INDIRECT ESTIMATION OF SOIL WATER CONTENT IN PLOUGH LAYER 
BASED ON TOPSOIL SPECTRUM

Zhai Haoran¹, Li Xican²*, Zhong Hao³

¹ College of Information Science and Engineering, Shandong Agricultural University, Tai’an – 894562458@qq.com
² College of Information Science and Engineering, Shandong Agricultural University, Tai’an – kxc@sdau.edu.cn
³ College of Information Science and Engineering, Shandong Agricultural University, Tai’an – 260919837@qq.com

KEY WORDS: Soil Water content; Hyperspectral Remote Sensing; Indirect Estimation; Estimation Model

ABSTRACT:

In order to estimate the soil water content of the plough layer using the surface soil spectrum, based on the water content data and hyperspectral data of 85 surface layers and ploughed soil samples in Jiayang County, Shandong Province, 10 transformation methods such as square root are used. The spectral reflectance of the topsoil is transformed, and the estimation factor is selected according to the principle of maximal correlation. Then, according to the correlation between the water content of the surface layer and the plough layer, two indirect estimation models are constructed. Six kinds of modeling methods such as multiple linear regression, partial least squares regression and support vector machine were used to estimate the soil water content of the cultivated layer, and the effectiveness of the two indirect estimation models was compared and analyzed. The results show that the decision coefficient $R^2$ of the first mode is 0.8138, the average relative error is 9.6061% and the decision coefficient $R^2$ of the second mode is 0.8115, the average relative error is 12.1403%. The researches show that it is feasible and effective to estimate the soil water content of the plough layer by using the surface soil spectrum. It provides a new idea for the rapid diagnosis of soil moisture and nutrients by using optical satellite remote sensing technology.

1. INTRODUCTION

Soil water content is an important part of soil and an important indicator for land resource evaluation and crop growth monitoring. It has an irreparable influence on crop growth and development (He Haibing, 2016). In particular, surface soil water content is sensitive to micro-meteorological responses, which can be used to effectively monitor soil and crop drought conditions (Zhao Wenzhi, 2006). However, traditional soil water content determination methods are difficult to meet the needs of rapid detection of soil moisture. Remote sensing technology is widely used in the estimation of soil moisture by its advantages of large area, non-contact and timeliness (Ma Chunya, 2018).

At present, domestic and foreign scholars have conducted a lot of research on the spectral estimation of soil water content. Studies have shown that soil reflectivity decreases with increasing water content within a certain range, which provides a basis for soil water inversion (Etienne Muller, 2001; LOBELL D B, 2002; Kaleita A L, 2005; Pate M, 2015). The soil moisture has obvious absorption peaks near the 1400 nm, 1900 nm and 2200 nm bands. The quantitative estimation of soil water content can be achieved by extracting spectral absorption parameters at these wavelengths (Lu Yanli, 2018; Lesaignounx A, 2009/2010; Huijun L, 2018). SANTRA and HUMMEL studied the variation of soil reflectance spectra under different water content conditions in the laboratory, and carried out soil moisture prediction (SANTRA P, 2009; HUMMEL J W, 2001). Wang Xiao found through the soil samples and spectral data collected in Hengshan County, Shaanxi Province, that the soil water content spectral characteristics were sensitive at 1350–1400 nm and 1880–1920 nm when the soil water content was 10% (Wang Xiao, 2012). He Ting extracted the spectral characteristics of the two absorption valleys of 1450 nm and 1925 nm of loess, sand and wind sand, and established a prediction model of soil water content (He Ting, 2006). It has been found that spectral data transformation is beneficial to the mining and extraction of spectral information, and thus improves the accuracy and stability of the prediction model. Kemper performed a first-order differential treatment on the original spectral reflectance of the soil, which improved the monitoring accuracy of soil heavy metals and clay content (Kemper T, 2002). Jia Xueqin and others have significantly improved the accuracy of the model by first-order differential transformation of the logarithm of the original spectrum (Jia Xueqin, 2018). Yin Yebiao and others through the soil spectrum to the envelope and then first-order differential, the predicted water content $R^2$ reached 0.855 (Yin Yebiao, 2011). With the deepening of research, the method of estimating soil water content by using soil spectral characteristics is gradually enriched, and statistical models such as multiple linear regression (Yan Changrong, 2008; Jia Jitang, 2013; Yu H, 2017) and partial least squares regression (Peng Sengmin, 2010; Raj A, 2018) gradually move to support vector machines (Bordoni M, 2018; Wang Haijiang, 2018; Li Binnan, 2019), BP neural network (Guo Qingchun, 2012; Hassan-Esfahani L, 2015) and other mechanism models developed. Jin Huing established a linear regression model by analyzing the spectral absorption characteristics of black soil, and the
2. EXPERIMENTAL DATA ACQUISITION

2.1 Experimental Area Overview

Jiyang County is located in the lower reaches of the Yellow River, the southern part of the Lubei Plain (latitude 36°41'—37°15’ north, east longitude 116°52'—117°27’ east). Jiyang County is developed on the alluvial mother material of the Yellow River, and the soil layer is deep; the land use types include cultivated land, garden land, forest land, unused land and some tidal flats. Jiyang County has 3 soil types, 5 subcategories, 9 soil genuses, and 84 soil types. The main soil type is fluvo-aquic soil. The tidal soil accounts for 99.41% of the soil area, the saline soil accounts for 0.30%, and the aeolian sandy soil accounts for 0.29%. The soil distribution types are sandy soil, loamy soil, clay soil, loamy soil and sandy soil from southwest to northeast. It belongs to the warm temperate monsoon climate. The annual average temperature is 12.8 °C, and the average annual precipitation is 586.9 mm.

2.2 Sample Collection and Spectrometry

Before sampling, according to the collected terrain, landform, land use status and other information related to the test area, the distribution of sampling points is reasonably determined. Samples were randomly placed in each sampling area according to the predetermined route. At each sample point, about 0.5 cm of topsoil and 0-20 cm of ploughed soil were collected, respectively, 85, and the soil samples were sealed, numbered sequentially. During the sampling process, the soil type of each soil sample was recorded and simultaneously positioned using a handheld GPS instrument.

Based on the data of 85 soil samples from Jiyang County, Shandong Province, this paper constructs two estimation models of soil moisture content based on surface soil spectrum, and realizes indirect estimation of soil moisture content in cultivated soil, aiming to utilize optical satellites. The remote sensing technology provides a technical reference for quickly estimating soil moisture content such as soil moisture in the plough layer.

3. DATA PROCESSING AND MODELING METHOD

3.1 Data Preprocessing

Using the ViewSpecPro spectral data processing software, the spectral curve of each soil sample was corrected for breakpoints, and then the average of 10 spectral curves was calculated and used as the spectral reflectance of the corresponding sample. Because of the various factors, the spectral curve has a lot of small noise. In order to reduce the influence of glitch noise and make the spectral curve more stable, the spectral curve is smoothed by a 9-point weighted moving average method.

In order to prevent the abnormal samples from greatly affecting the experimental results, the spectral reflectance data of the soil samples are sorted according to the order of water content from small to large, and the scatter plots are drawn to observe the trend of the spectral reflectance curves and eliminate the abnormal samples. A total of 4 abnormal samples numbered 7, 45, 46, and 62 are excluded from the study. Then the remaining 81 samples are used for modeling analysis.

3.2 Spectral Transformation

Studies have shown that spectral data transformation has good effects in eliminating background and noise interference, reducing internal factors and amplifying spectral absorption characteristics, which is beneficial to the mining and extraction of spectral information, and thus improves the accuracy and stability of the prediction model. The original spectra were processed by 10 transformation methods such as square, logarithm, reciprocal, square root, first-order differential and their combination to improve the correlation between spectral data and water content. The comparative
analysis shows that the spectral reflectance of the topsoil is better after square root transformation, as shown in Figure 1.

![Figure 1. Correlation Coefficient between Water Content and Square Root Transformation of Reflectivity](image1.png)

It can be seen from Figure 1 that after the spectral reflectance of the soil is transformed by the square root, the correlation with the soil water content is significantly improved at different bands. There is a significant improvement at 500nm~700nm, 1450nm~1800nm and 1950nm~2450nm, and the correlation coefficient in the range of 1950nm~2450nm is up to 0.84.

### 3.3 Selection of Characteristic Factor

From the correlation coefficient curve shown in Figure 1, according to the principle of maximal correlation and the principle of dispersion as much as possible between the bands, the soil water content spectral feature inversion factor is selected, as shown in Table 2.

| Wavelength (nm) | Correlation Coefficient |
|-----------------|-------------------------|
| 1349            | −0.82                   |
| 1801            | −0.83                   |
| 1985            | −0.84                   |
| 1986            | −0.84                   |
| 2100            | −0.84                   |

It can be seen from Table 2 that the correlation coefficients at the bands 1349nm, 1801nm, 1985nm, 1986nm and 2100nm are relatively high, and the correlation coefficient is about 0.83, which has significant correlation and can be selected as the inversion factor.

### 3.4 Modeling Method

This paper constructs two indirect estimation models for the soil moisture content of plough layers.

The first model is an indirect estimation of the soil moisture content of the plough layer based on the spectral characteristics of the surface soil, that is, the inversion factor extracted from the surface spectral information and the soil moisture content of the plough layer are modeled to obtain the estimated soil water content of the plough layer.

The second model is based on the indirect estimation of the soil moisture content of the plough layer based on the relationship between the plough layer and the surface soil water content. Firstly, the surface soil water content estimation model is established based on the surface soil spectrum, and the estimated value of the surface soil water content is calculated. Then according to the measured data of soil surface layer and plough layer water content, the relationship between soil surface layer and plough layer water content is established, and then the estimated value of surface soil water content is substituted into the relationship, and the estimated value of water content of soil plough layer is calculated. The implementation process of two indirect estimation methods is shown in Figure 2.

![Figure 2. Two models of indirect estimation](image2.png)

For comparative analysis, this paper adopts linear modeling methods such as multiple linear regression (SLR), principal component regression (PCR) and partial least squares regression (PLSR), as well as support vector machine (SVM) and BP neural network (BP-NN). Nonlinear modeling methods such as RBF neural network (RBF-NN) are used to model and compare accuracy. The accuracy of the model is measured by the decision coefficient $R^2$ and the average relative error.

### 4. RESULTS AND ANALYSIS

#### 4.1 Correlation between Surface Soil and Cultivated Soil Water Content

Because the surface soil and the ploughed soil have great differences in temperature, humidity, light, ventilation, microbes and other conditions, and the soil moisture has infiltration phenomenon, the soil water content is not the same in the surface layer and the plough layer soil. Under normal circumstances, the surface soil is susceptible to external conditions such as sunlight and wind, and the water is easy to evaporate, so the water content of the ploughed soil is generally higher than that of the surface soil. A scatter plot is drawn using measured values of surface soil and tillage soil water content, as shown in Figure 3.

![Figure 3. Relationship between surface soil and cultivated soil water content](image3.png)
It can be seen from Figure 3 that there is a significant correlation between the surface soil water content and the soil moisture content of the plough layer, and the determination coefficient $R^2$ is 0.8424. It provides an important basis for the indirect estimation of soil moisture content in the stratified soil using surface soil spectroscopy.

4.2 Indirect Estimation Model of Tillage Soil Water Content Based on Surface Soil Spectrum

According to the estimation factors based on surface soil spectrum and the measured values of soil water content, six methods including multiple linear regression, principal component regression, support vector machine, BP neural network and RBF neural network were used. The multivariate linear regression experiment was carried out by using the statistical analysis software IBM SPSS Statistics 23; the Principal Component Regression and Least Squares Regression experiments were carried out by using the data processing software DPS7.05; the BP neural network and RBF neural network experiments were performed by using MATLAB2017b. A scatter plot is drawn from the estimated results of the six models and the measured values of the soil moisture content of the plough layer, as shown in Figures 4-9. The estimation accuracy of the six models is shown in Table 3.

![Figure 4. Estimated result of MLR](image1)

![Figure 5. Estimated result of PCR](image2)

![Figure 6. Estimated result of PLSR](image3)

![Figure 7. Estimated result of BP-NN](image4)

![Figure 8. Estimated result of SVM](image5)

![Figure 9. Estimated result of RBF-NN](image6)

| Modeling Method | Decisive Factor | Average Relative Error/% |
|-----------------|----------------|--------------------------|
| MLR             | 0.8115         | 12.1403                  |
| PCR             | 0.7164         | 16.4720                  |
| PLSR            | 0.6680         | 14.6546                  |
| BP—NN           | 0.7173         | 15.7129                  |
| SVM             | 0.7109         | 14.8668                  |
| RBF—NN          | 0.5506         | 17.5370                  |

It can be seen from Figures 4-9 and Table 3 that the estimation results of the six models are quite different, and the estimation accuracy is not the same. Except for the RBF neural network model, the determination coefficient $R^2$ of other models is about 0.7~0.8, among which the multivariate linear regression model has the highest estimation accuracy, the determination coefficient $R^2$ is 0.8115, and the average relative error is 12.1403%.

4.3 Indirect Estimation Results of The Second Model

According to the surface spectral characteristic factors of the
selected 81 samples, the second model was used to calculate the estimation results, and the results were analyzed with the measured values to obtain the following Table 4.

Table 4. Estimation accuracy comparison table of estimation mode two

| Modeling Method | Decisive Factor | Average Relative Error/% |
|-----------------|-----------------|-------------------------|
| MLR             | 0.8138          | 9.6601                  |
| PCR             | 0.6685          | 16.3818                 |
| PLSR            | 0.6690          | 16.3696                 |
| BP—NN           | 0.6781          | 14.1557                 |
| SVM             | 0.7949          | 12.7838                 |
| RBF—NN          | 0.6102          | 16.7429                 |

It can be seen from Table 4 that the results of multiple linear regression and support vector machine models in the six models are better, and the other models are less accurate.

4.4 Comparative Analysis of Results

It can be seen from the comparison between Table 3 and Table 4:

(1) The estimation accuracy of different models in the two patterns is different. The results of the multiple linear regression model in Pattern 1 are better, the decision coefficient $R^2$ reaches 0.8, and the estimation accuracy of the other models is slightly worse. The decision coefficient $R^2$ is about 0.55$-$0.71, and the multivariate linear regression model is the best. The results of multiple linear regression model and support vector machine model in Pattern 2 are better. The coefficient of determination $R^2$ is about 0.8. The estimation accuracy of other models is slightly worse. The coefficient of determination $R^2$ is about 0.6, and the results of multiple linear regression models are optimal.

(2) The optimal result of Pattern 1 and Pattern 2 is not much different. Relatively speaking, the average relative error of the estimated result of Pattern 2 is small, 9.6601%. The reason is that the field environment is complex and diverse, and the water content changes with time, temperature, humidity and other conditions, resulting in soil water content. There is an error in the estimation. Although the average relative error of the Pattern 1 results is slightly lower than that of the Pattern 2, the estimation accuracy is still high, so both patterns can be used to estimate the water content of the plough layer. Among the two indirect estimation methods, the multivariate linear regression model has the highest precision, indicating that there is a clear linear relationship between the soil water content and the spectral data, which is consistent with the existing research results.

(3) Traditional optical remote sensing satellites cannot detect the spectrum of cultivated soils, and it is impossible to promote and use ground remote sensing inversion results. By using the surface soil spectral information to indirectly estimate the soil moisture content of the plough layer, a joint point between the research results of aerospace and aerospace optical remote sensing technology and the ground hyperspectral research results is found, which can promotes the monitoring capacity of soil water content by using optical remote sensing technology.

5. CONCLUSION

Based on the water content data and hyperspectral data of 85 surface soils and ploughed soil samples in the tidal soil of Jiyang County, Shandong Province, two different patterns were implemented by data processing, spectral transformation, feature inversion factor selection and other steps, which are used to Indirect estimation of soil moisture in the plough layer. The experimental results show that the estimating accuracy of the soil water content of the six type models are different, and the multiple linear regression accuracy is better, while the optimal estimation results of the two indirect estimation patterns are similar, and the determination coefficient $R^2$ is 0.8115 and 0.8138, the average relative error was 12.1403% and 9.6601%, respectively. This indicates that it is feasible and effective to estimate the soil moisture content of the plough layer by using the surface soil spectrum. It provides a new idea for the rapid diagnosis of soil moisture and nutrients by using optical satellite remote sensing technology.

REFERENCES

Bordoni, M., Bittelli, M., Valentino, R., 2018: Soil Water Content Estimated by Support Vector Machine for the Assessment of Shallow Landslides Triggering: the Role of Antecedent Meteorological Conditions. Environmental Modeling & Assessment 23(4):333-352.

Chen Wenqian, Ding Jianli,Tan Jiao., Li Xiang.,2017: Analysis Vegetation of Hyperspectral Reflectance and Shallow Soil Water Content in Arid Area. Transactions of the Chinese Society for Agricultural Machinery 2017(09):1-16.

Etienne Muller, Henri Décamps,2001: Modeling Soil Moisture—reflectance. Remote Sensing of Environment 76(2):173-180.

Guo Qingchun, Wang Sujian,He Zhenfang.,2012: Study on Forecasting Model of Soil Water Content based on BP Artificial Neural Network. Shandong Agricultural Sciences 44(12):11-15.

Hassan-Esfahani, L., Torres-Rua, A., Jensen A., 2015: Assessment of Surface Soil Moisture Using High-Resolution Multi-Spectral Imagery and Artificial Neural Networks. Remote Sensing 7(3):2627-2646.

He Haibing, Wu Lianquan,Yang Ru, Ma Fuyu,Huang Yide.,2016: Photosynthesis Characteristics and Transpiration Efficiency of Rice Plants under Controlled Irrigation Technology in Arid Region. Transactions of the Chinese Society for Agricultural Machinery 47(9):186-193.

He Ting,Wang Jing, Chen Ye,Lin Zongjian.,2006:Spectral features of soil moisture. Acta Pedologica Sinica (06):1027-1032.

Huijun, L., Hongjun., S., Bo.,Z., 2018: Hyper-spectral Multiple Features Optimization Using Improved Firefly Algorithm. Remote Sensing and Application. HUMMEL, J W., SUDDUTH, K A., HOLLINGER, S E., 2001: Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor. Computers and Electronics in Agriculture 32:149-165.

Jia Jitang, Cheng Linlin, Yu Yang, Wang Pengfei, Ren Juntao, Meng Haocan.,2013: A Hyper-spectral remote sensing model
of soil moisture content based on multivariate statistical analysis. *Journal of Geomatics Science and Technology* 30(06):614-618.

Jia Xueqin, Feng Meichen, Yang Wude, Wang Chao, Sun Hui, Wu Gaihong, Zhang Song., 2018: Study on the spectral prediction model of soil moisture content based on SPA-MLR method. *Agricultural Research in the Arid Areas* 36(03):266-269+291.

Jin Huining, Zhang Xinle, Liu Huanjun, Kang Ran, Fu Qiang, Ning Donghao., 2016: Soil Moisture Predicting Model Based on Spectral Absorption Characteristics of the Soil. *Acta Pedologica Sinica* 53(03):627-635.

Kaleita. A L, Tian. L F, Hirschi M C., 2005: Relationship between Soil Moisture Content and Soil Surface Reflectance. *Transactions of the Asae* 48(48): 1979-1986.

Kemper. T., Sommer. S., 2002: Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environmental Science & Technology* 36(12): 2742-2747.

Lesaignoux. A, Fabre. S., Briottet. X., et al., 2010: Estimation of Soil Moisture Content of bare soils from their spectral optical properties in the 0.4 – 12 μm spectral domain. *Geoscience & Remote Sensing Symposium. IEEE, 2010.*

Li Binnan, Fan Guisheng., 2019: Prediction Model of Soil Moisture Characteristic Curve Based on Support Vector Machine. *Water Saving Irrigation* 2019(01):108-111+117.

LOBELL. D B., ASNER. G P., 2002: Moisture Effects on Soil Reflectance. *Soil Science Society of America Journal* 66 (3):722-727.

Ma Chunya, Wang Jinglei, Huang Xiuqiao., 2018: Progress of Retrieving Soil Moisture by Remote Sensing. *Water Saving Irrigation* 2018(08):70-74+78.

Pate. M., 2015: A Spectral Reflectance Analysis of Soil Moisture Content. *Dissertations & Theses- Gradworks 2015.*

Peng Sengmin, Lin Cunbin, Huang Jiaxin, Wang Fulin., 2010: The prediction of soil moisture model based on partial least-squares regression. *Journal of Agricultural Mechanization Research* 32(09):45-49.

Raj. A, Chakraborty. S., Duda. B M., et al., 2018: Soil mapping via diffuse reflectance spectroscopy based on variable indicators: An ordered predictor selection approach. *Geoderma* 314:146-159.

SANTRA. P., SAHOO. R N., DAS. B S., SAMAL. R N., PATTANAIK. A K., GUPTA V K., 2009: Estimation of Soil Hydraulic Properties using Proximal Spectral Reflectance in Visible, Near-infrared, and Shortwave-infrared (vis-nir-swir) Region. *Geoderma* 152(3/4): 338-349.

Wang Haijiang, Jiang Tianchi, YUNGER John A, et al., 2018: Hyperspectral Inverse Model for Soil Salt Ions Based on Support Vector Machine. *Transactions of the Chinese Society for Agricultural Machinery* 49(05):263-270.

Wang Xiao., 2012: Studies on soil moisture Hyper-spectrum characteristics and estimating model. *Shandong Agricultural University, 2012.*