Estimation of Available Phosphorus
in Man Interference Zone Based on Fractional Derivative

An-Hong Tian,[1] Cheng-Biao Fu,[1*] and Hei-Gang Xiong[2]

[1]College of Information Engineering, Qujing Normal University, Qujing, Yunnan, 655011, China
[2]College of Applied Arts and Science, Beijing Union University, Beijing, 100083, China

The fractional derivative has the advantages in terms of memory and globality, and it can overcome the shortcomings of the traditional integer differential algorithm. Moreover, the absorption characteristics of available phosphorus in soil in visible near-infrared bands are unclear, and the prediction model has a low precision. In this paper, we propose a novel method to improve the accuracy of the prediction model for available phosphorus content, which is based on the fractional derivative and stepwise multiple linear regression (SMLR). First, the relationship between the soil spectrum and the available phosphorus content under different fractional orders was studied. Secondly, spectrum dimensionality reduction based on sensitive bands was performed. Finally, the SMLR model was adopted to quantitatively predict the available phosphorus content, and the precision of different fractional order models was discussed. Simulation results revealed that the fractional derivative can describe the small differences in spectral data and increase the correlation between the soil spectrum and the available phosphorus content. The 1.4th-order model is the optimum fractional model. Thus, these results indicate that the fractional derivative could improve the accuracy of the estimation model for available phosphorus content.

1. Introduction

In recent years, spectrometry has become an important method for the rapid and nondestructive diagnosis of soil nutrients.[1] Soil spectral information is a comprehensive reflection of soil physical and chemical properties. At present, researchers are developing spectral measurement technology for the study of available phosphorus content in soil. In these studies, the prediction of available phosphorus content was based on the original or integer-order derivative (1st or 2nd) preprocessed spectrum. Wu *et al.* employed the first derivative to preprocess the original spectrum. The back propagation (BP) neural network regression method was used in the prediction model, and the estimation accuracy of available phosphorus content increased by 28.64% when using the BP regression method compared with the case of using the global regression approach.[2] Although the original or integer-order derivative preprocessed

---

spectrum can be used to predict the available phosphorus content in soil to a certain extent, the difference between spectral curves is large, some intermediate spectral information will be missed and lost, and high-frequency noise will be amplified, which in turn affects the prediction accuracy. However, in the case of using the fractional derivative, one can just take advantage of its intermediate information and fully exploit and utilize spectral data. The fractional derivative is a generalization of the integer derivative. Compared with the integer derivative, it has advantages in terms of background noise removal and effective information extraction. It has been widely used in image recognition and signal processing.

Nowadays, the fractional derivative is widely used in spectral data processing. Chen *et al.* proposed a new approach to predict the nitrogen content of a rubber tree using fractional calculus, which could better mine hidden details of original spectra than integer calculus.[3] Xia *et al.* presented a preprocessing method by fractional differential to analyze the correlation coefficient between raw spectral data and soil conductivity, which extensively excavated the potential information on the soil spectrum.[4] Hong *et al.* introduced a new combination method based on the fractional order derivative (FOD) and variable selection to estimate the soil organic matter (SOM), and their results showed that this novel method could improve the estimation model performance of SOM and the optimal model located in the 1.5-order derivative of the genetic algorithm.[5] Wang *et al.* employed the fractional derivative of the 0.1 interval to preprocess raw spectral data in the laboratory: two different modeling approaches were used to establish the estimation model of soil salinity, and the results showed that the most effective model was located in the 1.5-order derivative of random forest.[6] The above research shows that the fractional derivative can effectively extract potential information from spectral information and improve the accuracy of the prediction model. However, research on the prediction of available phosphorus content in soil using the fractional derivative is relatively rare. Available phosphorus is one of the important nutrients in the soil.[7] Its main role is to promote the transport, transformation, and accumulation of nutrients in crops and improve the ability of crops to resist cold and drought. Therefore, it is very important for crop field management, yield improvement, and farmland environmental protection.

Nevertheless, previous studies adopted the fractional differential method to analyze soil spectral characteristics, but soil spectra were determined in an ideal indoor environment: this method has not been used to study soil spectra in fields. Therefore, using the spectral data obtained in fields and the available phosphorus content measured in the laboratory as the research object, we explored the possibility of estimating the available phosphorus content in soil using the Grünwald–Letnikov (G–L) fractional derivative and stepwise multiple linear regression (SMLR). A spectrometer was used to determine field spectra from 9–23 May, 2017, and we obtained spectral data of soil samples by visible near-infrared spectroscopy in the wavelength range of 350–2500 nm. The bands that passed the 0.05 significance test were counted, and seven bands with a maximum absolute correlation coefficient greater than 0.5 were selected as sensitivity bands, which reduces the dimension of hyperspectral data. The SMLR model was used to estimate the available phosphorus content and we analyzed the accuracy index of different fractional models.

2.   Materials and Methods

2.1   Soil sample collection and field spectrum measurement

We took the desert soil in Xinjiang as the research object. Thirty soil sampling points were identified in the study area in May 2017, and the coordinates of each sampling point were located using the global positioning system (GPS), as shown in Fig. 1. In this study zone, human interference activities are high and the soil is reclaimed as a nursery forest. We collected topsoil soil samples (0–10 cm) and recorded the surrounding vegetation and conditions of the environment from which each sample was collected. The collected soil samples were brought to the laboratory and subjected to a series of pretreatments, including air drying, grinding, and sieving in a 1 mm mesh, and then sent to the Xinjiang Institute of Ecology and Geography for soil chemical analysis to measure the available phosphorus content. Moreover, the spectra of the soil samples from the field were measured using the Field Spec® 3Hi-Res spectrometer manufactured by American Analytical Spectral Devices. Field measurements required that the spectral reflectance of the soil samples should be determined on a clear and cloudless day.

2.2   Definition of fractional derivative and SMLR

The definition of the G–L fractional derivative is derived from the derivative of the nth-order continuous function,[8,9] which is obtained by extending the order n to any real order a. Let the function $f(x)$ be a continuous function, and there is a continuous derivative of order n, so the integer derivative of the 1st and 2nd orders can be defined as
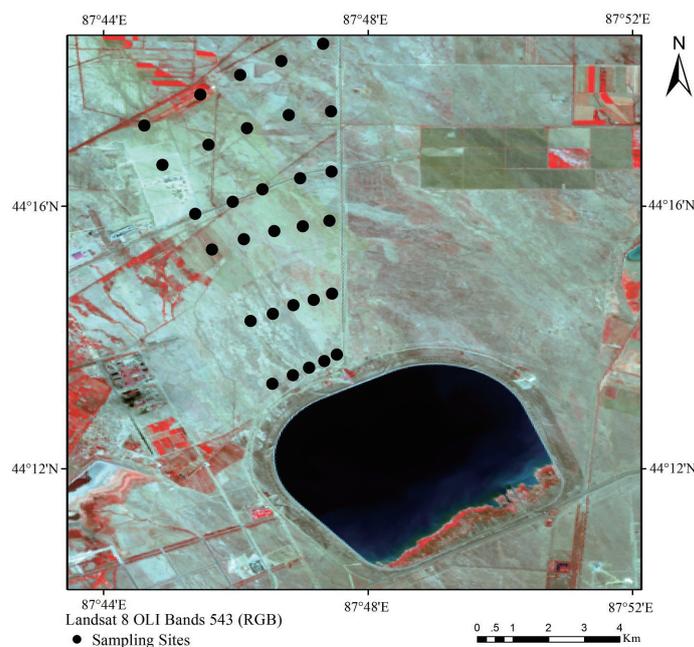
Fig. 1. (Color online) Soil sampling point locations.

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{1}{h}(f(x) - f(x-h)), \tag{1}$$

$$\frac{d^2 f(x)}{dx^2} = \lim_{h \to 0} \frac{1}{h^2}(f(x) - 2f(x-h) + f(x-2h)). \tag{2}$$

Then, the definition of the nth-order derivative is derived as

$$\frac{d^n f(x)}{dx^n} = \lim_{h \to 0} \frac{1}{h^n} \sum_{k=0}^{\infty} (-1)^k C_n^k f(x-kh), \, n > 0. \tag{3}$$

We extended the integer-order differential derivative of Eq. (3) to an arbitrary order, that is, we generalized the integer order n to the fractional order a and replaced the binomial coefficient with the Γ function. Therefore, the G–L fractional derivative can be defined as

$$^{GL}D^a f(x) = \lim_{h \to 0} \frac{1}{h^a} \sum_{k=0}^{\infty} (-1)^k C_a^k f(x-kh), \, a > 0, \tag{4}$$

where $f(x)$ represents a real function, $C_a^k = \dfrac{\Gamma(a+1)}{\Gamma(k+1)\Gamma(a-k+1)}$ represents a binomial coefficient, and $\Gamma(\cdot)$ represents a gamma function.

SMLR is an important method in multivariate statistical analysis, which is widely used in technology, society, economy, and natural science research. It is an algorithmic technique derived from multiple linear regression and automatically selects variables that are important for establishing regression equations from a large number of alternative variables. It is a process of establishing a regression equation based on the optimal combination of multiple independent variables. Its equation can be described as

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_j x_j + \cdots + \beta_n x_n + \varepsilon, \tag{5}$$

where Y is the available phosphorus content in soil, n is the number of independent variables used for modeling, x_i is the soil spectral reflectance of the ith waveband, ε is the random error of the regression equation, and β_i is the regression variance coefficient of the ith independent variable.

2.3 Model predictive ability evaluation

The indexes of coefficient determination (R^2), root mean square error (RMSE),[10] and ratio of performance to deviation (RPD) are used to evaluate the accuracy of the model's estimation ability.[11] RPD is divided into four types of estimation effect. When the RPD is less than 1.4, the estimation ability of the model is poor. When the RPD is between 1.4 and 1.8, the estimation ability of the model is average. When the RPD is between 1.8 and 2, the estimation ability of the model is good. When the RPD is greater than 2, the estimation ability of the model is very good.

3. Simulation Results

3.1 Effect of fractional derivative on correlation

Feature wavebands are the key to building the hyperspectral data model, and when the correlation between the spectral reflectance and the available phosphorus content is higher, the spectral response of the corresponding waveband is more sensitive, and the waveband is more likely to be chosen as the sensitive waveband. Therefore, before the quantitative analysis of available phosphorus content in soil under different fractional derivatives, we should first study the correlation between the soil spectral reflectance and the available phosphorus content to identify the sensitive wavebands for improving the estimation model accuracy. In this study, there were a total of 30 soil samples in the research area, and the 0.05 significance test was performed, so the significance level was ±0.361. Moreover, we programmed to achieve the fractional derivative between the 0.0th order and the 2.0th order under 0.2 steps in the MatlabR2015a software environment, as shown in Fig. 2. It can be seen that there are no
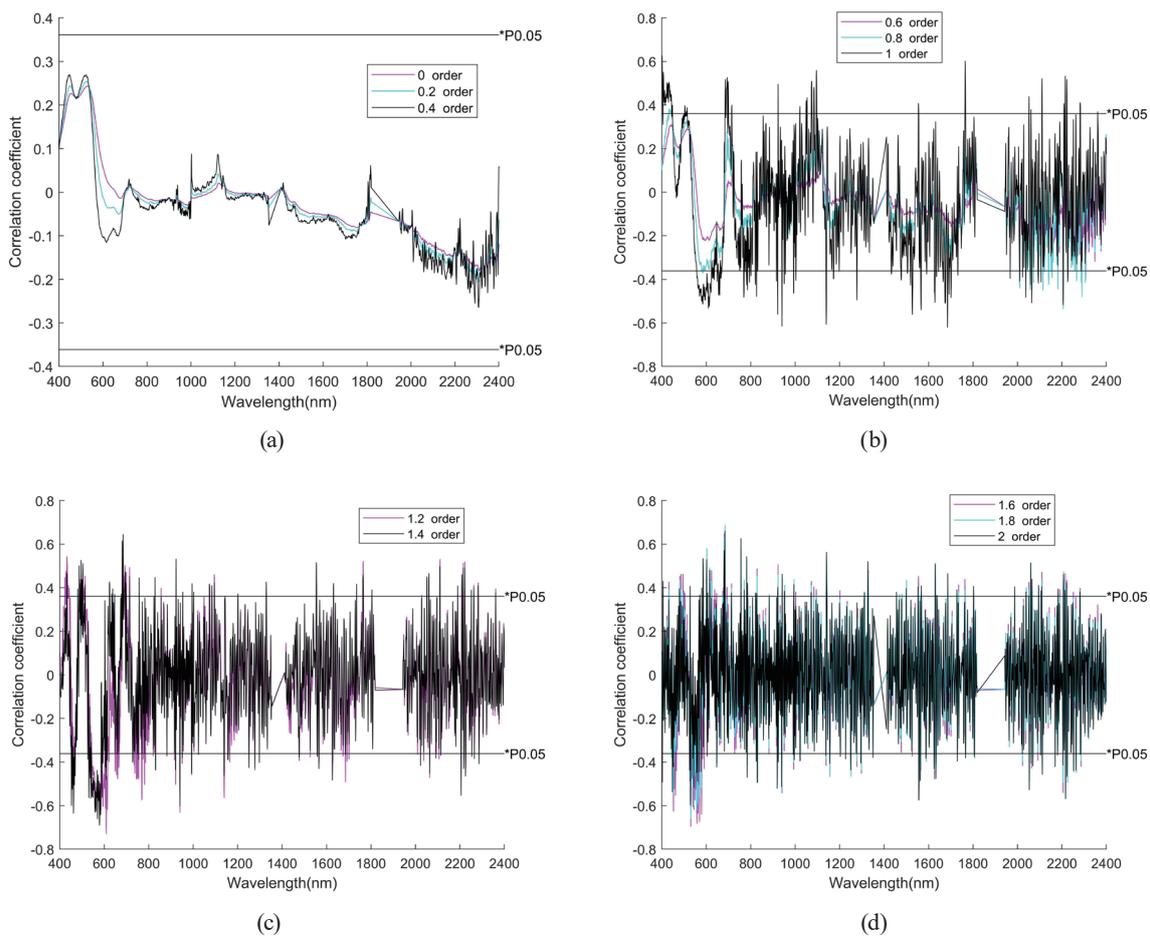
Fig. 2. (Color online) Correlation coefficient of different fractional derivatives: (a) 0th–0.4th, (b) 0.6th–1st, (c) 1.2nd–1.4th, and (d) 1.6th–2nd.

wavebands that passed the significance test between the 0.0th order and the 0.4th order, while the wavebands passed the significance test between the 0.6th order and the 2.0th order. In addition, the simulation results showed that the fractional derivative refines the change trend of the spectral curve and reduces the loss of differential information.

3.2 Choice of sensitive wavebands

The choice of sensitive bands plays an important role in building a reliable regression model. The maximum absolute correlation coefficient and waveband between the spectral reflectance and the available phosphorus content under different fractional derivatives is shown in Table 1. It can be concluded that the maximum correlation coefficient is 0.72862 for the 610 nm waveband located in the 1.2nd order, which is not located in the integer order of 1st or 2nd. Also, the maximum correlation coefficient is smaller than 0.5 between the 0th order and the 0.6th order, which means that the correlation between the soil spectrum and the available phosphorus content is very poor for the 0.0th, 0.2nd, and 0.4th orders. Moreover, the sensitive wavebands greater than 0.5 in Table 1 are selected to build the SMLR regression model, which are 400, 531, 579, 610, 684, 685, and 2206 nm.

3.3 Optimal fractional estimation model for available phosphorus content

A total of 30 soil samples were collected from the research area: 18 soil samples were chosen to build the SMLR model under different fractional derivatives, while 12 soil samples were selected to validate the established model. The regression equation of the SMLR model is shown in Table 2. Since the independent variables from the 0.0th order to the 0.4th order do not satisfy the conditions established by the regression model, there is no regression equation of the SMLR model from the 0.0th order to the 0.4th order. In Table 2, Y denotes the available phosphorus content, and R_{400}, R_{531}, R_{579}, R_{610}, R_{684}, R_{685}, and R_{2206} denote the spectral data in the wavebands of 400, 531, 579, 610, 684, 685, and 2206 nm, respectively. In the calibration set, R^2 is 0.547 in the 1st-order integer, while it is 0.415 in the 2nd-order integer. After conducting the fractional derivative process in the 1.2nd, 1.4th, and 1.6th orders, R^2 increases: the largest R^2 is 0.714 located in the 1.6th order, while the smallest RMSE is 2.35 located in the 1.6th order.

Table 1
Maximum correlation coefficient and waveband.

Different fractional derivative order	Correlation coefficient	Waveband	Different fractional derivative order	Correlation coefficient	Waveband
0.0	0.24368	530	1.2	0.72862	610
0.2	0.25463	524	1.4	0.68987	579
0.4	0.26964	524	1.6	0.69511	531
0.6	0.40394	2306	1.8	0.68717	685
0.8	0.53582	2206	2.0	0.66004	684
1.0	0.62727	400			

The estimated model verification indexes are RMSE, R^2, and RPD, and it is common to mainly consider the RPD to evaluate the model prediction performance. The precision evaluation of the SMLR estimation model under different fractional orders is also shown in Table 2. The fractional derivative has a higher RPD and a lower RMSE in the 1.2nd, 1.4th, and 1.6th orders; it shows that the prediction effects of these three types of fractional derivative are better than those of the others. Moreover, the best estimation model is located in the 1.4th order, which has a maximum RPD of 1.8653. As shown in Sect. 2.3, when the RPD is between 1.8 and 2, the estimation ability of the model is good. Therefore, this fractional estimation model can be adopted to predict the available phosphorus content in soil. In addition, the measured and predicted values for the verification set data from the 0.6th order to the 2.0th order are shown in Fig. 3.

Table 2
Precision evaluation of calibration and verification sets.

Order	Regression equation	Calibration set		Verification set		
		R^2	RMSE	R^2	RMSE	RPD
0.6	$Y = 18.496 - 1233.622 \times R_{2206}$	0.243	3.8321	0.0021	4.4520	0.4070
0.8	$Y = 11.523 - 4899.070 \times R_{2206}$	0.531	3.0168	0.0429	4.6976	0.7509
1.0	$Y = 14.717 - 4782.015 \times R_{2206} - 12337.541 \times R_{610}$	0.547	2.9660	0.3388	3.2970	0.8217
1.2	$Y = 11.842 - 63724.560 \times R_{610} + 26967.301 \times R_{685}$	0.588	2.8274	0.6961	2.3459	1.7718
1.4	$Y = 9.724 - 33071.386 \times R_{579} + 47488.149 \times R_{685} - 57543.268 \times R_{610}$	0.663	2.5572	0.6963	2.6488	1.8653
1.6	$Y = 13.332 - 65963.260 \times R_{531} + 48663.968 \times R_{685} - 62514.959 \times R_{579}$	0.714	2.3548	0.6238	3.1517	1.6942
1.8	$Y = 10.678 + 96021.784 \times R_{685}$	0.452	3.2621	0.4338	3.0178	0.9333
2.0	$Y = 10.915 + 108459.889 \times R_{684}$	0.415	3.3685	0.3904	3.1252	0.8312
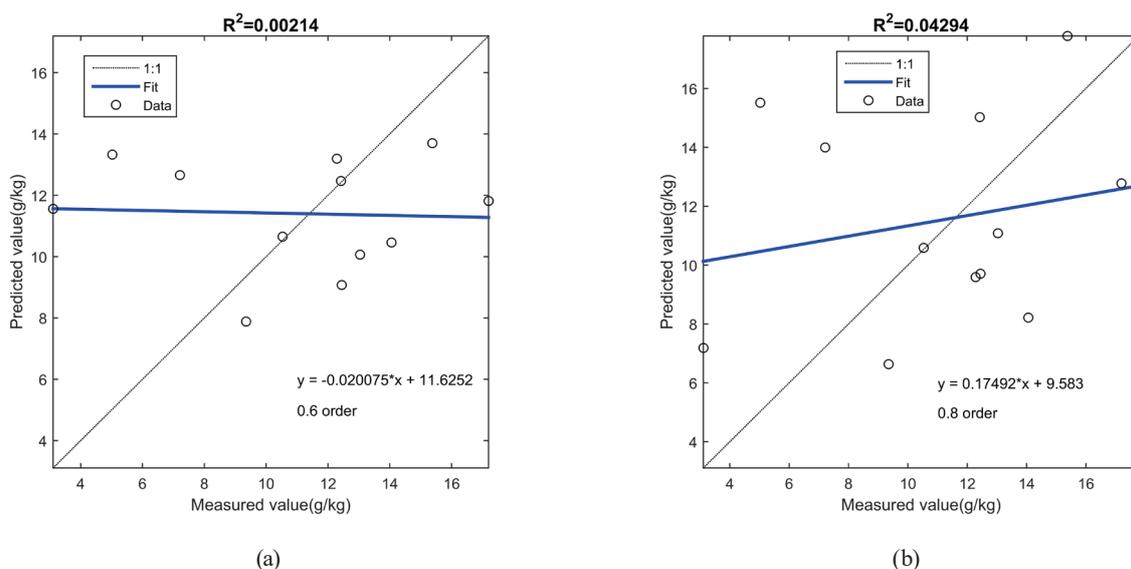
Fig. 3.   (Color online) Measured and predicted values for verification set data: (a) 0.6th order and (b) 0.8th order.
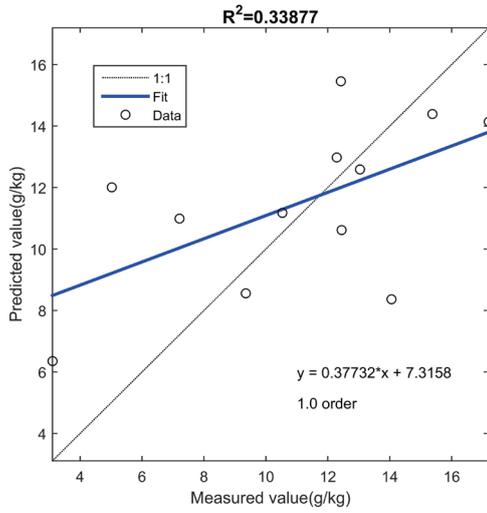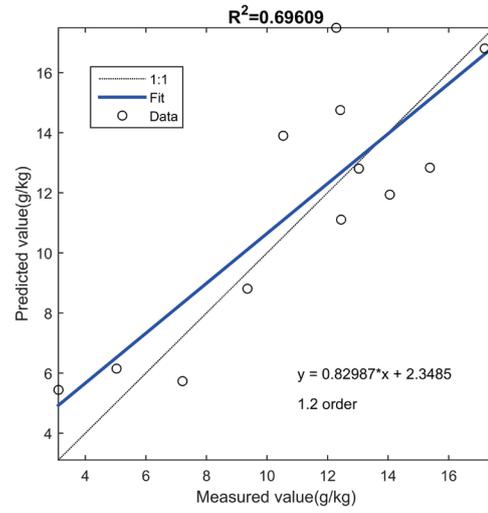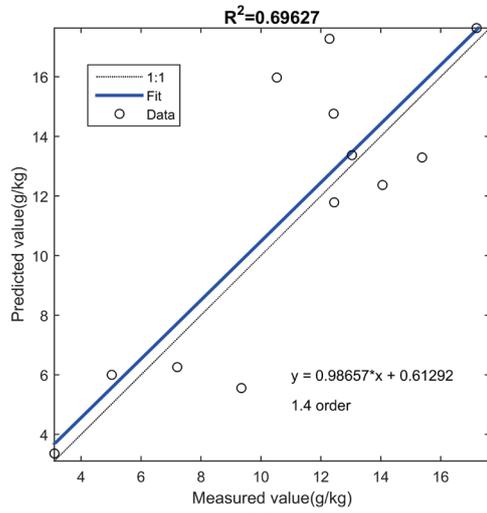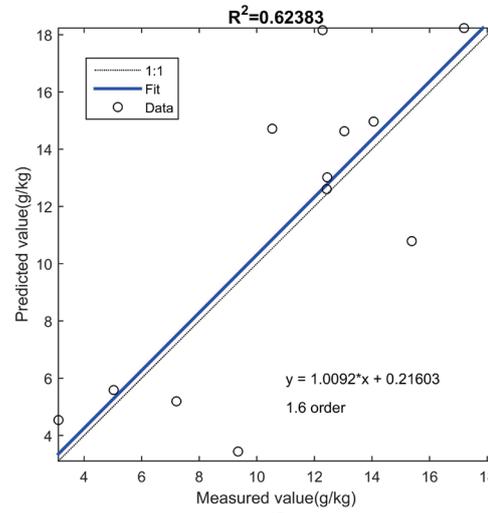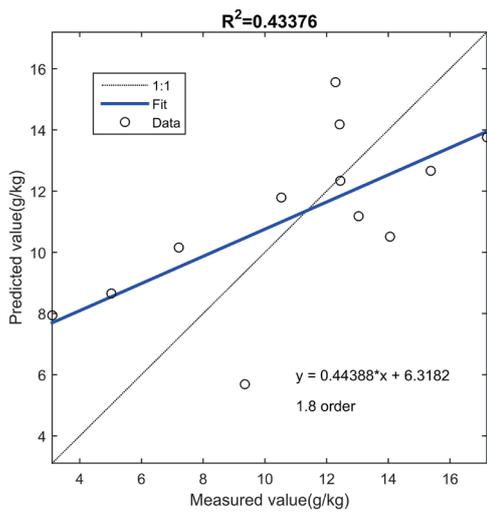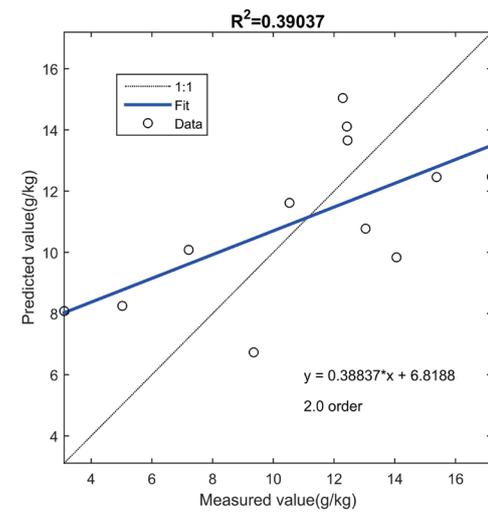
Fig. 3.   (Color online) (Continued) Measured and predicted values for verification set data: (c) 1st order, (d) 1.2nd order, (e) 1.4th order, (f) 1.6th order, (g) 1.8th order, and (h) 2nd order.

4.   Conclusions

The spectra of desert soils in Xinjiang were determined in the field environment, and the spectral data were preprocessed by the fractional differential method to improve the correlation between the soil spectral reflectance and the available phosphorus content. The maximum absolute correlation coefficient and its corresponding band information were determined, and the bands with a maximum absolute correlation coefficient greater than 0.5 were selected as sensitivity bands, which plays a significant role in reducing the dimension of hyperspectral data. Taking the available phosphorus content in soil as a dependent variable, we used the spectral reflectances of seven sensitive bands as independent variables to establish the SMLR model under different fractional orders. The seven sensitive bands for estimating the available phosphorus content in soil are 400, 531, 579, 610, 684, 685, and 2206 nm. Simulations showed that the best estimation model of available phosphorus content in soil is located in the 1.4th fractional order, and the estimation model can be defined as $Y = 9.724 - 33071.386 \times R_{579} + 47488.149 \times R_{685} - 57543.268 \times R_{610}$. The R^2, RMSE, and RPD of the verification set were 0.6963, 2.6488, and 1.8653, respectively, indicating that this model has good estimation accuracy for available phosphorus content.

Acknowledgments

References

1   A. Sidike, S. H. Zhao, and Y. M. Wen: Int. J. Appl. Earth Obs. **26** (2014) 156. https://doi.org/10.1016/j.jag.2013.06.002

2   Q. Wu, Y. H. Yang, Z. L. Xu, Y. Jin, Y. Guo, and C. L. Lao: Spectrosc. Spect. Anal. **34** (2014) 2102. https://doi.org/10.3964/j.issn.1000-0593(2014)08-2102-04

3   K. Chen, C. Li, and R. N. Tang: Ind. Crop. Prod. **108** (2017) 831. https://doi.org/10.1016/j.indcrop.2017.06.069

4   N. Xia, T. Tashpolat, K. Ardak, N. IIyas, J. L. Ding, F. Zhang, and D. Zhang: J. Spectrosc. **1** (2017) 1. https://doi.org/10.1155/2017/1236329

5   Y. S. Hong, Y. Y Chen, L.Yu, Y. F. Liu, Y. L. Liu, Y. Zhang, Y. Liu, and H. Cheng: Remote Sens. **10** (2018) 479. https://doi.org/10.3390/rs10030479

6   J. Z. Wang, J. L. Ding, A. Abulimiti, and L. H. Cai: PeerJ **6** (2018) e4703. https://doi.org/10.7717/peerj.4703

7   S. Y. Jia, X. L. Yang, G. Li, J. M. Zhang: Spectrosc. Spect. Anal. **35** (2015) 2516. https://doi.org/10.3964/j.issn.1000-0593(2015)09-2516-05

8   D. W. Brzeziński and P. Ostalczyk: Commun. Nonlinear Sci. **40** (2016) 151. https://doi.org/10.1016/j.cnsns.2016.03.020

9   M. S. Abdelouahab and N. E. Hamri: Mediterr. J. Math. **13** (2016) 557. https://doi.org/10.1007/s00009-015-0525-3

10 Y. J. Hou, T. Tashpolat, F. Zhang, S. Mamat, and N. IIyas: Spectrosc. Spect. Anal. **35** (2015) 350. https://doi.org/ 10.3964/j.issn.1000-0593(2015)02-0350-05

11 F. Xia, J. Peng, Q. L. Wang, L. Q. Zhou, and S. Shi: J. Infrared Millim. W. **34** (2015) 593. https://doi.org/10.11972/j.issn.1001-9014.2015.05.014.

About the Authors

An-Hong Tian 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, wireless navigation, and artificial intelligence. (tianah@mail.qjnu.edu.cn)

Cheng-Biao Fu received his B.S. and M.S. degrees from Chongqing University of Posts and Telecommunications, China, in 2005 and 2009, respectively. 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. (fucb@mail.qjnu.edu.cn)
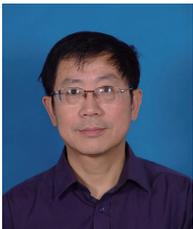
Hei-Gang Xiong received his B.S. degree from Xinjiang University, China, in 1982, and his M.S. and Ph.D. degrees from Peking University, China, in 1988 and 1991, respectively. Since 1994, he has been a professor at Xinjiang University. His research interests are in regional environmental sustainability and remote sensing. (heigang@buu.edu.cn)