Semi-Empirical Soil Organic Matter Retrieval Model With Spectral Reflectance

JING YUAN[1,2], **CHUNHUI HU**[1], **CHANGXIANG YAN**[1,3], **ZHIZHONG LI**[4], **SHENGBO CHEN**[5], **SHURONG WANG**[1], **XIN WANG**[1,2], **ZHENGYUAN XU**[5], **AND XUEPING JU**[1,2]

[1]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Center of Materials Science and Optoelectrics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[4]Shenyang Institute of Geology and Mineral Resources, CGS, Shenyang 110034, China
[5]College of Geoexploration Science and Technology, Jilin University, Changchun 130026, China

Corresponding author: Chunhui Hu (famous226@163.com)

ABSTRACT Rapid and accurate monitoring of soil organic matter (SOM) content is of great significance for precision fertilization of farmland. However, the SOM retrieval models are mainly established by statistical methods, which have limited application scope and incomplete theoretical foundation. Moreover, the accuracy of the SOM retrieval models remains raised. In this paper, for the first time, a semi-empirical SOM content retrieval model is constructed, which has certain theoretical basis, strong applicability and higher accuracy than before. Based on the Kubelka–Munk (KM) theory, the SOM retrieval model with the absorption coefficient k and scattering coefficient s related to SOM ($r = k/s$) is derived. The validity and reliability of the model are confirmed with validation set ($n = 26$) including three sorts of soils. Results show that the model can estimate SOM content in different sorts of soils with high prediction accuracy and good prediction ability (root mean square errors of prediction (RMSEP), coefficients of determination (R^2) and relative percentage deviation (RPD) values of 0.18%, 89.9% and 3.2, respectively) in the range of 552-950nm. The model provides an innovative method for predicting SOM content.

INDEX TERMS Soil organic matter retrieval, reflectance, semi-empirical model, KM.

I. INTRODUCTION

SOM is an important part of soil, and the SOM content is generally regarded as a criterion to assess soil fertility and an important indicator of soil degradation [1], [2]. Efficient and accurate monitoring of SOM content is of great significance for precision fertilization and rational utilization of land resources. The traditional determination of SOM content is mainly achieved by soil field sampling and chemical analysis. Although the precision is high, the cost is also high and it can't effectively monitor the degradation of soil [3]. The visible and near-infrared (VNIR) hyper-spectral technique can retrieve the SOM content by measuring the spectral reflectance of soil. Compared with the traditional method, it has many advantages such as low cost, high speed, no pollution, and real-time monitoring of soil component. It can greatly meet the development needs of precision agriculture, and has become a research hotspot [4]–[21].

At present, statistical model is the main method to establish SOM content retrieval model based on hyper-spectral characteristics of soil. Several multivariate methods for SOM prediction have been successfully utilized to develop a faster and higher-quality model, such as partial least-squares regression (PLSR), principal components regression (PCR), multiple linear regression (MLR), multivariate adaptive regression splines (MARS), support vector machines (SVM), PLS-SVM, bayesian model averaging (BMA) [22], weighted average partial least-squares(WAPLS), (GPR) [23], artificial neural networks (ANN), ordinary kriging (OK), regression kriging (RK) [24], random forest (RF) [25], [26], random forest kriging (RFK), gradient boosting (GB) [27]–[30] and gradient boosting kriging (GBK). The modeling results using these methods to predict SOM content are shown in Table 1, coefficients of determination (R^2): 0.17%-0.92%, root mean

The associate editor coordinating the review of this manuscript and approving it for publication was Wen-Sheng Zhao.

TABLE 1. Modeling results of different methods for predicting SOM content.

	Validation set			Author
	R^2 (%)	RMSEP (%)	RPD	
PCR	0.80	0.51	3.00	Andre, C.D. et al.(2018)
PLSR	0.81	0.49	3.12	Andre, C.D. et al.(2018)
	0.79	0.28	2.16	Said, N. et al.(2016)
	0.65	0.30	1.60	Yaser, O. et al.(2018)
	0.57	0.54	/	Lin, P.D. et al.(2018)
	0.86	0.52	2.64	Yin, Y.(2015)
MLR	0.79	0.52	2.93	Andre, C.D. et al.(2018)
	0.17	0.64	/	Panagiotis, T. et al.(2019)
MSR	0.50	0.62	/	Lin, P.D. et al.(2018)
	0.72	0.73	1.87	Yin, Y.(2015)
SVM	0.80	0.52	2.94	Andre, C.D. et al.(2018)
	0.75	0.26	2.00	Said, N. et al.(2016)
PLS-SVM	0.92	0.37	3.68	Yin, Y.(2015)
RF	0.77	0.55	2.77	Andre, C.D. et al.(2018)
	0.28	0.60	/	Panagiotis, T. et al.(2019)
RFK	0.31	0.58	/	Panagiotis, T. et al.(2019)
BMA	0.80	0.51	3.03	Andre, C.D. et al.(2018)
WAPLS	0.82	0.48	3.18	Andre, C.D. et al.(2018)
GPR	0.79	0.52	2.96	Andre, C.D. et al.(2018)
ANN	0.80	0.51	3.01	Andre, C.D. et al.(2018)
MARS	0.81	0.22	2.27	Said, N. et al.(2016)
OK	0.08	0.66	/	Panagiotis, T. et al.(2019)
RK	0.18	0.63	/	Panagiotis, T. et al.(2019)
GB	0.25	0.60	/	Panagiotis, T. et al.(2019)
GBK	0.28	0.59	/	Panagiotis, T. et al.(2019)

square errors of prediction (RMSEP): 0.22%-0.73%, relative percentage deviation (RPD): 1.60-3.68 [3], [31]–[35].

However, the SOM retrieval model established by statistical methods has limited application scope and incomplete theoretical foundation [36]. Moreover, the accuracy of the SOM retrieval model with statistical methods remains to be raised.

To solve this problem, a concise model which has a certain theoretical basis and strong applicability is presented. In previous studies, the diffuse reflectance in the KM model is usually regarded as a parameter that needs to be inverted or a constant for a given material and illumination wavelength. Nevertheless, further study finds that diffuse reflectance is

not only related to material and wavelength, but also to SOM content, since k and s of soil are both affected by SOM content, and diffuse reflectance is the function of k and s based on KM model. According to a frequently effective and commonly accepted assumption that the k and s of a mixed medium can be regarded as a simple sum function of the k and s weighted by their composition proportions [37]–[39], the k and s with SOM content are first signified, respectively. Then, the relationship between reflectance R and transformed reflectance r is derived based on the KM theory. Finally, the SOM retrieval model is further derived by relating R and SOM content with r = k/s. The model can estimate the SOM content with higher accuracy than before, and it provides an innovative method for predicting SOM content. The proposed model exhibits the following advantages:

(1) For the first time, a semi-empirical model for SOM content retrieval is constructed by introducing SOM content information into KM model. It has a certain theoretical basis, high prediction accuracy and extensive applicability.

(2) It is not restricted to a single band or wavelength. Properly calibrated, it is competent to describe the response of reflectance to SOM content over the full optical range.

The rest of this paper is organized as follows. Section II provides the description of the SOM retrieval model. The details of the soil samples and the measurement procedures are outlined in Section III. The measured reflectance spectra results and comparisons between the models and measurements are discussed in Section IV. Section V presents the conclusions of this paper.

II. SOM RETRIEVAL MODEL
A. KM THEORY
The KM [40] theory describes radiative transfer, considering a downward and an upward light propagation flux (I and J, respectively), in an absorbing and scattering medium, perpendicular to the layer (Fig.1).

The KM theory consists of two differential equations describing the light fluxes, $I(\lambda, z)$ and $J(\lambda, z)$, at a given wavelength, λ (nm), and at a depth in the layer, z (cm), with a light absorption coefficient, k (cm-1), and a light scattering coefficients s (cm-1):

$$\frac{dI(\lambda, z)}{dz} = -(k + s)I(\lambda, z) + sJ(\lambda, z) \quad (1)$$

$$\frac{dJ(\lambda, z)}{dz} = (k + s)J(\lambda, z) - sI(\lambda, z) \quad (2)$$

By analytically solving these equations, reflectance R can be obtained [37]:

$$R = \frac{(1 - \beta)^2[\exp(\alpha d) - \exp(-\alpha d)]}{(1 + \beta)^2 \exp(\alpha d) - (1 - \beta)^2 \exp(-\alpha d)} \quad (3)$$

where $\alpha = \sqrt{k(k + 2s)}$ $\beta = \sqrt{k/(k + 2s)}$

With increasing layer thickness, d, the reflectance reaches the infinite reflectance value, R_∞, which is used in diffuse
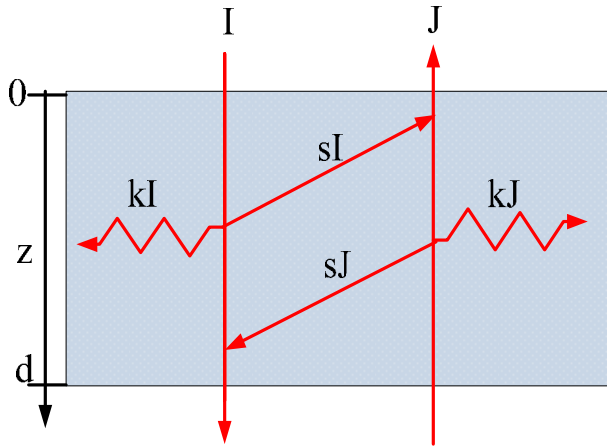
FIGURE 1. Visualization of the KM theory. I and J are the two light fluxes in opposite directions; k and s are the absorption and scattering coefficients, respectively.

reflectance spectroscopy because a further increase of the sample thickness does not affect the measured signal. In that case, the calculation of the infinite reflectance in (3) can be drastically simplified:

$$R_\infty = \frac{1-\beta}{1+\beta} = 1 + \frac{k}{s} - \sqrt{(\frac{k}{s})^2 + 2\frac{k}{s}} \qquad (4)$$

By solving (4) for transformed reflectance $r = k/s$, the so-called KM function is obtained

$$r = \frac{k}{s} = \frac{(1-R_\infty)^2}{2R_\infty} \qquad (5)$$

According to the KM theory, the relationship between infinite reflectance R_∞ and transformed reflectance r is derived.

B. SOM RETRIEVAL MODEL

For dry soil, reflectance, which is related to SOM, mainly depend on Fresnel reflectance R_i and diffuse scattering R_d [41]. The relationship can be expressed as:

$$R = R_i + R_d = R_i + (1 - R_i)\frac{(1-K_2)R_\infty}{1-K_2 R_\infty} \qquad (6)$$

where K_2 is the Fresnel reflectance for diffuse light that exits the material and transits a thin layer-air interface at the material surface. In general, K_2 is a function of surface roughness, refractive index and scattering angles. It is often assumed approximately equal to R_i or treated as a constant [41]. R_i is the Fresnel reflectance for light incident in air upon the target surface [42]:

$$R_i = \left(\frac{n_{soil} - n_{air}}{n_{soil} + n_{air}}\right)^2 \qquad (7)$$

where n_{soil} is refractive indices of soil (≈ 1.5) and n_{air} is refractive indices of air (≈ 1).

Equation (6) can be rearranged as:

$$R_\infty(R) = \frac{R - R_i}{R \cdot R_i + 1 - 2R_i} \qquad (8)$$

Combining (5) and (8) yields:

$$r(R) = \frac{k}{s} = \frac{(1-R_\infty)^2}{2R_\infty} = \frac{[1 - (\frac{R-R_i}{R \cdot R_i + 1 - 2R_i})]^2}{\frac{2(R-R_i)}{R \cdot R_i + 1 - 2R_i}} \qquad (9)$$

Equation (9) shows that reflectance R is affected by k and s, this is because they are functions of the soil particle characteristics (i.e., mineral composition, soil water content, nutrients, etc.) and the SOM. A frequently effective and commonly accepted assumption is that k and s of a mixed medium can be regarded as a simple sum function of k and s weighted by their composition proportions. Given this assumption, k and s of the soil surface can be described as:

$$k(\theta) = k_{other}(1 - \theta) + k_{SOM}\theta \qquad (10)$$

$$s(\theta) = s_{other}(1 - \theta) + s_{SOM}\theta \qquad (11)$$

where θ is SOM content, k_{SOM} and k_{other} are absorption coefficients of SOM and the other components, respectively; s_{SOM} and s_{other} are scattering coefficients of SOM and the other components, respectively. When SOM content is θ_1, absorption and scattering coefficients of the soil denoted as k_1 and s_1 can be described as:

$$k_1 = k_{other}(1 - \theta_1) + k_{SOM}\theta_1 \qquad (12)$$

$$s_1 = s_{other}(1 - \theta_1) + s_{SOM}\theta_1 \qquad (13)$$

Equations (10) and (11) can be described as:

$$k(\theta) = k_1(\frac{1-\theta}{1-\theta_1}) + k_{SOM}(\frac{\theta - \theta_1}{1-\theta_1}) \qquad (14)$$

$$s(\theta) = s_1(\frac{1-\theta}{1-\theta_1}) + s_{SOM}(\frac{\theta - \theta_1}{1-\theta_1}) \qquad (15)$$

Combining (14), (15), and (5) yields:

$$r(\theta) = \frac{k(\theta)}{s(\theta)} = \frac{k_1(\frac{1-\theta}{1-\theta_1}) + k_{SOM}(\frac{\theta-\theta_1}{1-\theta_1})}{s_1(\frac{1-\theta}{1-\theta_1}) + s_{SOM}(\frac{\theta-\theta_1}{1-\theta_1})} \qquad (16)$$

Next, the numerator and denominator on the right side of the (16) are simultaneously divided by the scattering coefficient s_1:

$$r(\theta) = \frac{k(\theta)}{s(\theta)} = \frac{r_1(\frac{1-\theta}{1-\theta_1}) + a_1(\frac{\theta-\theta_1}{1-\theta_1})}{(\frac{1-\theta}{1-\theta_1}) + a_2(\frac{\theta-\theta_1}{1-\theta_1})} \qquad (17)$$

With:

$$a_1 = \frac{k_{SOM}}{s_1} \qquad (18)$$

$$a_2 = \frac{s_{SOM}}{s_1} \qquad (19)$$

$$r_1 = \frac{k_1}{s_1} = \frac{(1-R_1)^2}{2R_1} \qquad (20)$$

where R_1 is the reflectance of the soil when SOM content is θ_1.

For remote sensing applications, one need to retrieve SOM content from reflectance data. As to such an application, equation (17) can be solved explicitly for SOM content as:

$$\theta(R) = \frac{r_1 - r(R) + [a_2 r(R) - a_1]\theta_1}{r_1 - r(R) + a_2 r(R) - a_1} \qquad (21)$$
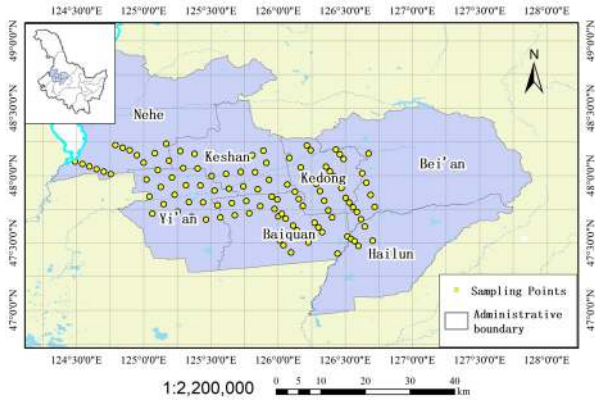
FIGURE 2. Study area with soil samples locations.

FIGURE 3. Experimental device.

With:

$$r_1 = \frac{k_1}{s_1} = \frac{(1 - R_1)^2}{2R_1} \tag{22}$$

$$r(R) = \frac{[1 - (\frac{R - R_i}{R \cdot R_i + 1 - 2R_i})]^2}{\frac{2(R - R_i)}{R \cdot R_i + 1 - 2R_i}} \tag{23}$$

III. EXPERIMENT

A. SOIL SAMPLE PREPARATION

In 2016, 109 soil samples, which include black soil, chernozem soil, and meadow soil, were equably collected in Qiqihar (124°26'15.86''E-126°45'22.44''E, 47°24'31.40'' N-48°14'25.33''N Hei Longjiang Province), showed in Fig. 2. The type of land is farmland. Krishnan *et al.* [43] found that the SOM content determines its role in the soil reflectance spectrum. When the SOM content is more than 2%, it plays a major role in describing the spectral reflectance characteristics of the soil; when the SOM content is less than 2%, the ability that SOM conceals the spectral reflectance properties of other soil constituents (such as iron and manganese) is diminished. The SOM content in the study area is high (greater than 2%). SOM is the dominant factor of soil reflectance spectrum characteristics in the study area, which provides a high-quality data basis for the SOM content retrieval model. The collected soil samples were further air-dried, and crushed to pass through a 1 mm sieve in order that stones, roots and the vegetation litter were avoided from soils.

B. SPECTRAL MEASUREMENT AND PRE-PROCESSING

The hyper-spectral reflectance data were acquired in a dark room using an ASD FieldSpec.3 Portable Spectrometer (Analytical Spectral Devices, Boulder, CO, USA). The main geometric parameters of the spectrometer set-up were illustrated as follows (Fig. 3): a 50 W halogen lamp as unique light source with a 30° incident angle was used to reduce the shadow effect caused by soil roughness. The lamp away from petri dish was set as 10 cm; the probe was mounted vertically about 5 cm above the dish, and the field angle of the probe was 1°. The soil depth is 1cm. In order to obtain the absolute reflectance, the r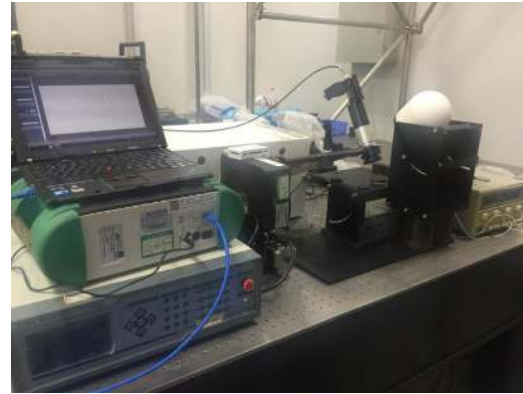eflectance was standardized using a white spectralon reference panel. The arithmetic average of 10 spectral curves collected from each soil sample was regarded as the actual reflectance spectrum data.

The reflectance of each spectrum was narrowed to 450–2500 nm. To eliminate the noise in the spectra, the study applied rlowess smoothing to the original reflectance spectra curve.

Three-quarter whole dataset were chosen by sample set partitioning based on joint x-y distance (SPXY) method [44] and used for the calibration set (n = 82). The remaining was used for the validation set (n = 26).The RMSEP, R^2 and RPD between the predicted and measured SOM in the processes of validation were selected to evaluate the model performance. The RMSEP, shown in (24), represents the mean absolute error of prediction that was calculated by the model between the observed estimators and the measured values.

$$RMSEP = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{24}$$

where y_i and $\hat{y}_i$ are the observed and predicted value, respectively; n is the number of samples with i = 1, 2, n.

The coefficients of determination R^2 which is the percentage of the total variation in the dependent values is defined in (25).

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

where $\overline{y}$ is the mean of the observed data.

RPD is also used for evaluation of the models' accuracy and is the standard deviation (SD) divided by RMSEP, and described by (26). According to [33], the estimations are "excellent" with RPD $\geqslant$ 2.5 and R^2 $\geqslant$ 0.80, "good" with RPD between 2 and 2.5 and R^2 $\geqslant$ 0.70, "moderate" with RPD between 1.5 and 2 and R^2 $\geqslant$ 0.60 and "poor" with RPD < 1.5 and R^2 < 0.60.

$$RPD = \frac{SD}{RMSEP} \tag{26}$$

TABLE 2. Statistical description of SOM contents CV: Coefficient of variation.

Dataset name	whole	calibration	validation
Number	108	82	26
Maximum (%)	5.86	5.86	5.24
Minimum (%)	2.93	2.93	3.07
Mean (%)	4.28	4.31	4.20
SD (%)	0.81	0.84	0.73
CV (%)	18.98	19.48	17.44

Generally, the larger R^2, RPD and the smaller RMSEP are indicators of a superior model.

The unknown parameter a_1 is the ratio of k_{SOM} to s_1, a_2 is the ratio of s_{SOM} to s_1. They needed to be acquired according to the calibration set based on least-squares algorithm. The best criterion for model parameter selection is to minimize the residual sum of squares between the simulated and the measured value. The optimization objective function is constructed as follows [21]:

$$\min \Delta R(\theta) = \sum (R_{measure} - R_{model})^2 \qquad (27)$$

where, $R_{measure}$ is the measured value for the laboratory, R_{model} is the theoretical value of the model. All data analyses were carried out in Matlab R2014b. (The Math Works Inc.: Natick, MA, USA).

IV. RESULTS AND DISCUSSION

A. STATISTICAL DESCRIPTION OF SOM

The summary statistics of SOM for the whole, calibration and validation sets are respectively provided in Table 2. The values of the mean, standard deviation (SD) and coefficient of variation (CV) from three sets are relatively similar. Generally speaking, the characteristic statistics of both the calibration and the validation set is similar to the whole set, indicating that they are well divided to represent the whole set.

B. REFLECTANCE SPECTRAL FEATURE OF SOIL

Figure 4 and figure 5 respectively show reflectance spectra and continuum removed reflectance of three sorts of soils measured in experiment The spectral profile shows three prominent absorption peaks at 1420nm, 1920nm, and 2200 nm, which are mainly caused by the hydroxyl group (OH) of free water at 1420 and 1920 nm and the Al–OH lattice structure in clay minerals at 2200 nm [45]. The shape of the soil spectral reflectance curves is consistent with the results of other studies [46], [47]. Fig. 4 and Fig. 5 show that the reflectance spectra of different sorts of soils are diverse, which is mainly in respect that soil particle characteristics (i.e., mineral composition, organic matter, nutrients, etc.) of different sorts of soils are distinct. The validity and reliability of the proposed model are confirmed with validation set including three sorts of soils. The diversity of soil optical properties exhibited in Fig. 4 and Fig. 5 illustrates how the proposed model works for different sorts of soils.
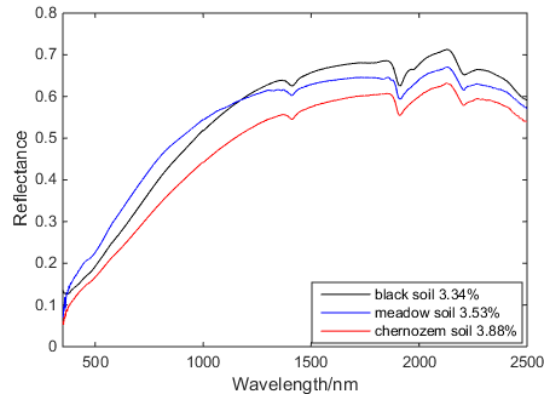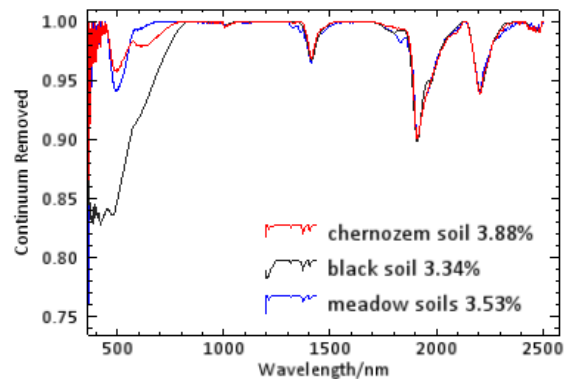

FIGURE 4. Reflectance spectra of three soils.

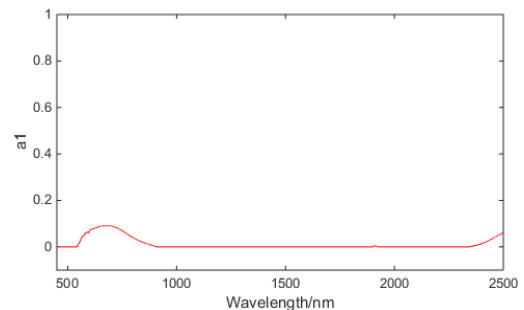**FIGURE 5.** Continuum removed reflectance spectra of three soils.

FIGURE 6. Parameter a_1 at different wavelength.

C. RETRIEVAL UNKNOWN PARAMETER a_1 AND a_2

θ_1 of soil is 2.95%. The unknown parameter a_1 and a_2 were acquired by the least-squares algorithm combining the calibration set, wavelength-by-wavelength, in the range of 450–2400 nm (Fig. 6 and Fig. 7).

D. SOM ESTIMATION

SOM content can be estimated with validation set by using the model mentioned in (21). The RMSEPs between estimated and measured SOM content were computed wavelength-by-wavelength in the range of 450–2500 nm Fig. 8 shows that the accuracy of the model is high,
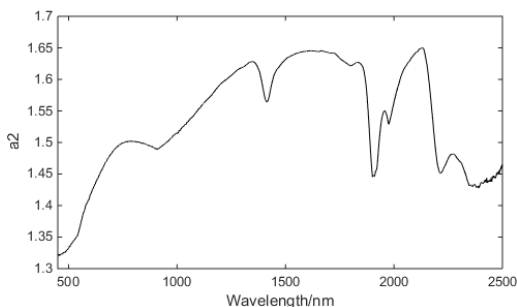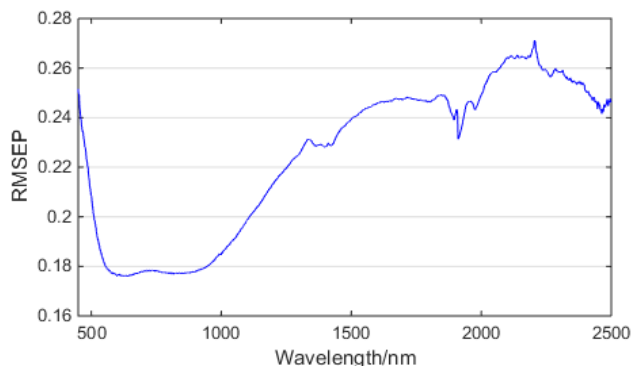
FIGURE 7. Parameter a_2 at different wavelength.

FIGURE 8. RMSEP at different wavelength.

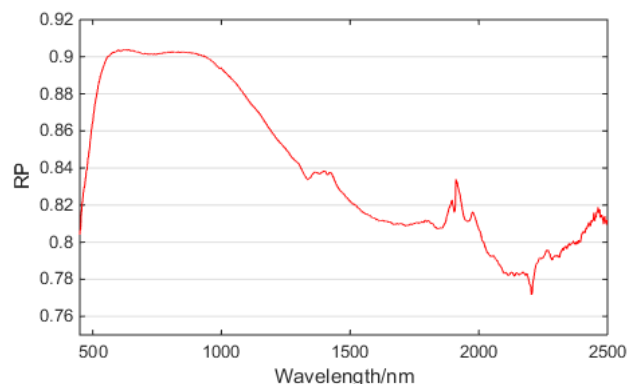**FIGURE 9.** R^2 at different wavelength.

FIGURE 10. RPD at different wavelength.

the RMSEPs are generally less than 0.27%. Especially in the range of 552—950nm, the RMSEPs are less than 0.18%.

The R^2s between estimated and measured SOM were computed in the range of 450–2500 nm, wavelength-by-wavelength (Fig. 9). The R^2s are generally more than 0.78, especially in the range of 552—950nm, the R^2s are more than 89.9%.

From the Fig. 10, the RPDs are generally greater than 2.2. In the range of 552—950nm, the RPDs are more than 3.2. Thus, the model has good prediction ability and it can be well applied to the prediction of SOM content of different sorts of soils.

According to [33], the estimations are "good" with the RPD value of 2.2 and the $R^2 \geqslant 78\%$ in the range of 450–2500 nm, the estimations are "excellent" with the RPD $\geqslant 3.2$ and the $R^2 \geqslant 89.9\%$ in VNIR bands. The
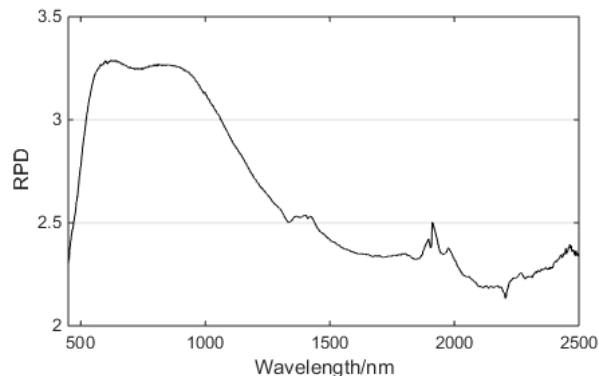
proposed model can be well applied to the prediction of SOM content in different sorts of soils with high prediction accuracy and good prediction ability (RMSEP, R^2 and RPD values of 0.18%, 89.9% and 3.2, respectively) in VNIR bands.

Fig. 8, 9 and 10 reveal that the VNIR bands provide the optimal bands in the solar domain (i.e. wavelength between 350 and 2500 nm) for remote sensing of SOM. It further verifies previous findings. Luan et al. found that saline-alkali SOM had a high correlation with the spectral reflectance at 560–750 nm and 760–1000 nm [48]. Liu et al. used the typical black earth area in Heilongjiang Province as the study area, and showed that the sensitive bands were 445–1380 nm, the significantly correlated spectral range was 545-1250 nm [49]. The results of this article show that in the range of 552-950nm, the model has the highest accuracy, the highest stability, and the best predictive ability. The research results can provide theoretical basis and technical support for sensor band setting.

To further verify the validity of the model, 26 soil samples were randomly selected from the whole sets as the validation sets, and the remaining samples were calibration sets. The unknown a_1 and a_2 in (21) were inverted using the validation set, and then the SOM was estimated by randomly selecting 26 verification samples. The RMSEPs between estimated and measured SOM were computed at six wavelengths corresponding to various bands of Landsat TM and ETM+ satellite, including band 1 (blue, 480 nm), band 2 (green, 560 nm), band 3 (red, 660 nm), band 4 (near infrared, 830 nm), band 5 (SWIR, 1650 nm), and band 7 (SWIR, 2210 nm). The above procedure was repeated 50 times, and the minimum value, the maximum value, and the average value of the RMSEPs were calculated at 6 wavelengths, respectively. The results (Table 3) show that the mean value of the RMSEPs ranges from 0.1950% to 0.2614%, which reveals the model can effectively invert the SOM content.

E. COMPARISON WITH PLSR

In this section, the model described in (21) was compared to PLSR method for estimating SOM content. Fig. 11 provides the RMSEPs of SOM retrieved, computed wavelength-by-wavelength in the range of 552-950nm, using the proposed

TABLE 3. Statistical description of the RMSEPs calculated 50 times.

Wavelength	SPXY (%)	Minimum (%)	Maximum (%)	Mean (%)
480nm	0.2255	0.1757	0.2811	0.2531
560nm	0.1792	0.1119	0.2449	0.2131
660nm	0.1765	0.1015	0.2373	0.2030
830nm	0.1773	0.1018	0.2309	0.1950
1650nm	0.2468	0.1770	0.2747	0.2470
2210nm	0.2673	0.1933	0.2975	0.2614
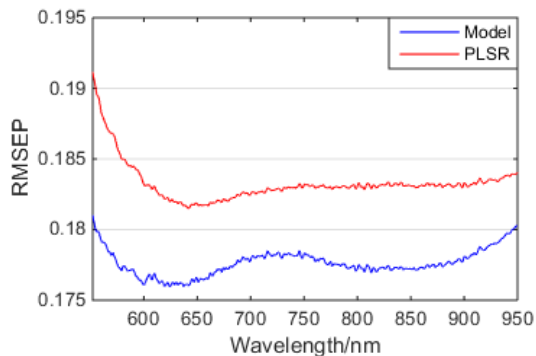
FIGURE 11. Comparison between the proposed model and PLSR method.

model and PLSR method. The RMSEPs of SOM retrieved using the proposed model are generally less 0.005% than PLSR method. Thus, the proposed model performs much better than PLSR method.

The model can estimate the SOM content with higher accuracy than before, and it provides an innovative method for predicting SOM content. The proposed model exhibits the following advantages:

(1) For the first time, a semi-empirical model for SOM content retrieval is constructed by introducing SOM content information into KM model. It has a certain theoretical basis, high prediction accuracy and extensive applicability.

(2) It is not restricted to a single band or wavelength. Properly calibrated, it is competent to describe the response of reflectance to SOM content over the full optical range.

A distinct constraint of the model is that it contains two unknown parameters and thus requires soil information a priori to be solved (i.e. calibration).

V. CONCLUSION

An innovative semi-empirical model is proposed in order to estimate SOM content. The main conclusions of this study are summarized below:

(1) For the first time, a semi-empirical model for SOM content retrieval is constructed by introducing SOM content information into KM model. Compared with the statistical model mostly used in the SOM estimation, the semi-empirical model has a certain theoretical basis and strong applicability

(2) The proposed model can be well applied to the prediction of SOM content in different sorts of soils with high prediction accuracy and good prediction ability (RMSEP, R^2 and RPD values of 0.18%, 89.9% and 3.2, respectively) in VNIR bands.

(3) The VNIR bands provide the optimal bands in the solar domain for remote sensing of SOM. In the range of 552-950nm, the model has the highest accuracy, and the best predictive ability.

This study is the first step toward focusing on the theoretical aspects of the model and its testing under well-controlled laboratory conditions. Future studies are underway to examine/extend the model for field and large-scale applications when facing challenges such as surface roughness, topographical features, shadow effects, etc.

REFERENCES

[1] G. P. Parling, D. Wheelerb, E.-T. Veselyc, and L. A. Schipper, "What is soil organic matter worth?" *J. Environ. Qual.*, vol. 35, no. 2, pp. 548–557, 2006.

[2] M. Ladoni, H. A. Bahrami, S. K. Alavipanah, and A. A. Norouzi, "Estimating soil organic carbon from soil reflectance: A review," *Precis. Agricult.*, vol. 11, no. 1, pp. 82–99, 2010.

[3] P. D. Lin, "Inversion of black soil organic matter content with field hyperspectral reflectance based on continuous wavelet transformation," *Res. Soil Water Conservation*, vol. 25, no. 2, pp. 46–57, 2018. doi: 10.13869/j.cnki.rswc.2018.02.007.

[4] T. Wang, J. Liang, and X. Liu, "Soil moisture retrieval algorithm based on TFA and CNN," *IEEE Access*, vol. 7, pp. 597–604, 2019.

[5] Y. Liu, Y. Liu, Y. Chen, Y. Zhang, T. Shi, J. Wang, Y. Hong, T. Fei, and Y. Zhang, "The influence of spectral pretreatment on the selection of representative calibration samples for soil organic matter estimation using Vis-NIR reflectance spectroscopy," *Remote Sens.*, vol. 11, no. 4, p. 450, Feb. 2019.

[6] W. Luo, Y. Zhang, A. Feizi, Z. Göröcs, and A. Ozcan, "Pixel super-resolution using wavelength scanning," *Light Sci. Appl.*, vol. 5, Dec. 2015, Art. no. e16060. doi: 10.1038/lsa.2016.60.

[7] J. Khan, P. Wang, Y. Xie, L. Wang, and L. Li, "Mapping MODIS LST NDVI imagery for drought monitoring in Punjab Pakistan," *IEEE Access*, vol. 6, pp. 19898–19911, 2018.

[8] T. He, J. Wang, Z. Lin, and Y. Cheng, "Spectral features of soil organic matter," *Geo-Spatial Inf. Sci.*, vol. 12, no. 1, pp. 33–40, 2009. doi: 10.1007/s11806-009-0160-x.

[9] G. Wang, W. Wang, Q. Fang, H. Jiang, Q. Xin, and B. Xue, "The application of discrete wavelet transform with improved partial least-squares method for the estimation of soil properties with visible and near-infrared spectral data," *Remote Sens.*, vol. 10, no. 6, p. 867, 2018. doi: 10.3390/rs10060867.

[10] M. Niels, C. R. Petersen, A. Barh, D. Jain, M. Jensen, G. Hannesschläger, P. Tidemand-Lichtenberg, C. Pedersen, A. Podoleanu, and O. Bang, "Real-time high-resolution mid-infrared optical coherence tomography," *Light Sci. Appl.*, vol. 8, p. 11, Jan. 2019. [Online]. Available: https://www.nature.com/articles/s41377-019-0122-5. doi: 10.1038/s41377-019-0122-5.

[11] A. Morellos, X.-E. Pantazi, D. Moshou, T. Alexandridis, R. Whetton, G. Tziotzios, J. Wiebensohn, R. Bill, and A. M. Mouazen, "Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy," *Biosyst. Eng.*, vol. 152, pp. 104–116, Dec. 2016.

[12] Z. Q. Sun, D. Wu, Y. Lv, and Y. Zhao, "Bidirectional polarized reflectance factors of vegetation covers: Influence on the BRF models results," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5687–5701, Oct. 2017.

[13] W. Yang, M. Deng, X. Yang, D. Wei, "Predictive soil pollution mapping: Hybrid approach for a dataset with outliers," *IEEE Access*, vol. 7, pp. 46668–46676, 2019.

[14] F. Lucà, M. Conforti, A. Castrignanò, G. Matteucci, and G. Buttafuoco, "Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy," *Geoderma*, vol. 288, pp. 175–183, Feb. 2017. doi: 10.1016/j.geoderma.2016.11.015.

[15] Y. Ding and S. Pau, "Circularly and elliptically polarized light under water and the Umov effect," *Light Sci. Appl.*, vol. 8, p. 32, Mar. 2019. doi: 10.1038/s41377-019-0143-0

[16] J. Liang and F. Zhu, "Soil moisture retrieval from UWB sensor data by leveraging fuzzy logic," *IEEE Access*, vol. 6, pp. 29846–29857, 2018.

[17] J. Yuan, "A semi-empirical model for reflectance spectral of black soil with different moisture contents," *Spectrosc. Spectr. Anal.* vol. 39, no. 10, Oct, 2019. DOI: 10.3964/j.issn.1000-0593(2019)10-00-05.

[18] F. Castaldi, A. Palombo, F. Santini, S. Pascucci, S. Pignatti, and R. Casa, "Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon," *Remote Sens. Environ.*, vol. 179, pp. 54–65, Jun. 2016.

[19] J. P. Wight, A. J. Ashworth, and F. L. Allen, "Organic substrate, clay type, texture, and water influence on NIR carbon measurements," *Geoderma*, vol. 261, pp. 36–43, Jan. 2016.

[20] M. Nocita, A. Stevens, G. Toth, P. Panagos, B. van Wesemael, and L. Montanarella, "Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach," *Soil Biol. BioChem.*, vol. 68, pp. 337–347, Jan. 2014.

[21] J. Yuan, X. Wang, C.-X. Yan, S.-R. Wang, X.-P. Ju, and Y. Li, "Soil moisture retrieval model for remote sensing using reflected hyperspectral information," *Remote Sens.*, vol. 11, no. 3, p. 366, 2019. doi: 10.3390/rs11030366.

[22] A. E. Raftery, "Bayesian model selection in social research," *Sociol. Methodol.*, vol. 25, pp. 111–164. Jan. 1995.

[23] C. K. I. Williams and D. Barber, "Bayesian classification with Gaussian processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1342–1351, Dec. 1998. doi: 10.1109/34.735807.

[24] T. Hengl, G. B. M. Heuvelink, and D. G. Rossiter, "About regression-Kriging: From equations to case studies," *Comput. Geosci.*, vol. 33, no. 10, pp. 1301–1315, Oct. 2007.

[25] J. Li, A. D. Heap, A. Potter, and J. J. Daniell, "Application of machine learning methods to spatial interpolation of environmental variables," *Environ. Model. Softw.*, vol. 26, pp. 1647–1659, Dec. 2011.

[26] M. Ließ, B. Glaser, and B. Huwe, "Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and Random Forest models," *Geoderma*, vol. 170, no. 1, pp. 70–79, Jan. 2012.

[27] T. Appelhans, E. Mwangomo, D. R. Hardy, A. Hemp, and T. Nauss, "Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania," *Spatial Statist.*, vol. 14, pp. 91–113, Nov. 2015.

[28] B. T. Pham, D. T. Bui, I. Prakash, and M. B. Dholakia, "Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS," *Catena*, vol. 149, pp. 52–63, Feb. 2017.

[29] O. Rahmati, H. R. Pourghasemi, and A. M. Melesse, "Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran," *Catena*, vol. 137, pp. 360–372, Feb. 2016.

[30] A, Shirzadi, H. Shahabi, K. Chapi, D. T. Bui, B. T. Pham, K. Shahedi, and B. B. Ahmad, "A comparative study between popular statistical and machine learning methods for simulating volume of landslides," *Catena*, vol. 157, pp. 213–226, Oct. 2017.

[31] C. D. Andre, R. S. D. Dalmolin, A. T. Caten, and S. Grunwald, "A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra," *Geoderma*, vol. 314, pp. 262–274, Mar. 2018. doi: 10.1016/j.geoderma.2017.11.006

[32] S. Nawar, H. Buddenbaum, J. Hill, J. Kozak, and A. M. Mouazen, "Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy," *Soil Tillage Res.*, vol. 155, pp. 510–522, Jan. 2016. doi: 10.1016/j.still.2015.07.021

[33] O. Yaser, S. Ghorbani-Dashtaki, H.-A. Bahrami, M. Abbasi, J. A. M. Dematte, E. Arthur, and P. Panagos, "Towards prediction of soil erodibility, SOM and $CaCO_3$ using laboratory Vis-NIR spectra: A case study in a semi-arid region of Iran," *Geoderma*, vol. 314, pp. 102–112, Mar. 2018. doi: 10.1016/j.geoderma.2017.11.014

[34] P. Tziachris, V. Aschonitis, T. Chatzistathis, and M. Papadopoulou, "Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters," *Catena*, vol. 174, pp. 206–216, Mar. 2019. doi: 10.1016/j.catena.2018.11.010

[35] Y. Yin, "Study on hyper-spectral models for predicting black soil organic matter content," M.S. thesis, Dept., Earth Explor. Inf. Technol. Jilin Univ., Changchun, China, 2015.

[36] Y. Liu and Y. M. Yao, "Research progress of soil moisture quantitative inversion by hyperspectral remote sensing," *Chin. Agricult. Sci. Bulletin*, vol. 32, no. 7, pp. 127–134, 2016.

[37] A. Ciani, K.-U. Goss, and R. P. Schwarzenbach, "Light penetration in soil and particulate minerals," *Eur. J. Soil Sci.*, vol. 56, no. 5, pp. 561–574, 2005.

[38] V. Barron and J. Torrent, "Use of the Kubelka—Munk theory to study the influence of iron oxides on soil colour," *J. Soil Sci.*, vol. 37, pp. 499–510, Dec. 1986.

[39] X. Hu and W. M. Johnston, "Concentration additivity of coefficients for maxillofacial elastomer pigmented to skin colors," *Dental Mater.*, vol. 25, pp. 1468–1473, Nov. 2009.

[40] P. Kubelka and F. Munk, "Ein Beitrag zur Optik der Farbanstriche," *Zeitschrift frTechnische Physik*, vol. 12, pp. 593–601, 1931.

[41] L. L. Hors, P. Hartemann, D. Dolfi, and S. Breugnot, "Phenomenological model of paints for multispectral polarimetric imaging," *Proc. SPIE*, vol. 4370, pp. 94–105, Sep. 2001.

[42] M. Sadeghi, S. B. Jones, and W. D. Philpot, "A linear physically-based model for remote sensing of soil moisture using short wave infrared bands," *Remote Sens. Environ.*, vol. 164, pp. 66–76, Jul. 2015.

[43] P. Krishnan, J. D. Alexander, B. J. Butler, and J. W. Hummel, "Reflectance technique for predicting soil organic matter," *Sci. Soc. Amer. J.*, vol. 44, no. 6, pp. 1282–1285, 1980.

[44] R. K. H. Galvão, M. C. U. Araujo, G. E. José, M. J. C. Pontes, E. C. Silva, and T. C. B. Saldanha, "A method for calibration and validation subset partitioning," *Talanta*, vol. 67, no. 4, pp. 736–740, Oct. 2005.

[45] D. Summers, M. Lewis, B. Ostendorf, and D. Chittleborough, "Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties," *Ecol. Indicators*, vol. 11, no. 1, pp. 123–131, 2011.

[46] L. Ramirez-Lopez, K. Schmidt, T. Behrens, B. van Wesemael, J. A. M. Demattê, T. Scholten, "Sampling optimal calibration sets in soil infrared spectroscopy," *Geoderma*, vols. 226–227, pp. 140–150, Aug. 2014.

[47] D. Li, X. Chen, Z. Peng, S. Chen, W. Chen, L. Han, and Y. Li, "Prediction of soil organic matter content in a litchi orchard of South China using spectral indices," *Soil Tillage Res.*, vol. 123, pp. 78–86, Jul. 2012.

[48] F. M. Luan, X. L. Zhang, H. G. Xiong, F. Zhangm, and F. Wang, "Comparative analysis of soil organic matter content based on different hyperspectral inversion models," *Spectrosc. Spectral Anal.*, vol. 1, no. 1, pp. 196–200. 2013.

[49] H. J. Liu, B. F. Wu, C. J. Zhao, and Y. S. Zhao, "Effect of spectral resolution on black soil organic matter content predicting model based on laboratory reflectance," *Spectrosc. Spectral Anal.*, vol. 32, no. 3, pp. 739–742, 2012.

JING YUAN was born in Changchun, Jilin, China, in 1993. She received the B.S. degree in optical information science and technology from Jilin University, China, in 2016. She is currently pursuing the Ph.D. degree in optical engineering with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun. Her current research interests include application of spectra, hyper-spectral remote sensing, and application of remote sensing in soil.

CHUNHUI HU was born in Xinyang, Henan, China, in 1986. He received the B.S. degree in optical information science and technology from Northwestern Polytechnical University, China, in 2008, and the Ph.D. degree in optical engineering from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China, in 2013, where he is currently a Research Assistant.

His current research interest includes installation and inspection of optical systems.

CHANGXIANG YAN was born in Honghu, Hubei, China, in 1973. He received the M.S. degree in engineering from Zhejiang University, Zhejiang, China, in 1998, and the Ph.D. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China, in 2001.

Since 2010, he has been the Director of the Space Optics Laboratory, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. His research interests include opto-mechatronics technology for space optical remote sensing instruments, multispectral and hyper-spectral spatial remote sensing imaging, polarization detection, and space surveillance.

ZHIZHONG LI was born in Datong, Shanxi, China, in 1961. He received the M.S. degree in geosciences from the China University of Geosciences, Beijing, China, in 1987.

He is currently the Director of the Shenyang Geological Survey Center of China Geological Survey, Shenyang Institute of Geology and Mineral Resources, Chinese Academy of Geological Sciences, and the Deputy Secretary of the Party Committee. His research interest includes geological application of remote sensing.

SHENGBO CHEN was born in Luoshan, Henan, China, in 1967. He received the M.S. degree in remote sensing geology from the Changchun Institute of Geology, China, in 1996, and the Ph.D. degree in earth exploration and information technology from Jilin University, Changchun, China, in 2000, where he is currently a Professor with the Faculty of Geo-Exploration Science and Technology.

His research interests include remote sensing information mechanism, inversion of land surface physical characteristic parameters, inversion of atmospheric trace gas composition, and lunar and planetary remote sensing

SHURONG WANG was born in Changchun, Jilin, China, in 1963. She received the B.S. degree in optics from the Harbin University of Science and Technology, Harbin, China, in 1983, and the M.S. degree from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, in 1993, where she has been a Researcher, since 1983. Her research interests include optical detection, ultraviolet radiometric measurement, and space ultraviolet spectrum remote sensing technology.

XIN WANG was born in Jilin, China, in 1992. He received the B.S. degree in photoelectric information engineering from the Changchun University of Science and Technology, China, in 2015. He is currently pursuing the Ph.D. degree in optical engineering with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. His current research interest includes application of polarization spectra.

ZHENGYUAN XU was born in Dongfeng, Jilin, China, in 1991. He received the B.S. degree in surveying and mapping engineering and the M.S. degree in architecture and civil engineering (surveying engineering) from the Jilin University of Architecture, China, in 2014 and 2018, respectively. He is currently pursuing the Ph.D. degree in geoscience science and technology with Jilin University, Changchun, China. His current research interests include hyper-spectral remote sensing, hyper-spectral analysis of black soil, and integrated satellite-space remote sensing monitoring of key black soil zones in the world.

XUEPING JU was born in Songyuan, Jilin, China, in 1991. He received the B.S. degree in control technology and instruments from Tianjin University, China, in 2014, and the Ph.D. degree in optical engineering with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China, in 2019. His current research interest includes polarization calibration.

• • •