PERFORMANCE OF PRETREATMENTS AND MULTIVARIATE METHOD ON THE HYPERSPECTRAL ESTIMATION OF SOIL MOISTURE CONTENT

YAN, X. B.[1] – WANG, Y. X.[1] – ZHANG, X.[1] – WANG, Z. G.[1] – YANG, S.[1] – LI, Y.[1] – YANG, C. B.[1] – FENG, M. C.[1] – SONG, X. Y.[1] – ZHANG, M. J.[1] – XIAO, L. J.[1] – FAHAD, S.[2] – YANG, W. D.[1*] – WANG, C.[1*]

[1]*College of Agriculture, Shanxi Agricultural University, Taigu, China*

[2]*Institute of Molecular Biology and Biotechnology, The University of Lahore, Pakistan*

**Corresponding authors*
e-mail/phone: sxauywd@126.com/+86-138-3483-5129 – W. D. Yang;
wcqxx2005@126.com/+86-153-0354-3914 – C. Wang

(Received 6th Jan 2022; accepted 21st Mar 2022)

Abstract. Soil moisture controls the exchange of energy between the land surface and the atmosphere and is a significant factor affecting plant growth and productivity. Hyperspectral monitoring of soil water fraction could provide a theoretical basis for real-time estimation of spatial and temporal variations in soil moisture. To quantitatively evaluate the hyperspectral monitoring of soil moisture content (SMC): the SMC and its corresponding spectral reflectance were measured in the laboratory. In addition, the original spectral data was pre-processed by single and multiple transformations to study the effect of spectral preprocessing methods on the quantitative evaluation of soil moisture. The successive projections algorithm (SPA) was used to extract the corresponding wavelengths of soil moisture and the spectral monitoring model was established by using the partial least squares (PLS). The results show that (1) SMC and spectral reflectance show an obvious negative correlation, and the spectral reflectance gradually decreases with the increase of SMC. (2) Appropriate pretreatment methods can improve the correlation between SMC and spectral reflectance and improve the accuracy of the SMC monitoring model, of which T19 (R^2 + SNV + FD) is the best spectral pretreatment method. (3) The optimal SMC monitoring model is T19-SPA-PLS (R^2v = 0.986, RMSEv = 1.824, RPD = 8.239). This study provided a reference for spectral data processing and an effective method for the accurate estimation of SMC using hyperspectral remote sensing.
Keywords: *soil moisture content, multivariate statistical analysis, spectral features, partial least squares, remote sensing*

Introduction

Soil surface moisture plays an important role in the exchange of water and heat energy between the land surface and the atmosphere (Shepherd et al., 2002). Soil moisture content (SMC) is an important factor affecting crop growth and development as well as an indicator of drought stress (Yuan et al., 2019). Real-time accurate determination of soil moisture is extremely difficult as traditional approaches cannot be used on such a large scale under field conditions (Zhang et al., 2020). Soil hyperspectral technology is characterized by a huge amount of information, fast operation without damaging and pollution concerns; and is preferred in soil water content estimation (Zeng et al., 2017; Yu et al., 2017; Cai et al., 2018). Monitoring SMC on different spatial scales is of great significance for formulating scientific and reasonable irrigation plans, realizing efficient utilization of water resources, and to improve crop yields with efficient irrigation systems (Hasan., 2014).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

Hyperspectral techniques could be used to monitor SMC as there is a strong correlation between soil moisture content and hyperspectral reflectance (Qi et al., 2017). To be more precise, spectral reflectance will decrease with the increase of soil moisture content within a certain range (1400~1900 nm range). Based on the excellent correlation between soil moisture and spectral reflectance, many studies have established soil moisture monitoring models (Sanchez et al., 2014; Brosinsky et al., 2014). However, the accuracy of these soil moisture monitoring models is still needed to be improved (Yuan et al., 2019). It is pertinent to mention here that spectral data collection is often affected by instrumental errors, environmental changes, and extraneous factors (Liu et al., 2004). Therefore, it is imperative to preprocess the spectral data before modeling to improve the accuracy of the model (Xu et al., 2016; Wu et al., 2020). The common soil spectral pre-treatment methods mainly include smooth denoising, normalization, first-order derivative and multivariate scattering correction, and so on (Rinnan et al., 2009). The use of spectral pretreatment methods can in turn improve the accuracy of monitoring models (Guo et al., 2014; Wu et al., 2018). It has been proved that the spectral pretreatments and the combination of appropriate calibrated models can generally make a positive contribution to the predictive model (Li et al., 2014). By contrast, there is no single or combined of pre-processing method suitable for all different soil situations (Dotto et al., 2018). Therefore, exploration of different preprocessing methods before model calibration can provide a new perspective for the improvement of model accuracy (Zhang et al., 2014; Wang et al., 2017).

Soil texture and type can have a significant impact on the accuracy of soil moisture content monitoring models. He (2006) showed that soil type can seriously affect the reflectance of soil spectra because different types of soils have different physicochemical characteristics. For example, sandy soils have greater spectral reflectance than loamy soils. Lu (2018) showed that the accuracy and stability of the SMC prediction model improved as the soil particle size became smaller. Many studies have shown that it is difficult to analyze the effect of SMC on spectral reflectance when mixing different types of soils for a study (Liu et al., 2014). Therefore, in order to avoid the influence of factors other than SMC on the spectral properties of soils, the spectral reflectance of different soil moisture contents was measured under laboratory conditions in this study with a single soil type as the study object. In this study, we adopted a variety of pretreatment methods to transform the original spectral data and then used SPA to extract the characteristic wavelength of the preprocessed spectral data. Finally, we constructed PLS models of SMC on the full spectrum and on the characteristic wavelengths, respectively. The main objective of this study was to explore the best spectral pre-treatment method to achieve accurate monitoring of SMC.

Materials and methods

Soil sample collection and moisture treatments

The soil samples were collected from Jinzhong City, Shanxi Province, China (37.6874° N, 112.7527° E). The soil is calcareous yellow-brown soil developed from loess parent material with a medium soil fertility (Wang et al., 2016). The soil was collected from 0 – 20 cm depth, homogenized, passed through a 2 mm sieve, and finally dried in an incubator at 105 °C for 24 h. Briefly, 10 soil samples of 100 g each were placed inside a box (diameter of 9 cm and a height of 2 cm). After its surface is scraped

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

flat, distilled water is slowly injected into the soil until soil saturation is achieved. Afterward, the soil samples were placed at room temperature to allow water evaporation naturally. Spectral information of soil samples was recorded regularly during the air-drying process and soil moisture content was determined on a gravimetric basis. A total of 120 soil spectral data were collected and we divided them into the calibration set (n = 80) and the validation set (n = 40).

Determination of soil moisture content

To avoid the error caused by water evaporation, the water content of the soil samples was estimated immediately after the soil spectral reflectivity data was collected. The SMC is expressed by θ and is calculated as follows:

$$\theta = (W_2 - W_0)/(W_0 - W_1) \times 100\% \quad \text{(Eq.1)}$$

where W_0 is the total weight of the dried soil sample and the black plastic box; W_1 is the weight of the black plastic box, and W_2 is the weight of the plastic box and the soil sample during air drying.

Spectral measurements

Spectral reflectance of soil was measured by using a Field Portable Spectrometer (ASD FieldSpec.3) inside a dark room. *Figure 1* shows the soil samples and experimental equipment. The device has a band range of 350~2500 nm, with a sampling interval of 1.4 nm (350~1000 nm) and 2 nm (1000~2500 nm): and a resampling interval of 1 nm, with a total of 2,151 bands. The light source of the spectrometer is a 50 W halogen lamp. When measuring the soil spectral reflectance, the light source was 30 cm away from the surface of the soil sample, and the zenith angle of the light source remained 15°. The sensor of the spectrometer was located 10 cm above the surface of the soil sample, the whiteboard calibration was carried out before each soil spectrometric determination. During the collection of spectral reflectance of soil, the frequency of our collection of spectral reflectance decreases as the moisture content of the soil decreases. When the SMC is high, the soil spectral reflectance is measured every 4 h because the water evaporates more quickly. When the SMC is low, the soil spectral reflectance is measured every 8 h because the soil evaporates more slowly. We set the scan time for a spectral curve to 5 s. While determining the spectral reflectance, four angles were measured for each soil sample. After an angle is determined, the soil sample was rotated 90 degrees to measure the next Angle, and therefore 10 spectral curves were collected from each angle. After the removal of abnormal soil spectra, the collected spectral data was averaged and processed as the final spectrum of the soil sample.

Spectral data preprocessing and analyses

We used ViewSpecPro software to remove outliers, rectify spectral breakpoints, and normalize the spectral data. Moreover, the Unscrambler X 10.4 software was used to preprocess spectral data, Matlab 2018 software was used to extract characteristic bands and establish models, and finally, the data was mapped using Origin 2021 software. Before the use of the spectral data, the spectral edge with low signal-noise ratio (350~399 nm and 2451~2500 nm) affected by the internal noise of the spectrometer was eliminated, and then the original spectral data were preprocessed with 20 methods in

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

total, including conventional mathematical transformation, standard normal transformation (SNV) and First derivative (FD) as shown in *Table 1*.

Figure 1. *Spectroscopic equipment and soil samples*

Table 1. *The list and details of different methods used during preprocessing of soil spectral data*

Shortened form	Pre-processing method	Shortened form	Pre-processing method
T0	R	T10	R + FD
T1	1/R	T11	1/R + FD
T2	Log R	T12	Log R + FD
T3	$\sqrt{R}$	T13	$\sqrt{R}$ + FD
T4	R^2	T14	R^2 + FD
T5	R + SNV	T15	R + SNV + FD
T6	1/R + SNV	T16	1/R + SNV + FD
T7	Log R + SNV	T17	Log R + SNV + FD
T8	$\sqrt{R}$ + SNV	T18	$\sqrt{R}$ + SNV + FD
T9	R^2 + SNV	T19	R^2 + SNV + FD

R is the original spectral reflectance

The introduction of the partial least squares method

Partial least squares (PLS) is a powerful multivariate statistical tool that has been widely used in many fields. PLS is an effective statistical method, which can compress a large number of related spectral data into several uncorrelated principal components, and mainly solve the regression problem from multiple dependent variables to multiple independent variables. It is especially suitable for the situation where the variables in the prediction matrix are more than the observed values, and the values of independent variables exist multicollinearity. Studies have shown that the PLS0 method can establish a linear equation between soil properties and spectra, and successfully use laboratory or field spectra for soil spectral analysis and soil characterization (Xu et al., 2020).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

Statistical analyses and model evaluation parameters

The determination coefficient (R^2): root mean square error (RMSE): and relative percent deviation (RPD) were selected to evaluate the prediction accuracy of the model. The values of R^2 reflect the stability of model establishment and subsequent verification. The closer R^2 to 1, the better the model is stable and has a higher fitting degree. By contrast, the smaller RMSE values represent the better predictive ability of the model. Finally, the RPD also reflects the predictive ability of the model. If $RPD < 1.4$, the stability of the model is poor; if $1.4 \leq RPD < 2$, the model can make a rough assessment of the samples; if $RPD \geq 2$, the model can make a good prediction of the samples (Guo et al., 2014).

Results

Descriptive statistical analysis for the SMC

In this study, a total of 120 samples were collected, and the samples were sorted from high to low based on the value of SMC, and then, we divided the calibration set and validation set by 2:1. As shown in *Table 2*, the number of samples in the calibration set was 80, including the maximum and the minimum values of SMC with 65.1% and 0.1%, respectively. The number of validations set samples was 40, the maximum value of 65% while, the minimum value of 0.3%. In addition, the correction coefficient (CV) of the calibration set and validation set were 73.7% and 74.6%, respectively. Since the data distribution of SMC is relatively uniform and the share of data in the calibration set is 2/3, the range, minimum, maximum, mean, SD, and CV values of the calibration set and the total sample are basically the same. The validation set, on the other hand, has a smaller number of samples, 1/3 of the total sample, and therefore exhibits slight differences in SD and CV data between the calibration set and the total sample. The CV of SMC is exceeded 36%, which indicates a high dispersion of SMC values.

Table 2. *Statistical characteristics of the SMC*

	Samples	**range**	**Min**	**Max**	**Mean**	**SD (%)**	**Skewness**	**Kurtosis**	**CV**
Calibration set	80	65.0%	0.1%	65.1%	20.1%	14.839	0.765	0.207	73.7%
Validation set	40	64.7%	0.3%	65.0%	20.1%	15.028	0.836	0.566	74.6%
Total sample	120	65.0%	0.1%	65.1%	20.1%	14.839	0.779	0.259	73.7%

SD is the standard deviation. Range represents the difference between the maximum and minimum values. Min. and Max. express the minimum value and maximum value of soil moisture content, respectively

Spectral response on SMC

Figure 2 shows the spectral curves of soils in different moisture ranges. It can be seen that there are three distinct absorption valleys near the 1400 nm, 1900 nm, and 2200 nm bands of the spectral curve, respectively. The absorption valley near the 1900 nm band is the largest, while the absorption valley near the 2200 nm band is the smallest. Moreover, with the increase of SMC, the absorption valley near the 2200 nm band became gradually less obvious. In addition, it can be seen that the spectral reflectance shows a gradual decrease as the soil moisture increases, and it decreases

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

sharply when the soil moisture range increases from 0~15% to 15~30%. It can be seen that the change in reflectance in the visible band (400-780 nm) is significantly smaller than that in the NIR band (780~2450 nm). And in the visible band, the spectral reflectance tends to increase significantly with the increase of wavelength, and the slope of the spectral reflectances decreases continuously with the increase of the soil moisture range.

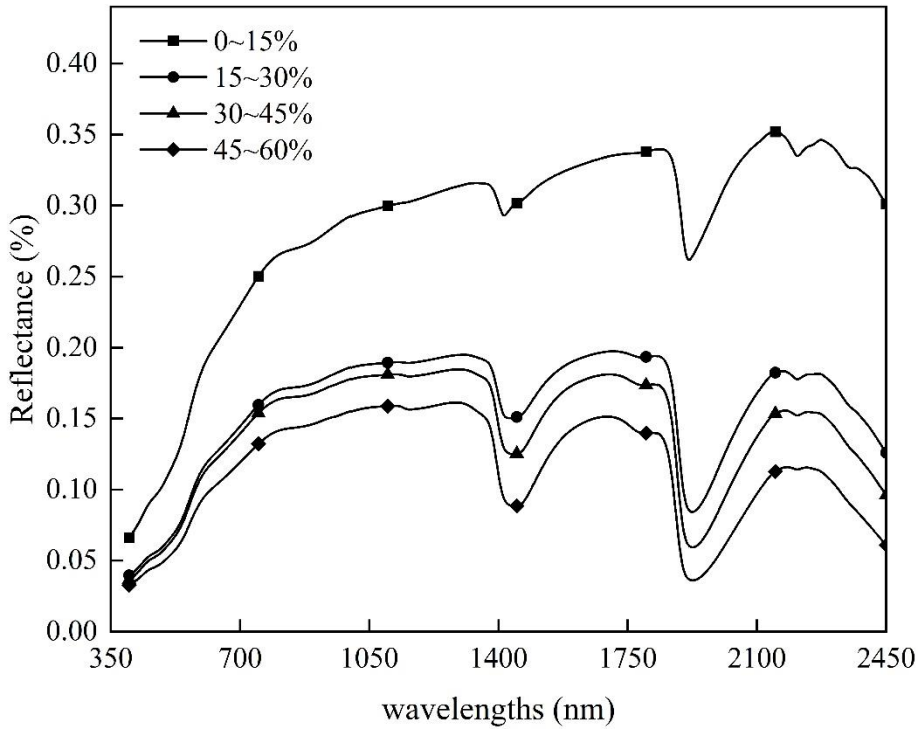

Figure 2. *Soil hyperspectral reflectance curves under different soil moisture ranges*

Correlation between soil spectral reflectance after pretreatment and soil moisture content

Figure 3a shows the correlation between the spectral reflectance and SMC after different pretreatments in the band of 400 nm to 2450 nm. *Figure 3b* shows the position of the highest correlation between the spectral reflectance and SMC with different pretreatments. It can be seen from *Figure 3* that the original spectral reflectance (T0) shows a significant negative correlation with the SMC, and its maximum absolute correlation coefficient (MACC) with SMC is 0.901. The MACC after T4 pretreatment was -0.838, which was significantly lower compared to the MACC of the original spectral reflectance with SMC. However, the MACC between the spectrum reflectance under other Pretreatment methods and the SMC improved to varying degrees. Among them, the MACC of spectral reflectance and SMC after T19 pretreatment showed the largest increase, with the MACC value of 0.991. It can be seen from *Figure 3b* that the MACC values of spectral reflectance and SMC after different pretreatments were distributed in three ranges, 1313~1495 nm, 1763~1943 nm, and 2188~2416 nm, respectively. As shown in the soil spectral reflectance curve in *Figure 2*, there are three obvious moisture absorption valleys in these three band ranges, therefore, these three bands must have a strong correlation with SMC.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

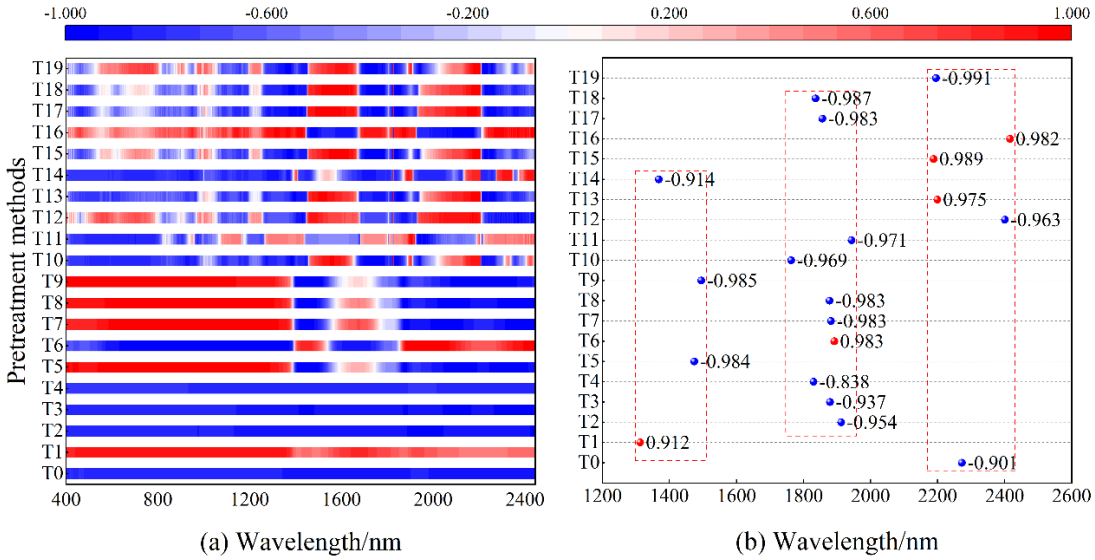

Figure 3. *Correlation of spectral reflectance with SMC after different pretreatments*

Establishment of SMC models based on the full spectrum.

With different pretreatment methods, we constructed full-spectrum PLS monitoring models for SMC, respectively. The models' performance were presented in *Table 3*. It can be seen that the accuracy of the SMC monitoring model decreases after applying the four classical mathematical variants from T1 to T4, as well as after applying FD preprocessing based on T1 to T4 (T11~T14). However, the accuracy of the SMC monitoring model improved after applying SNV preprocessing (T5): FD preprocessing (T10): SNV + FD preprocessing (T15): and applying SNV preprocessing based on T1~T4 (T6~T9). In addition, the accuracy of the SMC monitoring model improved after applying FD preprocessing based on T6~T9 (T16~T19). Comparing the SMC monitoring models constructed after applying different pretreatments, we can conclude that the SMC monitoring model constructed by applying the T19 pretreatment method is the best (R^2v = 0.987, RMSEv = 1.704, RPD = 8.819). To clearly show the PLSR model performance for SMC, the scatter plots and fitting lines of measured and predicted values for SMC were represented in *Figure 4*. As seen from the validation results, the data points in the scatter plot are near the 1:1 fit line, and the fit lines of the calibration and validation sets largely overlap with the 1:1 fit line. It indicates that the T19-PLS model has excellent monitoring ability for SMC.

Analysis of spectral variable selection results based on SPA

To screen out the characteristic wavelengths of SMC and further build the SMC monitoring model with lower complexity, as shown in *Figure 5*, SPA was used to screen the characteristic wavelengths under different pretreatments. It can be seen that the characteristic wavelengths under T0~T9 pretreatments are mainly distributed in four regions, which are 400~715 nm, 1256~1596 nm, 1799~1985 nm, and 2139~2278 nm, respectively. however, the characteristic wavelengths under T10~T19 pretreatment are mainly distributed in 402~585 nm, 805~1089 nm, 1805~1950 nm, and 2161~2278 nm. The reason for this phenomenon may be that T10~T19 applied the FD method, which led to the enhancement of the spectral features in the NIR short-wave region.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

Table 3. *Performance of SMC models constructed by using the full spectrum*

Model types	Calibration set		Validation set		
	R^2c	RMSEc	R^2v	RMSEv	RPD
T0-PLS	0.943	3.520	0.954	3.318	4.529
T1-PLS	0.845	5.815	0.735	11.080	1.356
T2-PLS	0.907	4.504	0.886	5.410	2.778
T3-PLS	0.930	3.907	0.953	3.258	4.613
T4-PLS	0.920	4.170	0.937	3.831	3.923
T5-PLS	0.981	1.998	0.983	1.944	7.730
T6-PLS	0.981	2.007	0.973	2.482	6.055
T7-PLS	0.974	2.376	0.977	2.265	6.635
T8-PLS	0.978	2.177	0.981	2.094	5.175
T9-PLS	0.983	1.936	0.983	1.943	7.734
T10-PLS	0.941	3.568	0.960	2.979	5.045
T11-PLS	0.936	3.741	0.908	5.262	2.856
T12-PLS	0.954	3.149	0.917	4.396	3.416
T13-PLS	0.927	3.991	0.932	4.066	3.696
T14-PLS	0.942	3.544	0.943	3.615	4.160
T15-PLS	0.988	1.617	0.979	2.149	4.157
T16-PLS	0.972	2.462	0.953	3.814	3.940
T17-PLS	0.989	1.538	0.982	2.009	7.480
T18-PLS	0.983	1.943	0.981	2.058	7.302
T19-PLS	0.999	0.517	0.987	1.704	8.819

R^2, RMSE, and RPD represent the determination coefficient, root means square error, and residual prediction deviation, respectively

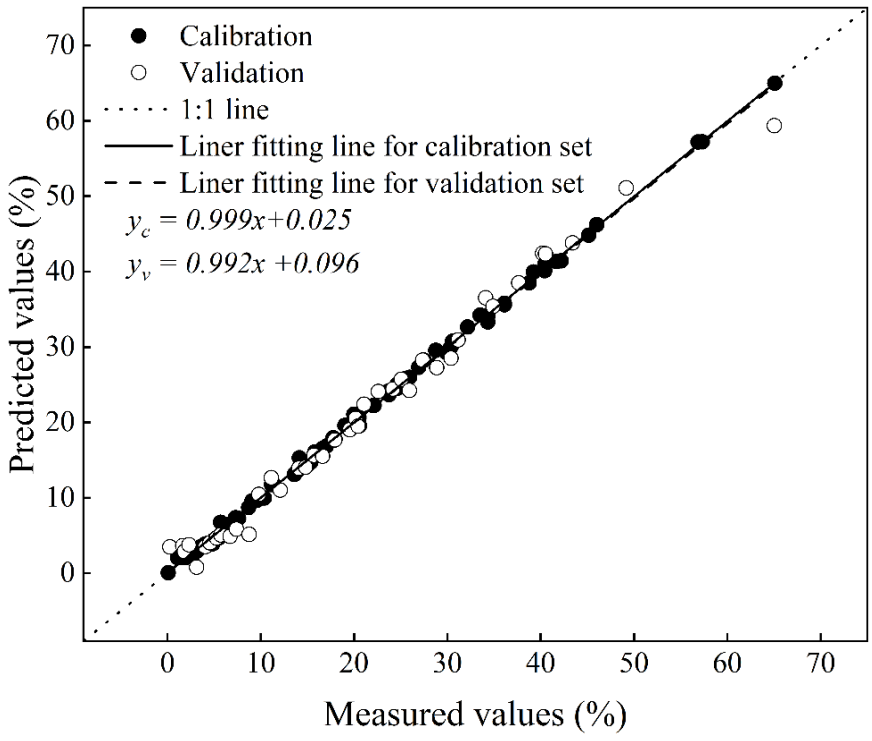

Figure 4. *Relationships between measured values and predicted values of SMC by using the T19-PLS model*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

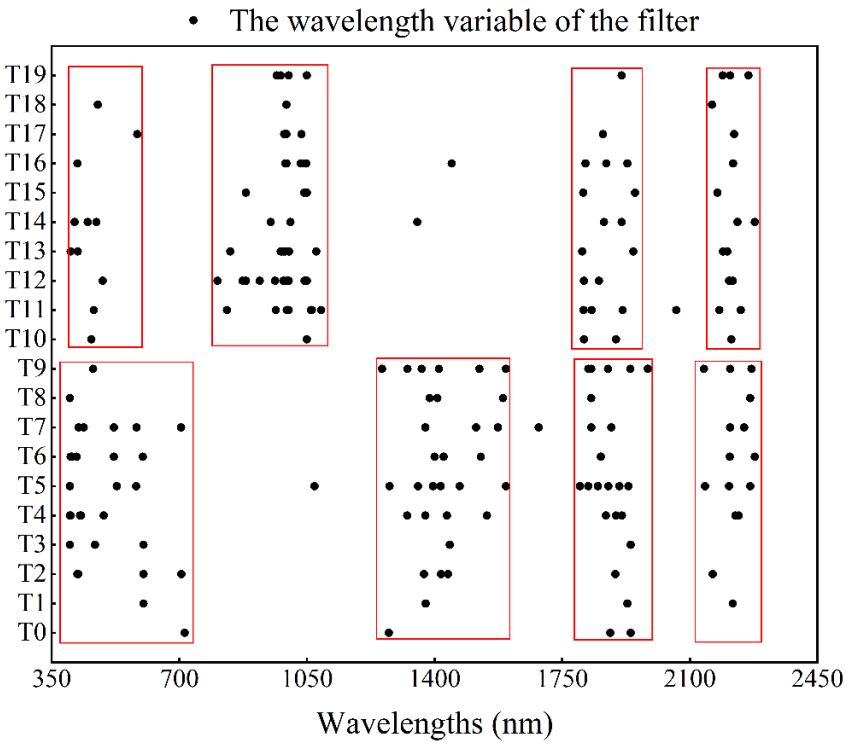

Figure 5. *Results of characteristic wavelengths filtered by SPA*

Establishment of SMC monitoring models based on characteristic wavelengths

Based on the above feature wavelengths screened using SPA under different pretreatments, we constructed SMC monitoring models based on the characteristic wavelengths, and the performance of these models is shown in *Table 4*. It can be seen that the accuracy of the T1-SPA-PLS, T2-SPA-PLS, T11-SPA-PLS, T13-SPA-PLS and T14-SPA-PLS models decreased compared with the T0-SPA-PLS model. However, the accuracy of all other models improved to different degrees, with the T19-SPA-PLS model having the highest accuracy ($R^2v = 0.986$, RMSEv = 1.824, RPD = 8.239). *Figure 6* is a scatter plot of the fit of the predicted and measured values of the T19-SPA-PLS model, from which it is clear that the performance of the T19-SPA-PLS model is outstanding. The fitted lines of the calibration and prediction set largely overlap with the 1:1 fit line. Comparing the models built from the full spectrum and the models built based on the characteristic wavelengths, it can be seen that T19 is the best spectral preprocessing method. Comparing the accuracy of the T19-PLS model and the T19-SPA-PLS model, it can be found that the accuracy of the two models is almost the same. However, compared to the model complexity, the T19-SPA-PLS model has few variables, so the model complexity is lower. Therefore, it can be concluded that the T19-SPA-PLS model is more valuable.

Discussion

Spectral response and character on SMC

The results of this study showed that the soil hyperspectral reflectance decreased with the increase of SMC, and there were two obvious absorption valleys near 1400 nm

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

and 1900 nm, which is the same as previous studies (Tan et al., 2021). SMC showed a significant negative correlation with spectral reflectance (Babaeian et al., 2015). Generally, when the mass moisture content of the soil is below field capacity, the increase of soil moisture causes the soil surface particles to absorb moisture first and then become covered in a thin layer of water in the form of film around particles, and then, the light over the surface of the soil particles and water film exhibit multiple reflections resulting in decreased soil reflectance patterns (Yang et al., 2019). This may explain why the spectral reflectance is higher when the SMC is in the range of 0~15% and lower when the SMC is in the range of 15~60%. On the other hand, in terms of the effect of SMC on soil color, when the SMC is low, the soil color is lighter and the spectral reflectance is higher, and as the SMC increases, the soil color becomes darker and the spectral reflectance decreases. When the SMC is in the range of 0~15%, the change of soil color changes significantly with the increase of moisture content, and when the SMC exceeds 15%, the effect of moisture content on soil color becomes smaller. Therefore, it may lead to the difference of spectral curve changes in different soil moisture ranges. In addition, in the short-wave infrared region, water has a significant effect on reflectance due to the absorption of water at wavelengths above 1000 nm, and the increase in water has a significant effect on the spectral reflectance of the soil (Lobell et al., 2002). Therefore, in the near-infrared spectral region, the spectral profile exhibits a huge variation phenomenon.

Table 4. *Performance of SMC models constructed by using the characteristic wavelengths*

Model types	Calibration set		Validation set		
	R^2c	RMSEc	R^2v	RMSEv	RPD
T0-SPA-PLS	0.881	5.102	0.891	5.127	2.931
T1-SPA-PLS	0.846	5.845	0.762	9.625	1.561
T2-SPA-PLS	0.901	4.653	0.868	5.877	2.557
T3-SPA-PLS	0.938	3.721	0.961	3.015	4.984
T4-SPA-PLS	0.912	4.375	0.938	3.758	3.999
T5-SPA-PLS	0.981	1.987	0.982	1.961	7.663
T6-SPA-PLS	0.966	2.721	0.955	3.125	4.809
T7-SPA-PLS	0.975	2.305	0.974	2.387	6.296
T8-SPA-PLS	0.976	2.371	0.974	2.412	6.231
T9-SPA-PLS	0.979	2.206	0.978	2.189	6.865
T10-SPA-PLS	0.900	4.751	0.946	3.658	4.108
T11-SPA-PLS	0.892	4.831	0.803	9.068	1.657
T12-SPA-PLS	0.987	1.759	0.977	2.351	6.392
T13-SPA-PLS	0.921	4.132	0.866	5.924	2.537
T14-SPA-PLS	0.856	5.568	0.879	5.204	2.888
T15-SPA-PLS	0.980	2.139	0.984	1.935	7.766
T16-SPA-PLS	0.988	1.654	0.981	2.106	7.136
T17-SPA-PLS	0.967	2.758	0.983	1.938	7.754
T18-SPA-PLS	0.971	2.514	0.982	2.038	7.374
T19-SPA-PLS	0.988	1.682	0.986	1.824	8.239

R^2, RMSE, and RPD represent the determination coefficient, root means square error, and residual prediction deviation, respectively

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

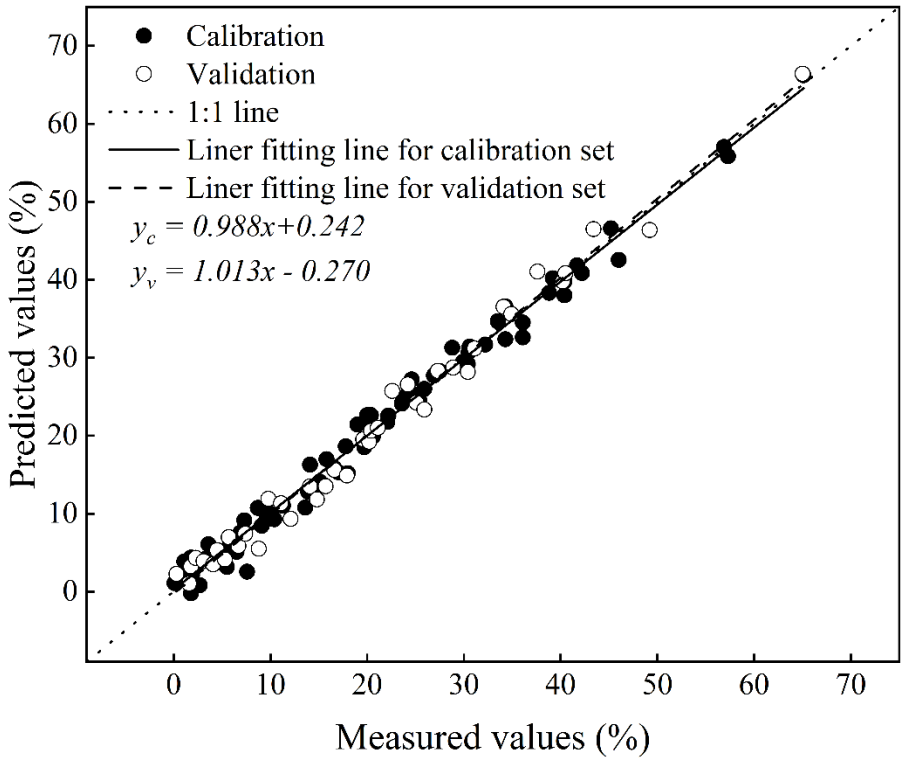

Figure 6. *Relationships between measured values and predicted values of SMC by using the T19-SPA-PLS model*

Correlation analysis of soil spectra data after different pretreatments with SMC

In the soil spectral pretreatment methods adopted in this study, when the SNV pretreatment method is used, the soil spectral characteristics before 1400 nm weakened whereas, soil spectral characteristics after 1400 nm were strengthened. After preprocessing the spectral data using the FD preprocessing method, the spectral characteristics of the soil greatly changed. The original weak spectral information was enhanced, and the spectral characteristics of some bands were highlighted (Wu et al., 2009; Sawut et al., 2014). From the positions of MACC about spectral reflectance and SMC under different pretreatments in *Figure 2,* the spectral characteristic regions of SMC are in three regions in the near-infrared band, where most of the MACC values appeared in the range of 1763~1943 nm, and in a previous study, it was also showed that the correlation between SMC and spectral reflectance was highest at 1969 nm, with the MACC value reached 0.94 (Zhai et al., 2020). Combined with the characteristic wavelengths of SMC screened using SPA in *Figure 5,* we found that the most obvious spectral characteristic region of SMC was near the 1900 nm band. In this study, compared with the original spectral reflectance (T0): except for T4, the correlation between soil spectral reflectance and soil moisture was improved after pretreatment. So, it can be concluded that suitable pre-processing methods can effectively enhance the soil spectral information and thus improve the correlation between soil spectral reflectance and SMC. For example, Yao (2011) used the spectral reflectance logarithm of the FD to estimate the SMC of black soil with a high prediction accuracy reaching 0.931. In this study, T19 was the best pretreatment method that maximized the correlation between SMC and spectral reflectance, and it is the superposition of three

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

preprocessing methods: square, SNV, and FD. The squared original spectral reflectance was processed to highlight more clearly the positions of the peaks and troughs of the spectral curve, enhancing the effective information of the spectrum. Then, the squared spectral reflectance is further processed using SNV, which reduces the multiplicative effect of the scattering variation in the NIR spectrum and achieves the denoising effect (Diwu et al., 2019). Finally, the FD method is then applied to eliminate the spectral background noise, and it makes the spectral features more obvious (Zhang et al., 2020). T19 Applying a combination of methods to preprocess the original spectral reflectance provides a new horizon for enhancing the spectral information.

Spectral monitoring on SMC

Constructing the quantitative monitoring model of SMC by using spectral technology is essential for exploring their potential relationship and realizing their practical application in the future. In this study, PLS was used to build the SMC monitoring models. The PLS combined the characteristics of principal component analysis and multiple linear regression and can be used to resolve the collinearity problem of hyperspectral reflectance (Kuang et al., 2015; Xu et al., 2020). *Tables 3* and *4* show the SMC monitoring models built using the full spectrum and the characteristic wavelengths, respectively. It can be seen that the accuracy of both types of models constructed based on T1 and T11 is low, which indicates that the T1 and T11 methods are not suitable for dealing with spectral reflectance. The SMC monitoring models with full-spectrum constructed by applying other preprocessing methods have a good performance due to the great advantage of PLS in handling multidimensional data, and the application of PLS compresses a large amount of spectral data into a few uncorrelated principal components, which well solves the covariance problem of spectral data (Wold et al., 2001; Kahaer et al., 2020). However, SMC monitoring models constructed based on the characteristic wavelengths also perform well. SPA can eliminate the redundant information in the spectral matrix and extract some characteristic wavelengths across the band (Wei et al., 2020). In this study, the SMC model constructed based on the T19 preprocessing method has the highest accuracy due to the squared reflectance of the original spectrum, which amplifies the features of the soil spectrum, and further application of SNV to eliminate the baseline shift and improve the signal-to-noise ratio (Fearn et al., 2009; Bi et al., 2016). The application of FD amplified the effective information of the spectrum, and these three preprocessing methods were superimposed to maximize the effective information of the original spectrum. Considering the accuracy and complexity of the model, we can conclude that the SMC monitoring model based on the characteristic wavelength is better and has a higher application value because this type of model not only has very high accuracy but also low complexity. The best SMC monitoring model in this study is the T19-SPA-PLS (R^2v = 0.986, RMSEv = 1.824, RPD = 8.239). Most of the previous studies only used a single pretreatment method to process the original spectral reflectance and then established a prediction model. For example, Jia (2018) preprocessed the original spectral reflectance by first-order differentiation, and the R^2v of the SMC prediction model constructed was 0.903. In comparison, the T19-SPA-PLS model preprocessed the original spectral reflectance three times, and after greatly enhancing the spectral information, the SPA was used to extract the characteristic wavelength of the enhanced spectrum, which greatly reduced the wavelength variables and decreased the complexity of the model. Therefore, the T19-SPA-PLS model can achieve more accurate spectral

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

monitoring of SMC compared with the models constructed in the previous studies. However, real field conditions are complex, and the collection of spectral reflectance can be affected by a variety of factors, such as soil texture, soil organic matter content, soil salinity, soil particle size, etc. Therefore, if complex field soils are collected for SMC monitoring studies, it is difficult to clarify the effect of SMC on spectral properties due to the differences in soil components. Therefore, this study was conducted under laboratory conditions simulating SMC, avoiding the interference of other factors besides soil moisture. Although the relationship between SMC and spectra was adequately elucidated, the applicability of the T19-SPA-PLS model in complex field environments needs further validation.

Conclusion

In this study, we investigated the effects of different pretreatment methods on the spectral characteristics of SMC, and then compared the full-spectrum SMC monitoring model with the SMC monitoring model built based on characteristic wavelengths, and finally constructed the best SMC monitoring model. It was found that SMC showed a significant negative correlation with spectral reflectance, and the response of SMC to the spectrum was most sensitive near the 1900 nm band. Proper preprocessing methods can improve the correlation between SMC and spectral reflectance, and improve the accuracy of the monitoring model as well. The SMC monitoring model based on the characteristic wavelengths screened by SPA is not only simple in model structure but also has high accuracy, among which, T19-SPA-PLS is the best SMC monitoring model (R^2v = 0.986, RMSEv = 1.824, RPD = 8.239). This study provided methodological and theoretical support for accurate monitoring of SMC by using hyperspectral. Future research should focus on the effects of soil properties interacting with soil moisture on the spectra, such as the interaction of soil grain size with moisture, soil organic matter content with moisture, and soil salinity with moisture. In addition, the quantitative effects of different soil components on the hyperspectral properties during the monitoring of SMC using hyperspectral should be explored.

Acknowledgements. This work was funded by the National Natural Science Foundation of China (31871571; 31371572): Outstanding Doctor Funding Award of Shanxi Province (SXYBKY2018040): Higher education Project of Scientific and Technological Innovation in Shanxi (2020L0132) and Scientific and Technological Innovation Fund of Shanxi Agricultural University (2018YJ17, 2020BQ32). The project was also supported by the Key Technologies R & D Program of Shanxi Province (201903D211002).

REFERENCES

[1] Babaeian, E., Homaee, M., Montzka, C., Vereecken, H., Norouzi, A. A. (2015): Towards retrieving soil hydraulic properties by hyperspectral remote sensing. – Vadose Zone Journal 14(3): 1-17.

[2] Bi, Y. M., Yuan, K. L., Xiao, W. Q. (2016): A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation. – Analytica Chimica Acta 909: 30-40.

[3] Brosinsky, A., Lausch, A., Doktor, D. (2014): Analysis of spectral vegetation signal characteristics as a function of soil moisture conditions using hyperspectral remote sensing. – Journal of the Indian Society of Remote Sensing 42(2): 311-324.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

[4] Cai, L. H., Ding, J. L. (2018): Prediction for soil water content based on variable preferred and extreme learning machine algorithm. – Spectroscopy and Spectral Analysis 38(7): 2209-2214.
[5] Diwu, P. Y., Bian, X. H., Wang, Z. F. (2019): Study on the selection of spectral preprocessing methods. – Spectroscopy and Spectral Analysis 39(9): 2800-2806.
[6] Dotto, A. C., Dalmolin, R. S. D., Caten, A. T. (2018): A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. – Geoderma 314: 262-274.
[7] Fearn, T., Riccioli, C., Garrido-Varo, A. (2009): On the geometry of SNV and MSC. – Chemometrics and Intelligent Laboratory Systems 96(1): 22-26.
[8] Guo, D. D., Huang, S. M., Zhang, S. Q. (2014): Comparative analysis of various hyperspectral prediction models of fluvo-aquic soil organic matter. – Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE) 30(21): 192-200.
[9] Hasan, S., Montzka, C., Ruediger, C. (2014): Soil moisture retrieval from airborne L-band passive microwave using high resolution multispectral data. – ISPRS Journal of Photogrammetry & Remote Sensing 91(5): 59-71.
[10] He, T., Wang, J., Cheng, H. (2006): Spectral Features of Soil Moisture. – Acta Pedologica Sinica (6): 1027-1032.
[11] Jia, X. Q., Feng, M. C., Yang, W. D. (2018): Study on the spectral prediction model of soil moisture content based on SPA-MLR method. – Agricultural Research in the Arid Areas 36(3): 266-269 + 291.
[12] Kahaer, Y., Tashpolat, N., Shi, Q. D. (2020): Possibility of Zhuhai-1 hyperspectral imagery for monitoring salinized soil moisture content using fractional order differentially optimized spectral indices. – Water 12(12): 1-29.
[13] Kuang, B. Y., Tekin, Y., Mouazen, A. M. (2015): Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content. – Soil & Tillage Research 146: 243-252.
[14] Li, J. M., Ye, X. J., Wang, Q. N. (2014): Development of prediction models for determining N content in citrus leaves based on hyperspectral imaging technology. – Spectroscopy and Spectral Analysis 34(1): 212-216.
[15] Liu, W. D., Baret, F., Zhang, B. (2004): Using hyperspectral data to estimate soil surface moisture under experimental conditions. – Remote Sensing 8: 434-442.
[16] Liu, Y., Ding, X., Liu, H. J. (2014): Quantitative analysis of reflectance spectrum of black soil as affected by soil moisture for prediction of soil moisture in black soil. – Acta Pedologica Sinica 51(5): 1021-1026.
[17] Lobell, D. B., Asner, G. P. (2002): Moisture effects on soil reflectance. – Soil Science Society of America Journal 66(3): 722-727.
[18] Lu, Y. L., Bai, Y. L., Wang, L. (2018): Spectral characteristics and quantitative prediction of soil water content under different soil particle sizes. – Scientia Agricultura Sinica 51(9): 1717-1724.
[19] Qi, H., Xiu, J., Liu, Z. (2017): Predicting sandy soil moisture content with hyperspectral imaging. – International Journal of Agricultural and Biological Engineering 10(2): 175-183.
[20] Rinnan, A., Van, D. B. F., Engelsen, S. B. (2009): Review of the most common pre-processing techniques for near-infrared spectra. – Trends in Analytical Chemistry 28(10): 1201-1222.
[21] Sanchez, N., Piles, M., Martinez-Fernandez, J. (2014): Hyperspectral optical, thermal, and microwave L-Band observations for soil moisture retrieval at very high spatial resolution. – Photogrammetric Engineering & Remote Sensing 80(8): 745-755.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

[22] Sawut, M., Ghulam, A., Tiyip, T. (2014): Estimating soil sand content using thermal infrared spectra in arid lands. – International Journal of Applied Earth Observation & Geoinformation 33: 203-210.
[23] Shepherd, A., Mcginn, S. M., Wyseure, G. (2002): Simulation of the effect of water shortage on the yields of winter wheat in North-East England. – Ecological Modelling 147(1): 41-52.
[24] Tan, Y., Jiang, Q. G., Yu, L. F. (2021): Reducing the moisture effect and improving the prediction of soil organic matter with VIS-NIR spectroscopy in black soil area. – IEEE Access 9: 5895-5905.
[25] Wang, C., Feng, M. C., Yang, W. D. (2016): A new method to decline the SWC effect on the accuracy for monitoring SOM with hyperspectral technology. – Spectroscopy and Spectral Analysis 35(12): 3495-3499.
[26] Wang, X. P., Zhang, F., Kung, H. T. (2017): Spectral response characteristics and identification of typical plant species in Ebinur lake wetland national nature reserve (ELWNNR) under a water and salinity gradient. – Ecological Indicators 81: 222-234.
[27] Wei, L. F., Pu, H. C., Wang, Z. X. (2020): Estimation of soil arsenic content with hyperspectral remote sensing. – Sensors-Basel 20(14): 1-16.
[28] Wold, S., Sjostrom, M., Eriksson, L. (2001): PLS-regression: a basic tool of chemometrics. – Chemometrics & Intelligent Laboratory Systems 58(2): 109-130.
[29] Wu, C. Y., Jacobson, A. R., Laba, M. (2009): Alleviating moisture content effects on the visible near-infrared diffuse-reflectance sensing of soils. – Soil Science 174(8): 456-465.
[30] Wu, L. G., Wang, S. L., He, J. G. (2018): Study on soil moisture mechanism and establishment of model based on hyperspectral imaging technique. – Spectroscopy and Spectral Analysis 38(8): 2563-2570.
[31] Wu, T. H., Yu, J., Lu, J. X. (2020): Research on inversion model of cultivated soil moisture content based on hyperspectral imaging analysis. – Agriculture-Basel 10(7): 292.
[32] Xu, C., Zeng, W. Z., Huang, J. S. (2016): Prediction of soil moisture content and soil salt concentration from hyperspectral laboratory and field data. – Remote Sensing 42(8): 1-20.
[33] Xu, L., Wang, Z., Hu, J. (2020): Estimation of soil salinity under various soil moisture conditions using laboratory based thermal infrared spectra. – Journal of the Indian Society of Remote Sensing (4): 1-11.
[34] Yang, X. G., Yu, Y., Li, M. Z. (2019): Estimating soil moisture content using laboratory spectral data. – Journal of Forestry Research 30(3): 1073-1080.
[35] Yao, Y., Wei, N., Tang, P. (2011): Hyper-spectral characteristics and modeling of black soil moisture content. – Transactions of the Chinese Society of Agricultural Engineering 27(8): 95-100.
[36] Yu, L., Hong, Y. S., Zhu, Y. X. (2017): Removing the effect of soil moisture content on hyperspectral reflectance for the estimation of soil organic matter content. – Spectroscopy and Spectral Analysis 37(7): 2146-2151.
[37] Yuan, J., Wang, X., Yan, C. X. (2019): A semi-empirical model for reflectance spectral of black soil with different moisture contents. – Spectroscopy and Spectral Analysis 39(11): 3514-3518.
[38] Zeng, W. Z., Lei, G. Q., Zhang, H. Y. (2017): Estimating root zone moisture from surface soil using limited data. – Ecological Chemistry & Engineering S 24(4): 501-516.
[39] Zhai, H. R., Li, X. C., Zhong, H. (2020): Hyperspectral indirect estimation model of soil water content in cultivated layer. – Chinese Agricultural Science Bulletin 36(11): 86-91.
[40] Zhang, D., Tashpolat·Tiyip, Zhang, F. (2014): Application of fractional differential in preprocessing hyperspectral data of saline soil. – Transactions of the Chinese Society of Agricultural Engineering 30(24): 151-160.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732

[41] Zhang, X. G., Kong, F. C. (2020): Using hyperspectral imagery to estimate soil moisture and calculate evapotranspiration from coastal saline soil. – Journal of Irrigation and Drainage 39(9): 14-19.
[42] Zhang, Y., Tan, K., Wang, X. (2020): Retrieval of soil moisture content based on a modified Hapke photometric model: a novel method applied to laboratory hyperspectral and Sentinel-2 MSI data. – Remote Sensing 12(14): 2239.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2003_27172732