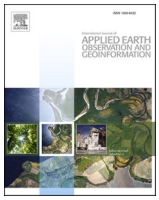
International Journal of Applied Earth Observation and Geoinformation 124 (2023) 103493

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Prediction of soil organic matter by Kubelka-Munk based airborne hyperspectral moisture removal model

Depin Ou [a], Kun Tan [b,c,*], Jie Li [d], Zhifeng Wu [a], Liangbo Zhao [e], Jianwei Ding [f], Xue Wang [b,c], Bin Zou [g]

[a] School of Geography and Remote Sensing, Guangzhou University, Guangzhou 510006, China
[b] Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China
[c] Key Laboratory of Spatial-Temporal Big Data Analysis and Application of Natural Resources in Megacities (Ministry of Natural Resources), East China Normal University, Shanghai 200241, China
[d] South China Institute of Environmental Sciences, Ministry of Ecology and Environment of the People's Republic of China, Guangzhou 510655, China
[e] Institute of Remote Sensing Satellite, China Academy of Space Technology, Beijing 100094, China
[f] The Second Surveying and Mapping Institute of Hebei, Shijiazhuang, 050037, China
[g] School of Geosciences and Info-Physics, Central South University, Changsha 410083, China

ARTICLE INFO

Keywords:
Airborne hyperspectral imagery
Soil organic matter
Kubelka-Munk
Moisture removal model
Sensitive band

ABSTRACT

Obtaining high-precision soil organic matter (SOM) spatial distribution information is of great significance for applications such as precision agriculture. But in the current hyperspectral SOM inversion work, soil moisture greatly influences the representation of the sensitive information of SOM on the spectrum. Therefore, a Kubelka-Munk theory based spectral correction model for soil moisture removal is proposed to improve the spectral sensitivity of SOM. Firstly, the soil moisture content was obtained by the use of a Kubelka-Munk based physical soil moisture model and an unmixing method. Then, the spectral correction model for soil moisture removal was implemented based on the quantitative description of the Beer-Lambert law. The results show that the proposed spectral correction model for soil moisture removal can significantly enhance the expression of the sensitive spectral features of SOM, especially for the short-wave infrared range. After moisture removal, the imaging spectral data were used for inversion, using the sensitive band at 0.69 μm and a support vector machine regression (SVR) modeling method. The Kubelka-Munk moisture removal model for soil moisture removal can improve the accuracy of SOM inversion by at least 22% comparing with the 0.69 μm original spectral inversion model, with R^2 of 0.42. Moreover, the proposed model can also improve the accuracy of SOM inversion by 19% for the SVR statistical regression method, with R^2 of 0.69. Finally, the SOM distribution maps based on sensitive band model and SVR regression method were analyzed. Findings show that the two methods have high consistency, but the statistical method obtains better details of the SOM spatial distribution, due to its higher accuracy.

1. Introduction

With the high spatial and spectral resolution, airborne hyperspectral images are now widely used for classification (Wang et al., 2019) and quantitative analysis (Ou et al., 2021). The estimation of soil composition content using hyperspectral data is also emerging as a hot research topic (Ben-Dor et al., 2019). Nowadays, accurate soil organic matter (SOM) content monitoring is crucial for intelligent agriculture. As the research has advanced, various SOM estimation methods based on hyperspectral data have been gradually developed (Nawar et al., 2016; Ou et al., 2022; Xu et al., 2020). However, most of the inversion models are based on statistical modeling methods, such as partial least squares regression (PLSR) and support vector machine regression (SVR) (Angelopoulou et al., 2019; Tan et al., 2021). Unlike indoor hyperspectral data, airborne hyperspectral data are more sensitive to environmental factors, such as angular variations during imaging, inconsistent lighting conditions, soil moisture (SM), and extremely mixed pixels (Ben-Dor, 2002; Kokhanovsky, 2019; Ou et al., 2022). Some physical or chemical properties, such as iron oxide content, soil particle size, and calcium carbonate content, are less variable for small-scale airborne

* Corresponding author at: Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China.
E-mail address: tankuncu@gmail.com (K. Tan).

https://doi.org/10.1016/j.jag.2023.103493
Received 18 April 2023; Received in revised form 11 September 2023; Accepted 15 September 2023
Available online 19 September 2023
1569-8432/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

hyperspectral image data, especially for arable land. However, the soil moisture of arable land is directly affected by topography, rainfall, snow melting and artificial factors, leading to the differences in SOM content distribution (Minasny et al., 2011). Therefore, in addition to the spectral variation caused by the difference in imaging conditions, the SOM and moisture content of soil can be considered as the main influencing factors for the variation of soil reflectance in a small arable area (Hong et al., 2018).

As the soil moisture content increases until it reaches saturation, the spectral reflectance shows a decreasing trend, after that, the soil reflectance is enhanced due to the effect of specular reflection (Weidong et al., 2002). When the SOM content is below 2%, the spectral reflectance of soils will show a significant decreasing trend (Al-Abbas et al., 1972). Many researchers found that the sensitive features for SOM mainly focused on the intervals of 0.60–0.75 μm and 1.73–2.43 μm (Angelopoulou et al., 2019; Ben-Dor et al., 1997; He et al., 2009). The accuracy of SOM content estimation directly from original soil hyperspectral information is easily affected by soil moisture (Hong et al., 2018; Minasny et al., 2011; Ogen et al., 2019). The variation of SOM content has a far less impact on the soil spectra comparing with the soil moisture. Moreover, water participates in redox reactions in soil and acts as a solvent for various ions and molecules, which influence the different spectrum properties. Therefore, to achieve SOM estimation with high performance by airborne hyperspectral data, the soil moisture effects need to be removed (Babaeian et al., 2019).

The most important prerequisite for removing the influence of moisture from hyperspectral data of soil is to obtain accurate soil moisture information. The inversion methods based on microwave remote sensing can effectively obtain soil moisture. However, the lack of radar data for simultaneous imaging and the low spatial resolution makes it difficult to meet the actual requirements. Moreover, the in-situ soil moisture data is also not available. Hence only the index-based models can be considered to obtain the soil moisture of a whole study area. The classical method VSDI (Zhang et al., 2013) model is simple to calculate, but the accuracy is relatively low compared to some of the other methods. The TOTRAM (Nemani et al., 1993) and other methods (Bablet et al., 2018; Sadeghi et al., 2017) are parameter complex. To date, Beer-Lambert law (Bach and Mauser, 1994) and Kubelka-Munk (K-M) (Kubelka, 1931) has been successfully applied to inversion of soil moisture. Kubelka-Munk physical model (Sadeghi et al., 2015) can achieve well performance with concise model parameters. Although many soil moisture inversion models have been developed, spectral correction models for soil moisture removal are still rare (Bablet et al., 2018; Yuan et al., 2019; Zhang et al., 2020). Minasny et al. (2011) used an EPO method to construct a spectral correction model to remove the effect of moisture, and the accuracy with the test set reached 0.84 when use the corrected spectra for SOM prediction. Yaron et al. (2019) used the nearest neighbor spectral correction (NNSC) method to build a wet-dry soil spectra transformation coefficient vector, and the accuracy of the soil clay content reached 0.69. Castaldi et al. (2015) used the PRISMA dataset and calculated the statistical variability between wet and dry soils to synthetically dry soil spectra, thus improving the prediction accuracy for clay. Despite the numerous studies have calibrated field-collected spectra against measured SOM, the variation in soil moisture will have significant impacts on the prediction of SOM (Bogrekci and Lee, 2006; Bricklemyer and Brown, 2010; Lobell and Asner, 2002). Some issues should be handled, which can be described as follows:

1) most available soil moisture correction models are applied to indoor spectral data. These correction models require measured dry soil spectra to build a linear or nonlinear correction model, which is ineffective when applied directly to imaging data;
2) Existing soil organic matter hyperspectral inversion methods can better obtain the spatial distribution of local areas. However, for a broad area mapping without sample sites, the generalization performance of traditional regression methodology is poor. It will also perform poorly in terms of explainability.

Therefore, developing a physical model-based soil organic matter inversion method is of great interest for radiative transfer theory interpretation and large-scale remote sensing application. The primary goals of the research are to rapidly and accurately invert soil organic matter composition using hyperspectral remote sensing data. It is necessary to develop a physical model-driven soil spectral correction model (such as Kubelka-Munk analytical radiative transfer model) to remove the influence of soil moisture on the spectrum. Moreover, the develop generalizable methods should be investigated for inverting soil moisture without the need for further measurement data.

2. Datasets and methodology

Fig. 1 shows the overall workflow of this study. Firstly, an unmixing method was used to extracted the water and soil abundance information from the airborne hyperspectral imagery. The driest and wettest soil spectra for the study area were then extracted based on the water and soil abundance information. And the soil moisture was inverted by the K-M physical model. Considering the effect of both the Fresnel effect and soil moisture, a spectral correction model for soil moisture removal based on K-M theory is proposed. Finally, sensitive bands were obtained from the corrected airborne hyperspectral data.

2.1. Datasets

2.1.1. Study area

Fig. 2 is the map of the study area, with an area of about 139 km^2 (ranges from 125.33°E–125.47°E and 43.22°N–43.33°N), located in the southeastern part of Yitong Manchu Autonomous County in western Jilin province, China. The study area is primarily low-lying hills with gentle undulations, with an average elevation of 305 m, a minimum elevation of 215 m, and a maximum elevation of 430 m. The slope of the study area is very extremely with an average slope of 5.18°. There are two gold mines in the study area. Natural villages are predominant, with only a few industrial plants. Agriculture is the mainstay of the economy, and the vast majority of the study area is cultivated with maize, with scattered plots cultivated with rice. Land use statistical information shows that 60 % of the area is agricultural land, 32% is forest, 6% is grassland, and only 1.1% impermeable layer.

The background survey data for the study area were obtained from the National Soil Information Service Platform (https://www.soilinfo.cn). According to the second Chinese soil census (1980–1990), the soil types in the study area are dominated by dark brown forest soil, albic soil, meadow soil, and black soil, while dark brown forest soil accounts for a significant proportion. The texture of the soil is sandy loam. The organic matter content is generally 30–50 g/kg, and the soil is slightly acidic to neutral, with pH 5.9–7.5. The agricultural land has been reclaimed and is rich in nutrients. As such, it is suitable for dry crops such as corn with a yield of more than 4500 kg per hectare.

2.1.2. Airborne hyperspectral data

A HyMap hyperspectral data (0.40 μm–2.50 μm) were acquired in our study area between April 18 and April 22, 2017. Before the aerial photography, the hyperspectral imaging system should be processed with rigorous radiometric calibration in the laboratory, so that it is capable of obtaining high-precision irradiance information of the ground features. The hyperspectral data preprocessing included radiometric calibration, geometric correction, atmospheric correction and seamless mosaicking correction. The MODTRAN4 atmospheric radiation transmission model (Berk et al., 1999) was used to perform atmospheric correction of all the orthophoto images. The bidirectional reflectance distribution function (BRDF)-based photometric correction algorithm was applied to handle the seamless mosaicking correction (Yu et al.,

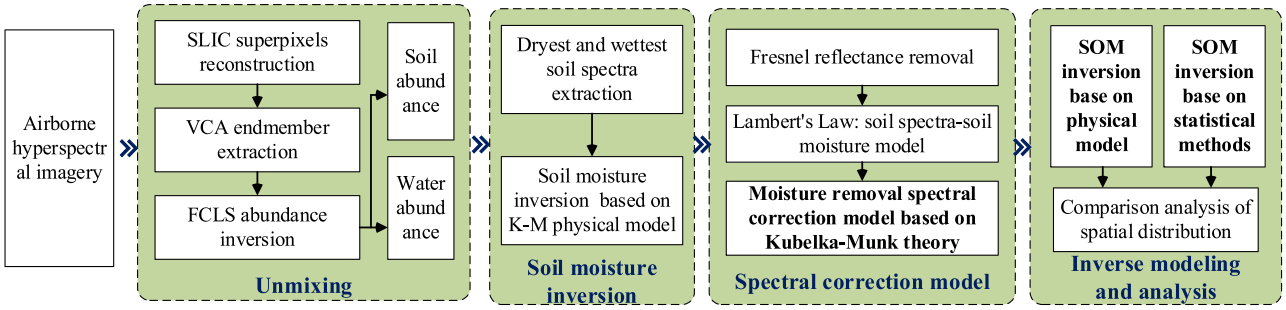


Fig. 1. The overall workflow of this study.

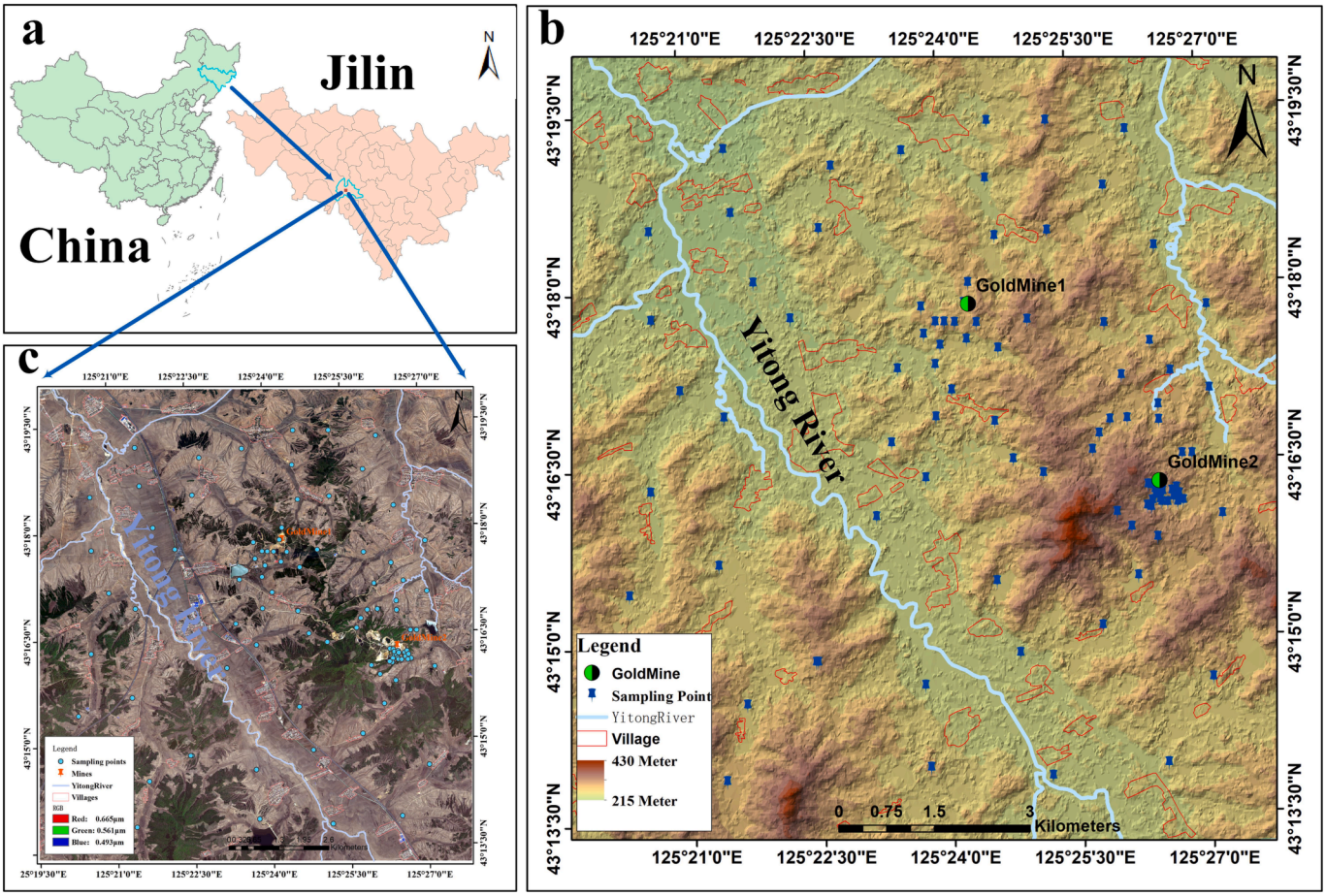


Fig. 2. Study area and field sampling points.

2017). All the processing was undertaken in HyMap-C™ processing software.

Finally, the preprocessed airborne hyperspectral image data were of 2734 × 2508 × 135 pixels in size, with a spectral resolution of 10–20 nm (VIS and NIR are 15 nm, SWIR1 and SWIR2 are 18 nm), and a spatial resolution of 4.5 m, as shown in Fig. 2. The absorption of water vapor near 1.40 μm (1.35–1.51 μm) and 1.90 μm (1.78–1.99 μm) causes anomalous reflectance of the spectral curve. Therefore, the bands in these two spectral ranges were removed, and 101 bands were finally retained. It has been published(Ben-Dor et al., 2019; Lekner and Dorf, 1988) that the variation of the spectra are influenced by several factors, especially soil components such as water, organic matter and iron oxides. As shown in Fig. 2, darker soils are mostly concentrated in the watershed, and therefore, they are more significantly affected by soil moisture and organic matter content.

2.1.3. Field sampling data

Field sampling in the study area was undertaken simultaneously with the airborne imaging, shown in Fig. 2. The distribution of the samples was homogeneous throughout the study area. Each sample was collected from topsoil with a depth of 5 cm, and high-precision coordinate information was recorded by real-time kinematic (RTK). Finally, ninety-three soil samples were acquired in the study area. The pretreatments of the soil samples including impurity removal, air drying, grinding, and sieving through a 100-mesh sieve were firstly conducted. Then, SOM content for ninety-three samples was determined by the potassium dichromate volumetric method. Table 1 is the SOM content statistics information, the mean value of SOM content is 30.50 g/kg, indicating relatively fertile soil. Theoretically, the standard normal distribution has zero skewness and kurtosis. Table 1 shows that the SOM's kurtosis is 0.69 and its skewness is 0.12, which indicates that the distribution of

Table 1
Soil organic matter content statistics.

Method	Num. of Samples	Max	Min	Mean	C.V	Skewness	Kurtosis
SOM(g/kg)	93	49.84	14.76	30.51	0.21	0.12	0.69

SOM content can be regarded as a normal distribution. The coefficient of variation is small, which indicates that the spatial distribution of SOM content has not been disturbed by human activities.

2.2. Unmixing method for driest and wettest soil spectra extraction

The K-M physical model (Sadeghi et al., 2015) was chosen to estimate the surface soil moisture in the study area. In comparison with the other models, the inversion methods based on the K-M physical model are simpler to implement and can obtain a higher accuracy. The K-M physical model based soil moisture inversion method developed by Sadeghi (2015) is shown in Equations (1) – (2):

$$\theta = \frac{STR - STR_d}{STR_w - STR_d} \tag{1}$$

$$STR = \frac{(1 - R_{swir})^2}{2R_{swir}} \tag{2}$$

where θ is the soil moisture, and STR is R_{swir} converted to the K-M space, the 12th band (with a central wavelength of 2202.4 nm) in Sentinel 2A was used as R_{swir} in Sadeghi (2015). STR_d and STR_w are the K-M space spectra of the driest soil and the wettest soil in the study area, respectively. The model describes the linear relationship between surface soil moisture and the R_{swir} under K-M space, which is easily solvable for the whole region.

Soil moisture content extraction based on the K-M physical model method requires a dry soil spectrum with a known local minimum moisture content and a wet soil spectrum with a local maximum moisture content. In fact, we don't have the field samples. And, it is still difficult to directly extract reliable dry and wet soil spectra from hyperspectral imagery. As the soil in the imaging spectra is a mixture of soil and moisture, an unmixing method is introduced to extract the dry and wet soil spectra. Firstly, superpixels were generated using the simple linear iterative clustering (SLIC) method proposed by Achanta et al. (2012). To reduce the time consumption, vertex component analysis (VCA) (Nascimento and Dias, 2005) was used to extract the endmembers. Finally, the fully constrained least squares (FCLS) (Heinz and Chein-I-Chang, 2001) method was applied to extract the soil abundance map. The method used to extract the study area's dry and wet soil spectra were as follows.

I. Soil abundance = 1, dry soils, manually select a dry soil spectrum as the driest soil spectrum.
II. Soil abundance + water abundance = 1, wet soil dataset.
III. Selected a soil spectrum with the highest moisture content in the wet soil dataset as the wettest soil spectrum.

2.3. Spectral correction model for soil moisture removal

The K-M model describes the relationship between two radiation fluxes I and J with opposite directions, whose spectra are related to the absorption coefficient k, scattering coefficient s, and thickness × of the material, as shown in Equations (3) and (4):

$$-dI = -kI2dx - sI2dx + sJ2dx \tag{3}$$

$$dJ = -kJ2dx - sJ2dx + sI2dx \tag{4}$$

When the thickness × tends to infinity, the sample's transmittance tends to 0, and the reflectance of sample R can be expressed using Equation (5):

$$R = 1 + \frac{k}{s} - \sqrt{\left(\frac{k}{s}\right)^2 + 2\frac{k}{s}} \tag{5}$$

When $r = \frac{k}{s}$,

$$r = \frac{k}{s} = \frac{(1-R)^2}{2R} \tag{6}$$

Equation (5) is the well-known K-M model. For airborne imaging hyperspectral data, the thickness of the soil can be considered as infinitely thick, so that the soil spectrum can be directly converted to the K-M model space using Equation (6), which is expressed as the ratio of the absorption and scattering coefficients.

If a completely non-scattering material exists, s = 0, and Equation (3) and Equation (4) are equivalent to:

$$-\frac{dI}{dx} = -kI \tag{7}$$

$$\frac{dJ}{dx} = -kJ \tag{8}$$

Equation (9) is obtained according to Kortüm Gustav's (2012) derivation:

$$R = R_g e^{-2kx} \tag{9}$$

Equation (9) can be converted to the well-known Lambert's law, where R_g is the background spectrum, which can be regarded as a special case of the K-M model. Bach and Mauser (1994) described the relationship between the soil spectra and the active thickness of water layer using the Beer-Lambert law, as shown in Equation (10):

$$R = R_0 * e^{-\varepsilon * l} \tag{10}$$

where R is the spectrum of wet soil, R_0 is the spectrum of dry soil, ε is the absorption coefficient of water (each band has a unique coefficient), and l is the active thickness of water layer of the soil.

The Fresnel reflectivity effects are not considered in the derivation of the K-M model (Philpot, 2010). Fresnel reflectance describes the reflection behavior of light between different media. The Fresnel effect is enhanced with increasing soil moisture, so it needs to be removed. At the air–soil interface, the Fresnel reflectance is proportional to the soil moisture content:

$$R_i = \left(\frac{n_{water} - n_{air}}{n_{water} + n_{air}}\right)^2 * \theta \tag{11}$$

where R_i is the Fresnel reflectance; n_{water} is the refractive index of water, which is about 1.33; and n_{air} is the refractive index of air, which is about 1. Equation (12) is the soil spectrum after removing the Fresnel reflectance:

$$R_t = R - R_i \tag{12}$$

where R_t is the corrected spectrum, R is the original spectrum, and R_i is the Fresnel reflectance.

Since SOM and other materials also have a large effect on the soil spectrum, the direct use of the soil moisture obtained from Equation (1–2) (instead of the soil active thickness of water layer l) to obtain ε and R_0 will result in a large deviation. Therefore, it can first be assumed that the dry soil spectra are the same for all the samples, which means that R_0 is also the same. The two soil spectra R_{t1} and R_{t2} after removing the Fresnel reflectance are then randomly selected and substituted into Equation (13) to obtain the initial ε:

$$\varepsilon = \frac{\ln \frac{R_{t1}}{R_{t2}}}{\theta_2 - \theta_1} \tag{13}$$

The derived ε is substituted into $A = e^{-\varepsilon^*\theta}$, and by fitting the intermediate variable A to the soil spectrum R_t in exponential form, we can obtain R_f:

$$R_t = R_f * e^{\alpha * A} = R_f * e^{\alpha * e^{-\varepsilon^*\theta}} \tag{14}$$

where ε is the water absorption coefficient obtained by Equation (13) for any two soil spectra, θ can be calculated from Equation (9), and R_f and α are obtained by fitting operations.

When $\theta = 0$, $R_t = R_f * e^{\alpha}$, which means that $R_f * e^{\alpha}$ is the theoretical reflectance of dry soil. Therefore, all the spectra should be stretched to this theoretical spectrum to achieve soil moisture removal and obtain the final spectral correction model for soil moisture removal:

$$R_{rw} = R_t * \frac{R_f * e^{\alpha}}{R_f * e^{\alpha * e^{-\varepsilon^*\theta}}} = R_t * e^{\alpha * \left(1 - e^{-\varepsilon^*\theta}\right)} \tag{15}$$

where R_{rw} is the corrected spectrum. The correction coefficients are theoretically wavelength-independent, but due to the influence of environmental factors in the actual imaging, the moisture inversion varies greatly at different wavelengths. According to the relevant literature (Sadeghi et al., 2017; Sadeghi et al., 2015), the moisture inversion accuracy is highest at 2.2 μm, so the correction coefficient calculated at this wavelength can be considered the most accurate, which means that the factor of 2.2 μm can be directly used for the whole spectrum.

2.4. Evaluation and analysis methods

In this paper, the evaluation indices are coefficient of determination (R^2), root mean squared error, (RMSE), mean absolute error (MAE), ratio of performance to inter-quartile distance (RPIQ), and F-test (F).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{16}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \tag{17}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{18}$$

$$RPIQ = \frac{Q3 - Q1}{RMSE} \tag{19}$$

In this study, Advanced Spaceborne Thermal Emission and Reflection Global Digital Elevation Model Version 2 (ASTER GDEM 2) data, with a spatial resolution of 30 m, were used to generate a stream network (Tachikawa et al., 2011). A series operation consisting of pit-filling, flow direction grid, accumulation and threshold processing are conducted to obtained the stream network. The conventional techniques pre-processing with a pit-filling algorithm(Jenson and Domingue, 1988) and flow direction calculation with the D-8 algorithm (O'Callaghan and Mark, 1984) were included to obtain the DEM hydrological analysis. All of the processing was performed using the ArcGIS 10.4.

2.5. Model and parameter information

In this experiment, the samples were all sorted according to the SOM content, and every three samples were treated as a group. All soil samples were divided into a training set and testing set, which was composed according to the 2:1 segmentation rule at each soil samples group. Finally, sixty-two samples are selected into the training set and thirty-one samples are selected into the test set. Feature selection is required to avoid data redundancy before the SOM estimation task, which can remove irrelevant features and reduce the difficulty of the feature learning task. There are many feature selection methods, such as the competitive adaptive reweighted sampling (CARS) algorithm (Li et al., 2009), genetic algorithms (GAs) (Leardi, 2000), and Pearson correlation coefficient based feature selection methods (Tan et al., 2021). In the comparison experiments, the full-band (101 features, ALL) and CARS band selection methods were used as feature comparisons. PLSR (Geladi and Kowalski, 1986), SVR(Drucker et al., 1997) and Random Forest regression (RF) (Liaw and Wiener, 2002) statistical regression methods were used as regression method comparisons. In addition to the above, SemiDNN (Ou et al., 2021), WT-RF (Gu et al., 2019) and CR-PLSR (Yu et al., 2015) from published literature were used as comparative experiments.

The search space of the default hyperparameter of PLSR was [2, 20]. The kernel function applied in SVR was radial basis function (RBF). The search space of penalty parameter C was [2^-5, 2^20], and gamma was [2^-20,2^20]. The n_estimators in the random forest regression model are 30 and the rest of the parameters are defaults. Different methods were used for the comparative experiments to verify the effectiveness of the spectral correction model for soil moisture removal. In the feature comparison, all the original 101 spectral features (ALL + PLSR/SVR), the original spectral features selected by the CARS method (CARS + PLSR/SVR), random forest method and the spectral features selection after soil moisture removal by the CARS method (Removed + CARS + PLSR/SVR) were chosen for the comparison experiments. The experiments were conducted by randomly selecting two soil spectra and using Equation (13) to calculate the initial ε, which was approximately 1.4880.

3. Results and analysis

3.1. Effect of moisture on soil spectra

Fig. 3a show the scatter plot and exponential fit of the soil moisture θ to soil spectral reflectance after Fresnel correction. Fig. 3b shows the scatter plot and exponential fit of $e^{-\varepsilon^*\theta}$ to soil spectral reflectance after Fresnel correction. It can be observed that the direct exponential fit of the soil moisture θ obtained by the K-M physical model inversion method to the soil reflectance can reach an determination coefficient of 0.9820. However, when the θ value is lower, the fitted curve deviates more from the true curve (as shown in the red box), which significantly impacts the correction of the soil sample spectra to the dry soil condition. In contrast, the results were also calculated after $e^{-\varepsilon^*\theta}$ was fitted to the reflectance, where the coefficient of determination can reach 0.9995. When $\theta = 0$, the ideal reflectance obtained by fitting directly with the moisture content information is 0.27, while the ideal reflectance obtained by fitting after $e^{-\varepsilon^*\theta}$ processing is 0.30, indicating that this tiny precision difference has a significant impact on the performance.

After obtaining ε, α, and R_f (for example, in this experiment ε = 1.4884, R_f = 0.0545, and α = 1.7059), soil moisture removal can be performed by Equation (15). Fig. 4 shows the raw soil spectrum extracted by the hyperspectral imaging, with a minimum SOM content of 14.76 g/kg and a maximum of 49.84 g/kg in the soil samples. Fig. 5 shows the corrected soil spectra obtained by the spectral correction model for soil moisture removal proposed in this paper. The comparative analysis in Figs. 4 and 5 demonstrates that the soil moisture content has a stronger influences on the soil spectral information than the SOM content. The soil reflectance decreases as the soil moisture increases. Soil spectra after soil moisture removal show a decreasing trend with increasing SOM content, as shown in Fig. 5. For example, the soil spectra with the lowest SOM content has the highest reflectance after the soil moisture removal, while the soil spectra with the highest SOM has the lowest reflectance. For soil spectra with only minor differences in SOM content, they are more difficult to distinguish because the correction

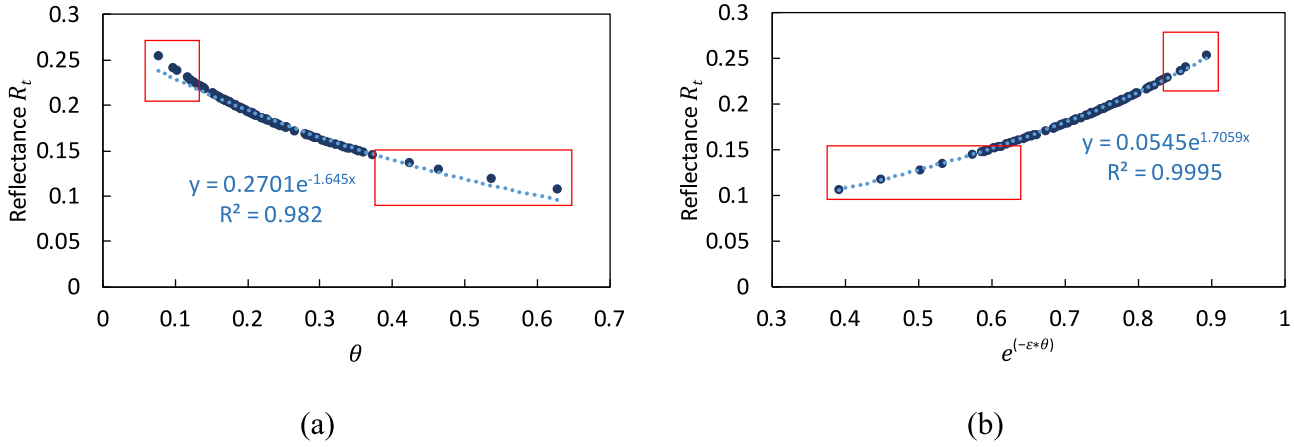


Fig. 3. (a) The scatter plot between the soil moisture information θ and the spectra after Fresnel correction. (b) After calculation of the initial coefficient ε, the scatter plot between $e^{-\varepsilon^*\theta}$ and the spectra after Fresnel correction.

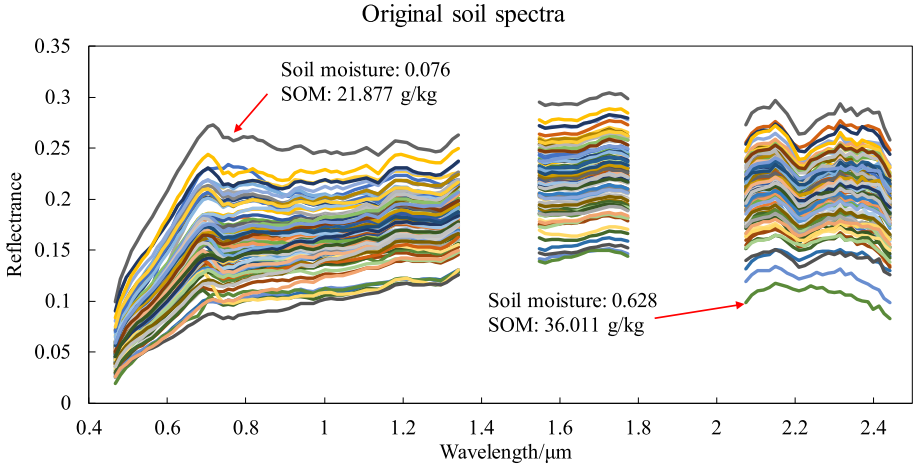


Fig. 4. Spectra of the original soil samples.

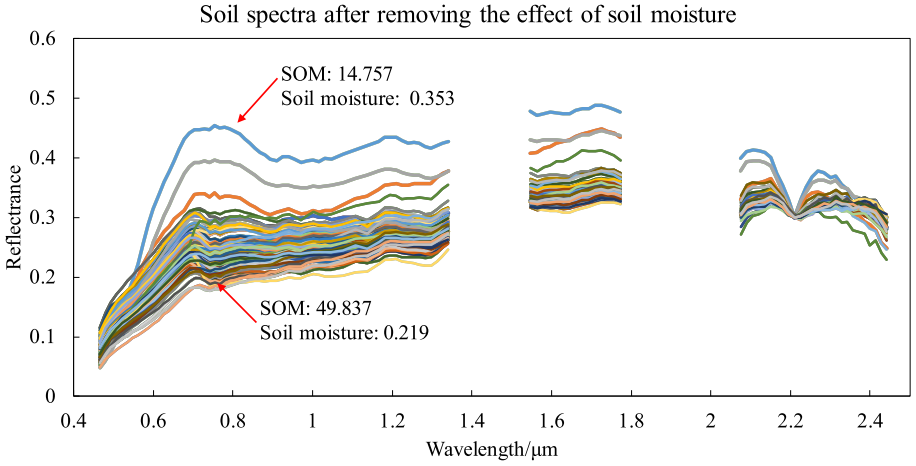


Fig. 5. Spectra of the soil samples after soil moisture removal.

factor is utilized only by a factor of 2.21 μm. However, the distinguishability is superior when the soil spectra have more significant gradients of SOM content.

3.2. Soil organic matter physical inversion model

The primary purpose of soil moisture removal is to investigate the relationship between SOM content and soil spectral reflectance. Fig. 6 shows a comparison of the Pearson correlation coefficients between the

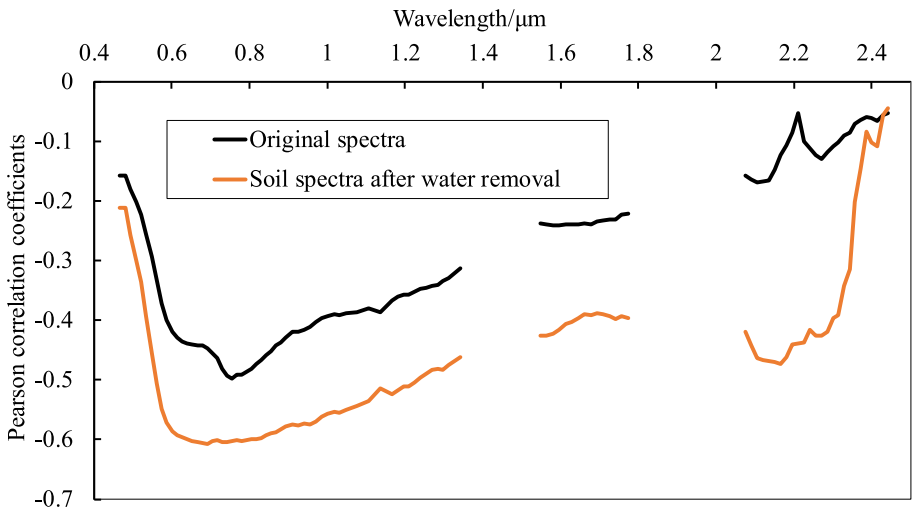


Fig. 6. Comparison of the Pearson correlation coefficients between the SOM content and soil spectra before and after soil moisture removal.

SOM content and soil spectra before and after soil moisture removal. Clearly, the SOM shows a negative correlation with the spectral reflectance, while the correlation after soil moisture removal is significantly enhanced in the vast majority of bands, especially around 2.21 μm. In the visible spectral range, both the original spectra and the spectra after removing the influence of moisture have a strong correlation, especially in the range of 0.68–0.73 μm. The spectra after soil moisture removal show the highest correlation of − 0.61 at 0.69 μm. At 0.69 μm, the correlation of the spectra after soil moisture removal improves from − 0.45 to − 0.61, which is an improvement of 16%. At 2.17 μm, the correlation of the spectra after removing the effect of moisture also reaches a peak, with an improvement from − 0.12 to − 0.47, representing a significant increase of 35%. It is particularly notable that, at around 2.21 μm, the correlation between the SOM content and soil spectra after soil moisture removal is substantially enhanced. This can be attributed to the fact that, since the moisture effects in the bands around 2.21 μm are similar, the correlation coefficients should also be similar, which can substantially enhance the expression of the spectral features of SOM. In the rest of the bands, such as the visible band, the correlation coefficients, due to the effect of moisture, are different from those in the short-wave infrared band, so the enhancement of the spectral expression of SOM is relatively low.

Therefore, SOM inversion can be performed using the single-band information after removing the effect of moisture. Hence, a SOM sensitive band inversion model was constructed in the form of the Beer-Lambert law, as shown in Equation (20):

$$SOM = SOM_0 * e^{-\varepsilon * R} \quad (20)$$

where SOM_0 is the organic matter content when the hyperspectral reflectance is 0, ε is the coefficient, R is the sensitive band reflectance, and both SOM_0 and ε can be obtained by fitting.

Since the correlation between the spectra and SOM at 0.69 μm and 2.21 μm showed peaks, these two bands were mainly used to bulid the SOM inversion model by using Equation (20). The results are listed in Table 2 and shown in Fig. 7. At both 0.69 μm and 2.17 μm, the correlation between the original spectra and SOM is extremely low, with the determination coefficient at 0.69 μm being only 0.20, while that at 2.17 μm shows no correlation. From the scatter plot of the spectral reflectance after soil moisture removal, the best inversion results can be obtained from the sensitive band inversion model based on the Beer-Lambert law. The SOM inversion model constructed at 0.69 μm can reach a determination coefficient of 0.42, while that at 2.17 μm can reach 0.30, which is a significant improvement in inversion accuracy compared with the original spectrum. Therefore, to a certain extent, the information at 0.69 μm can be used as a sensitive waveband for SOM inversion. The inversion model also shows that the SOM is negatively correlated with the soil spectrum, which is consistent with reality.

Table 2
Organic matter content inversion by the single-band method.

Wavelength	Type	Model	R^2	MAE	RMSE
0.69 μm	Original	SOM = 48.111*exp (−2.960*R)	0.20	4.40	5.70
	Correction	SOM = 75.761*exp (−3.733*R)	**0.42**	4.00	5.02
2.17 μm	Original	SOM = 34.484*exp (−0.684*R)	0.01	4.78	6.34
	Correction	SOM = 3281.5*exp (−14.27*R)	**0.29**	4.29	5.62

3.3. Soil organic matter statistical inversion model

From the perspective of the inversion accuracy, directly applying SOM inversion by the sensitive bands is less accurate, and the highest determination coefficient for the inversion model is 0.42. From the sensitive band inversion model, it is known that the inversion results for samples with a high SOM content will be low compared with the actual content, so that more bands information should be considered. The optimal waveband combination obtained by the CARS method is shown in Fig. 8, with 24 feature bands. From the selected bands, it is apparent that most of the bands are located in the visible position, especially in the range of 0.60–0.90 μm.

Table 3 lists the final accuracies, where the determination coefficient for the training set reaches 0.68, while that for the testing set reaches 0.69. Fig. 9 shows the scatter plots of the regression models obtained by the PLSR and the SVR. There are no overfitting of the PLSR method and SVR method, but the model built by the SVR method obtains a higher model accuracy and is concentrated around the 1:1 line in both the training and testing sets.

3.4. Analysis for the spatial distribution of soil organic matter

Fig. 10 shows the SOM spatial distribution results for the 0.69 μm sensitive band inversion model, while Fig. 11 shows the spatial distribution results for the SVR regression model. By overlaying the stream network generated by the digital elevation model (DEM) hydrology analysis, it shows the areas with high SOM are concentrated around the stream network. Generally speaking, the stream network derived from the DEM calculation is oriented toward valleys and lowlands. Therefore, the aggregation of SOM is mainly due to the transportation and

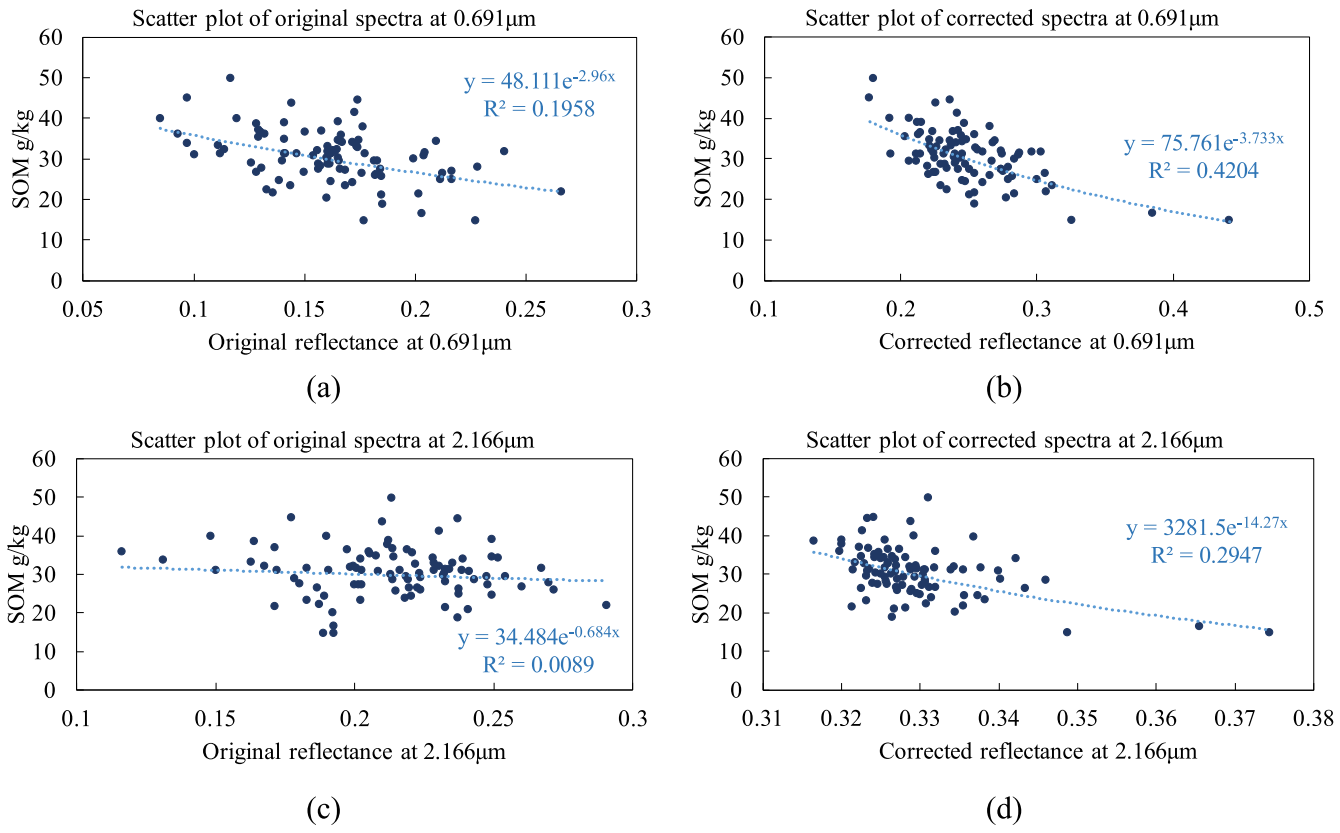


Fig. 7. Scatter plots of soil organic matter and the original and modified soil spectral reflectance at 0.69 μm and 2.17 μm. (a) Scatter plot of the original spectra at 0.69 μm. (b) Scatter plot of the corrected spectra at 0.69 μm. (c) Scatter plot of the original spectra at 2.17 μm. (d) Scatter plot of the corrected spectra at 2.17 μm.

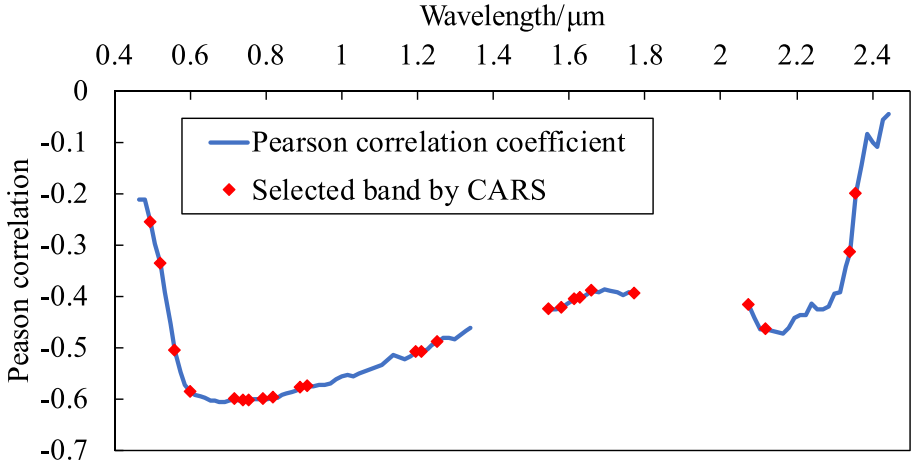


Fig. 8. Band selection by the CARS method.

Table 3
Accuracy comparison of the various methods.

Method	Num. of features(Bands)	Training set (62 samples)				Testing set (31 samples)			
		R^2c	RMSEc	MAEc	RPIQc	R^2p	RMSEp	MAEp	RPIQp
Removed + CARS + PLSR	24	0.46	4.58	3.64	1.64	**0.55**	4.42	3.44	1.70
Removed + CARS + SVR	24	0.68	3.51	2.38	2.14	**0.69**	3.68	2.82	2.04

aggregation by surface water and other factors. On the other hand, arable soil is loose and is easily eroded, so that SOM is easily migrated through the flow of rain and snow meltwater, etc. In both maps, the SOM content is higher in the flat cultivated land on the east side of the Yitong River, while the cultivated land with steep slopes shows a lower SOM content. The gold mining has a certain impact on the soil nutrient loss,

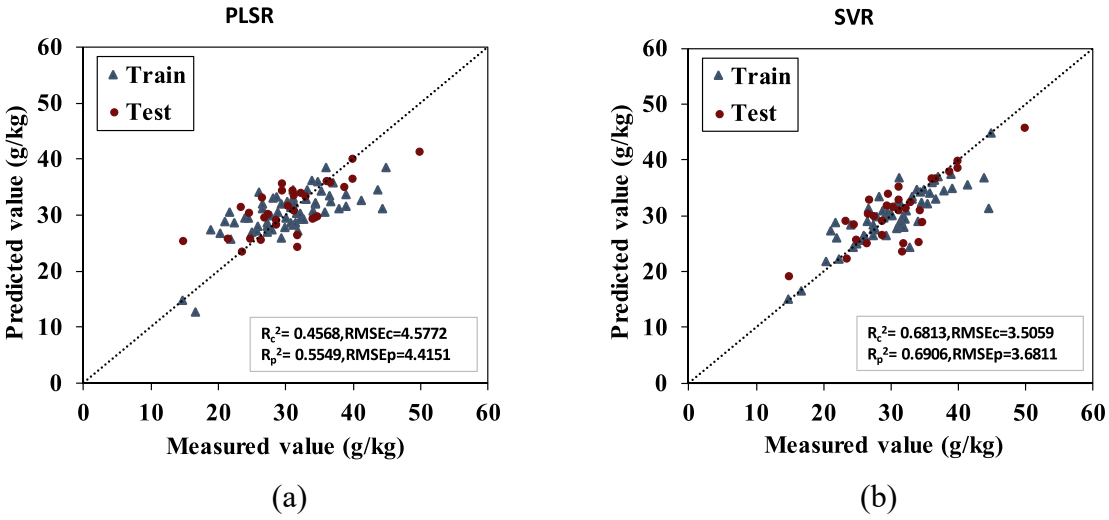


Fig. 9. Scatter plots of the predicted and measured values obtained by the PLSR and SVR methods.

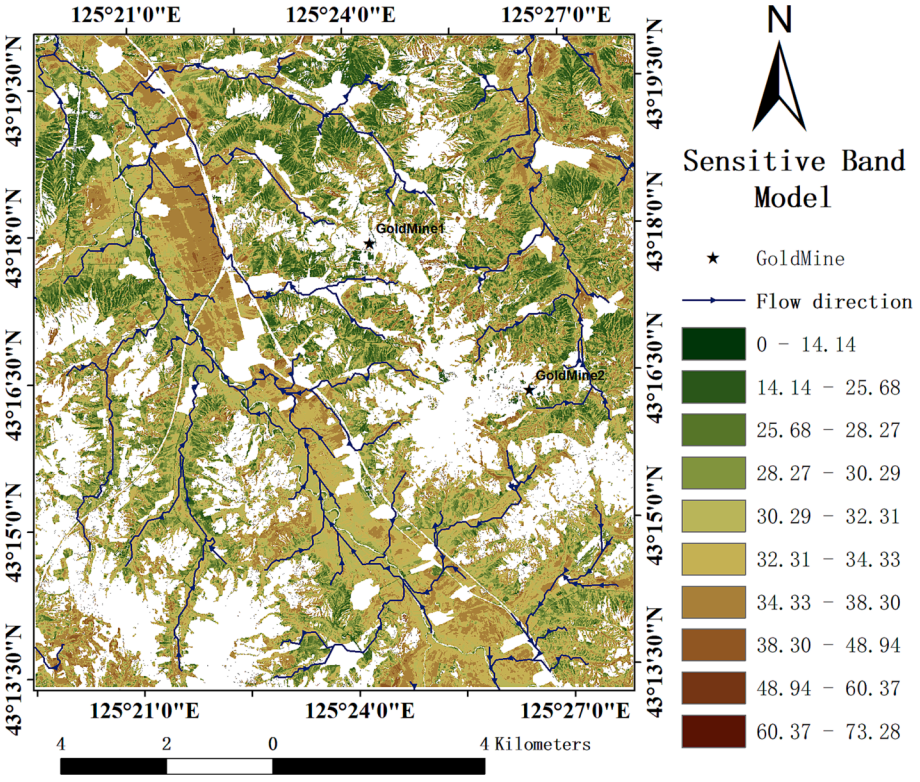


Fig. 10. Soil organic matter content inversion results by the 0.69 μm sensitive band model.

hence the SOM content is low around the two gold mining areas, but the influence is relatively weak. The location of the soil with a lower SOM content, as shown in the green part of the figure, has textural characteristics that are related to the topographic characteristics. This means that soils on cultivated land with a steeply sloping topography have a lower SOM content. The annual cultivation activities lead to great changes in the SOM content of the surface soil, especially in sloping cultivated land, where soil erosion and soil degradation are caused by the looseness of the soil and the topographic relief.

Comparing the two maps, it can be noted that the overall distribution trends of the two maps are consistent, but there is large variability in the details. Combined with the kernel density estimation plot, the sensitive band inversion method results in lower values of estimated SOM content for soils with higher actual SOM content. The spatial map obtained using the multi-band regression method has more detailed and more accurate information. For example, in the soil around the Yitong River, the SOM obtained by the sensitive band model tends to be more consistent, while the SOM obtained using the regression method is more detailed, as shown in Fig. 12. Field investigation found that the Yitong River carries a lot of fine sand, and the soil around the riverbank is seriously degraded. The results obtained by the regression method have a lower SOM content, which is more realistic. From the SOM distribution map around the stream network, the SOM content obtained by the regression method is higher, and both methods show the phenomenon of aggregation to the low-lying areas.

In summary, the results obtained by the two methods are consistent

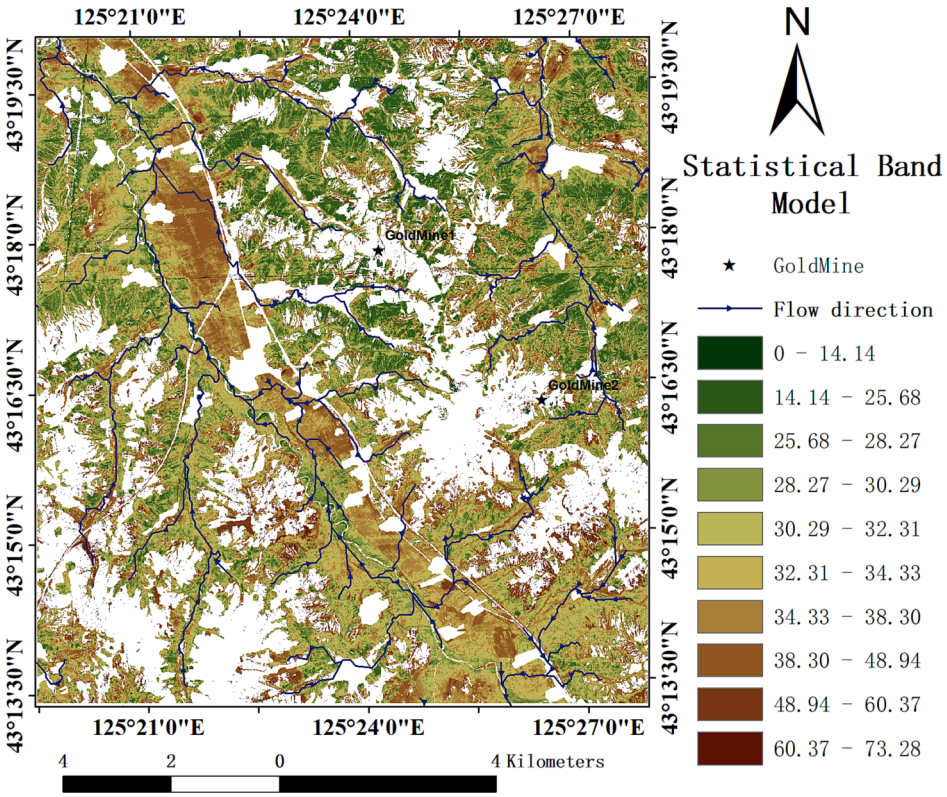


Fig. 11. Soil organic matter content inversion results by the SVR regression model.

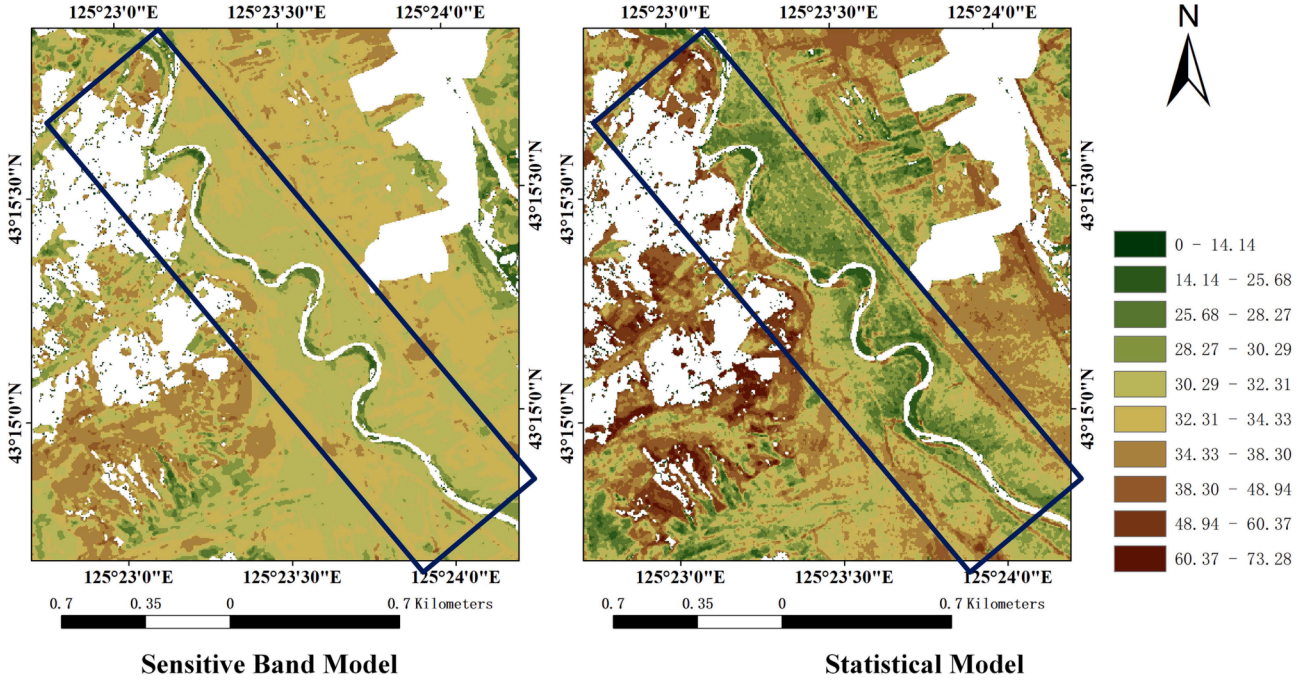


Fig. 12. Comparison of the organic matter content obtained by the two methods in the Yitong River area.

in the overall spatial distribution trend characterization of SOM content. However, from the viewpoint of accuracy, the regression-based inversion method obtains a higher accuracy with more detailed information, and can better characterize the realistic spatial distribution of SOM.

4. Discussion

4.1. Comparative experimental accuracy analysis

Table 4 provides an accuracy comparison of the various methods. When comparing the accuracy of the regression model learning using

Table 4
Accuracy comparison for the various methods.

Method	Num. of features(Bands)	Training set (62 samples)					Testing set (31 samples)				
		R^2c	RMSEc	MAEc	RPIQc	F	R^2p	RMSEp	MAEp	RPIQp	F
0.69 μm inversion	1	**0.42****	4.00	5.02	1.58	55.43	–	–	–	–	–
ALL + PLSR	101	0.38	4.95	3.98	1.56	21.47	0.37	5.17	4.07	1.46	28.21
CARS + PLSR	11	0.37*	4.98	3.90	1.55	26.86	0.38*	5.12	3.97	1.47	21.80
Removed + CARS + PLSR	24	0.46**	4.58	3.64	1.64	55.36	**0.55****	4.42	3.44	1.70	32.47
ALL + SVR	101	0.45	4.66	3.37	1.63	41.28	0.47	4.73	3.65	1.52	31.20
CARS + SVR	11	0.51*	4.38	3.17	1.73	41.79	0.50*	4.61	3.48	1.56	41.84
Removed + CARS + SVR	24	0.68**	3.51	2.38	2.14	136.34	**0.69****	3.68	2.82	2.04	51.99
Random Forest	30	0.95**	1.75	1.40	4.86	4524.21	**0.57****	4.20	3.27	1.86	50.24
Semi-DNNR(Ou et al., 2021)	17	0.96**	1.15	0.69	6.59	7945.82	0.71**	3.52	2.59	2.04	101.33
WT-RF(Gu et al., 2019)	20	0.48**	4.46	3.57	1.71	71.43	0.47**	4.84	3.51	1.51	31.01
CR-PLSR(Yu et al., 2015)	22	0.70**	3.37	2.63	2.27	143.18	0.55**	4.44	3.52	1.65	35.97

ALL: all original features. Removed: soil moisture removal spectral correction. *: $p < 0.1$. **: $p < 0.05$.

the PLSR method, it can be found that the method after moisture removal (Removed + CARS + PLSR) obtains the second highest accuracy, with a determination coefficient of 0.55 on the test set. Compared with the CARS + PLSR method, the accuracy is improved by at least 17.71%. In the model obtained using the SVR regression model, the accuracy of the Removed + CARS + PLSR method is at least 19.44% better than that of the CARS + PLSR method. SemiDNN achieves the highest accuracy. Although the RF method is capable to obtain high accuracy on the training set, the accuracy on the testing set is relatively low and therefore suffers from overfitting. WT-RF using wavelet transform and random forest regression has poor performance on this dataset. The main reason is that the pre-processing by the wavelet transform method may lead to a sensitivity decrease of the SOM features on the reflectance spectra. While the CR-PLSR method can improve the inversion accuracy through the continuum removal pre-processing method. However, compared to the proposed moisture removal model, our proposed method has more advantageous, and the accuracy is improved by at least 14%.

Therefore, after removing the effect of moisture, the spectra can significantly improve the feature expression of SOM, and can improve the model inversion accuracy by at least 19.44% or more. After removing the effect of moisture, the SOM inversion model can also reach an accuracy of 0.42 when using only one feature band, which is better than the (ALL + PLSR) and (CARS + PLSR) methods. Although the accuracy of the proposed method will be slightly lower than that of deep learning method, it has an less time consumption in the training process. The regression inversion model (Removed + CARS + SVR) after moisture removal achieves an accuracy of 0.69 on the testing set, which is consistent with the results of the published papers. For example, Levi et al. (2022) constructed a soil quality index (SQI) based on AisaFENIX hyperspectral airborne imagery to predicted SOM with a final accuracy of 0.72. Wu et al. (2023) developed a semi-empirical soil multi-factor radiative transfer model for soil organic matter prediction, which had an accuracy of 0.66 for GF-5 satellite imagery and 0.68 for Hymap airborne imagery. Majeed et al. (2023) achieved an accuracy of 0.70 for soil organic matter prediction model constructed on the AVIRIS-NG hyperspectral data. Whether on the satellite-based or airborne hyperspectral image data, the accuracy of the proposed model is similar to that of the existing soil organic matter inversion models, which demonstrates the high reliability of the proposed model. The mapping results of single-band based reversion model shows a high consistence with the SVR model. Therefore, the single-band inversion model has the advantages of fast and reliable in soil organic matter estimation and mapping.

Fig. 13 shows the two-dimensional kernel function density estimation of the organic matter inversion model in the sensitive band of 0.691 μm, the SVR statistical method and the SemiDNN method. The results obtained by the three methods have high consistency in their distributions, and all of them have the highest kernel function density estimates

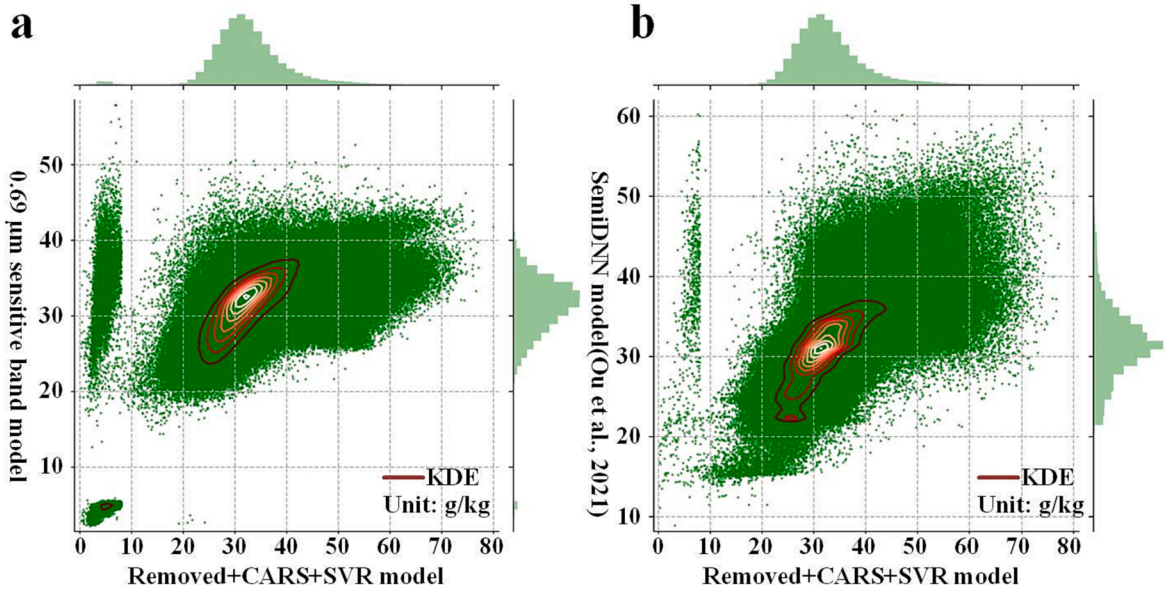


Fig. 13. Comparison of the kernel function density estimation for the soil organic matter inversion results. a. between the sensitive band model and the SVR regression model. b. between SVR regression model built in this manuscript and the SemiDNN regression model.

at a 31 g/kg content. The sensitive band model has fewer high values (greater than40 g/kg), while the SVR method has a more consistent distribution with the Gaussian distribution properties. However, the high-value interval predicted by the SVR method would be low compared with the sensitive band inversion model. Besides, it is indicated that though the sensitive band inversion model is enough for the whole-area trend analysis, a higher-precision SVR regression model should be introduced to reveal more detailed characteristics of spatial distribution. In general, the organic matter inversion model in the sensitive band of 0.691 μm exhibits the discontinuities in the low SOM value region.

4.2. Shortcomings and outlook

Despite the good performance of our proposed model, the experimental conditions described in this study have some influence on the model performance. Firstly, inconsistent sampling strategies lead to inaccurate model predictions. Fig. 13 demonstrates that the single-band inversion model fails to capture the distribution characteristics of SOM in the range of 5–10 g/kg. Even though the minimal SOM in the research area is 14 g/kg, it is concentrated near the two mining zones and is located consistent with the sampling sites, which is caused by the sampling conditions and the sampling schemes aiming to analyze the impact of mining activities on the SOM. The spatial distribution of the samples is not extensive enough to draw strong inferences in the entire research region. Although the SVR method can solve this issue, it is essential to further expand the sample size in low-density areas for single-band inversion method. Moreover, the model employs a single 2200 nm band for inversion of soil moisture and selects only one spectral curve for acquiring dry and wet soils, resulting in a low accurate soil moisture result in some regions.

The study area is arable land, which can be considered homogeneous soil, and the substances such as soil particle size can be considered constant. Moreover, the low content of soil heavy metals can hardly impact the spectrum (Wu et al., 2007). Therefore, the factors such as soil particle size or heavy metals are not considered in constructing the mechanism modelling based on an application-friendly model. Hence, the spectral correction due to iron oxides and other soil heavy metals should be further considered in our future works.

5. Conclusion

In this paper, we have provided a new spectral correction model based on K-M and Beer-Lambert models to remove the influence of soil moisture, which greatly improves the spectral feature expression of SOM. The soil spectra processed based on the Kubelka-Munk moisture removal model can effectively attenuate the influence of soil moisture on the spectra and highlight the feature information of SOM components in the spectra. The correlation between the spectra and SOM after removing the effect of moisture was improved by 16% at 0.69 μm, while it was improved by 35% at 2.166 μm. The inversion model for SOM obtained by the sensitive band at 0.69 μm can acquire the highest inversion accuracy, with a R^2 of 0.42. Using the spectral information after the removal of the soil moisture influence, combined with the CARS method for feature selection, and finally using the SVR model to construct the inversion model, the test set coefficient of determination reached as high as 0.69. The accuracy of the inversion model was improved by at least 19.44% or more, compared with the original spectra. This indicates that the spectral correction model for soil moisture removal can significantly improve the feature information of the SOM components in soil spectra, and the method is both effective and convenient. Spatial mapping of the results of sensitive band inversion and SVR regression was performed and analyzed. The results showed that the spatial distribution maps obtained by the two methods had a high similarity, and the overall distribution trends were consistent. However, the details of the SVR regression model are better presented.

CRediT authorship contribution statement

Depin Ou: Methodology, Data curation, Writing – original draft. **Kun Tan:** Conceptualization, Resources, Writing – review & editing. **Jie Li:** . **Zhifeng Wu:** Methodology, Writing – review & editing. **Liangbo Zhao:** Investigation. **Jianwei Ding:** Supervision. **Xue Wang:** Investigation. **Bin Zou:** Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This research was supported in part by the National Natural Science Foundation of China (No. 42171335, 42201385), and by China Postdoctoral Science Foundation (Grant No. 2022M720866, 2023T160218), the Shanghai Municipal Science and Technology Major Project (No. 22511102800) and the National Civil Aerospace Project of China (No. D040102).

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. 34 (11), 2274–2282.

Al-abbas, A.H., Swain, P.H., Baumgardner, M.F., 1972. Relating organic matter and clay content to the multispectral radiance of soils. Soil Sci. 114 (6), 477–485.

Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., Bochtis, D., 2019. Remote sensing techniques for soil organic carbon estimation: A review. Remote Sens. (Basel) 11, 676.

Babaeian, E., Sadeghi, M., Jones, S.B., Montzka, C., Vereecken, H., Tuller, M., 2019. Ground, proximal, and satellite remote sensing of soil moisture. Rev. Geophys. 57 (2), 530–616.

Bablet, A., Vu, P., Jacquemoud, S., Viallefont-Robinet, F., Fabre, S., Briottet, X., Sadeghi, M., Whiting, M.L., Baret, F., Tian, J., 2018. MARMIT: A multilayer radiative transfer model of soil reflectance to estimate surface soil moisture content in the solar domain (400–2500 nm). Remote Sens. Environ. 217, 1–17.

Bach, H., Mauser, W., 1994. Modelling and model verification of the spectral reflectance of soils under varying moisture conditions, Proceedings of IGARSS'94-1994 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 2354-2356.

Ben-Dor, E., 2002. Quantitative remote sensing of soil properties. Adv. Agron. 75, 173–243.

Ben-Dor, E., Inbar, Y., Chen, Y., 1997. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. Remote Sens. Environ. 61, 1–15.

Ben-Dor, E., Chabrillat, S., Demattè, J., Thenkabail, P., Lyon, J., Huete, A., 2019. Characterization of soil properties using reflectance spectroscopy. CRC Press, Boca Raton, Hyperspectral remote sensing of vegetation.

Berk, A., Anderson, G.P., Bernstein, L.S., Acharya, P.K., Dothe, H., Matthew, M.W., Adler-Golden, S.M., Chetwynd Jr, J.H., Richtsmeier, S.C., Pukall, B., 1999. MODTRAN4 radiative transfer modeling for atmospheric correction, Optical spectroscopic techniques and instrumentation for atmospheric and space research III. International Society for Optics and Photonics 348–353.

Bogrekci, I., Lee, W., 2006. Effects of soil moisture content on absorbance spectra of sandy soils in sensing phosphorus concentrations using UV-VIS-NIR spectroscopy. Trans. ASABE 49, 1175–1180.

Bricklemyer, R.S., Brown, D.J., 2010. On-the-go VisNIR: Potential and limitations for mapping soil clay and organic carbon. Comput. Electron. Agric. 70 (1), 209–216.

Castaldi, F., Palombo, A., Pascucci, S., Pignatti, S., Santini, F., Casa, R., 2015. Reducing the influence of soil moisture on the estimation of clay from hyperspectral data: A case study using simulated PRISMA data. Remote Sens. (Basel) 7, 15561–15582.

Drucker, H., Burges, C.J., Kaufman, L., Smola, A.J., Vapnik, V., 1997. Support vector regression machines. Adv. Neural Inf. Proces. Syst. 155–161.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1–17.

Gu, X., Wang, Y., Sun, Q., Yang, G., Zhang, C., 2019. Hyperspectral inversion of soil organic matter content in cultivated land based on wavelet transform. Comput. Electron. Agric. 167, 105053.

He, T., Wang, J., Lin, Z., Cheng, Y.e., 2009. Spectral features of soil organic matter. Geospatial Information Science 12 (1), 33–40.

Heinz, D.C., Chein-I-Chang, 2001. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. 39 (3), 529–545.
Hong, Y., Yu, L., Chen, Y., Liu, Y., Liu, Y., Liu, Y., Cheng, H., 2018. Prediction of Soil Organic Matter by VIS–NIR Spectroscopy Using Normalized Soil Moisture Index as a Proxy of Soil Moisture. Remote Sens. (Basel) 10, 28.
Jenson, S.K., Domingue, J.O., 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. Photogramm. Eng. Remote Sens. 54, 1593–1600.
Kokhanovsky, A., 2019. Springer Series in Light Scattering. Springer, New York.
Kortüm, G., 2012. Reflectance spectroscopy: principles, methods, applications. Springer Science & Business Media, New York.
Kubelka, P., 1931. Ein Beitrag zur Optik der Farbanstriche (Contribution to the optic of paint). Z. Tech. Phys. 12, 593–601.
Leardi, R., 2000. Application of genetic algorithm-PLS for feature selection in spectral data sets. J. Chemom. 14 (5-6), 643–655.
Lekner, J., Dorf, M.C., 1988. Why some things are darker when wet. Appl. Opt. 27, 1278–1280.
Levi, N., Karnieli, A., Paz-Kagan, T., 2022. Airborne imaging spectroscopy for assessing land-use effect on soil quality in drylands. ISPRS J. Photogramm. Remote Sens. 186, 34–54.
Li, H., Liang, Y., Xu, Q., Cao, D., 2009. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Anal. Chim. Acta 648 (1), 77–84.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2, 18–22.
Lobell, D.B., Asner, G.P., 2002. Moisture effects on soil reflectance. Soil Sci. Soc. Am. J. 66 (3), 722–727.
Majeed, I., Purushothaman, N.K., Chakraborty, P., Panigrahi, N., Vasava, H.B., Das, B.S., 2023. Estimation of soil and crop residue parameters using AVIRIS-NG hyperspectral data. Int. J. Remote Sens. 44 (6), 2005–2038.
Minasny, B., McBratney, A.B., Bellon-Maurel, V., Roger, J.-M., Gobrecht, A., Ferrand, L., Joalland, S., 2011. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. Geoderma 167, 118–124.
Nascimento, J.M.P., Dias, J.M.B., 2005. Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE Trans. Geosci. Remote Sens. 43 (4), 898–910.
Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., Mouazen, A.M., 2016. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. Soil Tillage Res. 155, 510–522.
Nemani, R., Pierce, L., Running, S., Goward, S., 1993. Developing satellite-derived estimates of surface moisture status. J. Appl. Meteorol. Climatol. 32 (3), 548–557.
O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. Computer vision, graphics, and image processing 28 (3), 323–344.
Ogen, Y., Faigenbaum-golovin, S., Granot, A., Shkolnisky, Y., Goldshleger, N., Bendor, E., 2019. Removing Moisture Effect on Soil Reflectance Properties: A Case Study of Clay Content Prediction. Pedosphere 29 (4), 421–431.
Ou, D., Tan, K., Lai, J., Jia, X., Wang, X., Chen, Y.u., Li, J., 2021. Semi-supervised DNN regression on airborne hyperspectral imagery for improved spatial soil properties prediction. Geoderma 385, 114875.
Ou, D., Tan, K., Wang, X., Wu, Z., Li, J., Ding, J., 2022. Modified soil scattering coefficients for organic matter inversion based on Kubelka-Munk theory. Geoderma 418, 115845.
Philpot, W., 2010. Spectral reflectance of wetted soils. Proceedings of ASD and IEEE GRS 2, 1–12.
Sadeghi, M., Jones, S.B., Philpot, W.D., 2015. A linear physically-based model for remote sensing of soil moisture using short wave infrared bands. Remote Sens. Environ. 164, 66–76.
Sadeghi, M., Babaeian, E., Tuller, M., Jones, S.B., 2017. The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. Remote Sens. Environ. 198, 52–68.
Tachikawa, T., Kaku, M., Iwasaki, A., Gesch, D.B., Oimoen, M.J., Zhang, Z., Danielson, J. J., Krieger, T., Curtis, B., Haase, J., 2011. ASTER global digital elevation model version 2-summary of validation results. NASA.
Tan, K., Ma, W., Chen, L., Wang, H., Du, Q., Du, P., Yan, B., Liu, R., Li, H., 2021. Estimating the distribution trend of soil heavy metals in mining area from HyMap airborne hyperspectral imagery based on ensemble learning. J. Hazard. Mater. 401, 123288.
Wang, X., Tan, K., Du, Q., Chen, Y.u., Du, P., 2019. Caps-TripleGAN: GAN-Assisted CapsNet for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens. 57 (9), 7232–7245.
Weidong, L., Baret, F., Xingfa, G.u., Qingxi, T., Lanfen, Z., Bing, Z., 2002. Relating soil surface moisture to reflectance. Remote Sens. Environ. 81 (2-3), 238–246.
Wu, Y., Chen, J., Ji, J., Gong, P., Liao, Q., Tian, Q., Ma, H., 2007. A Mechanism Study of Reflectance Spectroscopy for Investigating Heavy Metals in Soils. Soil Sci. Soc. Am. J. 71 (3), 918–926.
Wu, F., Tan, K., Wang, X., Ding, J., Liu, Z., 2023. A novel semi-empirical soil multi-factor radiative transfer model for soil organic matter estimation based on hyperspectral imagery. Geoderma 437, 116605.
Xu, X., Chen, S., Xu, Z., Yu, Y., Zhang, S., Dai, R., 2020. Exploring Appropriate Preprocessing Techniques for Hyperspectral Soil Organic Matter Content Estimation in Black Soil Area. Remote Sens. (Basel) 12, 3765.
Yu, L., Hong, Y., Geng, L., Zhou, Y., Zhu, Q., Cao, J., Nie, Y., 2015. Hyperspectral estimation of soil organic matter content based on partial least squares regression. Transactions of the Chinese Society of Agricultural Engineering 31, 103–109.
Yu, J., Yan, B., Liu, W., Li, Y., He, P., 2017. Seamless Mosaicking of Multi-strip Airborne Hyperspectral Images Based on Hapke Model, International Conference on Sensing and Imaging. Springer, pp. 285–292.
Yuan, J., Wang, X., Yan, C.-X., Wang, S.-R., Ju, X.-P., Li, Y., 2019. Soil moisture retrieval model for remote sensing using reflected hyperspectral information. Remote Sens. (Basel) 11, 366.
Zhang, N., Hong, Y., Qin, Q., Liu, L.u., 2013. VSDI: a visible and shortwave infrared drought index for monitoring soil and vegetation moisture based on optical remote sensing. Int. J. Remote Sens. 34 (13), 4585–4609.
Zhang, Y., Tan, K., Wang, X., Chen, Y., 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 Sens. (Basel) 12, 2239.