Monitoring Soil Nitrate Nitrogen Based on Hyperspectral Data in the Apple Orchards

Yu Wei1, Xicun Zhu1,2, Cheng Li1, Lizhen Cheng1, Ling Wang1, Gengxing Zhao1, Yuanmao Jiang3

1College of Resources and Environment, Shandong Agricultural University, Tai’an, China
2Key Laboratory of Agricultural Ecology and Environment, Shandong Agricultural University, Tai’an, China
3College of Horticulture Science and Engineering, Shandong Agricultural University, Tai’an, China

Email: wy18763890707@163.com

Abstract
This paper is aimed to monitor the soil nitrate nitrogen content in the apple orchards rapidly, accurately and in real time by making full use of the effective information of soil spectra. The 96 air-dried soil samples of the apple orchards in Qixia county, Yantai city, Shandong province were used as the data source. Spectral measurements of soil samples were carried out by ASD Fieldspec 3 in the darkroom, and the content of the soil nitrate nitrogen was determined by chemical method. Then the hyperspectral reflectance of soil samples were preprocessed by Multivariate Scatter Correction (MSC) and First Derivative (FD), the correlation analysis was carried out with the soil nitrate nitrogen content. The sensitive wavelength of soil nitrate nitrogen was screened. Finally, the Support Vector Machine (SVM) model for the soil nitrate nitrogen content was established. The results showed that the selected sensitive wavelengths were 617 nm, 760 nm, 1239 nm, 1442 nm, 1535 nm, 1695 nm, 1776 nm, 1907 nm and 2088 nm. Hyperspectral monitoring model was established by SVM, in which the prediction set R² was 0.959, RMSE was 0.281, RPD was 3.835; the correction set R² was 0.822, RMSE was 0.392, RPD was 2.037. The SVM model could be used to monitor the soil nitrate content accurately.

Keywords
Hyperspectrum, Nitrate Nitrogen Content, Support Vector Machine, Sensitive Wavelength

1. Introduction
Soil nitrate nitrogen is the best nitrogen source for crops; it has non-volatility and can be directly absorbed by plants. It can promote the absorption of potassium, calcium, magnesium and other cations, and limit the absorption of harmful sub-
stances. It plays an important role in the nutrient absorption process of crops. In the north of China, the soil is mainly dominated by nitrate nitrogen, and its content reflects the short-term of the soil nitrogen supply directly. Therefore, the real-time monitoring of the soil nitrate nitrogen content has important practical significance in crop growth status, crop yield and soil pollution prevention. The traditional chemistry method of soil nitrate nitrogen—ultraviolet spectrophotometry, which has high measuring precision, is time-consuming and laborious and is difficult to analyze the results in a timely manner. In recent years, the emerging hyperspectral remote sensing technology has the advantages of high resolution, strong band continuity and large spectral information, which provides technical support for rapid, non-destructive and accurate monitoring of the soil nitrate nitrogen content [1].

Using hyperspectral technology combined with different pretreatment methods and modeling methods, the soil nitrogen content (mainly focused on soil total nitrogen) was monitored by domestic and foreign scholars. Yet monitoring soil nitrate nitrogen content is rare. In early 21st century, Chang et al. [2] used the near-infrared spectroscopy modeling and analysis technology to monitor the total nitrogen content of 744 soil samples from the United States, with an accuracy of 0.85 and a good estimate of total nitrogen. Confalonieri et al. [3] obtained near-infrared spectroscopy data of different soil types using Fourier transform near-infrared spectrometer, then the content of soil component including the nitrogen content was well predicted and analyzed. Lee et al. [4] used PLSR and SMLR to model 165 surface soil samples and 697 soil profile samples from the United States corn growing areas. The results indicated that the characteristic band of soil nitrogen was about 510 nm, and the content of total nitrogen in soil was successfully predicted. Genot et al. [5] used the improved PLSR model to establish the local model of the soil total nitrogen in Wallonia of Belgium with an accuracy of more than 0.9 and the model had a good prediction accuracy. The content of total nitrogen in different soil types was predicted by the domestic scholars. Yujianfei et al. [6] estimated the soil total nitrogen content using the first derivative pretreatment method combined with the partial least-squares regression model. The model accuracy was 0.88. Lu Yanli et al. [7] estimated the total nitrogen content in black soil using the first derivative of the logarithm of soil reflectance and the normalized spectral index and $R^2$ reached 0.863, 0.829 respectively. Zhang Jianjuan et al. [8] studied the quantitative relationship between total nitrogen and hyperspectral reflectance of five types of soils in central and eastern China. Based on this, the partial least squares (PLS) model, BP neural network (BPNN) model were established. The models were of high precision and could be used to estimate the soil total nitrogen content. Different soil hyperspectral data pretreatment methods and different modeling methods will affect the prediction accuracy of the model. Peng Jie et al. [9] studied the four main types of cultivated soils in Hunan Province, analyzed the correlations between different spectral indices and the soil total nitrogen content and established the linear regression model and the stepwise multiple regression model.
The results showed that the accuracy of the multiple stepwise regression model established by \((1/\log)\)' was the highest. Yang Chao et al. [10] carried out different pretreatments of soil near-infrared spectral data and established the PLSR model to monitor soil total nitrogen content. The results showed that different pre-treatment methods would greatly affect the accuracy of the model and the smoothing +MSC+OSC method was the best. Viscarra Rossel et al. [11] applied the discrete wavelet transform to the spectra of 876 soil samples in Australia. The models of stochastic forest (RF), support vector machine (SVM) and artificial neural network (ANN) had been established, and the results showed that the SVM model performed the best. Ji Wenjun et al. [12] predicted the SOM content of 441 paddy soils in Zhejiang Province using PLSR, RF, SVM, ANN and PLSR-ANN models, the conclusion also showed that the SVM model had the best effect.

The above studies showed that it was feasible to monitor the soil nitrogen content using hyperspectral technology. Previous studies mainly focused on the quantitative relationship between soil total nitrogen content and hyperspectral reflectance, but monitoring of the soil nitrate nitrogen content is rarely reported. The main reason is that the content of soil nitrate nitrogen is lower than that of total nitrogen, so it is difficult to monitor the soil nitrate nitrogen by using hyperspectral technique. Based on the correlation analysis between hyperspectral data of the apple orchards soils in Qixia and nitrate nitrogen content measured by chemical experiments, the Support Vector Machine (SVM) model was established, in order to provide theoretical basis and technical support for monitoring the soil nitrate nitrogen content by using hyperspectral techniques.

2. Material and Methods

2.1. Study Area

The study area was located in Qixia City (120°33′-121°15′, 37°05′-37°32′), which is situated in the center of Jiaodong Peninsula (Figure 1). The total area is 201,600 hm², and the area of orchards is 43,300 hm². Qixia City has four distinct seasons and plenty of sunshine. Its climate is warm and semi-humid monsoon. The annual average temperature is 11.3°C, the annual average rainfall is about 650 mm, the frost-free period is 207 d, and the total sunshine duration is 2690 h. Qixia belongs to mountain hilly terrain. The parent material is mostly acid rock, granite, gneiss. The soil type is mainly brown soil.

2.2. Soil Samples Collection and Preparation

The samples were collected from 32 orchards in 3 streets and 12 townships in Qixia city (Figure 1). The sampling time was 20 - 23 October 2010 and the sampling depth was 0 - 20 cm. Three trees were selected randomly from each orchard, and each tree was taken as a sampling point. Soil samples were collected in the four directions vertically from the east to the west of the crown edge of the selected fruit tree, and then mixed. A total of 96 soil samples were obtained. The soil samples were naturally air-dried, crushed, decontaminated, ground, sieved
through a 1 mm sieve and mixed thoroughly. 100 g soil samples were taken by quartile method, respectively [13], into a dish (2 cm) for the measurement of soil nitrate nitrogen content and soil spectrum. The collected samples were randomly divided into two groups. The prediction set (72) was used for modeling, and the calibration set (24) was used for model checking.

2.3. Soil Spectral Date and Nitrate Nitrogen Content Determination

2.3.1. Soil Spectral Date Measurements

Soil spectral data were measured by ASD FieldSpec 3. The spectral range is 350 ~ 2500 nm. The sampling interval was 1.4 nm and the spectral resolution was 3.5 nm during 350 ~ 1000 nm; the sampling interval was 2 nm and the spectral resolution was 10 nm during 1000 ~ 2500 nm. When the spectrum was output, the resampling interval was 1 nm and the total number of output bands was 2151 [14]. Spectral measurements were carried out in a dark room. The sample dishes were filled with soil samples and placed on black rubber mats with a reflectivity of 0 approximately and a thickness of 3 cm. The light source used a 50 w halogen lamp, the probe’s field of view was 25°, the light source incidence angle was 45°, the distance of the probe to the soil surface was 15 cm and the light source to the soil center distance was 30 cm. During the measurement, the standard white plate was calibrated every 20 minutes, the sample was rotated three times and rotated about 90°. The soil sample curve was obtained in four directions. Each

Figure 1. The geographical location of Qixia and distribution of sample points.
soil sample was measured 10 times and the arithmetic mean was taken as the reflection spectrum data of the soil sample [15]. The edge region (350 - 399 nm and 2451 - 2500 nm) with larger spectral noise was removed, and only the band (400 - 2450 nm) was reserved for the follow-up study.

2.3.2. Soil Nitrate Nitrogen Content Analyses

After the data collection was completed, the content of soil nitrate nitrogen was determined by ultraviolet spectrophotometry [16]. 10 g air-dried soil samples were accurately weighed and placed in a 100 ml stoppered Erlenmeyer flask adding with 0.1 g of CaSO₄ and 50 ml of distilled water. In the (25 ± 1)°C conditions, the flask was shaken on the shaker for 10 min at a rate of 150 r/min and placed 10 min. Then the supernatant was filtered with dry filter paper. 10 mL of the filtrate was pipetted into a 25 mL cuvette. The absorbance was measured at 220 nm and 275 nm respectively by the UV-2102 ultraviolet-visible spectrophotometer. The mass fraction of nitratenitrogen in each soil sample were calculated according to the calibration standard curve (correlation coefficient was 0.995). The results of the determination of the nitrate nitrogen content in 96 soil samples were shown in Table 1.

2.4. Soil Spectral Date Preprocessing

In order to reduce the spectral data affected by laboratory optical field variation and sample in homogeneity, spectral data were analyzed by Multivariate Scatter Correction (MSC) combined with First Derivative (FD), Savitzky-Golay Smoothing combined with First Derivative(FD), and Savitzky-Golay smoothing, Multiple Scattering Correction combined with First Derivative (FD) processing. Through the comparative analysis of the late prediction results, we can see that the multispectral correction plus First Derivative (FD) is the best. It is showed that MSC combined with FD transform method had the best effect.

2.4.1. Multivariate Scatter Correction (MSC)

MSC was proposed by Martens et al. [17], is a data processing method which is commonly used in a multi-band calibration modeling at this stage. MSC is mainly used to eliminate scattering effects caused by the uneven distribution of particles and particle size and eliminate the phenomenon of baseline shift and migration caused by scattering of near-infrared spectroscopy, thereby enhancing the Signal to Noise Ratio of the original absorbance spectrum and enhancing the spectral absorption information associated with component content [18].

The algorithm of MSC was as follows [19]:

1) Calculate the average spectrum of the calibration set samples $x$ (1 × m) (ideal spectrum);

Table 1. Statistics of nitrate nitrogen content of soil samples.

<table>
<thead>
<tr>
<th>Number of samples</th>
<th>Mean/mg·kg⁻¹</th>
<th>Standard deviation/mg·kg⁻¹</th>
<th>Maximum/mg·kg⁻¹</th>
<th>Minimum/mg·kg⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>96</td>
<td>13.938</td>
<td>11.312</td>
<td>52.607</td>
<td>1.321</td>
</tr>
</tbody>
</table>
2) Linear regression of $x_i$ and $\bar{x}$, $x_i = a_i + b_i$, find $a_i$ and $b_i$;  

(1)

3) $x_{i,\text{MSC}} = (x_i - a_i) / b_i$  

(2)

(Among, $\bar{x}$ was the average spectrum of the calibration set samples; $x_i$ was the average value of the $i$th sample spectrum; $x_{i,\text{MSC}}$ was the mean value of the $i$th sample spectrum after MSC treatment; $i = 1, 2, ..., n$, $n$ was the number of the calibration set samples; $l$ was the unit vector of $1 \times m$; $m$ was the number of wavelength points.)

2.4.2. First Derivative (FD)

Differential transformation can eliminate the effect of baseline drift and high frequency noise and amplify the subtle peak-to-valley variations in the original spectrum, making it easier to identify and analyze the spectral curve inflection points and the wavelength positions at the maximum and minimum reflectivities. [20] The low order differential of the spectrum is less sensitive to the effect of noise [21], where FD can effectively remove the influence to the sample spectrum which is caused by the partial linear or similar linear noise spectrum and background [22]. The FD equation [23]:

$$R'(\phi) = \frac{R(\phi_i - \phi_{i-1})}{2\Delta\phi}$$  

(3)

(Among, $\phi_i$ was the wavelength value of each band; $R$ was the first derivative of the reflectance at wavelength $i$; $\Delta\phi$ was the interval of wavelength $\phi_i$ to wavelength $\phi_{i-1}$, where the larger $\Delta\phi$ was, the smoother the spectral curve was. The value of $\Delta\phi$ which was too large will lead to the removal of useful spectral information, which was too small will not achieve the desired effect, The value of $\Delta\phi$ in this paper was set to 2.)

2.5. Model Establishment and Verification

Support Vector Machine (SVM) was used to establish the spectral model of soil nitrate nitrogen content. SVM is a new machine learning method based on statistical learning theory and structural risk minimization [24]. It uses the limited sample information to find the best compromise between model complexity and learning ability and has good generalization ability. It uses the kernel function to reduce the computational complexity of mapping low-dimensional space vector into high-dimensional space, and has been successfully applied in small sample, nonlinear, high-dimensional pattern recognition and other fields [25].

The accuracy and estimation capability of the model were tested by the coefficient of determination ($R^2$), the root mean square error (RMSE) and the relative analysis error (RPD). The larger $R^2$ was, the smaller the RMSE was, the better the estimation accuracy of the prediction model was. In addition, when $\text{RPD} > 2$, the model had excellent predictive ability; when $1.4 < \text{RPD} < 2$, the model could make a rough prediction of the sample; when $\text{RPD} < 1.4$, the model could not estimate the sample [26].
\[ R^2 = \frac{\sum (\hat{y}_i - \bar{y}_i)^2}{\sum (y_i - \bar{y})^2} \]  \hspace{1cm} (4)

\[ RMSE = \sqrt{\frac{1}{n} \sum^n_{i=0} (y_i - \hat{y}_i)^2} \]  \hspace{1cm} (5)

\[ RPD = \frac{STDEV (y_i)}{\sqrt{\frac{1}{n} \sum^n_{i=0} (\hat{y}_i - y_i)^2}} \]  \hspace{1cm} (6)

(Among, \( y_i \) was the measured value; \( \hat{y}_i \) was the predicted value; \( \bar{y} \) was the average of the measured value; \( \bar{\hat{y}}_i \) was the average value of the predicted value; \( n \) was the number of samples.) \cite{27}

3. Results

3.1. Spectral Characteristics of Soil Samples with Different Nitrate Nitrogen Content

All soil samples were divided into four groups averagely according to the level of the soil nitrate nitrogen content, and then all the soil spectral curves in each group were averaged to obtain the spectral reflectance curves under different nitrate nitrogen content (Figure 2). It could be seen from Figure 2 that the variation trend of the soil spectral curves of different nitrate nitrogen content was basically similar. In the range of 400 nm to 2450 nm, there were three distinct water absorption valleys (1400 nm, 1900 nm and 2200 nm), which may be related to the water content in the apple orchards soils \cite{28}. In the visible range, the reflectance of soil samples was low, but the reflectance increased rapidly with the increase of the wavelength. In the near-infrared range, the spectral curve began to flatten. The spectral curve began to show an overall downward trend after 2200 nm, which reflected the typical soil spectral characteristics. With the increase of the nitrate nitrogen content, the general reflectance of soil spectral reflectance was decreasing, which was close to the conclusion that spectral reflectance was decreasing with the increase of organic matter content in previous stu-

**Figure 2.** The spectral reflectance of soil samples with different nitrate nitrogen content.
dies [12]. However, the spectral reflectance of the soil was abnormally high in the range of 10.835 mg·kg$^{-1}$ to 18.097 mg·kg$^{-1}$, which did not follow the general trend of decreasing, indicating that when the soil nitrate nitrogen content in this range, the spectral reflectance of the soil may be affected by other factors, thus affecting the accuracy of the nitratenitrogen spectral prediction model.

### 3.2. Soil Nitrate Nitrogen Sensitive Spectral Band Selection

Figure 3 showed the correlation between the soil nitrate nitrogen content and the original reflectance and the correlation between the soil nitrate nitrogen content and the reflectance after MSC. It could be seen from Figure 3, the coefficient after MSC had been significantly improved, eliminating the scattering effects of the uneven of particle and particle size effectively.

The FD treatment was carried out on the spectral reflectance after MSC, to obtain the correlation between the soil nitrate nitrogen and the spectral reflectance after the transformation, as shown in Figure 4. According to the maximum principle of correlation coefficient, the sensitive wavelengths of soil nitrate nitrogen were selected, which were 617 nm, 760 nm, 1239 nm, 1442 nm, 1535 nm, 1695 nm, 1776 nm, 1907 nm and 2088 nm.

### 3.3. Establishment and Verification of Spectral Inversion Model for Soil Nitrate Content

The nonlinear SVM model was established using the data after MSC combined with FD. The radial basis function (RBF) was chosen as kernel function of the SVM model, and the penalty coefficients C and kernel coefficients RBF which had a great influence on the result were determined by several tests. The rest of the parameters adopted the default values. After several tests, when the C value was set to 10 and gamma value was set to 0.4, the effect was the best. The $R^2$ of the prediction set of the SVM model was as high as 0.9585, the RMSE was 0.281,
Figure 4. Correlation of soil nitrate nitrogen content with reflectance after MSC and FD.

and the RPD was 3.835 (more than 2), which indicated that the model had a good prediction accuracy.

In order to make the effect of model prediction more intuitionistic, the measured values and predicted values of the calibration set of the SVM model were plotted as a scatter plot, as shown in Figure 5.

It could be seen from Figure 5 that the $R^2$ of the calibration set of the SVM model was 0.822, the RMSE was 0.392 and the RPD was 2.037, which showed that the model had good stability and reliability.

4. Discussion

In terms of analyzing the spectral characteristics of soil samples with different nitrate nitrogen content, the spectral reflectance of the soil was abnormally high in the range of 10.835 mg∙kg$^{-1}$ to 18.097 mg∙kg$^{-1}$, which did not follow the general trend of decreasing. The spectral reflectance of the soil may be influenced by other factors when the soil nitrate nitrogen content in this range, and further study was needed. In terms of selecting the sensitive wavelength, a series of sensitive wavelengths (617 nm, 760 nm, 1239 nm, 1442 nm, 1535 nm, 1695 nm, 1776 nm, 1907 nm, 2088 nm) were selected, in which 617 nm, 760 nm, 1239 nm, 1442 nm, 1695 nm 1776 nm were consistent with the best bands of predicting soil total nitrogen in the visible and near-infrared regions by the previous studies [9] [29] [30] [31]. In addition, two sensitive wavelengths of 1535 nm and 1907 nm in the near-infrared long-wave region were newly discovered, which were considered as the important wavelengths of soil nitrogen content distinguishing from soil total nitrogen. In terms of the establishment and verification of spectral inversion model for soil nitrate content, Viscarra Rossela [32] used the spectral analysis method to predict and analyze various physical and chemical parameters of the soil, in which the linear PLSR model for soil nitrate content was established with low precision. SVM is a non-linear regression model, which is more accurate than the linear model established by previous researchers. In this paper, the hyperspectral monitoring model of soil nitrogen content established by SVM
had a good fitting effect. However, further study is needed when the selected penalty coefficients $C$ and kernel coefficients $RBF$ are applied to other applications. This study took the apple orchards soil in Qixia county of Shandong Province as the research object, and soil type and soil texture in different regions are different. If the monitoring model of soil nitrate nitrogen which was established is applicable to other soil types in other regions, further research is needed.

5. Conclusion

Using the ASD Fieldspec 3 to obtain the hyperspectral reflectance information of soil samples, a method for rapid monitoring of soil nitrate nitrogen content in the apple orchards in Qixia county was established. 72 soil samples were selected as the prediction set, a hyperspectral monitoring model by SVM could well predict the soil nitrate nitrogen content using MSC combined with FD pretreatment method, and the $R^2$ was 0.9585, the RMSE was 0.281, and the RPD was 3.835. While the $R^2$ of the calibration set was 0.822, the RMSE was 0.392, and the RPD was 2.037. The results showed that the SVM model had a high accuracy in predicting soil nitrate nitrogen content.

References


Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.
A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
Providing 24-hour high-quality service
User-friendly online submission system
Fair and swift peer-review system
Efficient typesetting and proofreading procedure
Display of the result of downloads and visits, as well as the number of cited articles
Maximum dissemination of your research work

Submit your manuscript at: http://papersubmission.scirp.org/
Or contact as@scirp.org