



OPEN

Study on hyperspectral estimation model of soil organic carbon content in the wheat field under different water treatments

Chenbo Yang¹, Meichen Feng^{1✉}, Lifang Song¹, Chao Wang¹, Wude Yang¹, Yongkai Xie², Binghan Jing¹, Lujie Xiao¹, Meijun Zhang¹, Xiaoyan Song¹ & Muhammad Saleem¹

Hyperspectral remote sensing technology can be used to monitor the soil nutrient changes in a rapid, real-time, and non-destructive manner, which is of great significance to promote the development of precision agriculture. In this paper, 225 soil samples were studied. The effects of different water treatments on soil organic carbon (SOC) content, and the relationship between SOC content and spectral reflectance (350–2500 nm) were studied. 17 kinds of preprocessing algorithm were performed on the original spectral (R), and the five allocation ratios of calibration to verification sets were set. Finally, the model was constructed by partial least squares regression (PLSR). The results showed that the effects of water treatment on SOC content were different in different growth stages of winter wheat. Results of correlation analysis showed that the differential transformation can refine the spectral characteristics, and improve the correlation between SOC content and spectral reflectance. Results of model construction showed that the models constructed by second-order differential transformation were not good. But the ratio of standard deviation to the standard prediction error (RPD) values of the models were constructed by simple mathematical transformation (T0–T5) and first-order differential transformation (T6–T11) can reach more than 1.4. The simple mathematical transformation (T0–T2, T4–T5) and the first-order differential transformation (T6–T10) resulted in the highest RPD in mode 5 and mode 2, respectively. Among all the models, the model of T7 in mode 2 reach the highest accuracy with a RPD value of 1.9861. Therefore, it is necessary to consider the data preprocessing algorithm and allocation ratio in the process of constructing the hyperspectral monitoring model of SOC.

Soil organic matter is the material derived from living organisms in the soil, of which 60–80% of the carbon component is called SOC. It is well known that SOC is a vital factor to improve soil fertility, although the content is small relative to the soil^{1–3}. However, the previous methods⁴ to obtain SOC content are time-consuming and laborious. Hyperspectral technology can solve this problem. In recent years, with the development of hyperspectral remote sensing technology, due to its characteristics of rapid, real-time, and non-destructive monitoring of targets⁵, it has been become an important technology for many scholars to study the soil properties^{6–8}.

The spectral reflectance refers to the ratio of reflected flux to the incident flux at a certain wavelength⁹. Various molecules in the soil absorb and reflect light at different wavelengths to form soil spectral characteristic curve. The visible (350–780 nm) and near-infrared (780–2500 nm) bands have been widely used in agricultural research^{10,11}. In the process of quantitative analysis of chemical measurement values and hyperspectral data, analysis result is mainly affected by three aspects: the expression ability of hyperspectral data^{12–17}, the data reasonable allocation of calibration to validation sets¹⁸, and the scientific selection of model construction method^{19–23}. The expression ability of hyperspectral data refers to reducing the noise of original spectral data through different spectral transformation algorithms, and improving the correlation between the spectral data and the chemical measurement values^{24,25}. The data reasonable allocation of calibration to verification sets will be affected by the amount of data and the degree of dispersion, which is also a segment that is easy to be ignored in the process of model construction²⁶. The scientific selection of the modeling method is determined by the quantitative relationship between the chemical measurement values and the spectral data. At present, there are many kinds of methods

¹Agronomy College, Shanxi Agricultural University, Taigu 030801, Shanxi, China. ²Institute of Geography Science, Taiyuan Normal University, Jinzhong 030619, Shanxi, China. ✉email: fmc101@163.com

to construct hyperspectral monitoring model. Some simple linear modeling algorithm include stepwise multiple linear regression (SMLR), partial least squares regression (PLSR), and principal component regression (PCR), etc. Nonlinear modeling algorithm include artificial neural network (ANN), support vector machine (SVM), and random forests (RF)^{27–29}, etc. Among them, PLSR is a kind of linear regression model which has been widely used and has a good prediction effect^{30–33}.

Previous studies have shown that hyperspectral monitoring has a good ability to predict SOC content^{34,35}. PLSR has been used to monitor SOC in many studies, but the results were different. Amin et al.³⁶ constructed a hyperspectral monitoring model of SOC content based on PLSR in Azerbaijan, and found that the model constructed after Savitzky-Golay smoothing can reach the highest accuracy with R^2 and RPD are 0.85 and 2.54, respectively. Yu et al.³⁷ constructed a hyperspectral monitoring model of soil organic matter content by using PLSR after preprocessing the spectral data, and the results showed that the model based on continuous removal (CR) preprocessing had the best accuracy. Ji et al.¹⁸ used a variety of modeling methods to predict soil organic matter content based on different data allocation ratios, it was found that when the ratios of calibration to validation sets were different, the accuracy of model constructed by different modeling methods was different. But the data allocation ratios used in this study were few, and hyperspectral data were not preprocessed. It can be seen that previous researchers have done some research on data preprocessing and sample allocation ratio in the process of building the SOC model. From the results, it is necessary to consider these two factors in modeling. However, there are few studies considering these two factors at the same time. Based on this, we conducted this study.

According to previous studies, different data preprocessing algorithm, data allocation ratio, and modeling method all have an effect on the accuracy of the model. In this paper, from the perspective of the data allocation ratio, different mathematical transformations of spectral data were carried out, and finally constructs the model by using PLSR. The purposes of this study are: (1) Study the effect of mathematical transformation algorithm and data allocation ratio on the accuracy of hyperspectral monitoring model of SOC content, (2) Find the best spectral preprocessing algorithm and data allocation ratio to predict SOC content.

Materials and methods

Experimental design. The experimental field located in the Experiment Station of Shanxi Agricultural University. Each experimental plot is all a square of 9 m², with a total of 15 plots. The water content was controlled according to the percentage of the maximum soil field capacity. For example, when the maximum field capacity of soil is 0.2 g·g⁻¹ soil, 80% of the maximum field capacity is 0.16 g·g⁻¹ soil. The soil water content was kept at the same level from the returning green stage of winter wheat. There were five irrigation levels: 80%, 60%, 45%, 35%, and 30% of maximum soil field capacity. The experiment was set up in a completely randomized design with three replications. The winter wheat cultivar ‘Zhongmai 175’ was planted. The soil drill was used to collect topsoil of 0–20 cm in critical growth period such as jointing stage, booting stage, heading stage, flowering stage, filling stage, and maturity stage of winter wheat. Use the 5-point sampling method to collect soil samples and mix them for standby. The experiment was carried out for two years, sampling 8 times and 7 times in 2018 and 2019, respectively. The soil texture for the experimental field is calcareous cinnamon soil developed from loess parent material with medium fertility. The average total nitrogen content, the total phosphorus content, and the total potassium content are 0.70 g·kg⁻¹, 1.32 g·kg⁻¹, and 22.13 g·kg⁻¹, respectively. All methods were performed in accordance with Chinese guidelines and legislation.

Indexes measurement. Before the measurement, the animal, plant residues, and other impurities were removed from samples. After drying in the shade at 25–30 °C for one week, passed through a 0.154 mm sieve. Then measured soil spectral reflectance and SOC content.

Soil reflectance measurement. The Field-spec 3 hyperspectral radiometer produced by ASD company in the United States was used for soil spectral measurement. The wavelength acquisition range is 350–2500 nm, among the spectral sampling intervals of 350–1000 nm and 1000–2500 nm are 1.4 nm and 2 nm, respectively, and the spectral resolutions are 3 nm and 10 nm, respectively. Before the measurement, each soil sample with a flat surface was placed in a culture dish. Use the plant probe with its own light source for measurement, and the measurement area is 3.14 cm². Three points were measured for each sample, and each point was measured 10 times. Before each measurement, the whiteboard was used for calibrating.

SOC content measurement. SOC content was measured by the Walkley–black method³⁰.

Data processing and analysis. There are 17 common mathematical algorithms were selected to preprocess the original spectral data of soil, as it showed in Table 1.

In this paper, 225 samples were divided into calibration set and verification set by concentration gradient method³⁸, including 5 samples allocation ratios (calibration set: verification set): mode 1(2:2), mode 2(3:2), mode 3(4:2), mode 4(5:2), mode 5(6:2).

In this paper, PLSR models were constructed in MATLAB 2010, and leave one out cross-validation method was selected to improve the stability of the model. The coefficient of determination (R^2), root mean square error (RMSE), and RPD were used to evaluate the accuracy of the model.

This paper also uses View Spec Pro 6.0 to preprocess soil spectral data, Microsoft Excel 2010, Unscrambler 9.7, and SPSS Statistics 20 for data processing, Origin 2016 for mapping.

Simple mathematical transformation		First-order differential transformation		Second-order differential transformation	
Code	Algorithm	Code	Algorithm	Code	Algorithm
T0	R	T6	R'	T12	R''
T1	$1/R$	T7	$(1/R)'$	T13	$(1/R)''$
T2	$\text{Log } R$	T8	$(\text{Log } R)'$	T14	$(\text{Log } R)''$
T3	$1/\text{Log } R$	T9	$(1/\text{Log } R)'$	T15	$(1/\text{Log } R)''$
T4	\sqrt{R}	T10	$(\sqrt{R})'$	T16	$(\sqrt{R})''$
T5	$1/\sqrt{R}$	T11	$(1/\sqrt{R})'$	T17	$(1/\sqrt{R})''$

Table 1. Soil spectral preprocessing mathematical algorithms. R represents the original spectrum, the same below.

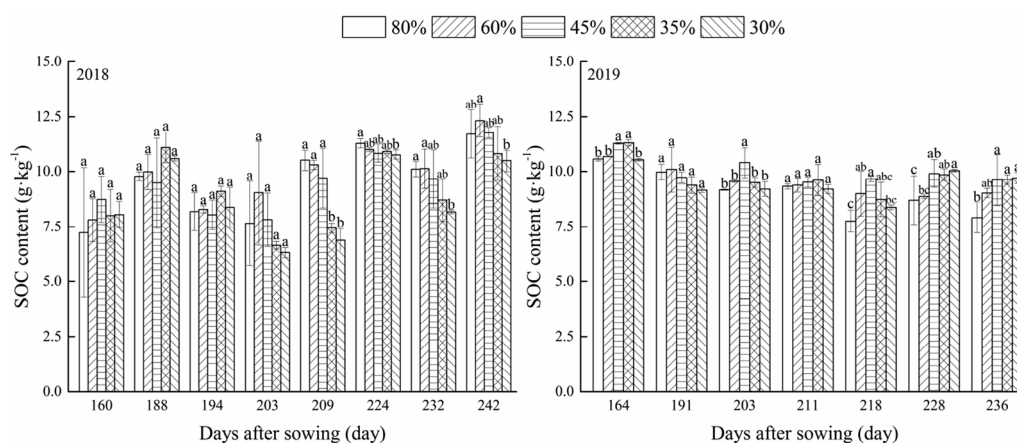


Figure 1. Changes of SOC content under different water treatments.

Results and analysis

Change of SOC content. It can be seen from Fig. 1 that the change trends of SOC content are different with the advance of the winter wheat growth process in the two-year experiment. In 2018, there is an obvious increase–decrease–increase trend, while in 2019 is slowly decreasing to a stable trend. We compare the sampling stages that have close days after sowing in two-year. 160 days and 203 days after sowing in 2018 correspond to 164 days and 203 days after sowing in 2019, respectively. SOC content increased first and then decreased with the aggravation of drought stress, but their significance of difference was different. Although the change trends of 188 days and 194 days after sowing in 2018 and 191 days after sowing in 2019 were different, there was no significant difference in SOC content among the three growth stages. SOC content in 209 days, 224 days, 232 days, and 242 days after sowing in 2018 basically decreased with the aggravation of drought stress, while the change trends of 211 days, 218 days, 228 days, and 236 days after sowing in 2019 were opposite.

Descriptive statistical analysis. 225 samples were analyzed in this paper, it can be seen from Table 2 that the standard deviation of the SOC content was small. The samples have a certain negative skewness, and the data distribution trend was slower than the standard normal distribution, but basically conformed to the normal distribution. In different allocation modes, the minimum and maximum values of the total data set were all assigned to the calibration sets. The average value and standard deviation of the calibration sets and the verification sets were close. This means that the data dispersion was similar. At the same time, the calibration sets and verification sets in the five modes all basically conformed to normal distribution, which can be used for subsequent processing.

Changes of original spectral reflectance. Since 93% of the 225 samples used in this paper were concentrated in the middle of 6–12 $\text{g}\cdot\text{kg}^{-1}$, it was necessary to analyse the soil spectral reflectance of SOC content at different levels (6, 7, 8, 9, 10, 11 $\text{g}\cdot\text{kg}^{-1}$). It can be seen from Fig. 2 that the changing trends of soil spectral reflectance with different SOC content levels were basically consistent. The reflectance increases with the increase of wavelength, and obvious absorption valleys were formed at near 1400 nm, 1900 nm, and 2200 nm. With the increase of SOC content, soil spectral reflectance decreased gradually, but the change range was uneven. The spectral curves of two samples with the SOC content of 6 and 7 $\text{g}\cdot\text{kg}^{-1}$ were close, and those of 8, 9, and 10 $\text{g}\cdot\text{kg}^{-1}$ were close to each other.

	Num	Min (g·kg ⁻¹)	Max (g·kg ⁻¹)	Ave	SD	Skewness	Kurtosis
Total	225	4.296	13.013	9.226	1.678	-0.288	-0.098
Cal-set of Mode 1	113	4.296	13.013	9.222	1.702	-0.314	-0.008
Ver-set of Mode 1	112	4.631	13.009	9.230	1.661	-0.263	-0.151
Cal-set of Mode 2	135	4.296	13.013	9.227	1.682	-0.295	-0.065
Ver-set of Mode 2	90	4.632	13.009	9.225	1.682	-0.283	-0.085
Cal-set of Mode 3	150	4.296	13.013	9.226	1.678	-0.287	-0.096
Ver-set of Mode 3	75	4.632	13.009	9.227	1.690	-0.296	-0.026
Cal-set of Mode 4	161	4.296	13.013	9.232	1.688	-0.291	-0.046
Ver-set of Mode 4	64	5.041	12.721	9.211	1.666	-0.290	-0.156
Cal-set of Mode 5	169	4.296	13.013	9.224	1.691	-0.301	-0.040
Ver-set of Mode 5	56	5.041	12.829	9.232	1.655	-0.253	-0.209

Table 2. Descriptive statistical analysis of test samples. Num, Min, Max, Ave, SD, Cal-set, and Ver-set are number, minimum, maximum, average, standard deviation of samples, calibration set, and verification set, respectively.

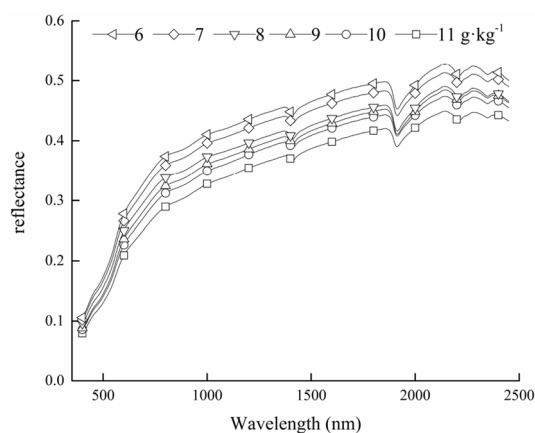


Figure 2. Spectral reflectance changes of soils with different SOC contents.

Changes of preprocessing spectral reflectance. In this paper, one sample (SOC content is 9.222 g·kg⁻¹) with the closest average content among all samples was taken as an example, and spectral preprocessing was performed according to Table 1. Mapping with directly measured spectral data, as shown in Fig. 3. Figure 3 showed that compared with the original spectral curve (T0), different preprocessing algorithms have different influences. Except for T2 and T4, the changing trends of other spectral preprocessing curves all had an obvious change. Among different types of preprocessing algorithms, the first-order differential transformation and the second-order differential transformation all have an obvious effect on refining spectral characteristics.

Correlation analysis between SOC content and different preprocessing spectrum. It can be seen from Fig. 4 that the correlations between SOC content and the spectral reflectance of T0–T5 were within ± 0.15 , while the correlations of T6–T17 were improved. Among them, the correlations of T12–T17 after second-order differential transformation were basically within ± 0.4 , and that of T6–T11 after first-order differential transformation were within ± 0.6 . In addition, the correlations of T6 and T10 in some wave bands can reach ± 0.7 to ± 0.8 . Although differential transformation can improve the correlation between the spectral reflectance and SOC content, the second-order differential transformation has a slight improvement only. However, the first-order differential transformation can improve the correlation greatly. The results showed that the differential transformation may enlarge the spectral characteristics in some wave bands which have a great correlation with SOC content. And the effect of first-order differential transformation was better than second-order differential transformation.

Construction and validation of PLSR model. In this study, the PLSR method is used to construct the model, and the model accuracy is shown in Figs. 5 and 6. Figure 5 showed that, compared with the simple mathematical transformation, the accuracy of the models constructed by the first-order differential transformation was increased, but the accuracy of the models constructed by the second-order differential transformation was

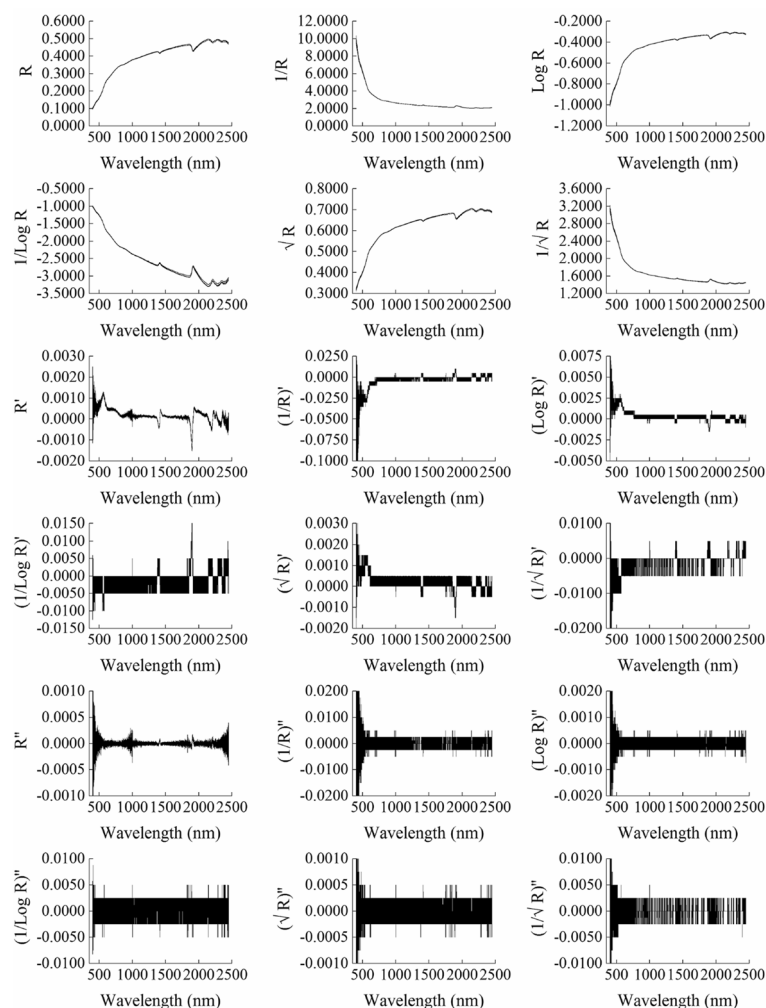


Figure 3. Spectral curves of different preprocessing mathematical algorithms.

reduced. Among all the models, the accuracy and stability of T6–T8 and T10–T11 after first-order differential transformation were basically reached the highest in mode 3, and T9 in mode 5 was also higher.

Figure 6 showed that the validation accuracy of the models constructed by the second-order differential transformation was low in all modes. Combined with Fig. 5, it can be seen that the second-order differential transformation was not conducive to construct hyperspectral monitoring models of SOC content by PLSR. The models constructed by simple mathematical transformation and first-order differential transformation achieved the highest verification accuracy in mode 5 and mode 2, respectively. At the same time, the verification accuracy of each model constructed by first-order differential transformation was higher than that of the corresponding simple mathematical transformation. Combined with Fig. 5, it can be seen that the models accuracy can be improved by first-order differential transformation. Different from Fig. 5, the accuracy of the models constructed by simple mathematical transformation and first-order differential transformation reached the lowest in mode 3. It can be seen that it is necessary to consider both the sample allocation ratio and mathematical transformation algorithm when constructing the hyperspectral monitoring model of SOC content. Among all the models constructed in this study, T7 and T11 reached the highest accuracy in mode 2 and mode 5, RPD are 1.9861 and 1.9217, respectively.

Discussion

From Fig. 1, we can see that the change trends of SOC content with the advance of the winter wheat growth process are different in the 2-year experiment. The change trend of SOC content shows an opposite trend with the aggravation of drought stress appears in the later stage of the 2 years experiment. This may be due to the different climatic conditions lead to the difference in the growth of winter wheat in the 2-year experiment. In fact, the growth and development conditions of winter wheat in the first year was better than in the second year. Severe high temperature and drought weather occurred in the experimental plot from March to April of 2019. During this period, the winter wheat is in the stage from returning green to jointing, and the water demand is high. The high temperature leads to water shortage of crops and affects the later growth³⁹. In May 2019, the temperature

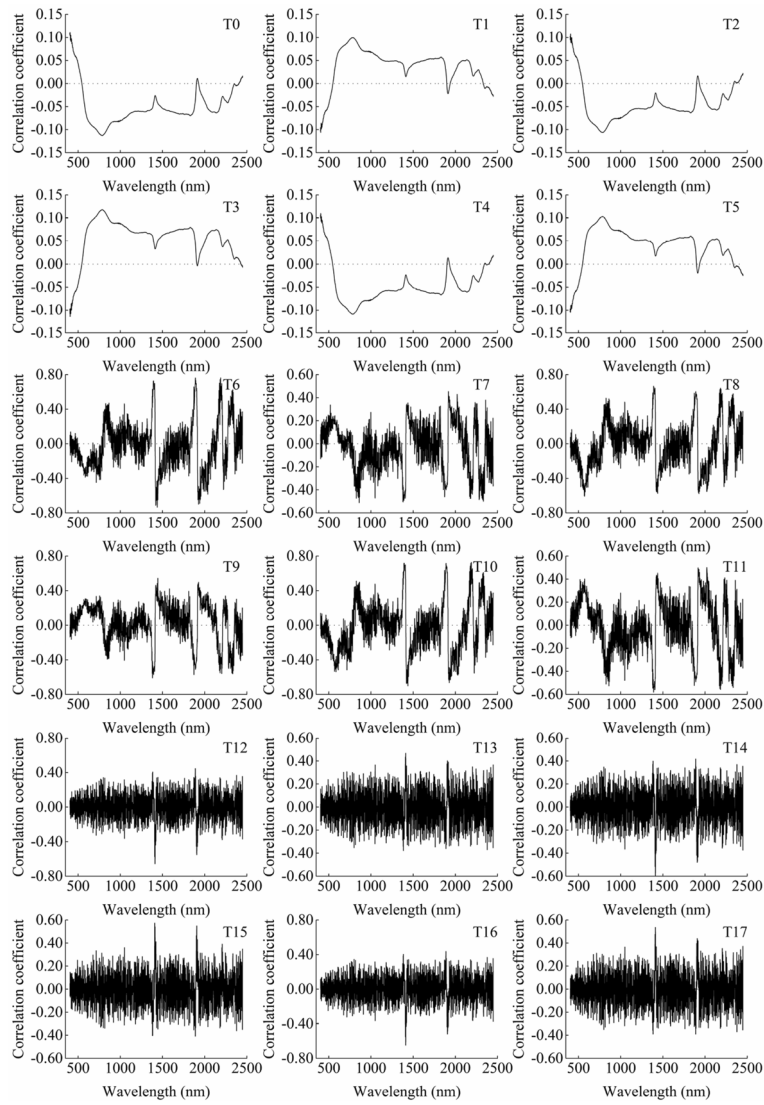


Figure 4. Correlation of different mathematical transformation spectra and SOC content. T0: R; T1: 1/R; T2: Log R; T3: 1/Log R; T4: \sqrt{R} ; T5: $1/\sqrt{R}$; T6: R²; T7: (1/R)²; T8: (Log R)²; T9: (1/Log R)²; T10: (\sqrt{R})²; T11: (1/ \sqrt{R})²; T12: R³; T13: (1/R)³; T14: (Log R)³; T15: (1/Log R)³; T16: (\sqrt{R})³; T17: (1/ \sqrt{R})³, the same below.

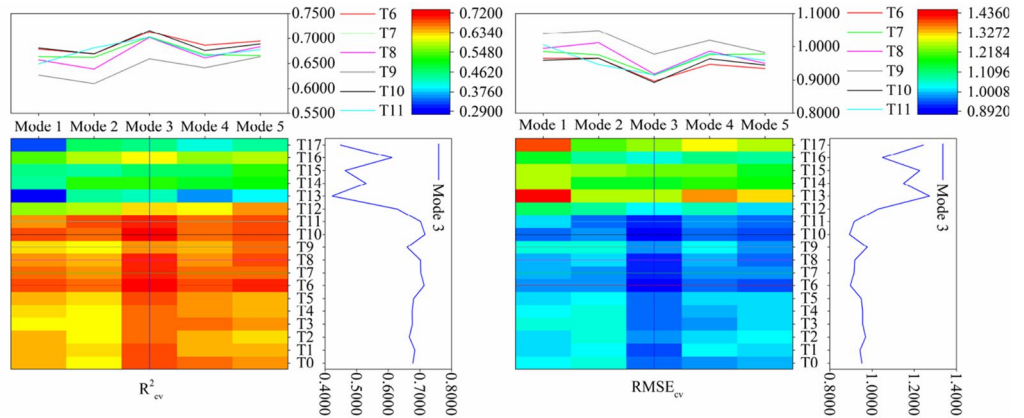


Figure 5. Construction of PLSR models in different modes.

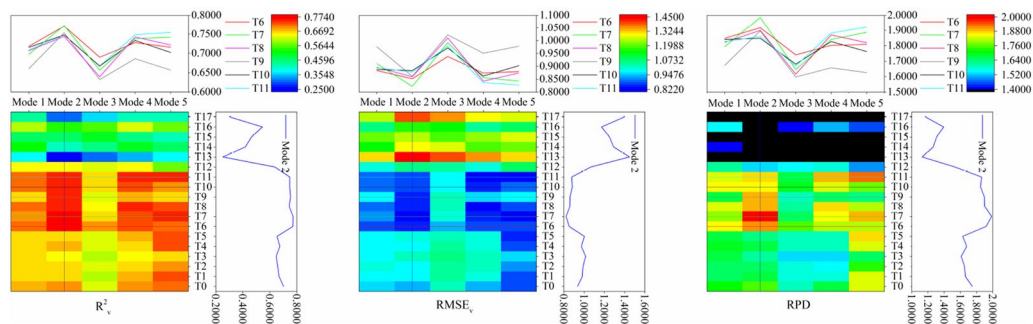


Figure 6. Validation of PLSR models in different modes.

of the experimental site fluctuated greatly, frost disaster occurred in the early stage, and high temperature above 35 °C appeared in the later stage. The extreme temperature further affected the growth of winter wheat⁴⁰.

In this paper, the change of the spectral reflectance of soil with different SOC content have been compared. It was found that the spectral reflectance of soil with different SOC content increased gradually with the increase of wavelength, and obvious absorption valleys were formed near 1400 nm, 1900 nm and 2200 nm. Physical studies show that the spectrum in the visible region is caused by the outer electron transition, while the near-infrared spectrum is mainly affected by molecular vibration, which can reflect the composition and structure of molecules. This is the basic principle of quantitative analysis of target materials by using hyperspectral technology. It ensures that almost no two substances have the same spectral characteristics, but also ensures that the same kind of substances have certain similar spectral characteristics⁴¹. These absorption valleys are caused by water molecules, OH⁻ and minerals⁴². At the same time, it was also found that the overall value of soil spectral reflectance will decrease with the increase of SOC content, which was consistent with the previous research results⁴³. This may be due to the higher content of organic carbon molecules absorbed more light, which may also be one of the principles of quantitative monitoring SOC content with hyperspectral.

Compared with the original spectral curve, 17 mathematical transformations amplify the spectral characteristics to varying degrees. And the effect of differential transformation is the most obvious. Previous researchers generally believe that differential transformation has a good effect in refining spectral features, which is consistent with the results of this paper^{44,45}.

In terms of improving the correlation between SOC content and spectral data, simple mathematical transformation (T1–T5) did not significantly improve compared with T0. T2 and T4 were basically the same as T0, while T1, T3, and T5 only changed the positive and negative correlation at the some wavelengths. The first-order and second-order differential transformation improved the correlation between spectral reflectance and SOC content to varying degrees. And the effect of first-order differential transformation was better than second-order differential transformation. At the same time, the results of model construction showed that the RPD value of the models constructed by simple mathematical transformation and first-order differential transformation reached more than 1.4, which were higher than that of the second-order differential transformation. It can be seen, from the mathematical characteristics of the first-order differential, that the spectrum characteristic of SOC is not only limited to some individual wavelengths, but also related to the increase and decrease of some continuous wave bands. At the same time, it is necessary to properly preprocess the spectral data before constructing the hyperspectral monitoring model.

Extracting sensitive band or using full band is a common way to construct hyperspectral quantitative monitoring model of SOC content. Some studies have shown that 435 nm, 500 nm, 965 nm, 1409 nm, 1910 nm, 2170 nm, 2260 nm, and 2380 nm are sensitive bands of SOC^{42,46}. Although the model constructed by using sensitive bands can be more concise and efficient, it can be seen from Fig. 4 that the spectral characteristics of SOC may also be related to the change trend of spectral curve. In this case, it is necessary to consider using the full band. PLSR can effectively deal with the complex relationship of high-dimensional data and provide a quantitative model. Therefore, this paper uses PLSR as the model construction method.

In this paper, it was found that the R^2_{cv} of the models constructed by the second-order differential transformation were mostly below 0.6, and the $RMSE_{cv}$ were higher, the accuracy of model construction was lower than that of simple mathematical transformation and first-order differential transformation. At the same time, the model validation results also showed that the accuracy of models constructed by second-order differential transformation was lower than that of simple mathematical transformation and first-order differential transformation, except for individual models. The accuracy of the models with different sample allocation ratios was different under the same mathematical transformation. Except T3 and T11 reached the highest RPD in mode 1 and mode 5, respectively. Other simple mathematical transformation and first-order differential transformation reached the highest RPD in mode 5 and mode 2, respectively. Generally speaking, among the different mathematical transformation algorithms, the models constructed by the first-order differential transformation have the best accuracy. The model construction and verification accuracy was the best in mode 3 and mode 2, and reached the highest in T10 and T7, respectively.

In this study, R^2_{cv} and $RMSE_{cv}$ were obtained by cross-validation that were used as the indicators to evaluate the accuracy of model construction. Therefore, with the increase of the number of samples in the calibration

set, the more time it takes and the higher the requirement for a computer to construct the model using PLSR. In practice, mode 2 can save more time and computer memory. Therefore, in the process of constructing a hyperspectral monitoring model of SOC content, the first-order differential transformation combined with a sample allocation ratio of 3:2 (calibration set: validation set) is conducive to improve the model accuracy and stability.

Conclusion

The prediction accuracy of the hyperspectral monitoring model of SOC content is greatly affected by soil spectral preprocessing mathematical algorithms and the ratio of the calibration set to validation set. Different mathematical transformation preprocessing can enlarge or reduce some characteristic information of the original spectral data. Reasonable data allocation ratio can reduce the excessive redundancy of calibration set. But different mathematical transformation algorithms have different requirements for the ratio of the calibration set to verification set. Therefore, in order to improve the accuracy of the model, it is necessary to comprehensively consider the mathematical transformation algorithm and sample allocation ratio.

This study showed that the effects of drought stress on SOC content were different in different growth stages of winter wheat. And the higher the SOC content, the lower the spectral reflectance. The first-order differential transformation and the second-order differential transformation have a greater role in amplifying spectral information. And they can also significantly improve the correlation between SOC content and spectral reflectance, the maximum correlation can reach ± 0.8 . The prediction accuracy of the hyperspectral monitoring model of SOC content can be effectively improved by first-order differential transformation and the sample allocation ratio of calibration set: verification set = 3:2. Among them, the model constructed by the first reciprocal transformation and then the first-order differential transformation has the highest accuracy, and the RPD of the verification model is 1.9861.

Received: 16 May 2021; Accepted: 3 September 2021

Published online: 20 September 2021

References

- Aryal, D. R., De Jong, B. H., Ochoa-Gaona, S., Esparza-Olguin, L. & Mendoza-Vega, J. Carbon stocks and changes in tropical secondary forests of southern Mexico. *Agr. Ecosyst. Environ.* **195**, 220–230 (2014).
- Aryal, D. R., De Jong, B. H., Ochoa-Gaona, S., Mendoza-Vega, J. & Esparza-Olguin, L. Successional and seasonal variation in litterfall and associated nutrient transfer in semi-evergreen tropical forests of SE Mexico. *Nutr. Cycl. Agroecosys.* **103**(1), 45–60 (2015).
- Aryal, D. R. *et al.* Soil organic carbon depletion from forests to grasslands conversion in Mexico: A review. *Trop. Agric.* **8**, 181 (2018).
- Gao, W., Yang, J., Ren, S. R. & Liu, H. L. The trend of soil organic carbon, total nitrogen, and wheat and maize productivity under different long-term fertilizations in the upland fluvo-aquic soil of North China. *Nutr. Cycl. Agroecosys.* **103**, 61–73 (2015).
- Qi, H., Paz-Kagan, T., Karnieli, A., Jin, X. & Li, S. Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data. *Soil Till Res.* **175**, 267–275 (2018).
- Dong, X., Tian, J., Zhang, R., He, D. & Chen, Q. Study on the relationship between soil emissivity spectra and content of soil element. *Spectrosc. Spect. Anal.* **37**(02), 557–564 (2017).
- Kemper, T. & Sommer, S. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environ. Sci. Technol.* **36**(12), 2742–2747 (2002).
- Panigrahi, N. & Das, B. S. Canopy spectral reflectance as a predictor of soil water potential in rice. *Water Resour. Res.* **54**(4), 2544–2560 (2018).
- Peddle, D. R., White, H. P., Soffer, R. J., Miller, J. R. & LeDrew, E. F. Reflectance processing of remote sensing spectroradiometer data. *Comput. Geosci.* **27**(2), 203–213 (2001).
- Ben-Dor, E. *et al.* Using imaging spectroscopy to study soil properties. *Remote Sens. Environ.* **113**, S38–S55 (2009).
- Rossel, R. A., Walvoort, D. J., Mcbratney, A. B., Janik, L. J. & Skjemstad, J. O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **131**(1), 59–75 (2006).
- Cheng, H. *et al.* Estimating heavy metal concentrations in suburban soils with reflectance spectroscopy. *Geoderma* **336**, 59–67 (2019).
- Ding, J., Yang, A., Wang, J., Sagan, V. & Yu, D. Machine-learning-based quantitative estimation of soil organic carbon content by VIS/NIR spectroscopy. *PeerJ* **6**(3), e5714 (2018).
- Gobrecht, A., Bendoula, R., Roger, J.-M. & Bellon-Maurel, V. A new optical method coupling light polarization and vis-NIR spectroscopy to improve the measurement of soil carbon content. *Soil Till Res.* **155**, 461–470 (2016).
- Gu, X., Wang, Y., Song, X. & Xu, X. *The Inversion Model of Soil Organic Matter of Cultivated Land Based on Hyperspectral Technology. Remote Sensing for Agriculture, Ecosystems, and Hydrology XVII* (International Society for Optics and Photonics, 2015).
- Nawar, S., Buddenbaum, H., Hill, J., Kozak, J. & Mouazen, A. M. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil Till Res.* **155**, 510–522 (2016).
- Yu, X., Liu, Q., Wang, Y., Liu, X. & Liu, X. Evaluation of MLSR and PLSR for estimating soil element contents using visible/near-infrared spectroscopy in apple orchards on the Jiaodong peninsula. *CATENA* **137**, 340–349 (2016).
- Ji, W. J., Li, X., Li, C. X., Zhou, Y. & Shi, Z. Using different data mining algorithms to predict soil organic matter based on visible-near infrared spectroscopy. *Spectrosc. Spect. Anal.* **32**(09), 2393–2397 (2012).
- Douglas, R. K., Nawar, S., Alamar, M. C., Mouazen, A. M. & Coulon, F. Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-NIR spectroscopy and regression techniques. *Sci. Total Environ.* **616**, 147–155 (2018).
- Mouazen, A. M. & Al-Asadi, R. A. Influence of soil moisture content on assessment of bulk density with combined frequency domain reflectometry and visible and near infrared spectroscopy under semi field conditions. *Soil Till Res.* **176**, 95–103 (2018).
- Rossel, R. A. & Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **158**(1), 46–54 (2010).
- Nawar, S. & Mouazen, A. M. Comparison between random forests, artificial neural networks and gradient boosted machines methods of on-line vis-NIR spectroscopy measurements of soil total nitrogen and total carbon. *Sensors* **17**(10), 2428 (2017).
- Wang, J., Chen, Y., Chen, F., Shi, T. & Wu, G. Wavelet-based coupling of leaf and canopy reflectance spectra to improve the estimation accuracy of foliar nitrogen concentration. *Agr. Forest Meteorol.* **248**, 306–315 (2018).
- Hong, Y. *et al.* Combining fractional order derivative and spectral variable selection for organic matter estimation of homogeneous soil samples by vis-NIR spectroscopy. *Remote Sens.* **10**(3), 479 (2018).

25. Sorenson, P. T. *et al.* Monitoring organic carbon, total nitrogen, and pH for reclaimed soils using field reflectance spectroscopy. *Can J Soil Sci.* **97**(2), 241–248 (2017).
26. Gomez, C., Rossel, R. A. V. & Mcbratney, A. B. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma* **146**(3–4), 403–411 (2008).
27. Shi, T. Z. *et al.* Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy. *Plant Soil.* **366**(1–2), 363–375 (2013).
28. Stenberg, B., Rossel, R. A. V., Mouazen, A. M. & Wetterlind, J. Chapter five-visible and near infrared spectroscopy in soil science. *Adv. Agron.* **107**, 163–215 (2010).
29. Uddin, M. P., Mamun, M. A. & Hossain, M. A. PCA-based feature reduction for hyperspectral remote sensing image classification. *IETE Tech. Rev.* **5**, 1–21 (2020).
30. Cambule, A. H., Rossiter, D. G., Stoorvogel, J. J. & Smaling, E. M. A. Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park Mozambique. *Geoderma* **183**, 41–48 (2012).
31. Kawamura, K. *et al.* Vis-NIR spectroscopy and PLS regression with waveband selection for estimating the total C and N of paddy soils in Madagascar. *Remote Sens.* **9**(10), 1081 (2017).
32. Leone, A. P., Viscarra-Rossel, R. A., Amenta, P. & Buondonno, A. Prediction of soil properties with PLSR and vis-NIR spectroscopy: Application to mediterranean soils from southern Italy. *Curr. Anal. Chem.* **8**(2), 283–299 (2012).
33. Wang, S., Chen, Y., Wang, M., Zhao, Y. & Li, J. SPA-based methods for the quantitative estimation of the soil salt content in saline-alkali land from field spectroscopy data: A case study from the Yellow River irrigation regions. *Remote Sens.* **11**(8), 967 (2019).
34. Barnes, E. M. *et al.* Remote- and ground-based sensor techniques to map soil properties. *Photogramm. Eng Rem S.* **69**(6), 619–630 (2003).
35. Priori, S. *et al.* Field-scale mapping of soil carbon stock with limited sampling by coupling gamma-ray and vis-NIR spectroscopy. *Soil Sci Soc Am J.* **80**(4), 954–964 (2016).
36. Amin, I., Fikrat, F., Mammadov, E. & Babayev, M. Soil organic carbon prediction by vis-NIR spectroscopy: Case study the Kur-Aras plain Azerbaijan. *Commun. Soil Sci. Plan.* **51**(6), 726–734 (2020).
37. Yu, L. *et al.* Hyperspectral estimation of soil organic matter content based on partial least squares regression. *Trans. CSAE.* **31**(14), 103–109 (2015).
38. Liu, Y. F., Lu, Y. N., Guo, L., Xiao, F. T. & Chen, Y. Y. Construction of calibration set based on the land use types in visible and near-infrared (VIS-NIR) model for soil organic matter estimation. *Acta Pedol. Sin.* **53**, 332–341 (2016).
39. Zhou, X. M. & Zhang, T. Analysis of the April 2019 atmospheric circulation and weather. *Meteor. Mon.* **45**(7), 1028–1036 (2019).
40. Guan, L. & Zhang, T. Analysis of the May 2019 atmospheric circulation and weather. *Meteor. Mon.* **45**(8), 1181–1188 (2019).
41. Li, X., He, Y. & Wu, C. Non-destructive discrimination of paddy seeds of different storage age based on Vis/NIR spectroscopy. *J. Stored Prod. Res.* **44**(3), 264–268 (2008).
42. Boško, M. & Bensa, A. Prediction of soil organic carbon using VIS-NIR spectroscopy: Application to Red Mediterranean soils from Croatia. *Eurasian J. Soil Sci.* **6**(4), 365–373 (2017).
43. McCarty, G. W., Reeves, J. B. III., Reeves, V. B., Follett, R. F. & Kimble, J. M. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci. Soc. Am. J.* **66**(2), 640–646 (2002).
44. Gholizadeh, A. *et al.* Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features. *Soil Water Res.* **10**(4), 218–227 (2015).
45. Wang, X., Xue, L., He, X. W. & Liu, M. H. Vitamin C content estimation of chilies using Vis/NIR spectroscopy. *Int. Conf. Electr. Inf. Control Eng.* **2011**, 1894–1897 (2011).
46. Lee, K. S. *et al.* Wavelength identification and diffuse reflectance estimation for surface and profile soil properties. *Am. Soc. Agric. Biol. Eng.* **52**(3), 683–695 (2009).

Acknowledgements

This work was supported by National Natural Science Foundation of China (31871571, 31371572), Outstanding Doctor Funding Award of Shanxi Province (SXYBKY2018040), Scientific and Technological Innovation Fund of Shanxi Agricultural University (2018YJ17), Applied Basic Research Project of Shanxi Province (201801D221299), and Science and Technique Innovation Project of Shanxi Agricultural University (2020BQ32).

Author contributions

Study concept and design: M.C.F., and Y.K.X. Data analysis and drafting of the manuscript: C.B.Y. Experimental participants: L.F.S., B.H.J., Y.K.X., and M.S. Critical revision of the manuscript for important intellectual content: M.C.F., C.W., L.J.X., M.J.Z., and X.Y.S. Obtained funding: W.D.Y., C.W., and L.J.X. Study supervision: M.C.F.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021