Geoderma 437 (2023) 116605

Contents lists available at ScienceDirect

Geoderma

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

A novel semi-empirical soil multi-factor radiative transfer model for soil organic matter estimation based on hyperspectral imagery

Fuyu Wu [a], Kun Tan [b,c,d,*], Xue Wang [b,c,d], Jianwei Ding [e], Zhaoxian Liu [e]

[a] Key Laboratory for Land Environment and Disaster Monitoring of NASG, China University of Mining and Technology, Xuzhou 221116, 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] School of Geographic Sciences, East China Normal University, Shanghai 200241, China
[e] Second Surveying and Mapping Institute of Hebei, Shijiazhuang 050037, China

ARTICLE INFO

Handling Editor: Budiman Minasny

Keywords:
Soil multifactor radiative transfer model
Hyperspectral imagery
Absorption and scattering coefficients
SOM concentration estimation

ABSTRACT

Soil organic matter (SOM) concentration is an important factor affecting soil quality, and rapid and wide-scale monitoring of SOM concentration is a key step toward sustainable agriculture. Hyperspectral technology is widely used in soil composition monitoring, due to its rich spectral information. The complex imaging environment of hyperspectral imagery and the mixed pixel problem have led to the current applications of soil condition estimation mostly using data-driven methods. However, the estimation process based on data-driven methods cannot be explained by radiative transfer theory. Therefore, in this paper, a semi-empirical soil multi-factor radiative transfer (SESMRT) model combining a soil radiative transfer model and data-driven model is proposed for SOM estimation based on hyperspectral imagery. The radiative transfer model in the SESMRT model fully considers the soil components, such as SOM, soil moisture, soil iron oxides, and particle size distribution, and achieves high-precision simulation of soil spectra by resolving the mechanism of the influence of soil components on spectra. The SOM spectra calculated by the radiative transfer model are then used to estimate the SOM concentration based on a data-driven model. The SOM spectra eliminate the interference of other factors in the spectrum and significantly enhance the correlation between the spectrum and SOM. Thus, the SOM concentration can be estimated with a high degree of accuracy ($R^2 = 0.660 \pm 0.034$, RMSE $= 3.923 \pm 0.236$, RPIQ $= 2.003 \pm 0.118$ for GF5 and $R^2 = 0.685 \pm 0.030$, RMSE $= 3.543 \pm 0.311$, RPIQ $= 2.309 \pm 0.207$ for HyMap). Finally, the comparison of the estimated results for 2017 and 2019 shows that the SOM concentration of the cultivated soils in the study area increased, while the opposite is observed for the soils around the mining areas.

1. Introduction

Soil is the key to food security, and soil quality directly affects the yield and quality of food production, especially the quality of cultivated soil (Lal 2004). Thus, rapid and extensive monitoring of soil organic matter (SOM) concentration in cultivated land is essential for food production security (Chen et al. 2022). Hyperspectral technology is widely used in soil composition estimation due to its rich spectral information (Gomez et al. 2012). With the release of hyperspectral images from a series of satellites and airborne platforms, how to estimate soil composition efficiently and accurately has become the focus of current research (Tian et al. 2013).

Most of the soil composition estimation methods based on hyperspectral data adopt data-driven models, and the concentration estimation is achieved by deeply mining the relationship between the spectra and soil composition through machine learning methods. The process of data-driven modeling can be divided into three parts: 1) spectral preprocessing; 2) feature selection; and 3) construction of a regression model. The spectral preprocessing includes filter noise removal (Tian et al. 2013), continuum removal (Bao et al. 2020), and spectral mathematical operation (Wang et al. 2018, Meng et al. 2021), which is aimed at reducing the spectral noise and improving the correlation between the spectra and soil components. Feature selection is aimed at reducing the redundancy of the spectral information and providing valid spectral data

* Corresponding author at: School of Geographic Sciences, East China Normal University, Shanghai 200241, China.
E-mail address: tankuncu@gmail.com (K. Tan).

https://doi.org/10.1016/j.geoderma.2023.116605
Received 10 January 2023; Received in revised form 3 July 2023; Accepted 5 July 2023
Available online 17 July 2023

for constructing regression models. The Pearson correlation coefficient (PCC) (Yang et al. 2022), competitive adaptive reweighted sampling (CARS) (Guo et al. 2021), and meta heuristic (MH) optimization algorithms (Sun et al. 2022) are often used for spectral feature selection. Constructing the regression model, which is a key step in the data-driven models, establishes the relationship between the spectral data and soil composition by learning the distribution of the spectral feature data. Soil composition estimation mostly involves the use of machine learning models, including partial least squares regression (PLSR) (Selige et al. 2006, Nanni et al. 2021), ensemble learning (EL) (Meng et al. 2021), support vector regression (SVR) (Guo et al. 2021), and deep neural networks (DNNs) (Ou et al. 2021) and convolutional neural networks (CNNs) (Wang et al. 2022). Although the data-driven models for soil parameter estimation based on hyperspectral imagery can achieve a good estimation accuracy, they cannot further explain the mechanism of the spectral features.

Soil radiative transfer models cannot easily describe the radiative transfer process accurately because of the complex and dense composition of the soil, and the low transmittance. Therefore, soil radiative transfer models tend to describe the main soil influencing factors. The main influencing factors of soil spectra include the soil-forming matrix (Demattê and da Silva Terra 2014), particle size distribution (PSD) (Janik et al. 2020), soil moisture (Zhang et al. 2020, Koch et al. 2021), SOM (Liu et al. 2018, Wang et al. 2020), soil iron oxides, and measurement conditions. The current research on soil radiative transfer models mainly focuses on three factors: 1) PSD; 2) soil moisture; and 3) SOM. Soil radiative transfer models can be broadly divided into three categories: 1) beam tracing (BT) models; 2) bidirectional reflection distribution function (BRDF) radiative transfer models; and 3) the multi-flux radiative transfer (MFRT) model. The BT models involve describing the reflection and refraction of optical radiation between the air and the soil medium, and the absorption inside the soil with Fresnel equations and Snell's law. Bablet et al. (2018) proposed the multilayer radiative transfer model of soil reflectance (MARMIT) based on the BT model, which considers wet soil as a combination of aqueous film and dry soil, and represents the radiative transfer process of wet soil by describing the refraction and scattering of light between water and the soil medium. Dupiau et al. (2022) proposed the MARMIT-2 model to address the problem of the large differences between the model assumptions and reality in the original MARMIT model. The MARMIT-2 model considers the mixing state of soil particles in the water layer and improves the previous water refractivity to mixed reflectivity, which is more consistent with the actual situation of soil radiative transfer. Sadeghi et al. (2018) described the effect of soil particle size on soil spectra using the BT model and extended the model from single particles to natural soils by an integral form (Norouzi et al. 2021). The BRDF model describes the distribution of incident and reflected light in hemispheric space. The Hapke model is the most widely used BRDF radiative transfer model for soil composition estimation, which decomposes the total radiance into the sum of single-scattering radiance and multiple-scattering radiance under the assumption of anisotropy. Based on the Hapke model, Jacquemoud et al. (1992) proposed the SOILSPECT model for the soil radiative transfer process, replacing the phase function in the Hapke model with the Legendre polynomial approximation of the phase function to explain the backward and forward scattering from the soil. Liang and Townshend (1996) divided the total radiance into single-scattering radiance, double-scattering radiance, and multiple-scattering radiance. The experimental results showed that the improved model significantly improves the calculation accuracy of the bidirectional reflectance. Ding et al. (2022) derived the relationship between single-scattering albedo and wavelength based on the Hapke model and proposed the Hapke-HSR model. The light soil spectra with different moisture concentrations were then simulated by the brightness shape moisture (BSM) model. The MFRT model ignores the multiple reflections and scatterings of optical radiation between media, and only considers the multi-directional fluxes of light entering the media and after absorption by the media, which greatly simplifies the difficulty of solving the radiative transfer model. The Kubelka-Munk (KM) model, which is a two-flux radiative transfer model with only downward and upward fluxes, is widely used for soil composition estimation. Sadeghi et al. (2015) constructed a soil moisture estimation model based on the KM model, in which the absorption and scattering coefficients of the soil are decomposed into the concentration weights of the absorption and scattering coefficients of dry soil, soil air, and soil moisture. The soil moisture concentration is then solved through the spectral reflectance of dry soil and moisture-saturated soil by explicitly using the model formulation. Ou et al. (2022) derived the soil thickness equation using the KM model, calculated the scattering coefficients of soil from soil spectra at different thicknesses, and implemented SOM estimation according to the scattering coefficients.

However, limited by the strict theoretical derivation of the soil radiative transfer models, recent research has focused on indoor controlled-variable experiments. Furthermore, hyperspectral imagery spectra are different from indoor spectra due to the complex imaging environment and the mixed pixel elements, which makes it difficult to extend the soil radiative transfer models to hyperspectral imagery. In this study, in order to enhance the correlation between the spectrum and SOM and to extend the radiative transfer model to hyperspectral imagery, the semi-empirical soil multi-factor radiative transfer (SESMRT) model is proposed to estimate SOM concentration. The main contributions of this paper are as follows:

1) A new semi-empirical soil multi-factor radiative transfer model consisting of a soil radiative transfer model and a data-driven model for SOM estimation based on hyperspectral imagery has been investigated.

2) The SESMRT model describes the soil radiative transfer process by considering the main factors affecting the spectra and simulates the imaging spectra with a high degree of accuracy. The accurate radiative transfer model provides theoretical support for the subsequent SOM estimation.

3) SOM spectra are calculated based on the soil radiative transfer model, which removes the interference of the remaining soil factors and enhances the correlation between the spectrum and SOM.

4) The SESMRT model utilizes a data-driven model to achieve a wide range of SOM concentration estimation through SOM spectra based on hyperspectral images from two different platforms.

2. Method

The SESMRT model describes the radiative transfer process of each soil factor through the soil radiative transfer (RT) model, analyzes the influencing mechanism on the spectrum, and extracts the SOM spectrum, which reduces the influence of other factors on the hyperspectral imagery spectrum, and finally achieves SOM estimation by combining with the data-driven model.

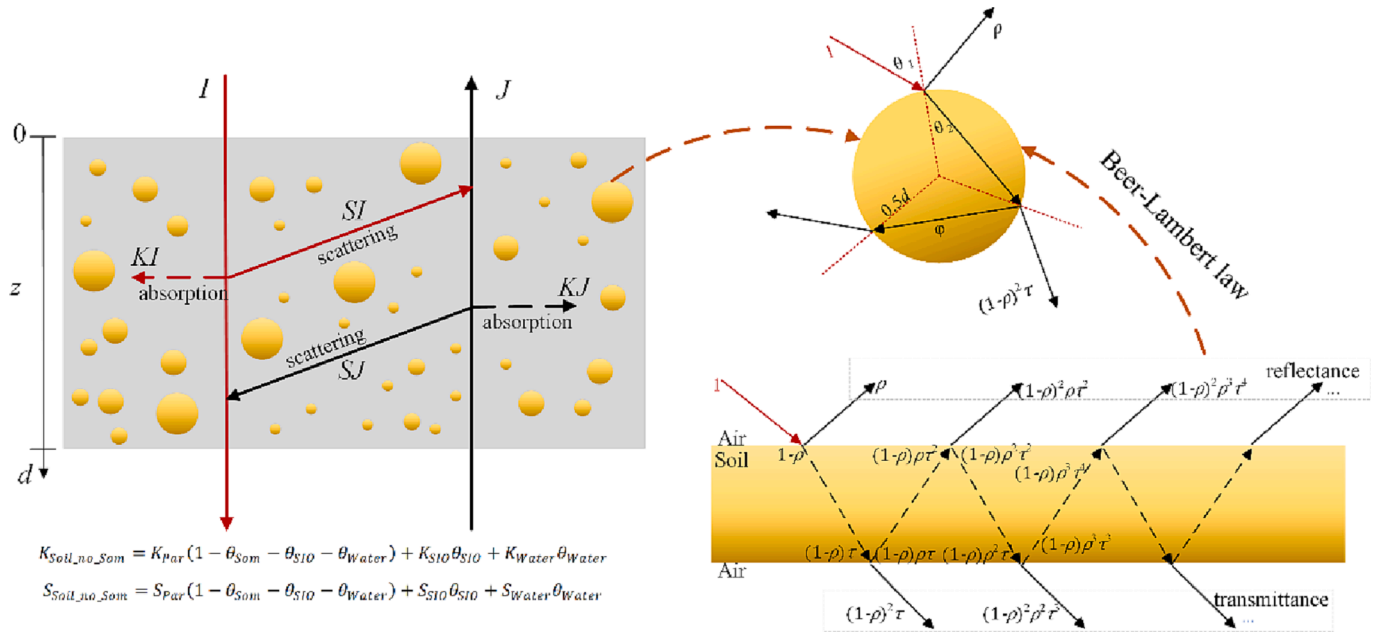
2.1. Calculation of SOM concentration based on the RT model

As shown in Fig. 1, the RT model is based on KM theory, which ignores the complex reflection and refraction and only considers the fluxes in the upper and lower directions (Ciani et al. 2005) to describe the soil radiative transfer process.

Although KM theory simplifies the soil radiative transfer process, its conclusions are derived strictly from the radiative transfer equation, which means that it has excellent physical meaning (Vargas and Niklasson 1997). As the soil thickness increases, the soil transmittance gradually becomes 0, and the soil reflectance can be solved by Eq. (1) (Christy et al. 1995):

$$R_\infty = 1 + \frac{K_{Soil}}{S_{Soil}} - \sqrt{\left(\frac{K_{Soil}}{S_{Soil}}\right)^2 + 2\frac{K_{Soil}}{S_{Soil}}} \quad (1)$$

where R_∞ is the soil reflectance; and K_{Soil} (mm^{-1}) and S_{Soil} (mm^{-1}) are



$$K_{Soil\_no\_Som} = K_{Par}(1-\theta_{Som}-\theta_{SIO}-\theta_{Water}) + K_{SIO}\theta_{SIO} + K_{Water}\theta_{Water}$$

$$S_{Soil\_no\_Som} = S_{Par}(1-\theta_{Som}-\theta_{SIO}-\theta_{Water}) + S_{SIO}\theta_{SIO} + S_{Water}\theta_{Water}$$

Fig. 1. Schematic representation of the radiation transmission process of the RT model.

the absorption and scattering coefficients of soil, respectively.

The absorption and scattering coefficients of mixed substances in KM theory are usually considered as the proportional weighted sum of the absorption and scattering coefficients of each substance. Thus, as shown in Eqs. (2) and (3), the soil absorption and scattering coefficients can be expressed by the proportional weighted sum of the absorption and scattering coefficients of the SOM, soil moisture, soil iron oxides, and soil particles characterizing the soil-forming parent material (Sadeghi et al. 2015).

$$K_{Soil} = K_{Par}(1-m_{Som}-m_{SIO}-m_{Moisture}) + K_{Som}m_{Som} + K_{SIO}m_{SIO} + K_{Moisture}m_{Moisture} \tag{2}$$

$$S_{Soil} = S_{Par}(1-m_{Som}-m_{SIO}-m_{Moisture}) + S_{Som}m_{Som} + S_{SIO}m_{SIO} + S_{Moisture}m_{Moisture} \tag{3}$$

where K_{Par} (mm^{-1}), S_{Par} (mm^{-1}), K_{Som} (mm^{-1}), S_{Som} (mm^{-1}), K_{SIO} (mm^{-1}), S_{SIO} (mm^{-1}), $K_{Moisture}$ (mm^{-1}), and $S_{Moisture}$ (mm^{-1}) represent the absorption and scattering coefficient of soil particles, SOM, soil iron oxides, and soil moisture, respectively. m_{Som} (%), m_{SIO} (%) and $m_{Moisture}$ (%) represent the concentration of SOM, soil iron oxides, and soil moisture, respectively.

The absorption and scattering coefficients of soil particles K_{Par} and S_{Par} are calculated by accumulating the reflectance r and transmittance t of single particles. According to Eqs. (4)–(8), the reflectance r and transmittance t of a single particle are obtained by converting the reflectance and transmittance of a flat plate using Beer-Lambert law, where the reflectance r_{Plat} and transmittance t_{Plat} of the flat plate are calculated by accumulating the reflection, absorption, and transmission of the light beams in different media several times (Tuckerman 1947).

$$r_{Plat} = \rho + (1-\rho)^2\rho\tau^2 + (1-\rho)^2\rho^3\tau^4 + \cdots = \rho + \frac{(1-\rho)^2\rho\tau^2}{1-\rho^2\tau^2} \tag{4}$$

$$t_{Plat} = (1-\rho)^2\tau + (1-\rho)^2\rho^2\tau^3 + \cdots = \frac{(1-\rho)^2\tau}{1-\rho^2\tau^2} \tag{5}$$

$$\tau = exp(-k\varphi) \tag{6}$$

$$r = \rho + \frac{(1-\rho)^2\rho exp(-2k\varphi)}{1-\rho^2 exp(-2k\varphi)} \tag{7}$$

$$t = \frac{(1-\rho)^2 exp(-k\varphi)}{1-\rho^2 exp(-2k\varphi)} \tag{8}$$

where ρ and τ represent the reflectivity and transmissivity of the light beam between two media; and k and φ are the linear absorption coefficient and length of the beam light inside the soil layer, respectively. According to the conclusions of Sadeghi et al. (2018), φ can be approximated as being equal to the particle size d(mm). Since there are soil particles of different particle sizes in the soil, the total reflectance r^* and transmittance t^* of soil particles can be obtained by accumulating the ratios of the different particle sizes.

$$r^* = \sum_i^n r(\rho, k, d_i)P_i \tag{9}$$

$$t^* = \sum_i^n t(\rho, k, d_i)P_i \tag{10}$$

where P_i is the probability of particle size d_i. Therefore, the absorption and scattering coefficients of soil particles K_{Par} and S_{Par} are calculated as the absorption and scattering at a single level multiplied by the number of levels per unit depth N (Banninger and Fluhler 2004).

$$K_{Par} = N^*(1-r^*-t^*) \tag{11}$$

$$S_{Par} = N^*r^* \tag{12}$$

Detailed soil PSD data are difficult to obtain for hyperspectral imagery applications, so soil texture data are used instead, and the average values of the clay, silt, and sand particle size ranges are used as the particle size d_i.

Fresnel reflections, caused by different media of soil and moisture, are often considered in the soil with soil moisture (Sadeghi et al. 2015), although they have weak influence on the reflectance.

$$R_{Fresnel} = \left(\frac{n_w - n_a}{n_w + n_a}\right)^2 m_{Moisture} \tag{13}$$

where n_w and n_a are the mean refractive index of water and air, respectively. Thus, the infinite reflectance of soil (R_∞) can be represented as:

$$R_\infty = R_{Fresnel} + 1 + \frac{K_{Soil}}{S_{Soil}} - \sqrt{\left(\frac{K_{Soil}}{S_{Soil}}\right)^2 + 2\frac{K_{Soil}}{S_{Soil}}} \tag{14}$$

The relationship between SOM content and soil spectrum can be derived from Eq.(2), Eq. (3) and Eq. (14):

$$m_{Som} = f(R_\infty,\ m_{SIO}, m_{Moisture}, P_i, d_i) = \frac{A_1}{A_2} \tag{15}$$

$$\begin{aligned} A_1 = {} & A_3^2(S_{Moisture} - S_{Par}){m_{Moisture}}^2 - 2A_3\{[(m_{SIO}/2 - 1/2)S_{Som} \\ & - S_{SIO}m_{SIO}/2]A_3 + (S_{Moisture} \\ & - S_{Par})R_\infty - S_{Moisture} + K_{Par} - K_{Moisture} + S_{Par}\}m_{Moisture}^2 + \{\{[(2m_{SIO} \\ & - 2)S_{Par} - 2S_{SIO}m_{SIO}]R_\infty - (2m_{SIO} - 2)S_{Par} + 2(K_{SIO} + S_{SIO} \\ & - K_{Par})m_{SIO} + 2K_{Par}\}A_3 + (S_{Moisture} - S_{Par})R_\infty^2 + 2(K_{Par} + S_{Par} \\ & - K_{Moisture} - S_{Moisture})R_\infty + S_{Moisture} - S_{Par}\}m_{Moisture} + [(1 \\ & - m_{SIO})S_{Par} + S_{SIO}m_{SIO}]R_\infty^2 + [(2m_{SIO} - 2)S_{Par} + 2(K_{Par} - K_{SIO} \\ & - S_{SIO})m_{SIO} - 2K_{Par}]R_\infty + (1 - m_{SIO})S_{Par} + S_{SIO}m_{SIO} \end{aligned} \tag{16}$$

$$\begin{aligned} A_2 = {} & A_3^2(S_{Par} - S_{Som})m_{Moisture}^2 - 2[(S_{Par} - S_{Som})R_\infty - K_{Par} \\ & - S_{Par} + K_{Som} + S_{Som}]A_3 m_{Moisture} + (S_{Par} - S_{Som})R_\infty^2 \\ & + 2(K_{Som} + S_{Som} - K_{Par} - S_{Par})R_\infty + S_{Par} - S_{Som} \end{aligned} \tag{17}$$

$$A_3 = \left(\frac{n_w - n_a}{n_w + n_a}\right)^2 \approx 0.02 \tag{18}$$

Since the soil multi-factor radiative transfer model assumes a consistent soil-forming parent material, soil samples need to be classified and then solved for SOM concentration based on each category in hyperspectral imagery applications.

The parameters including soil particle coefficients (ρ and k), absorption coefficients (K_{Som}, K_{SIO} and $K_{Moisture}$) and scattering coefficients (S_{Som}, S_{SIO} and $S_{Moisture}$) are obtained by genetic algorithm (GA) algorithm. The samples are randomly divided into a training set and a testing set in a ratio of 2:1, where the training set is used to obtain the model parameters and the testing set is used to verify the accuracy of SOM concentration simulated by RT model. The soil moisture varies rapidly due to rainfall and evaporation, so the soil moisture ($m_{Moisture}$) concentration is obtained by estimation of the KM model index. Soil iron oxide is relatively stable in the soil and the concentration (m_{SIO}) of each pixel in the hyperspectral image is obtained by interpolation using the sampling point assay data. Soil texture data is derived from published soil texture results (Liu et al. 2020).

2.2. Estimation of SOM concentration based on the SESMRT model

The SESMRT model is proposed in order to tackle the issue that the soil multi-factor radiative transfer model could not be applied on a large scale due to its assumptions. The SESMRT model reduces the interference of spectral features by calculating SOM spectra.

Due to the low spatial resolution of hyperspectral remote sensing images, the proportion of Fresnel reflections caused by soil moisture is very weak in each pixel. Meanwhile, as the SESMRT model is different from the RT model which requires a strict derivation formula, the Fresnel reflections caused by soil moisture are ignored in the SESMRT model, and only the absorption and scattering effects of soil moisture are considered to reduce the complexity and improve the robustness of the model. Thus, the soil reflectance can be solved by Eq. (1).

The soil spectra with SOM removals $R_{Soil\_no\_Som}$ are solved through the model parameters using Eqs. (19)–(21), where the absorption coefficient $K_{Soil\_no\_Som}$ and scattering coefficient $S_{Soil\_no\_Som}$ are calculated by removing the SOM factor.

$$R_{Soil\_no\_Som} = 1 + \frac{K_{Soil\_no\_Som}}{S_{Soil\_no\_Som}} - \sqrt{\left(\frac{K_{Soil\_no\_Som}}{S_{Soil\_no\_Som}}\right)^2 + 2\frac{K_{Soil\_no\_Som}}{S_{Soil\_no\_Som}}} \tag{19}$$

$$K_{Soil\_no\_Som} = K_{Par}(1 - m_{Som} - m_{SIO} - m_{Moisture}) + K_{SIO}m_{SIO} + K_{Moisture}m_{Moisture} \tag{20}$$

$$S_{Soil\_no\_Som} = S_{Par}(1 - m_{Som} - m_{SIO} - m_{Moisture}) + S_{SIO}m_{SIO} + S_{Moisture}m_{Moisture} \tag{21}$$

Eventually, as shown in Eq. (22), the SOM spectra R_{Som} is determined by subtracting the soil spectra with SOM removals from the original spectra.

$$R_{Som} = R_{Measure} - R_{Soil\_no\_Som} \tag{22}$$

Since the SOM concentration, which is the focus of the estimation application in this study, is an unknown factor, the SOM spectra cannot be calculated accurately in a large-scale application by Eqs. (19)–(21). To solve this problem, the absorption coefficient $K_{Soil\_no\_Som}$ and scattering coefficient $S_{Soil\_no\_Som}$ are calculated using Eqs. (23) and (24). Although the SOM concentration is transferred to the soil particles, it has less effect on the calculation of SOM spectra because the SOM percentage in soil is small, accounting for only about 3%, and the subsequent comparison results also show the small difference between the SOM spectra calculated in the two different ways.

$$K_{Soil\_no\_Som} = K_{Par}(1 - m_{SIO} - m_{Moisture}) + K_{SIO}m_{SIO} + K_{Moisture}m_{Moisture} \tag{23}$$

$$S_{Soil\_no\_Som} = S_{Par}(1 - m_{SIO} - m_{Moisture}) + S_{SIO}m_{SIO} + S_{Moisture}m_{Moisture} \tag{24}$$

Since some spectral information in SOM spectra is redundant and affects the accuracy of SOM estimation, the CARS algorithm is used for feature selection with the SOM spectra. Finally, the SVR regression algorithm is used for SOM concentration estimation, based on the selected spectral features. The SESMRT model parameters are solved in a similar way to the RT model, where the adaptation function of the GA algorithm in the SESMRT model is the *RMSE* of the real sample spectra and the model simulated spectra.

2.3. Evaluation methods

To evaluate the performance of the SESMRT model for SOM estimation based on hyperspectral imagery, as shown in Eqs. (25)–(28), the coefficient of determination (R^2), the root-mean-square error (*RMSE*), the mean absolute error (*MAE*), the residual prediction deviation (RPD) and the ratio of prediction performance to interquartile range (RPIQ) are introduced in this paper (Nocita et al. 2013, Chakraborty et al. 2017, Ou et al. 2021):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{25}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n}} \tag{26}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{27}$$

$$RPIQ = \frac{IQ}{RMSE} \tag{28}$$

where y_i is the measured value, $\hat{y}_i$ is the predicted value, $\hat{y}$ is the average of the measured value, n is the number of samples, and IQ is the interquartile distance.

3. Experimental data

3.1. Study area and soil composition measurement

The study area is located in Yitong Manchu Autonomous County, Jilin province, China, where the soil type is dominated by black soil, and the area is an important grain production base in China (Fig. 2). There are two gold mines in the study area, and the mining has led to serious contamination of the cultivated soils, especially those around the mining areas.

To validate the performance of the SESMRT model for SOM estimation and to analyze the variation of SOM concentration in the study area, 87 soil samples from the same locations were collected in 2017 and 2019, respectively. The soil samples were used to measure the concentration of SOM and soil iron oxides after air-drying, grinding, and sieving.

The SOM and soil iron oxide concentration were determined by the potassium dichromate method and inductively coupled plasma-mass spectrometry (ICP-MS), respectively. As shown in Table 1, the mean SOM concentration for 2019 is 22.3 g/kg, which represents a significant decrease compared with that in 2017 (30.8 g/kg).

Because iron oxides are relatively stable in the soil and the iron oxide concentration is mainly influenced by the soil-forming parent material, the kriging interpolation result for the soil iron oxides based on the 2019 chemical results was used in the GF5 and HyMap hyperspectral imagery application.

Table 1
Analysis and statistics of the SOM and soil iron oxide concentration (g/kg).

	Mean	Max	Min	Sd.
SOM (2017)	30.8	49.8	14.7	6.2
SOM (2019)	22.3	38.9	5.0	6.8
Soil iron oxide	23.9	36.7	14.0	4.3

3.2. Hyperspectral imagery

The GF5 hyperspectral imagery was acquired on 1 February 2019 by the Visible-shortwave Infrared Advanced Hyperspectral Imager (AHSI) on board the GF-5 satellite (Fig. 3a). The technical parameters of the AHSI sensor are listed in Table 2.

The calibration coefficients proposed by Niu et al. (2022) were used for the radiometric calibration, and the FLAASH model was used for the atmospheric correction. Finally, the GF5 hyperspectral imagery had 290 bands after removing the water vapor bands.

The HyMap hyperspectral imagery was acquired in early May 2017 by the HyMap-C airborne hyperspectral imaging spectrometer with a wavelength range of 450 nm to 2500 nm (Fig. 4a). Table 3 lists the technical parameters of the HyMap-C sensor.

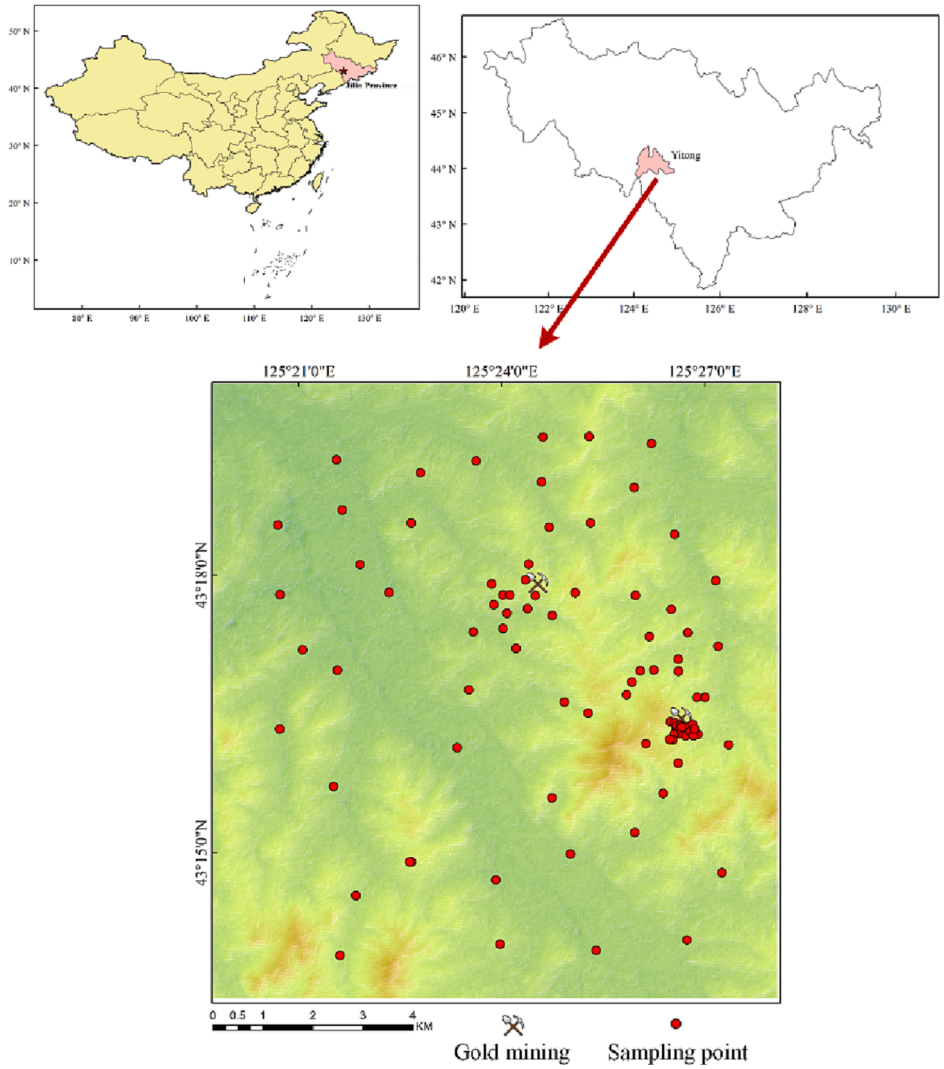


Fig. 2. Study area and sampling point locations.

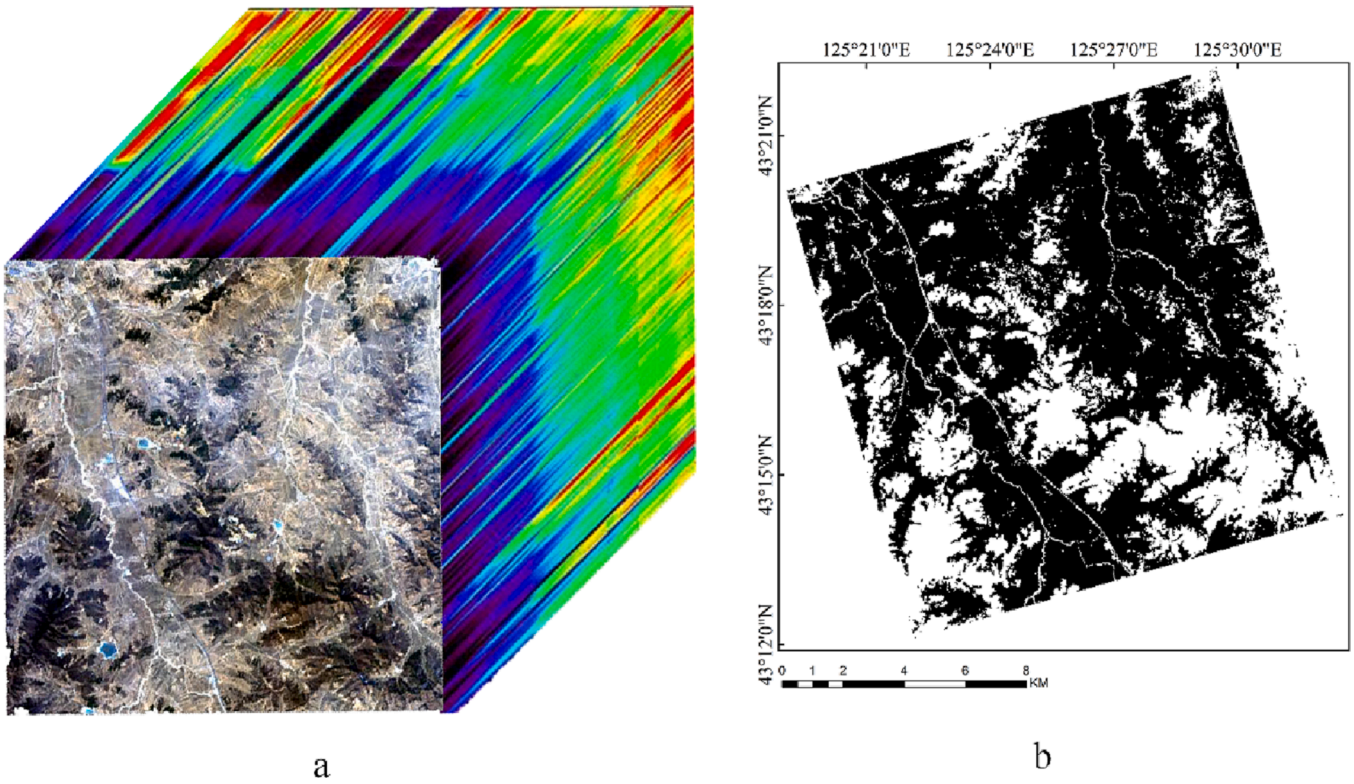


Fig. 3. A. 3d diagram of the gf5 hyperspectral imagery. b. soil information extracted from the gf5 hyperspectral imagery.

Table 2
Parameters of the GF5 AHSI sensor.

Technical indicator	Parameter
Spectral range	400–2500 nm
Spectral resolution	5 nm @ 400–1000 nm 10 nm @ 1000–2500 nm
Spatial resolution	30 *m*
Swath width	60 km
Satellite orbit	Sun-synchronous regressive orbit
Return period	51 days

The data acquisition time was from 11:00 to 15:00 to meet the solar altitude angle requirement. The altitude of the flight was 1800 m, and the spatial resolution was 4.5 m after geometric correction. The final image is a stitching of nine strips captured on the same day, and BRDF correction was processed using the HyMap-C sensor companion software (Lev1GUI). The Lev1GUI software uses a polynomial fitting algorithm to assess the pixel brightness across a scene and normalizes the values based on geometry of flight data to eliminate the significant radiometric differences between the different strips. The parameters acquired by an integrating sphere were used for the radiometric calibration, and atmospheric correction was performed using the

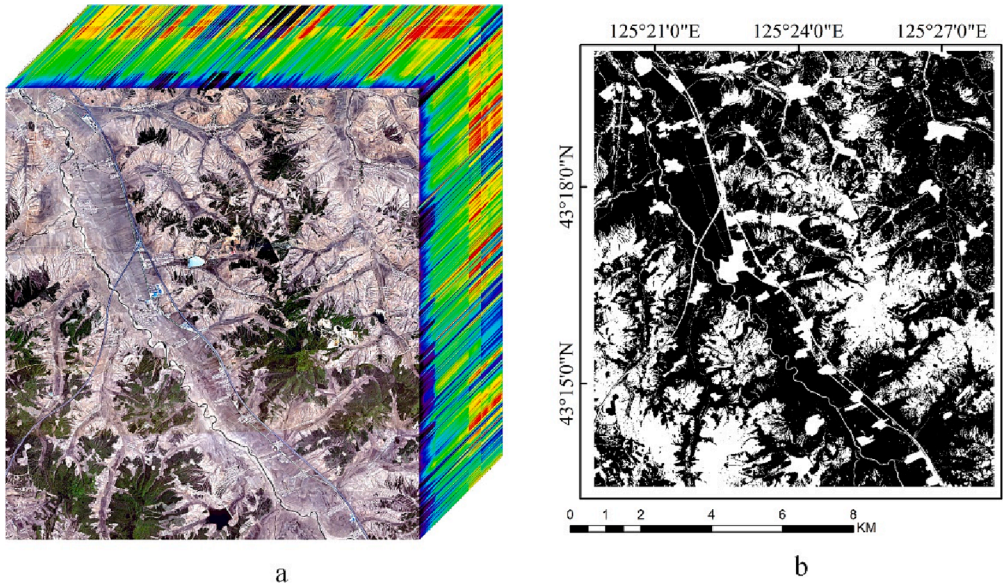


Fig. 4. A. 3d diagram of the hymap hyperspectral imagery. b. soil information extracted from the hymap hyperspectral imagery.

Table 3
Parameters of the HyMap-C sensor.

Technical indicator	Parameter
Spectral range	450–2500 nm
Spectral resolution	15 nm @ 450–905 nm, 880–1440 nm 18 nm @ 1400–1960 nm, 1950–2500 nm
Spatial resolution	4.5 m (the altitude of the flight is 1800 m)
Field of view	60°
Radiometric accuracy	95%
Signal noise ratio	VNIR: 1000:1 SWIR: 600:1
Working temperature	−10 to 40℃

MODTRAN4 atmospheric radiative transfer model, based on the atmospheric condition data. Finally, the HyMap hyperspectral imagery had 101 bands after removing the water vapor bands.

In the study area, the ground objects are generally classified into bare soil, vegetation, buildings, and water, and the soil information was extracted by the support vector machine (SVM) classification method (Fig. 3b and Fig. 4b).

3.3. Soil moisture estimation results

Soil moisture laboratory data do not accurately describe the effect of soil moisture on the spectrum as the concentration is unstable due to rainfall, artificial watering, and other factors. In this study, the soil moisture concentration was estimated directly from the hyperspectral imagery. The estimation method uses the KM model index, which has been shown to be superior to other methods for soil moisture concentration estimation from hyperspectral imagery:

$$m_{Moisture} = (STR - STR_D)/(STR_W - STR_D) \quad (29)$$

$$STR = (1 - R)^2/2R \quad (30)$$

where STR is the soil moisture-related parameter derived from the KM model; and STR_W and STR_D are the STR parameters indicating the wettest soil and driest soil in the study area, respectively. According to the conclusion of Sadeghi et al. (2015), the reflectance R at 2210 nm is generally used to obtain the highest estimation accuracy. Since STR_W and STR_D can be considered as a constant, Eq. (29) is a monotonically increasing function and Eq.(30) is a monotonically decreasing function between 0 and 1. Thus, when the soil moisture increases, the value of STR increases and R decreases, which satisfies the prior knowledge that the spectral reflectance decreases with increasing moisture. Soil moisture around the rivers in the study area is approximately saturated, and the continuous sunny weather at the time of imagery acquisition ensures the existence of dry soil. Thus, the STR calculated from the minimum and maximum value of soil reflectance at 2210 nm can be used as STR_W STR_D, respectively.

Fig. 5 shows the results of the soil moisture concentration estimation based on GF5 and HyMap hyperspectral imagery.

3.4. Soil particle size distribution and soil texture data

To determine the N value in Eqs.(11) and (12), the soil particle size distribution in the study area was measured, and the average particle size in the study area was determined to be 0.01 mm. According to the conclusion of Norouzi et al. (2021), the effective information depth of the soil spectrum is approximately 1 mm, so in this study, N value was set to 100 based on the average particle size.

As shown in Fig. 6, the soil texture data used in this study were the national soil texture data for China with a 90-m spatial resolution produced by Liu et al. (2020). According to the United States Department of Agriculture (USDA) classification system, the soil taxonomy is mainly silt, silt loam, and silt clay loam in the study area. The soil texture data were upsampled to the spatial resolution of the GF5 hyperspectral imagery (30 m) and HyMap hyperspectral imagery (4.5 m).

3.5. Soil classification data for RT model

Since the RT model assumes that the soil-forming parent matrices is the same, it is necessary to classify the soils in the study area and estimate the SOM concentration by category. Soil spectra, directly responded to the differences in soil-forming parent matrices, were used as the basis for clustering in this study. Finally, fourteen classes were determined according to the accuracy of organic matter calculation, and the SVM classifier was trained based on the clustering results. The soil classification results are shown in Fig. 7.

4. Result and analysis

4.1. Experimental results and analysis of RT model

Soil samples in each class are divided into a training set and a testing set in a ratio of 2:1, where the training set is used to fit the model parameters, and the SOM concentration is estimated by the band with the

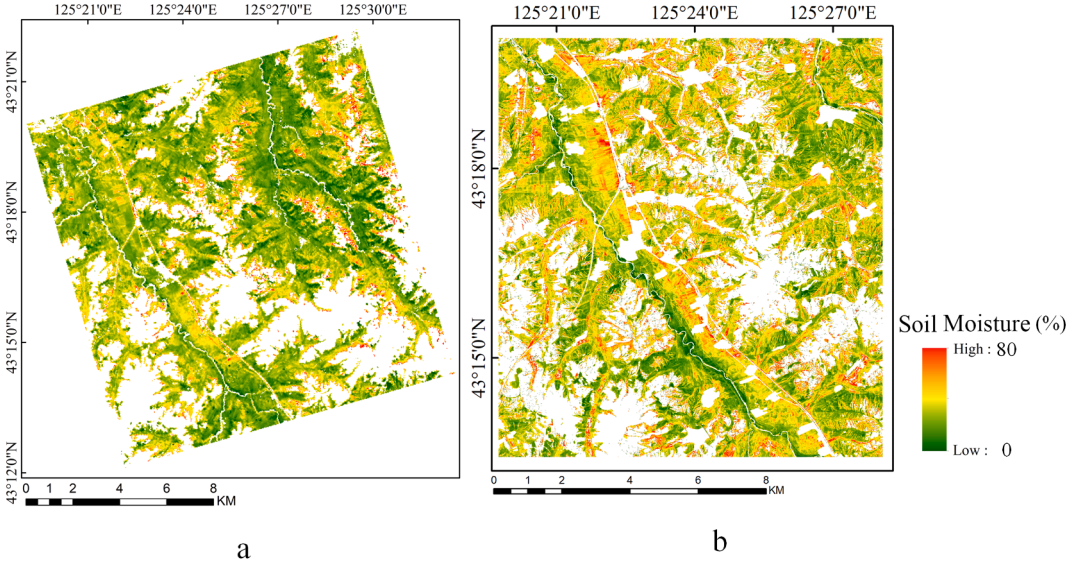


Fig. 5. Soil moisture content results. a GF5 data. b HyMap data.

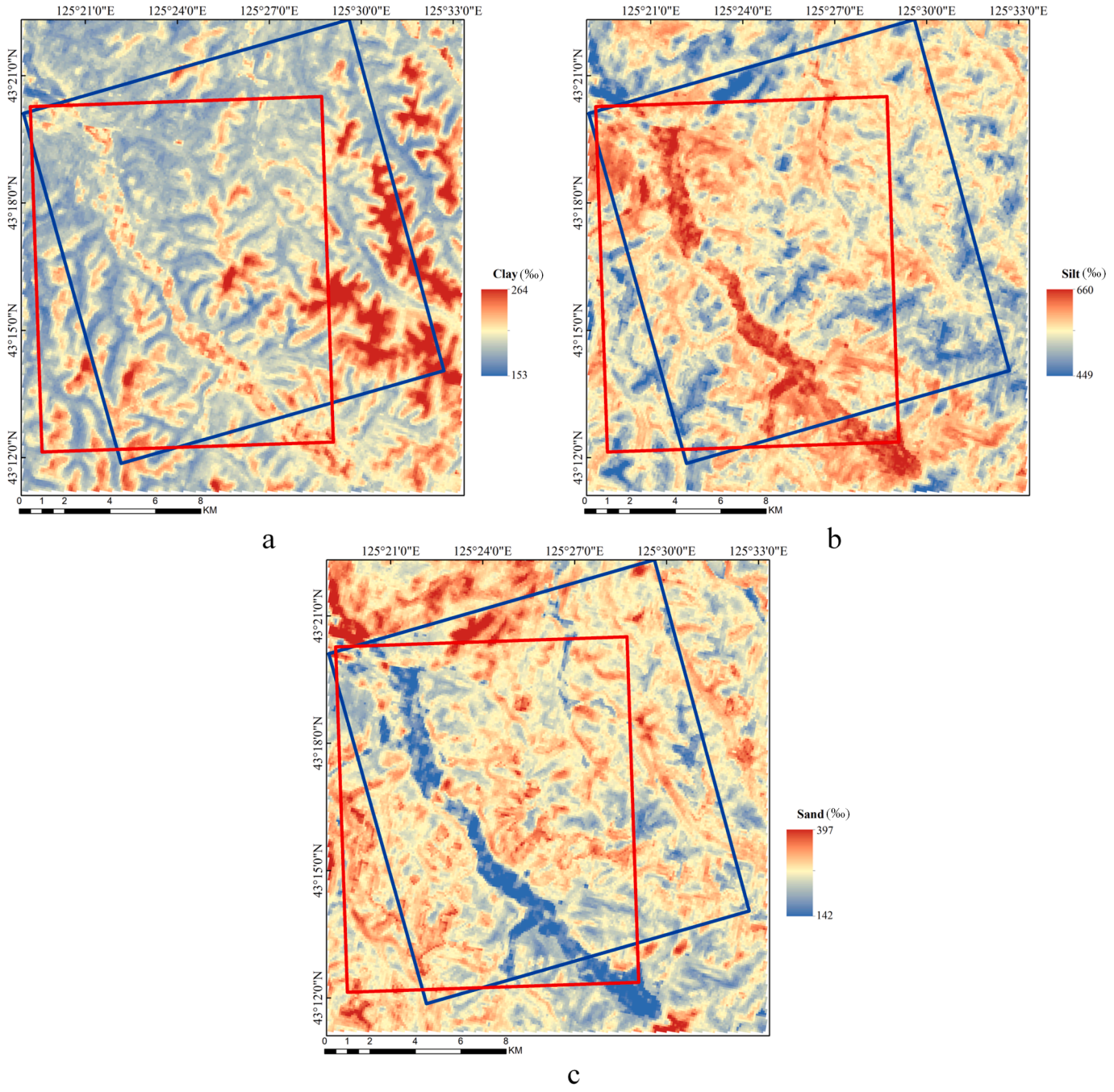


Fig. 6. Soil texture data. a. clay. b. silt. c. sand. The red area indicates the HyMap data range, and the blue area indicates the GF5 data range.

lowest RMSE on the training set. The chosen wavelengths and the correlation coefficient between wavelengths and SOM concentration are shown in Table 4.

The wavelengths used for estimating SOM concentration are basically focused on the range of 400–800 nm, 1100–1200 nm, and around 2200 nm. Table 5 lists the accuracy results of the SOM concentration calculated by RT model. The R^2 on the training set and testing set of GF5 data are 0.725 and 0.655 respectively, and the R^2 on the training set and testing set of HyMap data are 0.705 and 0.680 respectively. The R^2 on the training set and testing set of the two datasets is similar, and there is no obvious overfitting or underfitting. The *RMSE* and *MAE* are also consistent on the training and testing sets, indicating that the RT model can calculate the SOM concentration correctly by fitting the model parameters of each soil category.

Fig. 8 shows the scatter of the training and testing sets. The points in both the training and testing sets are in the region of the 1:1 line. The accuracy results with small differences in training and testing set indicate that the RT model can estimate the SOM concentration correctly by fitting the model parameters for each soil category through the training set.

4.2. Experimental results and analysis of SESMRT model

4.2.1. Spectral simulation accuracy

The model assumes consistent properties for the soil-forming parent material, soil iron oxides, SOM, soil moisture, and PSD, which guarantees the same model parameters for the soil compositions in the study area. Table 6 lists the accuracies of the spectral simulations. Both the GF5 and HyMap data can achieve an excellent spectral simulation

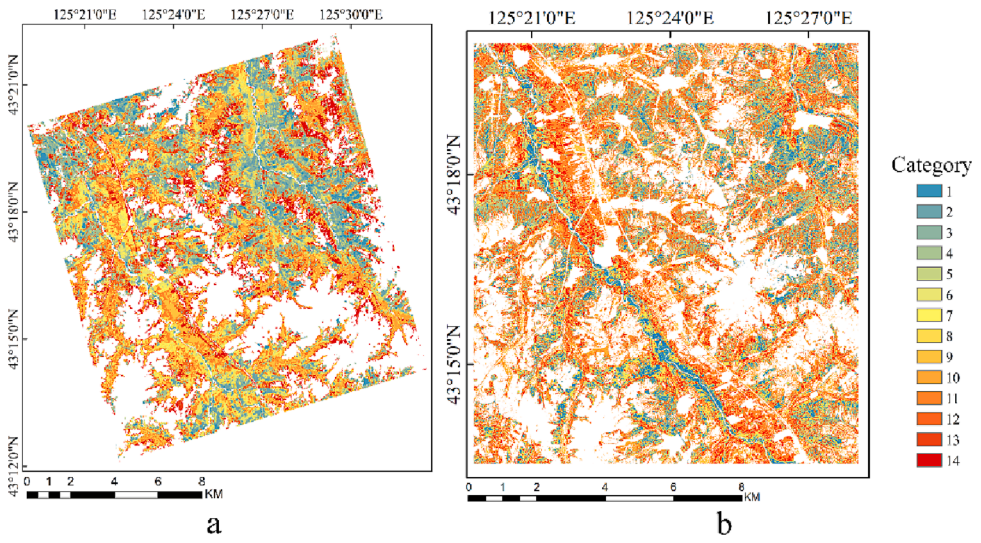


Fig. 7. Classification map of soil type. a. GF5 data. b. HyMap data.

Table 4
Correlation coefficient of wavelength.

	Wavelength (nm)	Correlation coefficient	Wavelength (nm)	Correlation coefficient
GF5	407	0.732	758	−0.644
	411	0.524	1122	0.777
	454	−0.733	1240	0.138
	471	0.459	1477	0.499
	522	0.358	1553	0.501
	574	−0.060	1679	0.001
	638	−0.239	1721	0.313
HyMap	534	0.576	1136	0.739
	613	0.192	1181	−0.128
	678	−0.667	1268	−0.310
	779	0.986	1297	−0.232
	792	−0.828	1564	−0.571
	986	−0.687	2074	0.647
	1107	0.356	2240	0.693

accuracy. The simulation accuracies for the training and testing sets are similar, indicating that the model parameters can correctly reflect the soil spectral features, which lays a solid base for the calculation of the SOM spectra. In addition, the spectral simulation accuracy of the GF5 data is higher than that of the HyMap data. The main reason for this is that GF5 data represent a single imaging of the sensor, and the spectra are more consistent, while the HyMap data are stitched together from multiple aerial strips, leading to spectral differences.

Fig. 9 shows the density scatter plots for the spectral simulation. The points are in the region of the 1:1 line, which indicates that the SESMRT model performs surprisingly well in simulating the soil spectra. In addition, there are differences in the spectral reflectance between the GF5 data and HyMap data. The radiation calibration parameters of the GF5 data use the alternative site calibration method, and the spectral reflectance is closer to the ground soil measured reflectance. In addition, the radiation calibration parameters of the HyMap data use the default parameters of the equipment, and the radiation-corrected spectral reflectance is lower than that of the ground-measured spectra.

A global sensitivity analysis was conducted using the Sobol method (Sobol, 2001) to assess the impacts of each soil component in the model by calculating their sensitivities to the model results through first and second-order variances. By using the SALib library (Herman and Usher, 2017), the global sensitivity indices were calculated based on the solved model parameters, considering soil properties (m_{Som}, m_{SIO}, $m_{Moisture}$, P_{clay}, P_{silt} and P_{sand}), as well as the reflectance of each band. The value of the global sensitivity index ranges from 0 to 1, with higher values indicating stronger sensitivity. Table 7 lists the global sensitivity values of each soil property. SOM plays a dominant role in the model during spectral simulation, whereas soil texture data has relatively minimal influence. These findings suggest that the simplified soil texture data has limited interference with the results of the soil spectral simulation.

4.2.2. SOM spectral reflectance

The model parameters determined by the spectral simulation were used to calculate the SOM spectral reflectance from Eqs. (19)–(24). The absorption coefficient $K_{Soil\_no\_Som}$ and scattering coefficient $S_{Soil\_no\_Som}$ calculated with Eqs. (20) and (21) are denoted as Scenario I, while those calculated with Eqs. (23) and (24) are denoted as Scenario II. Fig. 10 shows the average values of the spectral reflectance of the soil spectra with SOM removals $R_{Soil\_no\_Som}$ and the SOM spectra R_{Som} calculated by Scenario I and Scenario II, respectively. From Fig. 10 a and c, the effect on the spectra is relatively weak, due to the low SOM concentration, and the results for the soil spectra with SOM removals calculated by the two scenarios are similar, where the *RMSE* of the difference between the two scenarios is 0.0024 for the GF5 data and 0.0027 for the HyMap data.

The difference in *RMSE* between the two scenarios is 0.002 for the GF5 data and 0.003 for the HyMap data (Fig. 10 b and d). The difference in the wavelengths of the SOM spectra calculated by the two scenarios is similar to the soil spectra with SOM removals, where the difference between the calculated results of the HyMap data is significantly larger than that with the GF5 data. Negative values of SOM spectra exist mainly because: 1) SOM exhibits absorption effects in some bands, so the

Table 5
Accuracy results of SOM concentration calculated by RT model.

Dataset	Training set				Testing set			
	R^2	*RMSE*	*MAE*	*RPIQ*	R^2	*RMSE*	*MAE*	*RPIQ*
GF5	0.725	3.704	2.528	2.281	0.655	3.668	2.661	2.578
HyMap	0.705	3.463	2.680	2.084	0.680	3.722	2.827	1.926

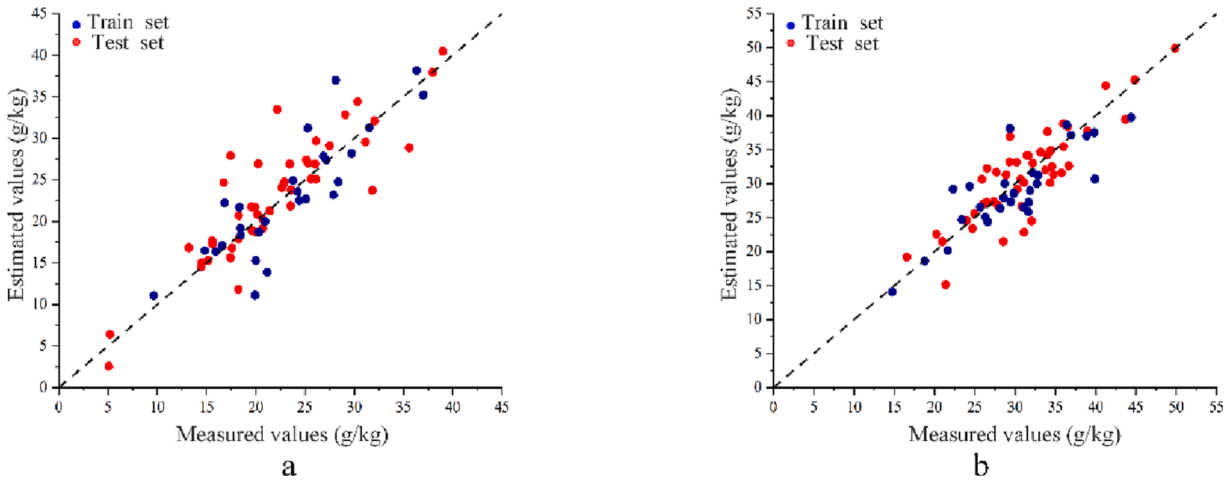


Fig. 8. Scatter plot of SOM concentration calculated by RT model. a. GF5 data. b. HyMap data.

Table 6
Accuracy of the spectral simulations.

Dataset	Training set			Testing set		
	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*
GF5	0.902	0.021	0.015	0.932	0.018	0.014
HyMap	0.814	0.021	0.016	0.860	0.017	0.013

spectra of soil with SOM removals will have increased reflectance in these bands, and the calculated SOM spectra will have negative values; and 2) errors from the spectral simulation. However, in general, as the results of both scenarios in calculating the SOM spectra are similar, the SESMRT model uses Scenario II to calculate the SOM spectra.

Fig. 11 shows the correlation between the SOM spectra and the original spectra with organic matter for both data. For HyMap data, the correlation between the SOM spectra calculated by the SESMRT model showed a noticeable improvement compared to the original spectra. For GF5 data, the correlation between the original spectra and SOM content is low, lacking statistical significance. Consequently, the correlation between the SOM spectra and content remains low as well.

In addition, the results show that the correlation between SOM spectra and organic matter is improved in both datasets, while maintaining similar wavelength trends in the original spectra, further indicating that the SOM spectra calculated based on the SESMRT model successfully extract the feature of SOM from the original spectra.

The correlation between the spectra and organic matter is also different in the two datasets, which can be mainly attributed to the different spatial resolutions and acquisition times. The acquisition time of the GF5 hyperspectral imagery was February, and some corn stalks were present in the cultivated soil. Thus, the serious mixed pixel problem brought by the 30-m spatial resolution of the GF5 hyperspectral imagery causes the correlation to not be significantly negative. The HyMap hyperspectral imagery was imaged in May, and the spatial resolution of 4.5 m has less influence from mixed pixels, so the correlation is similar to the indoor spectral correlation.

4.2.3. SOM estimation results

Fig. 11 shows the feature wavelengths of the SOM spectra and the original spectra selected by the CARS algorithm. The feature wavelengths of the HyMap data are mainly concentrated in the wavelengths of around 750 nm, 1250 nm, 1650 nm, and 2000–2300 nm, while the feature wavelengths of the GF5 data are concentrated in the wavelengths of around 400–700 nm, 1000 nm, 1250 nm, and 2000–2300 nm, which are often used to estimate SOM concentration. The wavelength ranges selected by the CARS algorithm for the original spectra and the SOM spectra are basically the same, and the correlation is significantly improved. Table 8 lists the relative weights in the CARS algorithm for each band used in the estimation model. The bolded bands indicate

Table 7
The global sensitivity values of each soil property.

	m_{Som}	m_{SIO}	$m_{Moisture}$	P_{clay}	P_{silt}	P_{sand}
Values	0.322	0.271	0.197	0.047	0.089	0.146

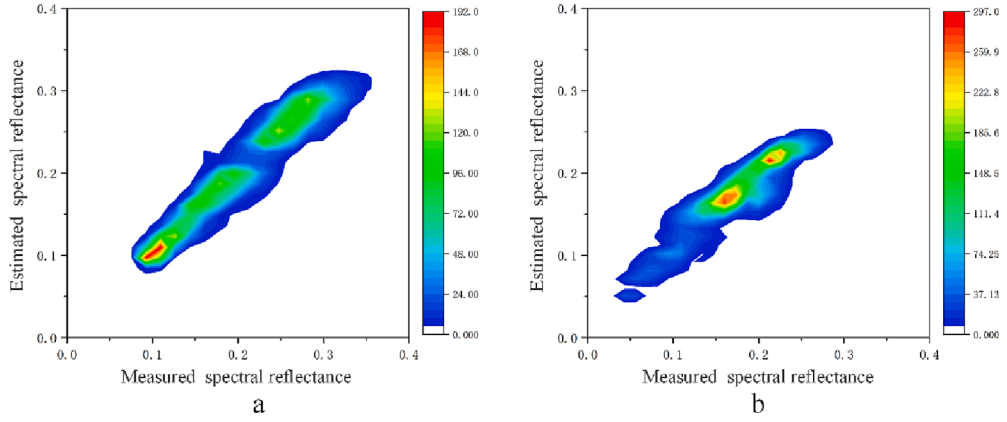


Fig. 9. Density scatter plots of the spectral simulation. a. GF5 data. b. HyMap data.

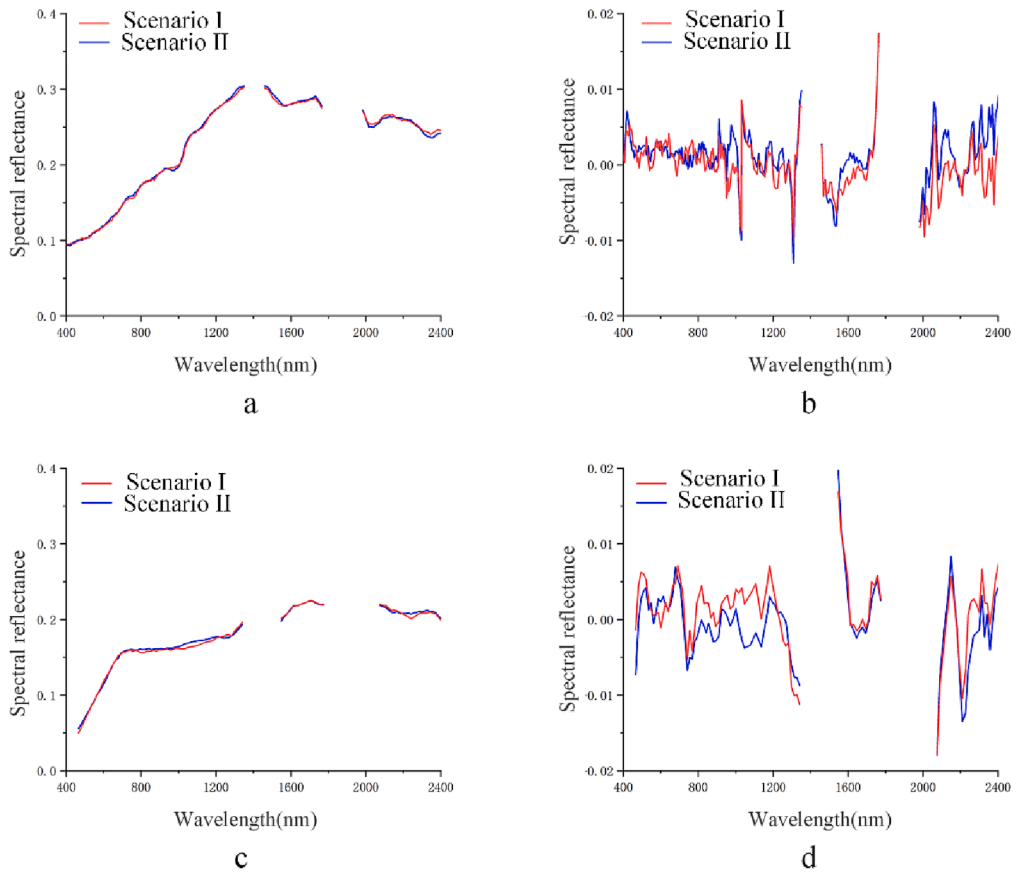


Fig. 10. Average spectra of the SOM and soil with SOM removals. a. Soil spectra with SOM removals for the GF5 data. b. SOM spectra of the GF5 data. c. Soil spectra with SOM removals for the HyMap data. d. SOM spectra of the HyMap data.

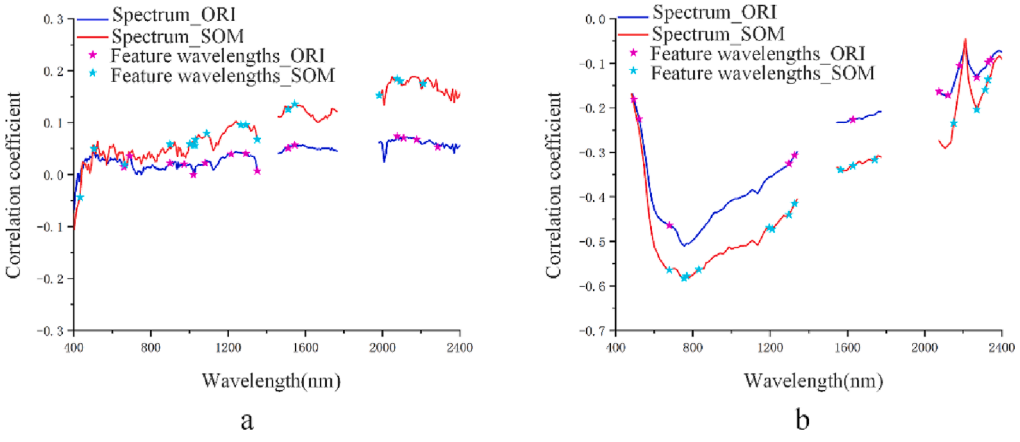


Fig. 11. Correlation of the spectra and SOM, and feature wavelength selected by CARS algorithm. a. GF5 data. b. HyMap data.

those with relative weights higher than 0.1. The bands with high weight are concentrated in the 678–1000 nm range. The bands with high weight for the original and SOM spectra are similar, indicating that the SESMRT model successfully extracts the intrinsic factors of SOM from the original spectra.

SVR was utilized to estimate the SOM concentration using the feature wavelengths selected by the CARS algorithm. The dataset was split into the training and testing datasets randomly, and the process was repeated ten times to obtain the mean and standard deviation of the SOM estimation accuracy. Table 9 lists the accuracies of the SOM estimation. For the GF5 data, the accuracy of the original spectra estimation is inferior, and the R^2, *RMSE*, *RPIQ* of the testing set is only 0.467 ± 0.068, 5.012 ± 0.313, 1.569 ± 0.097, while the R^2, *RMSE*, *RPIQ* of the SOM spectra is 0.660 ± 0.034, 3.923 ± 0.236, 2.003 ± 0.118. The R^2 value has shown an improvement of nearly 20 percentage points compared to the original spectra, and both the *RMSE* and *MAE* values are significantly lower compared to those of the original spectra. The HyMap data estimation accuracy is similar for the SOM spectra (R^2 = 0.685 ± 0.030, *RMSE* = 3.542 ± 0.311, *RPIQ* = 2.309 ± 0.207) and the original spectra (R^2 = 0.646 ± 0.043, *RMSE* = 3.711 ± 0.319, *RPIQ* = 2.192 ± 0.199). Although the improvement is not comparable to that with the GF5 data, the R^2 is improved by nearly 4 percentage points. As the HyMap data were acquired by an airborne flight with less interference from the

Table 8
The relative weights of feature based on the CARS algorithm.

	Band	Weight	Band	Weight
R	494 nm	0.002	2075 nm	0.030
	522 nm	0.030	2121 nm	0.024
	678 nm	**0.314**	2180 nm	0.088
	1297 nm	**0.119**	2270 nm	0.059
	1327 nm	**0.103**	2328 nm	0.097
	1629 nm	0.030	**2343 nm**	**0.102**
R_{Som}	467 nm	0.056	1327 nm	0.003
	678 nm	**0.132**	1564 nm	0.031
	755 nm	**0.146**	1629 nm	0.035
	767 nm	**0.139**	1741 nm	0.041
	831 nm	**0.116**	2151 nm	0.048
	1196 nm	0.029	2270 nm	0.045
	1210 nm	0.032	2314 nm	0.063
	1297 nm	0.015	2328 nm	0.069

atmospheric environment and mixed pixels, a better estimation accuracy can be obtained using the original spectra.

As shown in Fig. 12, the training set and testing set results for the original spectra and SOM spectra in both data sources are basically around the 1:1 line, indicating that the model estimates effectively, without overfitting, but the point dispersion of the original spectra is significantly higher than that of the SOM spectra. In addition, the predicted values of some of the original spectral testing set samples are significantly higher than the laboratory values, while the testing set of the SOM spectra is basically evenly distributed on both sides of the 1:1 line, indicating that the SOM spectra can estimate the organic matter concentration more accurately, without the problem of overall high estimation results. However, for the GF5 data, both the original spectra and SOM spectra show low prediction values for the data with high SOM concentrations, especially when the SOM concentration amounts to 35 g/kg. The main reason for this is that there was a sample with a 5 g/kg SOM concentration, and the lower SOM concentration affected the model construction. In contrast, the SOM concentration of the HyMap data is relatively concentrated, and there are no obvious outliers, so the estimation results of the HyMap data are better than those of the GF5 data.

4.3. Comparison of the SOM estimation results

4.3.1. SOM estimation results based on RT model and SESMRT model

The SOM concentration is grouped into six levels, according to the China second general soil survey (Table 10).

Fig. 13 and Fig. 14 depict the SOM estimation maps obtained by the

Table 9
Accuracy results of SOM concentration estimated by SESMRT model.

Dataset	Training set			Testing set				
	R^2	*RMSE*	*MAE*	*RPIQ*	R^2	*RMSE*	*MAE*	*RPIQ*
GF5_spectra_Ori	0.502 ± 0.051	4.797 ± 0.293	3.727 ± 0.326	1.645 ± 0.098	0.467 ± 0.068	5.013 ± 0.313	4.060 ± 0.272	1.569 ± 0.097
GF5_spectra_Som	0.647 ± 0.020	4.087 ± 0.127	2.795 ± 0.237	1.926 ± 0.059	0.660 ± 0.034	3.923 ± 0.236	3.062 ± 0.177	2.003 ± 0.118
HyMap_spectra_Ori	0.613 ± 0.024	3.858 ± 0.128	2.775 ± 0.132	2.058 ± 0.074	0.646 ± 0.043	3.711 ± 0.319	2.956 ± 0.310	2.192 ± 0.199
HyMap_spectra_Som	0.644 ± 0.038	3.671 ± 0.203	2.746 ± 0.268	2.170 ± 0.125	0.685 ± 0.030	3.543 ± 0.311	2.823 ± 0.177	2.309 ± 0.207

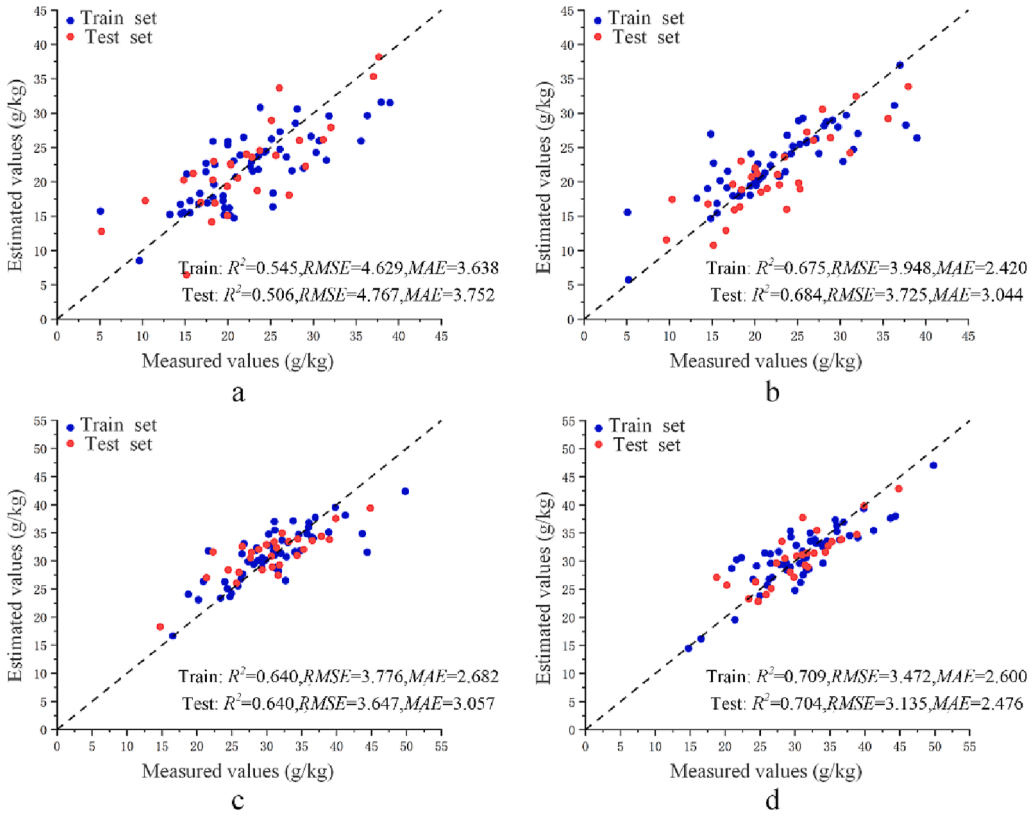


Fig. 12. Scatter plots of SOM concentration. a. Original spectrum of the GF5 data. b. SOM spectrum of the GF5 data. c. Original spectrum of the HyMap data. d. SOM spectrum of the HyMap data.

RT model. Fig. 13 shows that the SOM concentration in the study area is predominantly at level 2 and 3, with a large area of cultivated land around rivers in areas A and B being mainly level 1. However, there are more discrete points, especially around rivers and roads in area B. The main reason is that results calculated by the RT model are highly dependent on the soil classification results. Although the accuracy of the SOM concentration calculations was acceptable for both datasets, the mapping results had discrete anomalies within the study area due to the classification of the soils using spectral data. In addition, the mixed pixel problem associated with the 30 m spatial resolution of the GF5 hyperspectral imagery seriously affects the classification and calculation results.

Fig. 14 shows that the calculation results of the HyMap hyperspectral imagery, which are the same as those of GF5 hyperspectral imagery, and the areas with higher SOM concentration are the cultivated land around the river in areas A and B. The HyMap hyperspectral imagery with high spatial resolution has fewer discrete points than the GF5 hyperspectral imagery due to the weaker mixed pixel problem, and the mapping performance is better than that of the GF5 hyperspectral imagery.

Therefore, although the RT model can estimate the SOM concentrations, it is limited by the model assumptions and the impact of the mixed pixel problem, which makes it less applicability in large-scale monitoring.

Fig. 15 and Fig. 16 depict the SOM estimation results obtained within the study area by the SESMRT model. Fig. 15 shows that the SOM concentration estimated from the GF5 data is mainly Level 1 and Level 2, with the presence of Level 3 and Level 4 in 2019. The Level 1 and Level 2 standards are concentrated in the large areas of agricultural land around the rivers, because SOM is mainly derived from vegetation residue and root systems, and moisture is a key factor for vegetation growth, so the SOM concentration along the rivers is relatively elevated. The Level 3 and Level 4 standards are more often found in the soils in the mountainous areas, which are mostly natural soils influenced by the topography.

Fig. 16 shows that SOM is mainly Level 2 and Level 3, with some of the cultivated land around the rivers belonging to Level 1 in 2017. Similar to the result for 2019, the areas with high SOM concentrations are mainly distributed along the rivers, while the SOM concentration in the mountainous areas is less. Overall, the SOM concentration in the study area was higher in 2019 than in 2017.

The SESMRT model is based on the theoretical calculation of SOM spectra, and then uses a data-driven approach to construct the estimation model, which assumes the same soil-forming parent material throughout the study area. Thus, the estimation results of the SESMRT model are more aggregated than that of RT model, which promotes the application performance of the model. In conclusion, the RT is not practical for large scale applications due to strict modeling conditions, whereas the SESMRT model has better estimation performance for large scale applications.

4.3.2. SOM estimation results comparison based on GF5 and HyMap hyperspectral imagery

The comparison analysis is based on the SESMRT model mapping results. As shown in Fig. 17, the estimation map based on HyMap data was resampled to 30 m spatial resolution and was subtracted by GF5 estimation map. The areas of increased SOM concentration were mainly concentrated in the cultivated soils around the rivers, while the areas of decreased concentration were mainly concentrated in the natural soils around the mountains.

Table 11 lists the statistics for the SOM concentration percentage at each level in the estimated results for 2017 and 2019. Compared to the 2017 result, the percentage of Level 1 in the 2019 result increases from 16.03% to 19.79%, with an increase of nearly 3.7 percentage points, but the percentage of Level 2 and Level 3 soils decreases from 83.00% to 72.66%, and the percentage of Level 4 and Level 5 increases from 0.04% to 0.44%.

According to Fig. 15, Fig. 16 and Fig. 17, the increased areas of Level 1 are mainly concentrated in the cultivated soils around the rivers, while

Table 10
Classification standards for the SOM concentration.

SOM level	SOM concentration
Level 1	$m_{Som} \geq 40$ g/kg
Level 2	40 g/kg > $m_{Som} \geq 30$ g/kg
Level 3	30 g/kg > $m_{Som} \geq 20$ g/kg
Level 4	20 g/kg > $m_{Som} \geq 10$ g/kg
Level 5	10 g/kg > $m_{Som} \geq 6$ g/kg
Level 6	$m_{Som} < 6$ g/kg

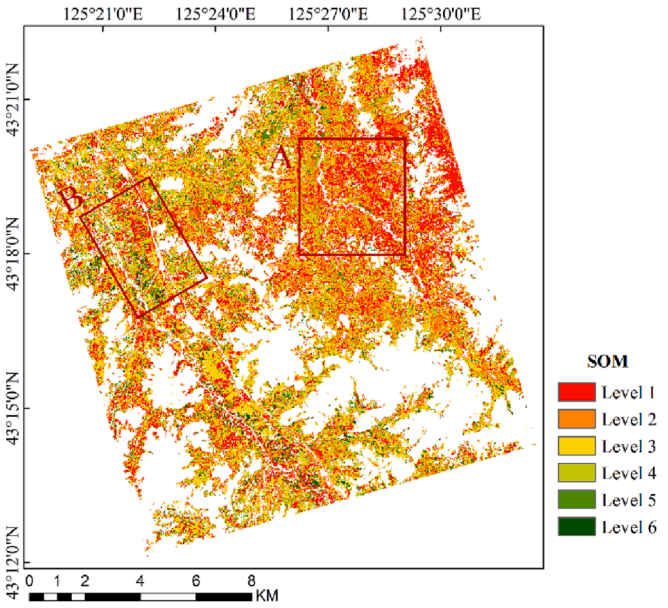


Fig. 13. SOM concentration results for the GF5 hyperspectral imagery based on RT model.

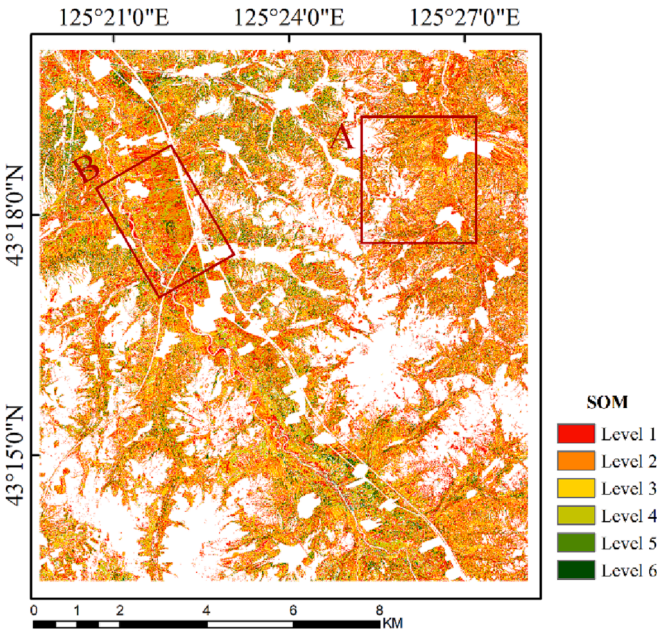


Fig. 14. SOM concentration results for the HyMap hyperspectral imagery based on RT model.

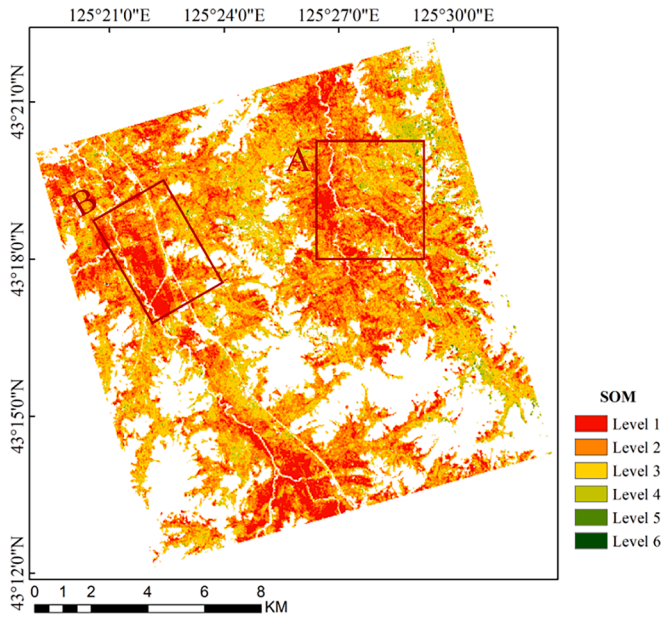


Fig. 15. SOM concentration results for the GF5 hyperspectral imagery based on SESMRT model.

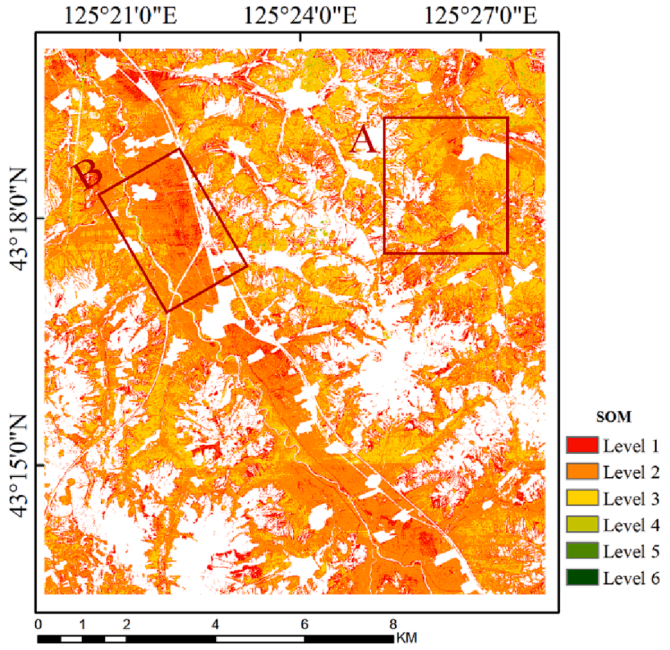


Fig. 16. SOM concentration results for the HyMap hyperspectral imagery based on SESMRT model.

Table 11
Proportions of the SOM levels.

Level	GF5 (2019)		HyMap (2017)	
	Area (KM^2)	Percentage	Area (KM^2)	Percentage
Level 1	29.75	19.79%	14.28	16.03%
Level 2	61.53	40.93%	52.43	58.86%
Level 3	47.69	31.73%	21.51	24.14%
Level 4	10.67	7.10%	0.83	0.93%
Level 5	0.49	0.33%	0.02	0.02%
Level 6	0.17	0.11%	0.02	0.02%
Total	150.31	100%	89.08	100%

the increased areas of Level 5 and Level 6 are mainly concentrated in the natural soils around the mountains. This indicates that, in recent years, the straw returned to the fields and other black soil protection measures have effectively increased the SOM concentration of the cultivated soils. However, the lack of protection measures for non-cultivated soils in the mountains and other non-cultivated lands, such as mining and washing areas, still affects the quality of the soil, resulting in a more serious loss of soil SOM concentration. Thus, in addition to the protection of cultivated soils, the follow-up of black soil protection measures should also cover the protection of non-cultivated soils.

5. Conclusions

The current hyperspectral imagery based soil composition estimation is mostly based on data-driven models, which are lacking in radiative transfer theoretical explanation. In this paper, we have described how the SESMRT model combining a radiative transfer model and data-driven model was applied to estimate SOM concentration based on hyperspectral imagery. The SESMRT model calculates the SOM spectra with radiative transfer theory to reduce the interference of other soil components in the soil spectra. After the optimization of the model parameters by a GA, the R^2 of the overall spectral simulation results for the GF5 data was 0.912, and that for the HyMap data was 0.828. Both data sources could simulate the soil spectra well and establish a solid base for the subsequent SOM spectral calculation. The correlation between the calculated SOM spectra and SOM concentration was significantly improved, compared with the original spectra, with the highest band correlation improved by about 0.1. Meanwhile, the band trend was similar to that of the original spectra, which further proves that the SESMRT model can successfully extract the SOM spectra. The SOM spectral features were extracted based on the CARS algorithm, and the SVR regression model was used to successfully estimate the SOM concentration in the study area. Compared with the original spectra, the SESMRT model shows an improved estimation accuracy for the hyperspectral imagery from the two different platforms. Based on the weight information of the soil organic matter spectral feature wavelengths, it can be observed that the wavelengths at 678 nm, 755 nm, 767 nm, and 831 nm have relatively high weights. This finding is significant for guiding the wavelength configuration of future multispectral satellites

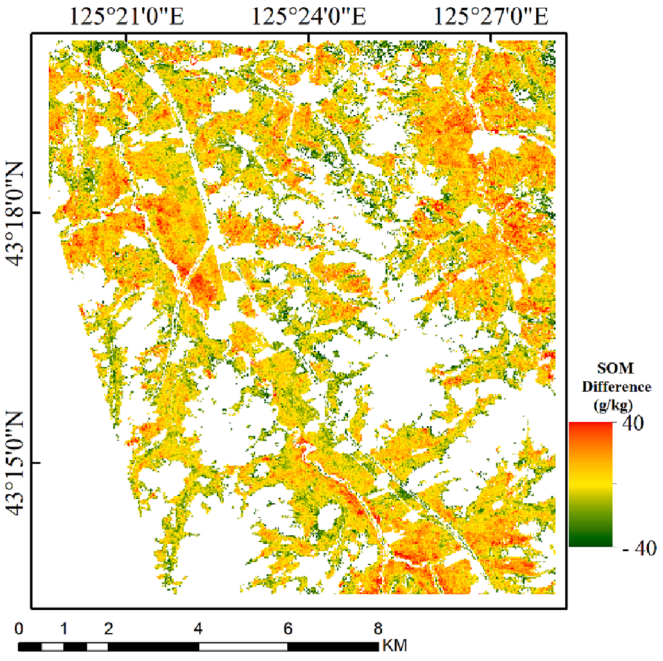


Fig. 17. Difference map of SOM concentration between 2019 and 2017.

aimed at soil organic matter monitoring. For the existing multispectral satellites, such as Sentinel-2 with Band 4 (650–680 nm), Band 7 (773–793 nm), Band 8 (785–900 nm), and Landsat with Band 4 (630–680 nm) and Band 5 (845–885 nm), their spectral bands are similar to those used in the study and can provide accurate spectral data for soil organic matter content estimation.

By comparing the estimation results for 2017 and 2019, the percentage of Level 1 had significantly increased by about 3.7 percentage points, and the increase was basically in the cultivated soil, which indicates that the series of black soil conservation measures have effectively improved the SOM concentration. However, in the non-cultivated soils around the mountains, there was a significant decrease in the SOM concentration level, compared to 2017, which means that the land currently damaged by previous mining and other actions still needs further protection. Moreover, compared with the statistical analysis of the laboratory results, where the SOM concentration in 2019 was lower than that in 2017, the estimated results based on hyperspectral imagery are more detailed and better able to reflect the trend of SOM changes in the different soil types, which further demonstrates the advantages of hyperspectral imagery in soil composition analysis.

The SESMRT model calculates SOM spectra by soil radiative transfer model, which reduces the impacts from other factors in the soil and significantly improves the estimation accuracy of SOM concentration. Compared to the RT model which calculates SOM concentration directly, the SESMRT model can simplify the conditions for model application and improve mapping performance by combining data-driven models. Restricted by the model assumptions, the SESMRT model parameters need to be determined using genetic algorithms in different study area. A large amount of subsequent research is required to determine empirical parameters for specific soil types and to reduce the complexity of the model. The problem of mixed pixel elements caused by the spatial resolution of the hyperspectral imagery seriously restricts the estimation accuracy. Especially for cultivated soils, the effect of crop straw on soil spectra is not negligible. Therefore, in the future, constructing a soil-vegetation hybrid radiative transfer model to extract pure soil spectra will be the key to allowing hyperspectral imagery to be applied in real time at a large scale.

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.

Acknowledgements

This research is supported in part by National Natural Science Foundation of China (No. 42171335), and the Shanghai Municipal Science and Technology Major Project (No. 22511102800) and the National Civil Aerospace Project of China (No. D040102).

References

Bablet, A., Vu, P.V.H., 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.

Banninger, D., Fluhler, H., 2004. Modeling light scattering at soil surfaces. IEEE Trans. Geosci. Remote Sens. 42 (7), 1462–1471.

Bao, Y., Meng, X., Ustin, S., Wang, X., Zhang, X., Liu, H., Tang, H., 2020. Vis-SWIR spectral prediction model for soil organic matter with different grouping strategies. Catena 195, 104703.

Chakraborty, S., Weindorf, D.C., Deb, S., Li, B., Paul, S., Choudhury, A., Ray, D.P., 2017. Rapid assessment of regional soil arsenic pollution risk via diffuse reflectance spectroscopy. Geoderma 289, 72–81.

Chen, S., Arrouays, D., Mulder, V.L., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. Geoderma. 409.

Christy, A.A., Kvalheim, O.M., Velapoldi, R.A., 1995. Quantitative analysis in diffuse reflectance spectrometry: a modified Kubelka-Munk equation. Vib. Spectrosc 9 (1), 19–27.

Ciani, A., Goss, K.U., Schwarzenbach, R.P., 2005. Light penetration in soil and particulate minerals. Eur. J. Soil Sci. 56 (5), 561–574.

Demattê, J.A.M., da Silva Terra, F., 2014. Spectral pedology: A new perspective on evaluation of soils along pedogenetic alterations. Geoderma 217–218, 190–200.

Ding, A., Ma, H., Liang, S., He, T., 2022. Extension of the Hapke model to the spectral domain to characterize soil physical properties. Remote Sens. Environ. 269, 112843.

Dupiau, A., Jacquemoud, S., Briottet, X., Fabre, S., Viallefont-Robinet, F., Philpot, W., Di Biagio, C., Hébert, M., Formenti, P., 2022. MARMIT-2: An improved version of the MARMIT model to predict soil reflectance as a function of surface water content in the solar domain. Remote Sens. Environ. 272, 112951.

Gomez, C., Lagacherie, P., Coulouma, G., 2012. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis–NIR data. Geoderma 189, 176–185.

Guo, P., Li, T., Gao, H., Chen, X., Cui, Y., Huang, Y., 2021. Evaluating calibration and spectral variable selection methods for predicting three soil nutrients using Vis-NIR spectroscopy. Remote Sens. (Basel) 13 (19), 4000.

Herman, J., Usher, W., 2017. SALib: An open-source Python library for sensitivity analysis. Journal of Open Source Software 2 (9), 97.

Jacquemoud, S., Baret, E., Hanocq, J.F., 1992. Modeling Spectral and Bidirectional Soil Reflectance. Remote Sens. Environ. 41 (2–3), 123–132.

Janik, L.J., Soriano-Disla, J.M., Forrester, S.T., 2020. Feasibility of handheld mid-infrared spectroscopy to predict particle size distribution: influence of soil field condition and utilisation of existing spectral libraries. Soil Res. 58 (6), 528.

Koch, M., Schodlok, M.C., Guggenberger, G., Stadler, S., 2021. Effects of water tension and surface roughness on soil hyperspectral reflectance. Geoderma 385, 114888.

Lal, R., 2004. Soil carbon sequestration to mitigate climate change. Geoderma 123 (1–2), 1–22.

Liang, S., Townshend, J., 1996. A Modified Hapke Model for Soil Bidirectional Reflectance. Remote Sens. Environ. 55 (1), 1–10.

Liu, Y.i., Shi, Z., Zhang, G., Chen, Y., Li, S., Hong, Y., Shi, T., Wang, J., Liu, Y., 2018. Application of Spectrally Derived Soil Type as Ancillary Data to Improve the Estimation of Soil Organic Carbon by Using the Chinese Soil Vis-NIR Spectral Library. Remote Sens. (Basel) 10 (11), 1747.

Liu, F., Zhang, G.-L., Song, X., Li, D., Zhao, Y., Yang, J., Wu, H., Yang, F., 2020. High-resolution and three-dimensional mapping of soil texture of China. Geoderma 361, 114061.

Meng, X., Bao, Y., Ye, Q., Liu, H., Zhang, X., Tang, H., Zhang, X., 2021. Soil organic matter prediction model with satellite hyperspectral image based on optimized denoising method. Remote Sens. (Basel) 13 (12), 2273.

Nanni, M.R., Demattê, J.A.M., Rodrigues, M., Santos, G.L.A.A.D., Reis, A.S., Oliveira, K. M.d., Cezar, E., Furlanetto, R.H., Crusiol, L.G.T., Sun, L., 2021. Mapping particle size and soil organic matter in tropical soil based on hyperspectral imaging and non-imaging sensors. Remote Sens. (Basel) 13 (9), 1782.

Niu, C., Tan, K., Wang, X., Han, B., Ge, S., Du, P., Wang, F., 2022. Radiometric Cross-Calibration of the ZY1-02D Hyperspectral Imager Using the GF-5 AHSI Imager. IEEE Trans. Geosci. Remote Sens. 60, 1–12.

Nocita, M., Stevens, A., Noon, C., van Wesemael, B., 2013. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. Geoderma 199, 37–42.

Norouzi, S., Sadeghi, M., Liaghat, A., Tuller, M., Jones, S.B., Ebrahimian, H., 2021. Information depth of NIR/SWIR soil reflectance spectroscopy. Remote Sens. Environ. 256, 112315.

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.

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., 2018. Particle size effects on soil reflectance explained by an analytical radiative transfer model. Remote Sens. Environ. 210, 375–386.

Selige, T., Böhner, J., Schmidhalter, U., 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma 136 (1–2), 235–244.

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simul 55 (1–3), 271–280.

Sun, W., Liu, S., Zhang, X., Li, Y., 2022. Estimation of soil organic matter content using selected spectral subset of hyperspectral data. Geoderma 409.

Tian, Y., Zhang, J., Yao, X., Cao, W., Zhu, Y., 2013. Laboratory assessment of three quantitative methods for estimating the organic matter content of soils in China based on visible/near-infrared reflectance spectra. Geoderma 202, 161–170.

Tuckerman, L.B., 1947. On the intensity of the light reflected from or transmitted through a pile of plates. J. Opt. Soc. Am. 37 (10), 818–825.

Vargas, W., Niklasson, G., 1997. Applicability conditions of the Kubelka-Munk theory. Appl. Opt. 36 (22), 5580–5586.

Wang, S., Guan, K., Zhang, C., Lee, DoKyoung, Margenot, A.J., Ge, Y., Peng, J., Zhou, W., Zhou, Q.u., Huang, Y., 2022. Using soil library hyperspectral reflectance and machine learning to predict soil organic carbon: Assessing potential of airborne and spaceborne optical soil sensing. Remote Sens. Environ. 271.

Wang, Y.-P., Lee, C.-K., Dai, Y.-H., Shen, Y., 2020. Effect of wetting on the determination of soil organic matter content using visible and near-infrared spectrometer. Geoderma 376, 114528.
Wang, X., Zhang, F., Kung, H.-t., Johnson, V.C., 2018. New methods for improving the remote sensing estimation of soil organic matter content (SOMC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR) in northwest China. Remote Sens. Environ. 218, 104–118.
Yang, Y., Shang, K., Xiao, C., Wang, C., Tang, H., 2022. Spectral Index for Mapping Topsoil Organic Matter Content Based on ZY1-02D Satellite Hyperspectral Data in Jiangsu Province. China. ISPRS International Journal of Geo-Information. 11 (2), 111.
Zhang, Y., Tan, K., Wang, X., Chen, Y.u., 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 (14), 2239.