Quantitative Prediction Study of Climate-sensitive Potassium with Hyperspectrum in Arid and Semi-arid Region of Northwestern China

F Hu \(^{1,2}\), Q Z Lin \(^1\), Q J Wang \(^*\) and Y J Wang \(^{1,2}\)

\(^1\) Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, 100094, China
\(^2\) University of Chinese Academy of Sciences, Beijing, 100094, China

E-mail: fanghu@ceode.ac.cn; qjwang@ceode.ac.cn

Abstract. Soil potassium content in arid and semi-arid region can reflect the conditions of the paleoclimate and it can be inverted by soil spectra. The relationship between soil spectra and soil potassium content was discussed in this research. Based on four reflectance transformations, single-variance analysis and multi-variances inversion model were built to invert potassium content. The results of single-wavelength inversions were very significant except for the reflectance model. The multi-variances models were good and accurately \((R^2 >0.674\) and RMSE \((\text{root-mean-square error}) <0.09)\). Then, the sensitive wavelengths of the potassium were chosen by using the higher correlation coefficients. The results of this study showed that both methods have a great potential for predicting soil potassium content. The sensitive wavelengths of the potassium content were at 2200 - 2300 nm which could be illustrated by the potassium-bearing minerals spectral absorption features.

1. Introduction

Soil is a key piece of the terrestrial ecosystem which is an important carrier of agricultural production. It can impact the global change processes through material exchange with other ecosystems \([1, 2]\) (the emission of greenhouse gases, the elements leaching and accumulation, et al.). Global change and its regional response have become an important issue in human lives. Potassium is an essential element for plants, and is considered the highest nutrient element content of soil \([3]\). Previous studies showed that soil potassium in arid and semi-arid region was sensitive to climate change and became a hot topic for researchers both domestic and international to study the climate change in the special area \([4, 5]\).

At present, the remote sensing technology has been widely used to the study of soil properties, as VIS-NIR reflectance have already been successfully applied to model organic matter \([6]\), moisture \([7]\), salinity \([8]\), total nitrogen \([9]\), and heavy metals \([10,11]\), that impact soil spectral characteristics mostly. Previous studies have shown that soil potassium contents of arid and semi-arid region can impact the climate \([12]\) and soil spectrum has the potential for inverting the potassium content \([13]\). But, the
research of quantitative prediction of potassium content with hyper-spectrum is rarely, and studies of some issues need to be urgently deepened.

In this paper, the quantitative inversion models of potassium content were built by single-variance analysis method and multi-variances method with soil spectral reflectance and its transformations. It is trying to give explanation from the analysis of the relationship between the potassium-bearing minerals and potassium.

2. Study Area
The study was carried out in alluvial-pluvial fan of Ejina Banner (see figure 1), the arid and semi-arid region of northwestern China. It is located on 39°53'-42°48' N, 97°10'-102°59' E and covers an area of 34,000 km2. It is composed by Gobi, low mountains, hills, deserts, lakes and Ruoshui oasis. The soil of this area was used to have an abundant supply of water, but now the climate here is becoming arid as the climate change, turning the area into Gobi and desert. Soil type is mainly gray-brown desert soil.

![Figure 1. The location and field sampling points of study area](image)

Green area indicates the study area, big yellow dots represent the field samples in 2011 and small dark green dots stand for sampling points in 2012.

3. Data and Methods

3.1. Data
Data sets in the study area were consist of soil reflectance, soil element contents obtained from field sampling data and potassium-bearing minerals from soil XRD analyse. Soil was sampled during two different periods: June-July 2011 and May 2012 (see figure 1). Soil samples were taken from the 0-20 cm depth. A total of 123 samples were taken in Ejina Banner, with section sample of 46 and plane sample of 87. Section samples were chosen to as the research object in this study. Spectral reflectance was collected over the 350-2500 nm wavelength domain with an ASD pro FR Portable Spectro-radiometer. The element content was measured by the American INNOV-X X-Ray fluorescence analyser. The potassium-bearing mineral content was acquired from soil X-Ray Powder Diffraction
analysis. All spectral data were smoothed with the five-spot triple smoothing algorithm, and transformed to first derivative reflectance (FDR), extracted band absorption depth (BD) from continuum removed spectral reflectance [14], corrected for light scattering by multiplicative signal correction (MSC) techniques.

3.2. Methods

3.2.1. Single-wavelength analysis. As a explore research of prediction of soil potassium content, first the correlation coefficients were calculated between the potassium contents and all wavelengths of the soil reflectance and its transformations, then the most correlation coefficient band was chose to build the inversion model of the soil potassium content. Four single-variable were used to build the linear and nonlinear predictive models which include linear, exponential, logarithmic function, power function and polynomial fitting function. The best prediction model of potassium content was selected from these models and thus all section samples were used. Descriptive analysis and difference significance test of statistics method were adopted in finally result analysis.

3.2.2. Multivariate analysis. Partial Least Squares Regression (PLSR) model is one of common tools in multivariate analysis. It is a method that combines the factor analysis and the regression analysis, it solved many difficult problems which is hard for ordinary multiple linear regression. The main idea of PLSR is to extract the latent predictor variables, accounting for as much of the variation of the dependent variable(s).

To quantify performance of the regression model, two model performance parameters were calculated: coefficient of determination (R^2), and root mean square error (RMSE). R^2 is used to verify the relationship between independent variable and dependent variable, and RMSE is used to evaluate the degree that prediction value deviates from the real value. The smaller RMSE is, and the larger R^2 is, the higher accuracy of model predictions and the better model robustness. The equation of RMSE is:

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

(1)

Y and Y' are measured and predicted values, n is the number of observations.

4. Results and Discussion

4.1. Single wavelength prediction model for soil potassium content

Figure 2 shows the correlation coefficient between potassium content and all wavelengths of soil reflectance and its transformation to all 42 section samples. The potassium content and reflectance spectra are positively correlated in the entire VNIR wavelengths, means that the soil reflectance increased with the potassium content (see figure 2 (a)).

The first derivate transformation of reflectance expanded spectral characteristic differences between the samples [14], and the correlation coefficient swing between positive and negative (see figure 2 (b)). The band depth variation and MSC transformation of the soil reflectance are conducive to the comparison of the spectral curve and the characteristic bands, the two curve peaks appear at the same wavelength, but the waveforms are opposite (see figure 2 (c) and figure 2 (d)). The maximum correlation coefficient all appear at 2300 nm.
Figure 2. Correlation coefficient between potassium content and soil reflectance and its transformations ((a) for reflectance, (b) for first derivative of reflectance, (c) for band absorption depth, (d) for multiplicative scatter correction).

The optimization prediction model results with the single variable regression method are presented in Table 1. In addition to the prediction model of reflectance, the significant of the rest model are 0. These showed a good prediction result (see Table 1). The maximum relative coefficient appears at about 2300 nm.

Table 1. Single variable regression prediction results and parametric statistics.

| Spectral indices | Wavelength (nm) | Equation | r     | Sigf |
|------------------|-----------------|----------|-------|------|
| R                | 2338            | K=1.2984*R0.6543 | 0.494 | 0.003|
| FDR              | 754             | K= 107*FDR2-6035.4*FDR+1.2634 | 0.683 | 0    |
| BD               | 2313            | K=193.82*BD2-30.54*BD+1.7193 | 0.784 | 0    |
| MSC              | 2083            | K=707.5*MSC2-661.23*MSC+155.13 | 0.715 | 0    |

Equation is the inversion formula between the potassium content and soil spectral in maximum correlation coefficient, r is the relative coefficient, Sigf is the statistical significance.

After analysis and comparison the potassium and potassium-bearing mineral, the correlation coefficient between the potassium content and the clay mineral content was 0.772 and p value was less than 0.01 which means the result was very significant. That is, the potassium content was shown very significant positive relation to the clay mineral. From the front research know that the spectral curve of potassium-bearing clay mineral (illite and kaolinite) has a small absorption peak around 2300 nm. This is consistent with the previous results.

4.2. PLSR model for soil potassium content

The PLSR model parameters are shown in Table 2. For all spectral indices, the $R^2$ values of calibration data sets are higher than the prediction data set. And all models have a lower RMSE not more than 0.09 and a lower $R^2$ not more than 0.674 (see Table 2). The results show that it is feasible to estimate soil potassium content by the measured reflectance and its transformation in the arid and semi-arid areas. The PLSR prediction model of the potassium content with laboratory soil spectrum is robust.
reliable and accurate. For different sample sets, the influence degree of the PLSR model accuracy is
different to variously transform indicators of reflectance spectra and it can’t be generalized.

Table 2. Performance statistics of PLSR models.

| Spectral      | Calibration | Prediction |
|---------------|-------------|------------|
| indices       |             |            |
| R             | 32          | 0.061      | 0.896      | 10          | 10          | 0.079      | 0.751      |
| FDR           | 32          | 0.040      | 0.955      | 7           | 10          | 0.090      | 0.674      |
| BD            | 32          | 0.057      | 0.911      | 10          | 10          | 0.079      | 0.754      |
| MSC           | 32          | 0.057      | 0.909      | 10          | 10          | 0.072      | 0.793      |

RMSEC is the root mean square error of calibration, RMSEP is the root mean square error of prediction, and PCs is the number of the principal components used in the PLSR model, N corresponds to the samples used.

The PLSR regression coefficient can reflect the critical bands of predict potassium content, figure 3 shows the PLSR regression coefficient varies with wavelength. The critical bands of the prediction of potassium content are 540nm, 570nm, 840nm, and 2140-2340 nm (see figure 3). 540nm and 840nm bands are absorption feature of Fe$^{3+}$ in soil and the two regression coefficients are negative, it indicated that soil potassium content had certain correlation with iron content. 2140-2340 nm are absorption peak for clay minerals, which contains a certain amount of potassium. The aforementioned prediction mechanism was verified by the significant wavelengths of potassium content inversion.

5. Conclusions

The potassium content has a positive relation with the soil reflectance in VNIR bands, and a negative relation in VNIR bands. The maximum correlation band with potassium content appears around 2300 nm, which is related to the absorption of potassium-bearing clay minerals.

Soil potassium content can be estimated accurately from reflectance measurements at the laboratory level. Quantitative retrieval models with PLSR method had a good reliability and high precision as a smaller RMSE and a greater $R^2$. The sensitive wavelengths of potassium content are 2200-2300 nm.
Quantitative prediction of potassium content is important for further study of climate change. Hyper-spectrum is a promising tool for the quantitative inversion by combining geochemistry and remote sensing.

Acknowledgement
This research was financially supported by the National Natural Science Foundation of China (No. 41171280), the National Science and Technology Support Program (No. 2012BAH27B05, No. 2012BAK15B05) and National Basic Research Program of China (973 Program, No. 2009CB723906).

References
[1] Sun B and Xie X L 2005 Adv. Earth Sci. 20 903-909
[2] Zhao Q G 2008 Acta Pedol. Sinica 45 771-777
[3] Cong R H, Li X K and Lu J W 2007 J. Huazhong Agric. Univ. 26 907-913
[4] Ma C M, Wang F B, Cao Q Y, Xia X C, Li S F and Li X S 2008 Chin. Sci. Bull. 53 3016-3027
[5] Fu C B and Ma Z G 2008 Chin. J. Atmos. Sci. 32 752-760
[6] Liu H J, Zhang Y Z and Zhang B 2009 Environ. Monit. Assess. 154 147-154
[7] Liu W D, Baret F, Gu X F, Tong Q X, Zheng L F and Zhang B 2002 Remote Sens. Environ. 81 238-246
[8] Farifteh J, Van der Meer F, Atzberger C and Carranza E 2007 Remote Sens. Environ. 110 59-78
[9] Chang C W and Laird D A 2002 Soil Sci. 167 110-116.
[10] Xu M X, Wu S H, Zhou S L, Liao F Q, Ma C M and Zhu C 2011 J. Infrared Millim. W. 30 109-114
[11] Wu D W, Wu Y Z and Ma H R 2010 Spectrosc. Spect. Anal. 30 1498-1502
[12] Chen W N, Gao S Y and Sun Z 1994 J. Desert Res. 14 22-30
[13] Hu F, Lin Q Z, Wang Q J and Wang Y J 2012 Remote Sens. Land Resour. 24 157-162
[14] He T, Wang J, Lin Z J and Cheng Y 2006 Geomatics. Inf. Sci. Wuhan Univ. 31 975-979