Hyper-spectral estimation of soil organic matter in apple orchard based on CWT

Ruiyan Li, Mingxiu Gao*, Zexin Xu, Jiafan Wang, Xiaoxin Xie

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

* Corresponding author: mxgao@sdau.edu.cn

Abstract. In the quantitative estimation by spectra of soil organic matter (SOM), there are great difficulties in the extraction of characteristic spectral information. It is difficult to effectively improve the correlation with SOM by ordinary spectral transformations, and the spectral estimation models are not high in accuracy and applicability. The purpose of this paper is to explore a more accurate and more suitable non-destructive evaluation model for apple orchard soil. In this paper, an apple orchard in Shuangquan Town, Changqing District, Jinan City, was selected as the study area. The continuous wavelet transform (CWT) was used to process the original soil spectra at multiple scales, and the effect of the estimation model on the SOM correlation at different resolutions was investigated. The results showed that the CWT treatment significantly improved the degree of spectral response to SOM in the orchard. Compared with the original spectra, the prediction accuracy of the model constructed by CWT was higher. After wavelet algorithm processing, the prediction ability of the model tends to increase with the decrease of spectral resolution. The best prediction accuracy was found at the spectral resolution of 20 nm, with the coefficient of determination $R^2 = 0.7547$ and the root mean square error $RMSE = 0.0042$; the validation accuracy $R^2 = 0.8015$ and the root mean square error $RMSE = 0.0039$. The model achieved a more satisfactory prediction result at the spectral resolution of 5 nm also. The results show that the broad-band spectral data processed by CWT can be used for accurate monitoring of SOM in apple orchards.

1. Introduction

The amount of soil organic matter (SOM) content is an important reference for evaluating the level of soil fertility\(^1\), and it profoundly affects the structural, infiltration, aeration, adsorption, and buffering properties of the soil. SOM in orchards plays an important role in the improvement of fruit quality while providing balanced nutrition for the growth and development of fruit trees. China is a major apple producing country and ranks first in the world in apple production. However, the average yield per unit area of apples in China is only half of that in Europe and the United States, and the amount of fertilizer input is 2-4 times higher than that in the corresponding countries. Excessive spending on inputs not only greatly increases production costs and reduces farmers’ income, but also causes great harm to the environment. At the same time, it weakens the soil buffering capacity and makes the orchard ecosystem fragile. The traditional method of measuring SOM is complicated, takes a long time to measure, and has poor repeatability and high cost. This makes it difficult to meet the demand for rapid and accurate measurement of SOM content. Visible/NIR spectroscopy is fast, non-destructive and non-polluting, and can measure soil samples in a short time and in large quantities at once. Many experts and scholars were
attracted to study it in recent years, and some certain results have been achieved\[2\]. Galvao and Vitorello demonstrated the existence of absorption peaks at 550-700 nm in the soil reflectance spectral curve due to SOM after laboratory analysis\[3-4\]. Yang performed partial least squares regression on various transformations of the original spectra of Hangzhou rice soil samples with first-order, the highest accuracy of PLS model estimation was obtained from the spectra transformed by the derivative pre-processing method\[4\].

To address the difficulties in the study of spectral estimation of SOM content in orchards, taking an apple orchard in Shuangquan Town as the study area, this study used the continuous wavelet transform (CWT) to process the raw soil spectra at multiple scales to study the effects of estimation models at different resolutions on SOM content correlation.

2. Materials and Methods

2.1. Overview of the study area
The study area is located in Shuangquan Town, Changqing District, Jinan City, Shandong Province, which belongs to the low mountainous hilly area in central Lu. It has a warm temperate semi-humid continental monsoon climate with short spring and autumn, long summer and winter, the average annual temperature of 14.7℃, average annual sunshine hours of 2616.8 h, and average annual precipitation of 671.1 mm, mainly in summer. The selected apple orchard covers an area of about 33 hm², and the soil type is brown soil. The planting spacing is about 1.5 m, the row spacing is about 4.5 m, the number of plants is about 100, the crown width of apple tree is 1-1.5 m. The planted apples are mainly Gala, which is a mature variety in August. The tree is 5 years old, and the root distribution is proportional to the tree crown width, and the vertical distribution is 0-50 cm.

2.2. Data sources

2.2.1. Soil sample data acquisition
Soil samples were collected on May 15, 2019. This is a critical period for regulating apple yield, quality, and tree vigor when young fruits are largely formed. Based on the checkerboard sampling method, 140 sampling plots were laid out in the apple orchard, and one fruit tree was selected as a sampling point for each sampling plot. The soil was collected as close to the capillary roots as possible, considering minimizing damage to the fruit trees. Soil samples were collected with a soil auger at a distance of about 30 cm horizontally from the roots of apple trees, 0-30 cm down from the soil surface. SOM was measured by the potassium dichromate volumetric method.

2.2.2. Spectral data acquisition
The ASD Fieldspec3 geo-spectrometer from Analytical Spectral Devices, USA, was used to obtain the original state soil spectral reflectance in a dark room with artificially controlled light conditions. The light source was a 50 W standard DC tungsten quartz halogen lamp. The ring knife holding the original state soil was placed on black rubber with an approximate reflectance of 0. The field of view of the probe was 25°, 15 cm from the soil sample, and the angle of incidence of the light source was 45°. For the measurements, each soil sample was rotated 3 times at 90° and 5 spectral curves were taken in each direction. The reflectance spectrum of the soil sample was obtained by arithmetic averaging of 20 curves. The measurements were corrected in time with a white plate with a reflectance of 1\[5\].

2.3. Research methods

2.3.1. Spectral resampling
Because the range accuracy of the spectrometer is good in the middle and poor at both ends, the measured raw spectral data has a large noise at the edge of the spectrum and a steep change at the 1000 nm curve. Therefore, ViewSpecPro was used to correct the collected spectral data for breakpoints. To
solve the problem of much noise at both ends of the spectrum, the data at 350-399 and 2451-2500 nm were removed after the corrected data were converted to spectral reflectance. To further improve the spectral signal-to-noise ratio, the spectral resolution was resampled to 5, 10, 20, 40, 80, and 160 nm.

2.3.2. Continuous wavelet transform
Wavelet analysis is used to achieve a more accurate description and separation of signal features by decomposing the signal data in the time and frequency domains and extracting more effective information from them. In this paper, CWT is used to decompose the soil spectral data at different scales with different spectral resolutions. The equations are as following:

\[ W_\lambda(a, b) = \int_{-\infty}^{\infty} f(\lambda) \varphi_{a,b}(\lambda) d\lambda \]  
(1)

\[ \varphi_{a,b}(\lambda) = \frac{1}{\sqrt{a}} \varphi\left(\frac{\lambda - b}{a}\right) \]  
(2)

In the above equation: \( f(\lambda) \) is the spectral reflectance; \( \lambda \) is the number of 400-2450 nm spectral bands; \( \varphi_{a,b} \) is the wavelet basis function; \( a \) is the scale factor; \( b \) is the translation factor. The wavelet coefficients \( W_{f(a,b)} \) contain two-dimensional data, band, and scale, respectively, to generate a matrix of the number of behavioral scales, listed as the number of bands. Soil spectral curve absorption characteristics are similar to the Gaussian function, so the Gaus 4 functions are chosen as the wavelet basis function. Soil spectral data with different spectral resolutions were subjected to CWT using MatlabR2018a. To reduce data redundancy, the decomposition scales of CWT were set to \( 2^0, 2^1, 2^2, ... 2^9 \), i.e. scales 1-10. The wavelet coefficients of the transformed 10 scales were correlated with the SOM to obtain the correlation coefficient matrix.

2.3.3. Extraction of characteristic spectral segments of SOM
To deeply study the characteristic spectral bands of SOM, the correlation coefficient method was used to analyze and extract the characteristic bands of SOM after various spectral transformations of the original soil reflectance. The commonly used Pearson coefficient was chosen for the correlation coefficient. The calculation formula was

\[ r_{ni} = \frac{\sum_{n=1}^{SW} (R_{ni} - \bar{R})(TSOM_{ni} - \bar{TSOM})}{\sqrt{\sum_{n=1}^{SW} (R_{ni} - \bar{R})^2 TSOM_{ni} - \bar{TSOM}^2}} \]  
(3)

In the above equation: \( r_{ni} \) is the correlation coefficient between SOM and input spectrum; \( i \) is the band number; \( R \) is the input spectrum; \( R_{ni} \) is the input value of the nth sample at the i-th band; \( \bar{R} \) is the mean value of the input spectral value of the i-th band soil sample; \( SOM_n \) is the SOM value of the nth soil sample; and \( \bar{SOM} \) is the mean value of the SOM of the soil sample.

In this paper, MATLABR2018a software is used to perform the screening of the characteristic bands based on the correlation coefficient method.

2.3.4. Model building and accuracy check
Based on the bands after CWT, the model is constructed using partial least squares (PLR) method. The correlations of the constructed models at different resolutions are compared based on the coefficient of determination (R²) and root mean square error (RMSE), and validated. PLR was chosen for the correlation model construction and accuracy evaluation of SOM at different resolutions because of its strong anti-interference capability and more objective evaluation results\(^6\). The full set of samples (140) was randomly divided into two groups, 100 samples in the modeling set (\( X_1=100 \)) and 40 samples in the validation set (\( X_2=40 \)), and modeled in MATLAB2018 using CWT-PLS.
3. Results and Discussion

3.1. Statistical characteristics of SOM content at the young fruit stage

Table 1 shows the descriptive statistics of soil SOM content in this apple orchard. Generally, the SOM content of a productive orchard should be higher than 0.5%-0.8%. The mean value of SOM content in this orchard is only 0.21% in the young fruit stage, and the coefficient of variation is small and the data are less discrete. According to the second soil census of China and the soil fertility index of Shandong Province, the SOM content of this orchard is all at a low level.

| Sample           | Number | Max   | Min   | Sd    | Mean  | Cv   |
|------------------|--------|-------|-------|-------|-------|------|
| Complete works   | 140    | 0.3896| 0.0762| 0.0542| 0.2055| 0.2637|
| Modeling set     | 100    | 0.3896| 0.0762| 0.0519| 0.2040| 0.2544|

3.2. Results of soil spectra at different resolutions after CWT

The results after CWT treatment of soil spectra at different resolutions are shown in figure 1. There are different relationships between different spectral resolutions and decomposition scales. Under the original resolution of the spectrum, the 1-6 scale spectral curve fluctuates gently, the reflectance is close to 0, and the spectral features are not obvious; the 7-10 scale fluctuates more and becomes a wavy line type with large undulations, among which the 8-10 scale fluctuation graph is parabolic. As the resolution gradually decreases, the 1-6 scale spectral curve starts to fluctuate and the fluctuation gradually becomes larger; the 7-10 scale spectral curve fluctuation gradually decreases and tends to be flat. As a whole, after CWT processing, the characteristics of the soil spectral curve will gradually weaken with the decrease of the spectral resolution; along with the increase of the scale, the wave peak located at the lower scale will gradually move to the short-wave direction.

![Figure 1 Results of soil spectra at different resolutions after CWT treatment](image)

3.3. Correlation between CWT results and SOM

The correlation coefficients of the spectra at different resolutions were analyzed with the SOM after continuous wavelet decomposition, and the results are shown in figure 2. The area with significantly higher correlation in the correlation coefficient matrix is the "spectral response interval". It can be seen that the overall correlation between the SOM content and the spectra at different spectral resolutions tends to increase and then decrease with decreasing spectral resolution. At the decomposition scale, the
spectral response interval of SOM at the original spectral resolution (1 nm) was mainly focused on the high scale decomposition, and the correlation was weak at the low scale decomposition. The response intervals gradually move to lower scales as the spectral resolution decreases, making the effective spectral information gradually prominent at lower scales. Since the reduction of spectral resolution causes the effective spectral information to be hidden, the information within the high decomposition scale is relatively homogeneous. Therefore, when analyzing the broad-band spectral data, the decomposition scale should be appropriately reduced to explore the hidden information in it.

Figure 2  Correlation coefficient of wavelet coefficient and SOM content under each spectral resolution

3.4. CWT-based modeling
Based on the analysis of the spectral data information at different resolutions processed with CWT, the PLS-based model for soil SOM content prediction was used, and the results are shown in table 2. In terms of the coefficient of determination, it is found that the model evaluation index constructed based on 20 nm is the best, and the model prediction accuracy and validation accuracy can reach 0.7547 and 0.8015, respectively; the root means square deviation can reach 0.0042 and 0.0039, both close to 0. On the whole, the prediction accuracy of the soil SOM estimation model based on CWT data is high, and it shows a fluctuating characteristic of decreasing, then increasing, and then decreasing with decreasing resolution. The stability of the model decreases then increases and then decreases as the spectral resolution decreases. It can be seen that the reduction of spectral resolution can effectively suppress the effect of high-frequency noise, but too low a resolution can hide some effective information in the spectrum. As can be seen from the table, with 80 nm resolution as the node, the effect of the model constructed at resolutions below this node becomes significantly worse, and the accuracy and stability are significantly reduced.

Table 2  Modeling results of SOM content under different resolution

| Spectral resolution /nm | Band /nm    | Modeling set R² | RMSE   | Validation set R² | RMSE   |
|-------------------------|-------------|-----------------|--------|-------------------|--------|
| 1                       | 224-284 451-691 | 0.7469          | 0.0044 | 0.7652            | 0.0041 |
| 5                       | 49-58 93-111 | 0.7416          | 0.0049 | 0.7395            | 0.0047 |
| 10                      | 26-29 48-56 | 0.7346          | 0.0044 | 0.7539            | 0.0045 |
| 20                      | 14 15 25-28 62 | 0.7547          | 0.0042 | 0.8015            | 0.0039 |
| 40                      | 8 13-15    | 0.7308          | 0.0046 | 0.7608            | 0.0047 |
| 80                      | 5 8 10     | 0.7107          | 0.0048 | 0.7599            | 0.0049 |
| 160                     | 3 6 7      | 0.5563          | 0.0054 | 0.6175            | 0.0050 |
4. Conclusions
The correlation between spectra and SOM increases and then decreases with the decrease of spectral resolution at different resolutions, and the highest correlation (\(|R|=0.75\)) is found at 20 nm resolution; the sensitive spectral interval keeps moving to the lower scale. The CWT method enhances the use of useful information in soil spectra and significantly improves the estimation of SOM; compared with the original spectral reflectance (1nm), the modeling by CWT substantially improves the prediction accuracy of SOM. The modeling accuracy increases, then decreases, and then decreases with the decrease of the spectral resolution after the CWT processing, and the model is gradually stabilized; the best model is estimated at the resolution of 20 nm with \(R^2=0.7547\) and RMSE=0.0042.

Acknowledgement
This paper was supported by the Shandong Province Major Science and Technology Innovation Project "Research and Demonstration of Precision Management Technology of Irrigation and Fertilization in Orchards" (2018CXGC0209).

References
[1] Chen H Y, Zhao G X, Li X C, Zhu X C, Sui L, Wang Y J. (2011) Hyper-spectral estimation of soil organic matter content based on wavelet transform[J]. Chinese Journal of Applied Ecology, 22(11): 2935-2942.
[2] Wang R J, Chen T J, Wang Y B, Wang L S, Xie C J, Zhang J, Li R, Chen H B. (2017) Soil near-infrared spectroscopy prediction model based on deep sparse learning[J]. Chinese Journal of Luminescence, 38(1): 109-116.
[3] Lin, P D, Tong, C J, Zhang, J K, et al. (2018) Inversion of black soil organic matter content with field hyperspectral reflectance based on continuous wavelet transformation[J]. Soil and water conservation research, 25(2): 46-52+ 57.
[4] Wang L S, Lu C P, Wang R J, Huang W, Guo H Y, Wang Y B, Lin Z D, Wang J, Jiang Q, Song L T. (2018) Optimization for Vis / NIRS prediction model of soil available nitrogen content [J]. Chinese Journal of Luminescence, 39(07): 1016-1023.
[5] Wang F Y, Shen Y, Zhang C, Liu B F, Zheng J Y. (2018) Sugar content nondestructive testing study for apple[J]. China Agricultural Informatics, 30(4): 101-108.
[6] Zhang J, Yuan S, Zhang J. (2020) Detection of sugar degree in strawberry based on wavelet transform-genetic algorithm-partial least squares [J]. Journal of Analytical Sciences, 36(1): 111-115.