether this method can be used to effectively estimate a large range of areas and a large number of samples still needs further exploration. In the future studies, a larger range of spectral data and a larger number of spectral data can be used to further improve the stability and universality of the estimation model.

5 CONCLUSION

In this study, SG smoothing was used to denoise the spectrum of soil samples, and the first-order derivative transformation was carried out, then equidistant sampling was used to divide the data into training set and verification set, and the estimation models of total nitrogen content in soil of PLSR, SVR and RF orchard were constructed respectively. The results showed that the R^2 of PLSR model was 0.822 and the RPD of PLSR model was 2.329, the PLSR model had the highest accuracy among the three models, and could play a good role in estimating the soil total nitrogen content of orchard. The R^2 of SVR model was 0.724, and the RPD was 1.940, which was lower than that of PLSR model, but it could also make a relatively accurate estimation of the soil total nitrogen content of orchard. The accuracy of the RF model was relatively low, with R^2 of 0.683 and RPD of 1.702, its estimation effect also reached the usable level. The hyperspectral estimation of total nitrogen content in orchard soil based on successive projections algorithm is feasible, which can provide certain technical support for fast acquisition of total nitrogen content in orchard soil.

ACKNOWLEDGEMENT

This research was supported by the National Key Research and Development Program of China (2017YFE0122500); Shandong major scientific and technological innovation project: Research demonstration and extension of orchard irrigation and fertilization in accurate management (2018CXGC0209); National Natural Science Foundation of China(41671346); Funds of Shandong “Double Tops” Program(SYL2017XTTD02); The Taishan Scholar Assistance Program from Shandong Provincial Government.

REFERENCE

- [1] Meihua Yang, Xiaomin Zhao. Study on soil total N estimation by Vis-NIR spectra with variable selection. *Scientia Agricultura Sinica*, 2014, 47(12): 2374-2383.
- [2] Gangfeng Zhao, Jun Li, Bingfeng Liu, et al. Monitoring model of leaf nitrogen content of winter wheat in guanzhong district by hyperspectral remote sensing. *Journal of Triticeae Crops*, 2012, 32(3): 530-536.
- [3] Lihua Xu, Deti Xie, Chaofu Wei, et al. Prediction of total nitrogen and total phosphorus concentrations in purple soil using hyperspectral data[J]. *Spectroscopy and Spectral Analysis*, 2013, 33(3): 723-727.
- [4] Dalal R C, Henry R J. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry[J]. *Soil Science Society of America Journal*, 1986, 50(1):120-123.
- [5] Sun J Y, Li M Z, Zheng L H, et al. [Real-time analysis of soil moisture, soil organic matter, and soil total nitrogen with NIR spectra]. [J]. *Spectroscopy & Spectral Analysis*, 2006, 26(3):426.
- [6] Reeves I J, Mccarty G. Quantitative analysis of agricultural soils using near infrared reflectance spectroscopy and a fibre-optic probe[J]. *Journal of Near Infrared Spectroscopy*, 2001, 9(1):25.
- [7] Ben-Dor E, Banin A. Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties[J]. *Soil Science Society of America Journal*, 1995, 59(2):364-372.
- [8] Songchao Chen, Lailei Feng, Shuo Li, et al. VIS-NIR spectral inversion for prediction of soil total nitrogen content in laboratory based on locally weighted regression[J]. *Acta Pedologica Sinica*, 2015, 52(2): 312 - 320.
- [9] Yanli Lu, Youlu Bai, Lei Wang, et al. Determination for total nitrogen content in black soil using hyperspectral data[J]. *Transactions of the CSAE*, 2010, 26(1):256-261.
- [10] Xiuying Liu, Li Wang, Qingrui Chang, et al. Prediction of total nitrogen and alkali hydrolysable nitrogen content in loess using hyperspectral data based on correlation analysis and partial least squares regression[J]. *Chinese Journal of Applied Ecology*, 2015, 26(7): 2107-2114.
- [11] Juanjuan Zhang, Yongchao Tian, Xia Yao, et al. Estimation of total nitrogen content in soil based on hyperspectral method[J]. *Journal of Nature Resources*, 2011, 26(5):881-890.
- [12] Yibing Lu, Wenqing Liu, Yujun Zhang, et al. An Adaptive Hierarchical Savitzky-Golay Spectral Filtering Algorithm and Its Application. *Spectroscopy and Spectral Analysis*[J]. 2019, 39(09):2657-2663.
- [13] Juanjuan Zhang, Yongchao Tian, Yan Zhu, et al. Spectral characteristics and estimation of organic matter contents of different soil types[J]. *Scientia Agricultura Sinica*, 2009, 42(9): 3154 - 3163.
- [14] Yan Li, Ranghui Wang, Yanlong Guan, et al. Prediction analysis of soil total nitrogen content based on hyperspectral[J]. *Remote Sensing Technology and Application*, 2017, 32(1):337-344.
- [15] Cloutis E A. Review Article Hyperspectral geological remote sensing: evaluation of analytical techniques[J]. *International Journal of Remote Sensing*, 1996, 17(12):2215-2242.
- [16] Bregman L M. Finding the Common Point of Convex Sets by the Method of Successive Projections[J]. *dokl.akad.nauk sssr*, 1965.
- [17] Fei Shen, Yitao Liao, Yibin Ying. Application of successive projections algorithm in quantitative analysis by near infrared spectroscopy[C]// China Agriculture Engineering Society. Commemorate the 30th Anniversary of the Establishment of China Society of Agriculture Engineering and China Agriculture Engineering Society 2009 Annual Conference (CSAE2009) Proceedings. Taigu, shanxi: China Agriculture Engineering Society, 2009.
- [18] Huansan Zhao, Xicun Zhu, Cheng Li, et al. Improving the Accuracy of the Hyperspectral Model for Apple Canopy Water Content Prediction using the Equidistant Sampling Method[J]. *Scientific Reports*, 2017, 7(1):11192.
- [19] Haikuan Feng, Fuqin Yang, Guijun Yang, et al. Estimation of chlorophyll content in apple leaves base on spectral feature parameters[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2018, 34(6): 182-188.
- [20] Shifei Ding, Bingjuan Qi, Hongyan Tan. A Survey of Support Vector Machine Theory and Algorithms[J]. *Journal of University of*

Electronic Science and Technology of China, 2011,40(001):2-10.

- [21] Dingcheng Wang, Tingjian Fang, Lifu Gao, et al. Support vector machines regression on-line modelling and its application[J]. Control and Decision, 2003, 18(1): 89-91, 95
- [22] Xinhai LI. Using “random forest” for classification and regression [J]. Chinese Journal of Applied Entomology, 2013, 50(4) :1190-1197.
- [23] Sims D A, Gamon J A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages[J]. Remote Sensing of Environment, 2002, 81(2-3):337-354.
- [24] Binbin Xu. Research on the spectral reflectance of soil section [J]. Soils, 2000, 32(6):281-287.
- [25] Hongzhi Gao, Qipeng Lu, Haiquan Ding, et al. Choice of characteristic near-infrared wavelengths for soil total nitrogen based on successive projection algorithm[J]. Spectroscopy and Spectral Analysis, 2009, 29(11):2951-2954.
- [26] Bin Chen, Xianglong Meng, Hao Wang. Application of successive projections algorithm in optimizing near infrared spectroscopic calibration model[J]. Journal of Instrumental Analysis, 2007, 26(1): 66 - 69.
- [27] Chunlan Zhang, Guijun Yang, Heli Li, et al. Remote sensing inversion of leaf area index of winter wheat based on random forest algorithm[J]. Scientia Agricultura Sinica, 2018, 51(5): 855-867.

AUTHORS



Li Zizhen, he is working on his Master degree in Agricultural Information major at College of Resources and Environment at Shandong Agricultural University.



Zhu Xicun, he is a professor at College of Resources and Environment at Shandong Agricultural University. His research direction is application of agricultural remote sensing and remote sensing technology.



Tian Zhongyu, he is working on his Master degree in Agricultural Information major at College of Resources and Environment at Shandong Agricultural University.



Li Meixuan, she is working on his Master degree in Agricultural Information major at College of Resources and Environment at Shandong Agricultural University.



Peng Yufeng, she is working on her Master degree in Cartography and Geographical Information Engineering major at College of Resources and Environment at Shandong Agricultural University.



Yu Xinyang, he is a lecturer at College of Resources and Environment at Shandong Agricultural University. His research direction is land use and remote sensing monitoring.



Bai Xueyuan, he is working on his Master degree in Land Resources Management major at College of Resources and Environment at Shandong Agricultural University.



Wang Ling, she is an associate professor at College of Resources and Environment at Shandong Agricultural University. Her research direction is quantitative remote sensing for agriculture.



Jiang Yuanmao, he is a professor at College of Horticulture Science and Engineering at Shandong Agricultural University. His research direction is mineral nutrition and fertilization of fruit trees.