

Quantitative Inversion Model of Total Potassium in Desert Soils Based on Multiple Regression Combined with Fractional Differential

An-Hong Tian,^{1,2} Jun-San Zhao,^{1*} Hei-Gang Xiong,³ and Cheng-Biao Fu^{1,2}

¹Faculty of Land Resource Engineering, Kunming University of Science and Technology,
Kunming, Yunnan 650093, China

²College of Information Engineering, Qujing Normal University,
Qujing, Yunnan 655011, China

³College of Applied Arts and Science, Beijing Union University,
Beijing 100083, China

(Received March 28, 2018; accepted August 8, 2018)

Keywords: fractional differential, total potassium content, multiple regression, desert soil, inversion model

Potassium is an important nutrient element for plant growth. The traditional integer-order differential transformation methods at first order and second order tend to reduce the accuracy of the total potassium content quantitative inversion model, and there are few reports on the use of the fractional differential algorithm for the prediction of soil total potassium content. In this paper, the use of the fractional differential algorithm to predict total potassium content is introduced. Field soils collected in the Xinjiang Uygur Autonomous Region from May 9 to 23, 2017, were used as the data sources. Firstly, we calculated the correlation between spectral reflectance and total potassium content for the original spectrum (R) and the root mean square spectrum (\sqrt{R}) under different fractional differential orders. Secondly, bands whose maximum absolute correlation coefficient was greater than 0.5 were selected as sensitive bands. R had seven bands: 562, 596, 1177, 2155, 2156, 2364, and 239 nm. \sqrt{R} had six bands: 596, 1177, 2155, 2156, 2364, and 2398 nm. Finally, a multiple regression analysis method was employed to quantitatively estimate the optimal model. The ratio of performance to deviation (RPD) evaluation index of a good model should be greater than or equal to 1.4. The simulation results showed that the optimal models for R and \sqrt{R} were the 0.8-order differential and the 0.6-order differential, respectively. The corresponding RPD values were 1.700182 and 1.783319, respectively. We found that the prediction model of \sqrt{R} was more accurate.

*Corresponding author: e-mail: jszht@kmust.edu.cn
<https://doi.org/10.18494/SAM.2018.2048>

1. Introduction

Precision agricultural variable fertilization depends on the understanding of soil nutrient distribution in farmland. Potassium is an indispensable nutrient element throughout the whole growth period of crops.⁽¹⁾ The traditional soil laboratory chemical measurement method is not suitable for the development of precision agriculture because it is time-consuming, labor intensive, and inefficient. Hyperspectral remote sensing technology can yield rapid, accurate, and extensive nondestructive estimation of soil nutrients.^(2,3) Some scholars have tried to quantitatively analyze the soil potassium content using spectral analysis technology. Confalonieri *et al.* used near-infrared reflectance spectroscopy to determine total organic carbon, total nitrogen, exchangeable potassium, and available phosphorus in different soils. Their results showed that the predictive effects of exchangeable potassium and available phosphorus were unsatisfactory.⁽⁴⁾ Malley *et al.* found that the correlation coefficient of the calibration dataset was relatively high, but the prediction accuracy of the validation dataset was not very high.⁽⁵⁾ Rossel *et al.* used different-wavelength spectra combined with partial least squares regression to estimate a variety of soil properties. Their results showed that near-infrared spectroscopy achieved higher precision for exchangeable potassium and the coefficient of determination (R^2) was 0.47, but the accuracy was much lower than that for other soil properties such as carbon and pH.⁽⁶⁾

These hyperspectral inversion models for total potassium are mainly based on the original spectral reflectance, reciprocal, logarithmic transformation, and its first-order and second-order differential spectral preprocessing methods. However, the traditional integer-order differential transformations at the first order and the second order ignore the fractional differential information, which may cause some information loss and difficulty in extracting effective spectral information, thus restricting the accuracy of the modeling.⁽⁷⁾ The fractional differential method is an extension of the concept of the integer differential. Regarding spectral analysis, Schmitt adopted the fractional derivative in the process of diffuse reflectance spectroscopy and found that it could effectively eliminate the effect on baseline drift, and separate overlapping peaks. At the same time, the order choice of the fractional derivative was more flexible, which provided a broader space for band selection.⁽⁸⁾ Zheng *et al.* used the Savitzky–Golay (SG) fractional derivative to preprocess near-infrared spectral based on corn, wheat, and diesel. He pointed out that the accuracy of infrared spectral data after fractional differential processing was improved compared with the integer-order operation.⁽⁹⁾ Zhang *et al.* applied the fractional differential in the pretreatment of hyperspectral data of saline soils, and indicated that it was desirable to use fractional differentials to excavate the potential information of soil spectra data.⁽¹⁰⁾

At present, there are few reports on the estimation of total potassium content in soils using the fractional differential algorithm. In this study, we selected desert soils in Xinjiang Fukang as the research target. The multiple regression analysis method based on the fractional differential algorithm was adopted to estimate the quantitative inversion total potassium content combined with field hyperspectral data and lab total potassium content data, which provided scientific support and an application reference for local precision agriculture.

2. Materials and Methods

2.1 Soil samples and field spectra collection

We conducted field surveys and soil sampling from May 9 to 23, 2017, at a total of 25 sample points. Location information is shown in Fig. 1(a). The sampling depth of soils was 0–10 cm, and the measurement of total potassium content was conducted by the Xinjiang Institute of Ecology and Geography. Hyperspectral data was acquired using ASD Field Spec[®] 3Hi-Res in a spectral range of 350 to 2500 nm and at a test time of 11:00 to 15:00 local time in cloudless and windless weather on a sunny day.

To reduce the influence of spectral noise, the edge bands with low signal-to-noise ratios (350–399 and 2401–2500 nm) and the bands near the moisture absorption band (1355–1410 and 1820–1942 nm) were eliminated. At the same time, the hyperspectral data was subjected to Savitzky–Golay smoothing to eliminate spectral noise, as shown in Fig. 1(b). The root mean square transform of original spectra was used to enhance the difference information between the spectra.

2.2 Multiple linear regression and fractional differential algorithm

Multiple linear regression is the most basic and most commonly used method in visible/near-infrared spectroscopy modeling. Its basic formula can be expressed by

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_j x_j + \cdots + \beta_n x_n + \varepsilon, \quad (1)$$

where y is the chemical composition of soils, x_j is the spectral reflectance value at the j -th band, β_j is the regression coefficient of the spectral reflectance value at the j -th band, and ε is the residual.

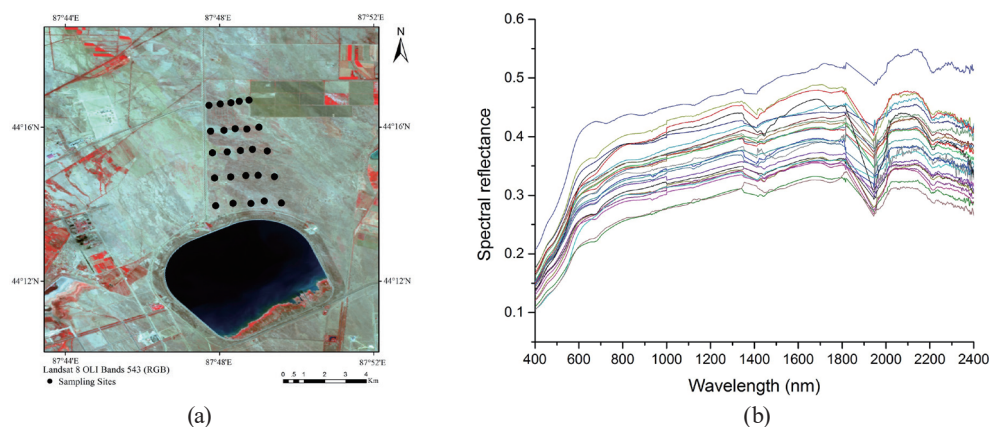


Fig. 1. (Color online) (a) Remote sensing map of sampling points and (b) spectral curves.

The common Grünwald–Letnikov (G–L) fractional differential⁽¹¹⁾ is defined as

$$d^\alpha s(x) = \lim_{h \rightarrow 0} \frac{1}{h^\alpha} \sum_{m=0}^{\frac{t-a}{h}} (-1)^m \frac{\Gamma(\alpha+1)}{m! \Gamma(\alpha-m+1)} s(x-mh). \quad (2)$$

Let the function $s(x)$ define a domain as $x \in [a, t]$, so that $h = 1$ (consistent with the resampling interval of 1 nm for the ASD Field Spec[®] 3Hi-Res spectrometer employed in this study) and $n = [t - a]$. Then, the ν -order fractional differential difference expression of function $s(x)$ is

$$\frac{d^\alpha s(x)}{dx^\alpha} \approx s(x) + (-\alpha)s(x-1) + \frac{(-\alpha)(-\alpha+1)}{2}s(x-2) + \dots + \frac{\Gamma(-\alpha+1)}{n! \Gamma(-\alpha+n+1)} s(x-n). \quad (3)$$

2.3 Model performance evaluation

The performance of the visible/near-infrared spectroscopy model is usually evaluated using R^2 , ratio of performance to deviation (RPD), and root mean square error ($RMSE$).⁽¹²⁾ A good prediction model should have R^2 and RPD as large as possible, and $RMSE$ as small as possible. When R^2 is greater than 0.8 and RPD is greater than or equal to 1.4, the model exhibits good predictive performance. $RPD < 1$ indicates that the model prediction ability is very poor, and the model is unreliable. $1 < RPD < 1.4$ indicates that the model prediction ability is poor, and only the high and low contents of the sample can be roughly estimated. When $1.4 < RPD < 2$, the model has satisfactory predictive ability and can estimate the contents of the sample. When $RPD \geq 2$, the model has a good quantitative prediction ability.

3. Experimental Results and Discussion

3.1 Modeling sensitive band selection

We calculated the maximum absolute correlation coefficient bands of two spectral transformations under different fractional differential orders (Table 1). The correlation between spectral reflectance and soil total potassium content was improved by \sqrt{R} at 0.6-, 0.8-, 1.2-, 1.6-, 1.8-, and 2nd-order differential transformations, and the remaining differential transformations reduced the correlation. On the basis of Table 1, bands whose maximum absolute correlation coefficient was greater than 0.5 were selected as sensitive bands. The original spectrum R had seven bands: 562, 596, 1177, 2155, 2156, 2364, and 2398 nm. The \sqrt{R} transform spectrum had six bands: 596, 1177, 2155, 2156, 2364, and 2398 nm.

Table 1
Maximum absolute value correlation coefficient and corresponding band.

Order	R		\sqrt{R}	
	maximum absolute	Band	maximum absolute	Band
0	0.436273	2399	0.419476	2399
0.2	0.474988	2398	0.457701	2398
0.4	0.537779	2398	0.520629	2398
0.6	0.599456	2364	0.604893	2364
0.8	0.613119	596	0.658464	596
1	0.702791	596	0.698334	596
1.2	0.712503	1177	0.721542	1177
1.4	0.711144	562	0.689548	1177
1.6	0.723515	2156	0.723707	2156
1.8	0.734595	2156	0.741477	2156
2	0.689147	2155	0.694199	2155

3.2 Establishment of multiple regression model based on fractional differential

The total number of samples was 25, and the numbers of samples in the calibration set and verification set were 15 and 10, respectively. Using the total potassium content as the dependent variable, the spectral reflectance of even sensitive bands for R and six sensitive bands of \sqrt{R} were used as independent variables, and 15 total potassium samples were adopted to establish a multiple regression model (Table 2). R^2 of the original spectrum in the first order and second order was 0.756794 and 0.394952, respectively. After calculation using the 0.4-, 0.6-, 0.8-, 1.6-, and 1.8-order fractional differentials, R^2 showed a certain improvement; the 0.6-order differential yielded the maximum improvement to 0.968126, and the corresponding $RMSE$ minimum was 0.194502. R^2 of the root mean square transformation in the first order and second order was 0.806077 and 0.736822, respectively. After calculation using 0.6-, 0.8-, 1.4-, 1.6-, and 1.8-order fractional differentials, R^2 showed a certain improvement; the 0.6-order yielded the largest improvement differential that reached 0.910981, and the corresponding $RMSE$ minimum was 0.325050.

After calculations using the 0.4-, 0.6-, 0.8-, 1.6-, and 1.8-order fractional differentials, the following multiple regression equations based on R can be defined.

The 0.4-order regression equation can be expressed as

$$\begin{aligned}
 Y = & 16.822 - 509.946 \times X_{562} + 412.949 \times X_{596} \\
 & + 448.143 \times X_{1177} - 4320.380 \times X_{2155} \\
 & + 3817.874 \times X_{2156} + 504.500 \times X_{2364} - 120.535 \times X_{2398}.
 \end{aligned} \tag{4}$$

The 0.6-order regression equation can be shown as

$$\begin{aligned}
 Y = & 18.717 - 1884.551 \times X_{562} + 1744.387 \times X_{596} \\
 & + 1086.512 \times X_{1177} - 3492.583 \times X_{2155} \\
 & + 2870.572 \times X_{2156} + 554.344 \times X_{2364} - 85.510 \times X_{2398}.
 \end{aligned} \tag{5}$$

Table 2
Multiple regression model accuracy of calibration dataset.

Order	R		\sqrt{R}	
	R^2	RMSE	R^2	RMSE
0	0.620731	0.670938	0.058936	1.056863
0.2	0.709743	0.586949	0.012481	1.155442
0.4	0.853964	0.416332	0.728539	0.567626
0.6	0.968126	0.194502	0.910981	0.325050
0.8	0.880573	0.376496	0.876839	0.382336
1	0.756794	0.537274	0.806077	0.479759
1.2	0.740234	0.555265	0.769490	0.523063
1.4	0.649099	0.645358	0.857238	0.411637
1.6	0.885461	0.368711	0.896732	0.350100
1.8	0.858124	0.410358	0.878534	0.379696
2	0.394952	0.847430	0.736822	0.558899

The 0.8-order regression equation can be described as

$$\begin{aligned}
 Y = & 22.527 - 3890.042 \times X_{562} + 5306.955 \times X_{596} \\
 & + 384.953 \times X_{1177} - 2557.774 \times X_{2155} \\
 & + 1701.045 \times X_{2156} + 356.483 \times X_{2364} + 49.625 \times X_{2398}.
 \end{aligned} \tag{6}$$

The 1.6-order regression equation can be expressed as

$$\begin{aligned}
 Y = & 23.149 - 2265.096 \times X_{562} - 7393.560 \times X_{596} \\
 & - 23296.602 \times X_{1177} - 1907.850 \times X_{2155} \\
 & + 3921.336 \times X_{2156} - 254.973 \times X_{2364} - 78.750 \times X_{2398}.
 \end{aligned} \tag{7}$$

The 1.8-order regression equation can be shown as

$$\begin{aligned}
 Y = & 23.591 - 5989.505 \times X_{562} - 12410.633 \times X_{596} \\
 & - 22785.443 \times X_{1177} - 1093.441 \times X_{2155} \\
 & + 4123.362 \times X_{2156} - 218.430 \times X_{2364} - 80.432 \times X_{2398}.
 \end{aligned} \tag{8}$$

Here, Y represents the total potassium content. X_{562} , X_{596} , X_{1177} , X_{2155} , X_{2156} , X_{2364} , and X_{2398} represent spectral reflectance values of the 562, 596, 1177, 2155, 2156, 2364, and 2398 nm bands.

After calculation using the 0.6-, 0.8-, 1.4-, 1.6-, and 1.8-order fractional differentials, the following multiple regression equations based on \sqrt{R} can be defined.

The 0.6-order regression equation can be described as

$$\begin{aligned}
 Y = & 19.153 + 657.644 \times X_{596} + 253.014 \times X_{1177} \\
 & - 4057.905 \times X_{2155} + 2950.064 \times X_{2156} \\
 & + 377.524 \times X_{2364} + 79.139 \times X_{2398}.
 \end{aligned} \tag{9}$$

The 0.8-order regression equation can be expressed as

$$\begin{aligned}
 Y = & 22.290 + 1600.325 \times X_{596} - 1218.430 \times X_{1177} \\
 & - 3052.278 \times X_{2155} + 1955.578 \times X_{2156} \\
 & + 245.908 \times X_{2364} + 149.643 \times X_{2398}.
 \end{aligned}
 \tag{10}$$

The 1.4-order regression equation can be shown as

$$\begin{aligned}
 Y = & 20.943 - 2307.233 \times X_{596} - 24037.384 \times X_{1177} \\
 & - 2745.460 \times X_{2155} + 4135.181 \times X_{2156} \\
 & - 232.746 \times X_{2364} - 35.343 \times X_{2398}.
 \end{aligned}
 \tag{11}$$

The 1.6-order regression equation can be described as

$$\begin{aligned}
 Y = & 22.621 - 4330.822 \times X_{596} - 25400.851 \times X_{1177} \\
 & - 2337.303 \times X_{2155} + 4806.951 \times X_{2156} \\
 & - 298.778 \times X_{2364} - 85.985 \times X_{2398}.
 \end{aligned}
 \tag{12}$$

The 1.8-order regression equation can be expressed as

$$\begin{aligned}
 Y = & 23.472 - 7444.538 \times X_{596} - 25011.298 \times X_{1177} \\
 & - 1340.174 \times X_{2155} + 5053.557 \times X_{2156} \\
 & - 308.412 \times X_{2364} - 106.964 \times X_{2398}.
 \end{aligned}
 \tag{13}$$

Here, Y represents the total potassium content. X_{596} , X_{1177} , X_{2155} , X_{2156} , X_{2364} , and X_{2398} represent spectral reflectance values of the 596, 1177, 2155, 2156, 2364, and 2398 nm bands, respectively.

3.3 Predictive model accuracy

Evaluation indicators of the predictive model are shown in Table 3. In the original spectral transformation, the RPD values were greater than 1.4 through the 0.4-, 0.6-, 0.8-, and 1.4-order fractional differential calculations, indicating that the multiple regression prediction models based on the four fractional differential processes can give quantitative estimates of soil total potassium content. In the root mean square transformation, the RPD values were greater than 1.4 after the 0.4-, 0.6-, and 0.8-order fractional differential calculations, indicating that the multiple regression prediction models based on these three fractional differential treatments can yield quantitative estimates of the total potassium content of the soils.

To obtain the best prediction model of soil total potassium content, a test was carried out with 10 samples. The 0.8- and 1.4-order of R and the 0.6- and 0.8-order of \sqrt{R} were taken as examples to analyze and verify the relationship between predicted values and measured values for total potassium content in the validation dataset (Figs. 2 and 3). In the original spectral

Table 3
Multiple regression model accuracy of validation dataset.

Order	R			\sqrt{R}		
	R^2	$RMSE$	RPD	R^2	$RMSE$	RPD
0	0.499215	0.795456	0.721053	0.063619	1.245733	0.283655
0.2	0.388472	0.871812	0.581725	0.004929	1.154239	0.309768
0.4	0.842945	1.820689	1.476974	0.839827	1.920983	1.440459
0.6	0.848096	1.493849	1.489513	0.847846	0.823432	1.788319
0.8	0.747722	0.768436	1.700182	0.678190	0.706277	1.453245
1	0.575259	0.723959	1.263356	0.711493	0.668999	1.467967
1.2	0.703224	0.656794	1.284884	0.600819	0.733024	0.985541
1.4	0.636047	1.575316	1.496350	0.276688	1.063404	0.925765
1.6	0.152691	1.302738	0.923140	0.159993	1.268080	0.913092
1.8	0.186358	1.215575	0.941686	0.147648	1.307531	0.930454
2	0.318387	1.181475	1.207771	0.604784	0.980351	1.469555

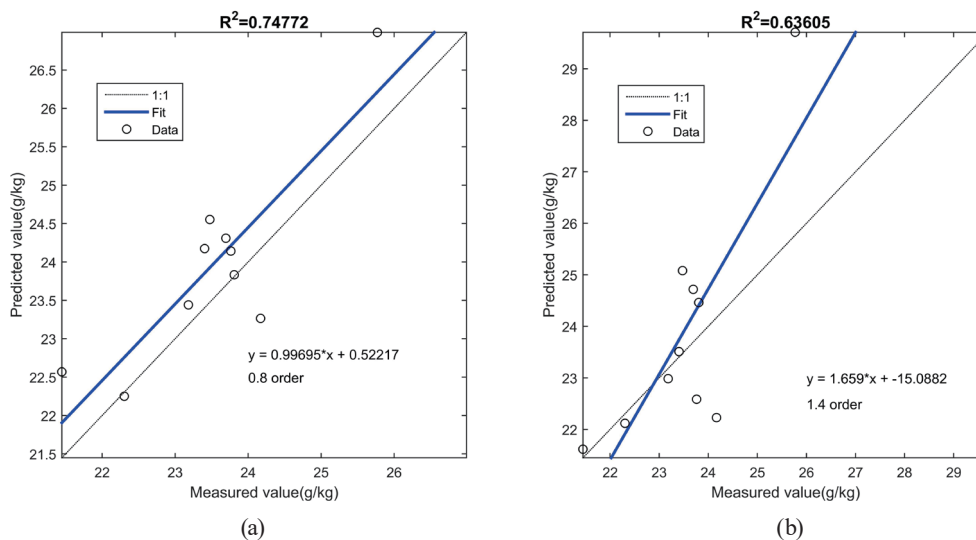


Fig. 2. (Color online) Relationship between predicted value and measured value of R : (a) 0.8- and (b) 1.4-order.

transformation, the sample data points of the 0.8-order verification set were all distributed on both sides of the 1:1 line. The model verified that the sample had higher R^2 , lower $RMSE$, and maximum RPD . However, in the root mean square transformation, the sample data points of the 0.6-order verification set were distributed on both sides of the 1:1 line. The model verified that the sample had the largest R^2 , low $RMSE$, and maximum RPD . At the same time, in the original spectral transformation, the corresponding maximum PRD at the 0.8-order was 1.700182. In the root mean square transformation, the corresponding maximum PRD at the 0.6-order was 1.783319. The multiple regression model based on the 0.6-order root mean square transformation was the best model for predicting soil total potassium content.

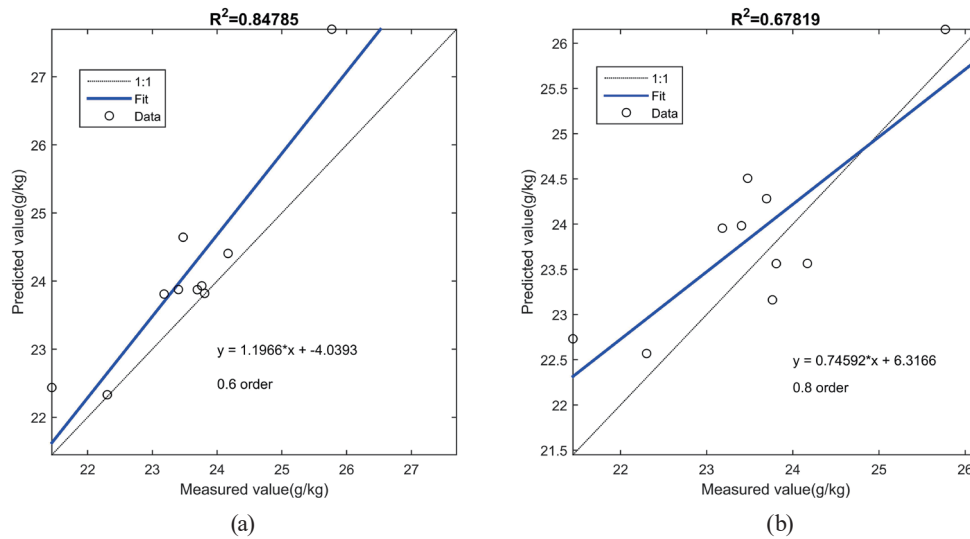


Fig. 3. (Color online) Relationship between predicted value and measured value of \sqrt{R} : (a) 0.6- and (b) 0.8-order.

4. Conclusions

We explored the effectiveness of the fractional differential algorithm for predicting total potassium content in desert soils. The 0.8-order differential model of R was the optimal model, and the 0.6-order differential model of \sqrt{R} was the best prediction model. These were used to quantitatively predict the total potassium content of desert soils. The R^2 and RPD for R and \sqrt{R} in the verification set were 0.747722 and 1.700182, and 0.847846 and 1.783319, respectively. It was shown that \sqrt{R} was better for improving the accuracy of the prediction model than R . This study fills the research gap of determining soil total potassium content using the fractional differential algorithm, and provided a novel idea for quickly and accurately measuring the total potassium content of desert soils in Xinjiang.

Acknowledgments

This work was supported by the financial support of the National Natural Science Foundation of China (41761081 and 41671198) and Yunnan Province Science and Technology Department and Education Department Project (2017FH001-067, 2017FH001-117, and 2016ZDX127), China.

References

- 1 G. Debaene, J. Niedzwiecki, A. Pecio, and A. Zurek: *Geoderma* **214** (2014) 114. <https://doi.org/10.1016/j.geoderma.2013.09.022>
- 2 C. H. Chen and C. H. Liou: *Sens. Mater.* **30** (2018) 365. <https://doi.org/10.18494/SAM.2018.1753>
- 3 P. Brunner, H. T. Li, W. Kinzelbach, and W. P. Li: *Int. J. Remote Sens.* **28** (2007) 3341. <https://doi.org/10.1080/01431160600928641>

- 4 M. Confalonieri, F. Fornasier, A. Ursino, F. Boccardi, B. Pintus, and M. Odoardi: *J. Near Infrared Spec.* **9** (2001) 123. <https://doi.org/10.1255/jnirs.299>
- 5 D. F. Malley, L. Yesmin, and R. G. Eilers: *Soil Sci. Soc. Am. J.* **66** (2002) 1677. <https://doi.org/10.2136/sssaj2002.1677>
- 6 R. A. V. Rossel, D. J. J. Walvoort, A. B. Mcbratney, L. J. Janik, and J. O. Skjemstad: *Geoderma* **131** (2006) 59. <https://doi.org/10.1016/j.geoderma.2005.03.007>
- 7 J. Z. Wang, J. L. Ding, A. Abulimiti, and L. H. Cai: *PeerJ* **6** (2018) e4703. <https://doi.org/10.7717/peerj.4703>
- 8 J. M. Schmitt: *Appl. Spectrosc.* **52** (1998) 840. <https://doi.org/10.1366/0003702981944580>
- 9 K. Y. Zheng, X. Zhang, P. J. Tong, Y. Yao, and Y. P. Du: *Chin. Chem. Lett.* **26** (2015) 293. <https://doi.org/10.1016/j.ccllet.2014.10.023>
- 10 D. Zhang, T. Tiyip, J. L. Ding, F. Zhang, I. Nurmemet, A. Kelimu, and J. Z. Wang: *J. Spectrosc.* **1** (2016) 1. <https://doi.org/10.1155/2016/1081674>
- 11 L. Galeone and R. Garrappa: *J. Comput. Appl. Math.* **228** (2009) 548. <https://doi.org/10.1016/j.cam.2008.03.025>
- 12 O. C. D. Neto, A. D. Teixeira, R. A. D. Leao, L. C. J. Moreira, and L. S. Galvao: *Remote Sens.* **9** (2017) 42. <https://doi.org/10.3390/rs9010042>

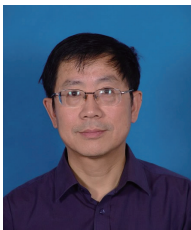
About the Authors



An-Hong Tian is a doctoral graduate student of the Faculty of Land Resource Engineering, Kunming University of Science and Technology. She received her B.S. and M.S. degrees from Chongqing University of Posts and Telecommunications, China, in 2007 and 2010, respectively. From 2011 to 2016, she was a lecturer at Qujing Normal University, China. Since 2017, she has been an associate professor at Qujing Normal University. Her research interests are in remote sensing, and artificial intelligence. (tianah@mail.qjnu.edu.cn)



Jun-San Zhao received his B.S. and M.S. degrees from Central South University, China, in 1985 and 1988, respectively. He received his Ph.D. degree from Wuhan University, China, in 2006. Since 2001, he has been a professor at Kunming University of Science and Technology. His research interests are in land resources, geographic information, and remote sensing. (jszht@kmust.edu.cn)



Hei-Gang Xiong received his B.S. degree from Xinjiang University, China, in 1982, and his M.S. and Ph.D. degrees from Peking University, China, in 1988 and 1991, respectively. Since 1994, he has been a professor at Xinjiang University. His research interests are in regional environmental sustainability and remote sensing. (heigang@bnu.edu.cn)



Cheng-Biao Fu is a doctoral graduate student of the Faculty of Land Resource Engineering, Kunming University of Science and Technology. Since 2015, he has been a lecturer at Qujing Normal University. His research interests are in information identification and processing, hyperspectral remote sensing, and Internet of Things. (fuch@mail.qjnu.edu.cn)