# Modeling Soil Cation Concentration and Sodium Adsorption Ratio Using Observed Diffuse Reflectance Spectra

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Canadian Journal of Soil Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>CJSS-2016-0002.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>30-Mar-2016</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Xiao, Zhenzhen</td>
</tr>
<tr>
<td></td>
<td>Li, Yi; Northwest A&amp;F University, China, Agricultural water resources</td>
</tr>
<tr>
<td></td>
<td>Feng, Hao</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Soil, Soil quality, Soil chemistry</td>
</tr>
</tbody>
</table>
Modeling Soil Cation Concentration and Sodium Adsorption Ratio Using Observed Diffuse Reflectance Spectra

Zhenzhen Xiao¹, Yi Li*¹,², Hao Feng²,³

¹College of Water Resources and Architecture Engineering, Northwest A& F University, Yangling, 712100, China; ²Institute of Water-Saving Agriculture in Arid Areas of China, Northwest A& F University, Yangling, 712100, China; ³Institute of Water and Soil Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, 712100, China.

Abstract

Spectral analysis is a useful tool for the rapid and accurate prediction of soil properties. Our objective is to select the best model for predicting the three soil cation concentrations ([Na⁺], [Mg²⁺] and [Ca²⁺]) and sodium adsorption ratio (SAR). Three methods were applied, i.e., stepwise multiple linear regression (SMLR), partial least squares regression (PLSR), and support vector machine (SVM). Estimation models for four soil properties were developed using three different spectral processing and transformation techniques, i.e., reflectance (Rₑ), logarithm of reciprocal Rₑ (LR), and standard normal variable of Rₑ (SNV) were used. A total of 36 models were established. Of these, 27 models for [Na⁺], [Mg²⁺] and [Ca²⁺] were not applicable for subsequent prediction, because the coefficients of determination (R²) were not high (0.224-0.689), and their relative percent deviations (RPD) were all smaller than the 1.4 threshold. However, the models for SAR~R using PLSR (R² =0.728 for calibration and 0.661 for validation, RPD=1.43), SAR~LR using SVM (R²=0.791 for calibration and 0.712 for validation, RPD=1.81), and SAR~SNV using SVM (R²=0.878 for calibration and 0.814 for validation, RPD=2.13) were valid for further prediction. Finally,
SAR~SNV using SVM was selected as the best model. There are intrinsic factors resulting in an unsatisfied model performance.

**Key words:** soil water-soluble cation; sodium adsorption ratio; spectral reflectance; stepwise multiple linear regression; partial least squares regression; support vector regression

**Abbreviations:** Cv-coefficients of variability, LR-logarithm of reciprocal $R_e$, NIR -near infrared, ONPLF-optimum number of PLSR latent factors, PLSR-partial least squares regression, $r$-Pearson correlation coefficient, $R^2$-coefficients of determination; $R_e$-reflectance, RMSE-root mean square error, RPD-relative percent deviations, SAR-sodium adsorption ratio, SMLR-stepwise multiple linear regression, SNV-standard normal variable of $R_e$, SOM-soil organic matter, SVM-support vector machines, Vis-visible.

**Introduction**

The prediction of the physico-chemical properties of degraded soils is very important for crop growth, low quality land improvement, and environmental sustainability (Childs and Hank, 1975; Metternicht and Zinck, 2003). Of the chemical soil properties, sodium adsorption ratio (SAR), estimated using the concentration of water-soluble, soil cations including $\text{Na}^+$, $\text{Mg}^{2+}$ and $\text{Ca}^{2+}$ (from here on out, represented by $[\text{Na}^+]$, $[\text{Mg}^{2+}]$ and $[\text{Ca}^{2+}]$), is a key indicator of soil alkalinity. One way of measuring $[\text{Na}^+]$, $[\text{Mg}^{2+}]$ and $[\text{Ca}^{2+}]$ is to use atomic absorption spectrophotometry. Another way to measure $[\text{Na}^+]$ is by using flame photometry, and to measure $[\text{Ca}^{2+}]$ and $[\text{Mg}^{2+}]$ using the titration method (Tavakkoli et al., 2015). Both have advantages and disadvantages. Atomic absorption spectrophotometry has a relatively high cost, while flame photometry and the titration method are both tedious and time consuming. The spectral analysis technique is a time-saving and convenient method, with no environmental pollution. Spectral analysis has potential as a rapid screening method to quantitatively characterize soil chemical properties (Ben-Dor et al., 1995; Aldabaa et al., 2015), and offers both cost and statistical power advantages (Cohen et al., 2005). Reflectance spectroscopy analysis of ground objects is a good means of predicting soil properties, and
contributes as a potential tool for obtaining continuous and large-scale, spatial-variability features of different soil properties (Islam et al., 2003; Davies, 2005; Cambule et al., 2012).

There have been a number of studies relating soil physico-chemical properties to reflectance-related features (Mahesh et al., 2014; Kim et al., 2014). Lü et al. (2013) concluded that \([\text{SO}_4^{2-}], [\text{Ca}^{2+}], [\text{Mg}^{2+}],\) and soil salt content are the main characteristic factors of soil salinization in the Manasi River Basin in Xinjiang, China. Liu et al. (2013) analyzed the quantitative relationship between soil salt content and soil organic matter (SOM) content, and reflectance-related transformation techniques in the Beiwucha town in Xinjiang, China using the stepwise multiple linear regression (SMLR) method. Their results indicated that the selected spectral models successfully predicted soil salt content and SOM. Several other studies also focused on using reflectance spectroscopy analysis to predict soil properties (Weng et al., 2010; Fan et al., 2015; Aldabaa et al., 2015), such as electrical conductivity (Moreira et al., 2015), pH (Terra et al., 2015), and other physical-chemical properties (Li et al. 2012; Zhang et al., 2016).

Since reflectance data are numerous and complex, the modeling approaches are crucial to achieve an accurate estimate of soil properties. Zhang and Li (2016) determined the most accurate model for predicting SOM against reflectance-related transformation techniques using the SMLR and the partial least square regression (PLSR) methods. Cohen et al. (2005) applied regression trees to model post-processed reflectance spectra with laboratory observations (pH, SOM, cation concentrations, extra-cellular enzyme activity, total N, C, and P). Soriano-Disla et al. (2014) applied PLSR to predict 31 soil properties (including \([\text{Na}^+], [\text{Mg}^{2+}]\) and \([\text{Ca}^{2+}]\)), determined by X-ray florescence in the mid-infrared spectral region. Principal component analysis was also used to simultaneously estimate several soil properties with acceptable accuracy (Chang et al., 2001). A variety of multivariate techniques, including multiple linear regression, PLSR, multivariate adaptive regression splines, support vector machines (SVM), random forests, boosted trees, and artificial neural networks were compared to estimate soil organic carbon, clay content, and pH.
(Viscarra-Rossel and Behrens, 2010). Of the different methods, PLSR has been the most popular. However, other multivariate methods accounting for nonlinear responses such as SVM (Zhai et al., 2012; Bao et al., 2014), wavelets, and artificial neural networks may perform better (Janik et al., 2009; Viscarra-Rossel and Lark, 2009; Viscarra-Rossel and Behrens, 2010; Viscarra Rossel and Webster, 2012). SVM is a machine-learning-algorithm that is extensively applied in nonlinear systems. Its approach is to map nonlinearity of samples to a high-dimensional feature space, and to seek a linear regression function in that high-dimensional space (Amari and Wu, 1999). It utilizes the information from a limited number of samples, and seeks the optimal compromise between the complexity and predictive power of the models. It is also good in performance generalization. Some advantages of SVM are solving problems with a small number of samples, nonlinearity, and pattern recognition in a high dimensional data space (Belousov et al., 2002; Chapelle et al., 1999). Due to the obtained models being regionally specific, the models need to be carefully calibrated and validated before they are applied to other regions.

Soil SAR is a comprehensive index containing respective contributions of [Na$^+$], [Mg$^{2+}$] and [Ca$^{2+}$]. In previous research, models of SAR based on diffuse reflectance spectroscopy analysis were limited. It was not clear whether the various approaches such as SMLR, SVM, and PLSR could accurately predict [Na$^+$], [Mg$^{2+}$], [Ca$^{2+}$], and SAR. It was also unknown whether the models would successfully perform when the different reflectance transformation techniques were applied to [Na$^+$], [Mg$^{2+}$], [Ca$^{2+}$], and SAR. The goals of this study are: i) to measure [Na$^+$], [Mg$^{2+}$], [Ca$^{2+}$], and SAR of the collected field soil samples from the Manasi River Basin, and to calculate their SAR values. In addition, this study aims to measure diffuse reflectance spectra in the visible (Vis) and the near infrared (NIR) (400-2500 nm) bands. Two spectral transformation techniques, including logarithm of reciprocal (LR=log (1/$R_e$), which was absorbance transformation) and standard normal variable (SNV), were applied to the raw observed reflectance ($R_e$) data, ii) to establish models of [Na$^+$], [Mg$^{2+}$], [Ca$^{2+}$], and SAR against the three transformation techniques (i.e.,
R, LR, and SNV), using three modeling methods (SMLR, PLSR, and SVM), and iii) to compare the performance of the models in determining the most appropriate model, and to find a proper method for predicting the studied soil properties. Factors that affect model performance will be discussed. This research provides a good reference for applying the spectral spectroscopy analysis technique in the prediction of soil SAR.

2.0 Materials and Methods

2.1 Field collection of soil samples and laboratory measurement of properties

The Anjihai irrigation district is a mid-west, economic development zone with an area of 7.84 km$^2$. It is located in the north-Tianshan Mountains, southwest of the Manasi River Basin in Xinjiang Uygur Autonomous Region, China. The district is located downstream of the Manasi river basin, where groundwater levels are shallow and highly saline. The annual mean precipitation is low (119 mm), but the potential evaporation is 950 mm, with an aridity index of 0.13 (Li et al., 2010). The district has a typical temperate, continental, arid climate (Zhao et al., 2015). Low quality, saline-alkali land accounts for 24.3% of the total arable land area in the region, which restricts the development of high-efficiency agriculture. Due to this, good crop management is important for producing successful yields. In the last 15 years, a saline-alkali demonstration project has been conducted in the region, and has achieved good results in improving soil quality.

The soils in the region were characterized as typical grey desert soils (Luo, 1985). The samples were collected from the 22nd of April to the 15th of May, 2013. Plant residues present on the soil surface were carefully removed before sampling. A total of 78 samples were randomly collected from within a 2 to 3 km range. The sampling depth was from 0 – 10 cm, and the sampling location is shown in Figure 1.

Each soil sample was put into a cloth bag and transported to the laboratory. All of the soil samples were air-dried, ground, and passed through a 2 mm sieve. The soil moisture of all the
samples following being air dried, was approximately 0.03 cm$^3$ cm$^{-3}$. Because there was very low soil water content in each sample, water content had very little effect on the results of spectral reflectance. SOM was measured using the potassium dichromate-dilution heat colorimetric method (Bao, 2000). The mean value of SOM was found to be 11.4 g kg$^{-1}$. The pipette method was used to measure particle size distribution of the samples. According to the USDA classification system, there were 46 loam soils, 12 sandy loam soils, 10 clay loam soils, 4 loamy sand soils, 3 silt clay loam soils, 1 sand, 1 sandy clay loam soil, and 1 clay soil.

The soil-water ratio was brought to 1:5 for all the crushed soil samples in preparation for the leaching liquor (Palacios-Orueta and Ustin, 1996). Each glass bottle containing the soil-water mixture was put on an oscillator for 30 minutes to thoroughly dissolve the soil salts. The bottles were then let to be settled for 24 hours, to obtain a clarified solution (Farifteh, 2008). Soil [Na$^+$], [Mg$^{2+}$], and [Ca$^{2+}$] were measured using an AA320N-Atomic Absorption Spectrophotometer (Shanghai Yanhe Instrument Co. Ltd.). SAR was calculated by (Hasheminejhad et al., 2013):

$$SAR = \frac{[Na^+]}{\sqrt{([Ca^{2+}]+[Mg^{2+}])/2}}$$

where the unit for SAR is (mmol L$^{-1}$)$^{1/2}$; and units of [Na$^+$], [Ca$^{2+}$], and [Mg$^{2+}$] are mmol L$^{-1}$.

### 2.2 Laboratory spectra reflectance measurement

Each sample was evenly placed in an opaque, aluminum container with a depth of 2 cm. Reflectance was observed outdoors at noon on a cloudless day, using a FieldSpec portable spectrometer (Analytical Spectral Devices Inc., Boulder, Co.). The wavelength range is 350-1830 nm, which covers the visible (350-780 nm) and near infrared (780-2500 nm) (Vis-NIR) bands. Sampling intervals are 1.4 and 2 nm corresponding to Vis-NIR bands respectively. 10 repetitive measurements were performed for each sample. The averages of the 10 observed reflectance curves for each soil sample were obtained, and used for further study.

### 2.3 Data pre-processing
The strong atmospheric vapor absorption affected the measured $R_e$ at 1350-1420 nm and 1800-1830 nm bands. $R_e$ values within these two wavelength ranges were excluded from the observed $R_e$ curve. Two spectral transformation techniques, including LR and SNV, were applied to the $R_e$ data (Buddenbaum and Steffens, 2012). LR is used to reduce the effects of noise, light, and background on $R_e$. SNV is adopted to eliminate the effects of solid particle size and optical path change of surface scattering on $R_e$. SNV is calculated by (Barnes et al., 1989):

$$SNV_i = \frac{R_{e,i} - \mu}{\sigma}$$  \hspace{1cm} \text{(2)}

where $\mu$ is the mean value of $R_e$, and $\sigma$ is the standard deviation of $R_e$.

2.4 Model establishment and validation

Soil samples were randomly divided into a calibration set for model establishment, and a validation set, each consisting of 58 and 20 samples, respectively, of the total 78 samples. The assignment of samples was random, but could be artifacts of this separation and affect model performance. With only 78 samples, there are potential difficulties in achieving good models, especially if the response variable range is limited. The four soil properties ([Na$^+$], [Mg$^{2+}$], [Ca$^{2+}$], and SAR) were selected as the dependent variables. The three reflectance-related transformation techniques, i.e., $R_e$, LR, and SNV, were selected as independent variables for the modeling. The modeling methods of SMLR, PLSR, and SVM were conducted. SMLR and PLSR are linear, and SVM is nonlinear. The optimum number of PLSR latent factors (ONPLF) is a rotation of the principal components for a slightly different optimization criterion (Aldabaa et al., 2015). ONPLF was selected on the basis of the number of factors with the smallest total residual validation Y-variance, or highest total explained validation Y-variance. The Epsilon-SVM is selected for the SVM type, and the linear function is selected as its kernel function.

The model establishment using the SMLR method was performed using SPSS 17.0 software, and the validation was performed in Microsoft Excel 2007. The software Unscrambler X 10.3 (CAMO AS Oslo, Norway) was utilized for modeling and validating the PLSR and SVM methods.
The leave-one-out cross-validation was chosen as the validation method for the PLSR and SVM methods. The main procedure for this method is to select a sample from the total 78 samples, and to establish PLSR (or SVM) models using the remaining 77 samples, and then predict the leave-out sample using the established models. This procedure was repeated for each sample (78 times). Each sample runs 77 times for model establishment and 1 time for model validation, resulting in 78 predicted values (Liu, 2013).

The subscript ‘C’ was chosen to distinguish the calibration processes for model establishment, and ‘V’ for the validation processes. Model performance was evaluated by comparing root mean square error (RMSE) and coefficient of determination ($R^2$). The smaller the RMSE and the greater the $R^2$, the more accurate the model is. The equations for $R^2$ and RMSE are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$  \hspace{1cm} (3)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i')^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$  \hspace{1cm} (4)

where $y_i$ and $\bar{y}_i'$ are the measured and predicted soil properties respectively, $\bar{y}$ and $\bar{y}'$ are the mean values of measured and predicted soil properties respectively, and $n$ and $n'$ are the total number of measured and predicted soil properties, respectively.

In addition, the relative percent deviation (RPD) was calculated to judge the predictive ability of the established models. When RPD is < 1.4, the predictive ability of the established model is very poor. When RPD is within the range of 1.4 and 2, the established model is applicable for coarse predictions. When RPD is > 2, the predictive ability of the established model is considered excellent (Chang et al., 2001). The RPD is estimated by:
2 Results

2.1 Variations and descriptive statistics of the soil chemical properties

The variations of the four studied soil chemical properties and the corresponding elevations at the sampling sites are illustrated in Figure 2. Variations in soil $[\text{Na}^+]$, $[\text{Mg}^{2+}]$, $[\text{Ca}^{2+}]$, and SAR were random, with some abnormally large values, which may affect the accuracy of the subsequent modeling. The elevations had small influences on the four soil chemical properties, with the $R^2$ values being smaller than 0.15. However, all the relationships were consistent in being negatively correlated, indicating that soil chemical properties were found in higher concentrations at low elevations.

Descriptive statistical characteristics of $[\text{Na}^+]$, $[\text{Mg}^{2+}]$, $[\text{Ca}^{2+}]$, and SAR are shown in Table 1. The skewness of a normally distributed series is 0, and its kurtosis is 3. As skewness is $>0$ (or $<0$), the series are positively (or negatively) skewed. Distribution with serial kurtosis $<3$ means the distribution is platykurtic, while kurtosis $>3$ means the distribution is leptokurtic. The descriptive statistical characteristics of $[\text{Na}^+]$, $[\text{Mg}^{2+}]$, $[\text{Ca}^{2+}]$ and SAR were generally consistent for the total, calibration, and validation subsets. Although small differences of statistical features were found between the validation, the calibration and the total sets, these would bring some slight errors in modeling processes. For the total set, among the three cation concentration properties ($[\text{Na}^+]$, $[\text{Mg}^{2+}]$ and $[\text{Ca}^{2+}]$), the maximum (28.85 mmol L$^{-1}$), average (4.75 mmol L$^{-1}$), and range (28.66 mmol L$^{-1}$) of $[\text{Na}^+]$ were the highest, but the standard deviations of $[\text{Mg}^{2+}]$ (0.94) were the smallest. The statistical features of soil properties varied according to different soil properties. Therefore, $\text{Na}^+$ was
the most abundant soil cation in the Anjihai irrigated district, which was different with Lü et al. (2013) for the same region, who found that $\text{Ca}^{2+}$ was the main positive cation. This was due to the different sampling area, depths and interval between this research and theirs. SAR values ranged from 0.23 - 11.83 (mmol·L$^{-1}$)$^{1/2}$. The skewness values of $[\text{Na}^+]$, $[\text{Mg}^{2+}]$, $[\text{Ca}^{2+}]$, and SAR were all positive, and were ranked as follows: $[\text{Mg}^{2+}] > [\text{Na}^+] > [\text{Ca}^{2+}] >$ SAR. $[\text{Na}^+]$ and $[\text{Mg}^{2+}]$ were both leptokurtic, as their kurtosis values were $> 3$, and $[\text{Ca}^{2+}]$ and SAR were both platykurtic, with kurtosis values $< 3$. The large skewness and kurtosis in the distribution of the soil properties, and the differences caused by the division of the subsets, may affect the accuracy of the subsequent modeling. The coefficients of variability (Cv) were all larger than 1, indicating a high variability in $[\text{Na}^+]$, $[\text{Mg}^{2+}]$, $[\text{Ca}^{2+}]$, and SAR (Nielsen, 1985). This could have a possible influence in decreasing the accuracy of the subsequent modeling.

2.2 The observed reflectance curves

Figure 3 shows the observed spectral reflectance $R_e$ for twelve typical soil samples, with different $[\text{Na}^+]$, $[\text{Mg}^{2+}]$, $[\text{Ca}^{2+}]$, and SAR values. Generally, the $R_e$ curves of different soil samples varied, but had similar patterns, i.e., as the wavelength increased, $R_e$ also increased. This increase was larger in the wavelength range of 340 - 590 nm, than in the range of 591 - 1810 nm, and there was generally a small valley at the wavelength of 1421 nm. The $R_e$ curve didn’t increase or decrease monotonically with the changes of any soil property, which may reduce accuracy of the subsequent modeling.

Figure 4 shows the Pearson correlation coefficient ($r$) between the measured soil properties and $R_e$ at various wavelengths. The correlations relating cation concentrations and SAR to $R_e$, LR and SNV were generally low (-0.29 < $r$ < 0.32 for $[\text{Na}^+]$, -0.35 < $r$ < 0.40 for $[\text{Ca}^{2+}]$, -0.30 < $r$ < 0.22 for $[\text{Mg}^{2+}]$, and -0.32 < $r$ < 0.43 for SAR). Variations in $r$ for $R_e$ and LR were generally opposite in signs but similar in values, showing the transformation effects of LR. Values of $r$ for
SAR had larger ranges. Curves of $r$ for relating soil properties to SNV fluctuated more than Curves relating to $R_c$ and LR. There were obviously large absolute values of $r$ for soil properties against SNV at wavelengths around 1121, 1284, 1422, and 1545 nm, this might be effective in improving modeling performance, especially for SAR against SNV.

2.3 Establishment and validation of models for $\left[\text{Na}^+\right]$, $\left[\text{Mg}^{2+}\right]$ and $\left[\text{Ca}^{2+}\right]$

The models’ performance varied between the four studied soil properties, and the three transformed $R_e$-related transformation techniques. The calibration and the validation results of different models relating $\left[\text{Na}^+\right]$, $\left[\text{Mg}^{2+}\right]$ and $\left[\text{Ca}^{2+}\right]$, and 3 spectral transformation techniques are shown in Table 2.

Table 2 shows that: (1) For the spectral models of $\left[\text{Na}^+\right]$, when using the SMLR method, $R^2_c$, $R^2_v$, and RPD values of the $\left[\text{Na}^+\right]$~$R_e$ model were 0.471, 0.442, and 0.76 respectively, which were slightly better than the $\left[\text{Na}^+\right]$~LR and $\left[\text{Na}^+\right]$~SNV models. Also, when using the PLSR method, $R^2_c$, $R^2_v$, and RPD values of $\left[\text{Na}^+\right]$~$R_e$ model were 0.585, 0.481 and 0.83 respectively, which were also larger than the $\left[\text{Na}^+\right]$~LR and $\left[\text{Na}^+\right]$~SNV models. Furthermore, when using the SVM method, $R^2_c$, $R^2_v$, and RPD values of the $\left[\text{Na}^+\right]$~SNV model were much larger than the $\left[\text{Na}^+\right]$~$R_e$ and $\left[\text{Na}^+\right]$~LR models, but performance of the models $\left[\text{Na}^+\right]$~$R_e$ and $\left[\text{Na}^+\right]$~LR was very poor. (2) When comparing the models of $\left[\text{Na}^+\right]$ against various spectral transformation techniques, the best method was the PLSR with generally high $R^2_c$, low $\text{RMSE}_c$, and higher RPD values. While for the models of $\left[\text{Mg}^{2+}\right]$ (or $\left[\text{Ca}^{2+}\right]$) against different spectral transformation techniques, the SMLR method gave the best performance. (3) In all 27 models, $R^2_c$, $R^2_v$, and RPD values of the 3 models using the SNV spectral index, i.e., $\left[\text{Na}^+\right]$ against SNV, $\left[\text{Mg}^{2+}\right]$ against SNV, and $\left[\text{Ca}^{2+}\right]$ against SNV, with the SVM method were the largest, while their RMSE values were small. This showed the absolute advantage of the SVM method, compared to the other two methods in model performance. (4) The optimum number of PLSR latent factors in the PLSR method was less than 7 for different $R_e$ transformation
techniques and various soil cation concentrations. (5) Unfortunately, although there were
differences in the performance of the established various models, which performed generally well,
their RPD values were all smaller than 1.4. This indicated that none of them had good predictive
ability, which limited their further application for the estimation of cation concentrations using
spectral analysis.

2.4 Establishment and validation of soil SAR models

Because the models had poor predictive ability for $[Na^+]$, $[Mg^{2+}]$ and $[Ca^{2+}]$, the performance
of the models for SAR should also be assessed, following similar modeling approaches to $[Na^+]$,$[Mg^{2+}]$ and $[Ca^{2+}]$. The calibration and the validation results of various established models relating
SAR to 3 spectral transformation techniques are presented in Table 3.

There were differences among model statistics including $R^2_c$, $RMSE_c$, $R^2_v$, $RMSE_v$, and RPD,
using the 3 regression methods for SAR. However, from the view of application in SAR prediction,
only two models, SAR against $R_c$ using the PLSR method and SAR~LR using the SVM method,
had rough predictive abilities, with RPD values > 1.4 and < 2.0. Their performance was generally
good, with large $R^2_c$ and $R^2_v$ values and small $RMSE_c$ and $RMSE_v$ values. The SAR~SNV model
using the SVM method, had good predictive ability with a RPD value > 2.0. The $R^2_c$ and $R^2_v$ values
were the largest, and the $RMSE_c$ and $RMSE_v$ values were the smallest, among all of the SAR models.
This model performed the best when compared to the other 35 models, including those in Table 3.
SAR~SNV was primarily chosen as the best model for predicting the studied soil properties, using
the spectral analysis technique. Performance of all the SAR models should be further validated by
comparing the observed and predicted SAR values, using each model. It is necessary to emphasize
that, the model performance assessment indicator RPD defined the ability of established models
very well.

Figure 5 shows the scatter plots of the observed and predicted SAR values in each model.
Using a step-by-step procedure, the most applicable model, i.e., SAR against SNV using the SVM
method was selected. It was more applicable to measure SAR using the spectral analysis technique, than to measure single cation concentrations. SAR was considered a more appropriate property to determine alkalization, than the single cation concentration of either [Na\(^+\)], [Mg\(^{2+}\)] and [Ca\(^{2+}\)] in the studied region. This conclusion has not been reported in previous research (Ben-Dor et al., 1995; Lü et al., 2013; Aldabaa et al., 2015). SAR was also found to be a good indicator for large-scale investigation of low-quality land areas, using the spectral and remote sensing techniques.

3 Discussion

3.1 The types of factors that affect model performance

It was shown, not only from previous research but also from this research, that there are many factors affecting model performance. These factors generally include two types, the first being intrinsic (physical, chemical, and biological) properties. These properties include soil composition (soil particle content, soil salt content, cation concentration, SOM, mineral elements, etc.); the spatial variability of the data; skewness and kurtosis; the variation pattern of soil properties against R, (increase or decrease with wavelength, correlations between them, whether or not there were abrupt changes in the curves), this factor is somewhat artificial as it is based on the skill of the observer; and the other uncertain factors, such as whether samples are undisturbed or disturbed, topography (including elevations, slope direction, and slope position), etc. The second type of factors is outer one and is not directly associated with the soil. It includes the approaches or methods applied for modeling, the data pre-processing (or transformation) techniques of R, the band selected for modeling, and the proficiency of the model operator. The natural soil properties (or the first type of factors) couldn’t be changed, but can be best characterized by successfully selecting the outer type of factors. A detailed discussion follows, on the two types of factors that affect model performance.

3.2 Factors of the first type (intrinsic) that affect model performance
Modeling performance varied when the studied soil properties were different. It is difficult to assess whether any of the physical, chemical, or biological properties were modeled better than the others, because soils are versatile and heterogeneous in nature. Even in studying the same soil property, the accuracy varied. A number of previous studies have focused on the spectral response characteristics of SOM, soil salt content, soil electrical conductivity, and soil particle sizes (McCarty and Reeves, 2002; Mashimbye et al., 2012; Nocita et al., 2014; Pang et al., 2014). Dunn et al. (2002) demonstrated that NIR can successfully determine exchangeable Ca and Mg, with a high level of accuracy. Chang et al. (2001) found that extractable Ca was successfully predicted by NIR, and extractable Mg and exchangeable Ca could also be predicted by NIR, but with less accuracy. In addition, Chang et al. (2001) found that exchangeable Na could not be predicted using the NIR and principal component analysis technique. Soriano-Disla et al. (2014) reviewed the performance of Vis, NIR, and mid-infrared reflectance spectroscopy for the prediction of different soil properties. Their table 3 presents median $R^2$ values for the validation results in different bands including Vis-NIR by comparing 90 literatures. They showed that median $R^2$ values for exchangeable Ca was 0.31 in Vis, 0.75 in NIR, and 0.80 in Vis-NIR; median $R^2$ values for total Ca was 0.78 in Vis-NIR; median $R^2$ values for exchangeable and total Mg was 0.83 and 0.61 respectively, in Vis-NIR; and median $R^2$ values for exchangeable and total Na were 0.61 and 0.81 respectively, in Vis-NIR. The RPD values were generally larger than 1.5, these were much different from our results. This research showed that all of the 27 established linear and nonlinear models for $[\text{Na}^+]$, $[\text{Mg}^{2+}]$ and $[\text{Ca}^{2+}]$, against the 3 $R_e$-related transformation techniques, had poor predictive ability, while 3 models of SAR showed better predictions. These are relatively new results, compared to previous research.

The complexity of soils that contains different physical, chemical, and biological compositions affects modeling performance. Viscarra-Rossel and Behrens (2010) pointed out that soil spectra are largely non-specific, quite weak, and broad, due to overlapping absorptions of soil constituents and their often small concentrations in soil. Soils with different soil textures have different spectra, due
to their range of moisture-holding capacity and particle size, which affects the accuracy of Vis and NIR spectroscopy. Viscarra-Rossel et al. (2006) found that more accurate predictions can be obtained for clay silt and sand content, using the mid-infrared region instead of the Vis-NIR. A limited number of samples, or samples from a limited number of sites of similar soil types, scoped the calibration to a certain type of soil, making it applicable only to the landscape being studied. In this research, field soils contained various contents of sand, silt, and clay (Figure 6), SOM, [Na$^+$], [Mg$^{2+}$], [Ca$^{2+}$], SAR, and other minerals that were not measured. All of these compositions affected $R_e$ to a different extent, which complicated the modeling results and increased the modeling uncertainty. Although the correlations of [Na$^+$], [Mg$^{2+}$], [Ca$^{2+}$], and SAR with particle content were generally low (Figure 6), it was consistently shown that the increases in clay and silt content, as well as the decreases in sand content, increased the values of [Na$^+$], [Mg$^{2+}$], [Ca$^{2+}$] and SAR. The clay content varied between 6.3-42.3%, the silt content varied between 10.3-51.2%, and the sand content varied between 12.6-83.4%. The large variability in particle size distribution increased the uncertainty of model performance to a great extent. The potential influences of soil texture on soil properties could also distract the correlations of soil properties to some extent.

Furthermore, the correlations between soil properties and $R_e$ at various wavelengths, affect modeling performance. The generally low correlations between the four studied soil chemical properties and $R_e$ (Figure 4), determined that the model has low efficacy. The generally inverse correlations between sand and silt against $R$ showed that the different soil particles had different effects on $R_e$. The simultaneous changes of various properties of soil composition decreased model performance in determining each soil property to some extent. This partially explains why there were unsatisfied modeling performances in predicting [Na$^+$], [Mg$^{2+}$] and [Ca$^{2+}$] in this research, even though different reflectance transformation techniques and modeling methods were applied.

Soil spectroscopy information can comprehensively reflect the soil properties. In this research, the pattern of $R_e$ with the wavelengths for [Na$^+$], [Ca$^{2+}$], [Mg$^{2+}$], and SAR mildly increased (Figure...
4), and there were no abrupt changes of \( R_e \) in any band range. The observed \( R_e \) curves of this research were very different from \( R_e \) of soil moisture (Whiting et al., 2004; Li et al., 2012), of Fe oxides (Scheinost, 1998), of soil mineral composition (Viscarra-Rossel and Behrens, 2010), and of heavy metals (Wu et al., 2007), but similar to the \( R_e \) curves for soil chemical properties (such as Ca, Mg, Fe, Mn, and K) in the same wavelength bands (Udelhoven et al., 2003), however, they obtained \( R^2 \) values as high as 0.91 and 0.94 for predictions of Ca and Mg respectively. It seemed that their spectra pre-treatment of minimum minus maximum normalization helped in improving modeling performance.

The distribution of the data in space, their variability, and descriptive statistics of skewness and kurtosis may also affect model performance. Although all of \([\text{Na}^+]\), \([\text{Ca}^{2+}]\), \([\text{Mg}^{2+}]\), and SAR had skewed distributions, the distribution of SAR had smaller skewness and kurtosis than the three cations. This may have contributed to the better performance of a few models for SAR than for \([\text{Na}^-]\), \([\text{Mg}^{2+}]\), and \([\text{Ca}^{2+}]\).

Additionally, the topography or elevations from where the samples were taken, would affect the soil properties, although not to a great extent (Figure 2). Other factors such as surface roughness may also affect soil properties (Soriano-Disla et al., 2014). Soil moisture should not affect \( R_e \) much in this research, because our samples were all air-dried, and the soil moisture in each sample was very low (around 0.03 cm\(^3\) cm\(^{-3}\)).

### 3.3 Factors of the second (outer) type that affect model performance

The second type of factors manifest artificially, leading to the selected modeling methods or approaches and having a large effect on the model performance. There are a variety of modeling methods such as PLSR, SMLR, SVM, principal component analysis, artificial neural networks (Daniel et al., 2003), multivariate adaptive regression splines (Friedman, 1991), PLSR with bootstrap aggregation (Viscarra-Rossel, 2007), and so on. In addition to the different linear methods (e.g., SMLR and PLSR), the nonlinear methods such as SVM and artificial neural network, are the most accurate models for estimating soil properties (Wang et al., 2014). SVM eliminated the bias of
scanned spectral curve, which may cause non-linear changes in $R_e$ (Chauchard et al., 2004).

Predictions of SOM, clay content, and pH using SVM with all Vis-NIR wavelengths, produced the smallest RMSE values (Viscarra-Rossel and Behrens, 2010). In this research, SVM also showed better performance than the SMLR and PLSR methods. The best model was discovered, and the most appropriate regression method was SVM, the most appropriate soil property was SAR, and the most appropriate spectral index was SNV. SVM showed its power to describe nonlinear relationships among soil properties of SAR that were generated from $[\text{Na}^+]$, $[\text{Mg}^{2+}]$ and $[\text{Ca}^{2+}]$, and the spectral transformation techniques. These results were found to be reasonable and are useful in the application of the soil property model, and in further investigation of low-quality soil control and management.

The transformation techniques of $R_e$ affect model performance. Based on the originally observed Vis-NIR $R_e$ data, Liu et al. (2013) applied 5 transformation techniques to $R_e$ including continuum-removed $R_e$, reciprocal values of $R_e$, logarithm of reciprocal values of $R_e$, the first order derivative of $R_e$, and the second order derivative of $R_e$, and performed SMLR for soil salt content and SOM. Their results showed that the model relating the continuum-removed $R_e$ to soil salt content, had the largest $r$ (0.76) and the smallest $RMSE$ (2.94) compared to the other 4 transformation techniques, and was selected as the best model. Becker et al. (2005) found that the 2nd-derivative technique was powerful for differentiating coastal wetland vegetation in the Great Lakes. In this research, transformation of SNV had the best predictive ability for SAR, followed by LR and $R_e$. All showed that the transformed $R_e$ index was better than the original $R_e$ in improving model performance. However, it is still difficult to determine which transformation technique was the best.

The bands that were selected for modeling also contribute to the modeling performance. The most commonly used bands for $[\text{Na}^+]$, $[\text{Mg}^{2+}]$, and $[\text{Ca}^{2+}]$ were mid-infrared region, NIR, and Vis-NIR, but the modeling performance was property-specific for different bands (Soriano-Disla et al., 2014). Models using the characteristic absorption band also performed well in the prediction of...
SOM and soil salt content (Rial et al., 2015; Liu et al., 2013). In this research, reflectance data in the effective spectral bands of Vis-NIR were used. Further study is needed to determine how the changes of band range affect accuracy of predicting $[\text{Na}^+]$, $[\text{Mg}^{2+}]$, $[\text{Ca}^{2+}]$, and SAR.

Moreover, operational proficiency during observation and modeling processes may also affect modeling accuracy. This research focused on field collected soil samples. The properties of these soil samples showed strong spatial variability with all $C_v > 1$ (Nielsen, 1985), decreasing the modeling performance. Model performance was excellent for the artificially configured soils ($R^2 > 0.99$ for the best models relating soil water content to normalized band depth for their four different soils), in which only soil property of water content changed (Li et al., 2012).

Nevertheless, factors that influence model performance vary. More research is need in the future to assess model performance comprehensively.

### 4. Conclusions

There ranges of descriptive statistics for the measured soil properties such as $[\text{Na}^+]$, $[\text{Mg}^{2+}]$, $[\text{Ca}^{2+}]$, and SAR, were different. The studied soil properties showed strong spatial variability and skewed distributions, which may have caused the generally unsatisfied modeling of soil $[\text{Na}^+]$, $[\text{Mg}^{2+}]$ and $[\text{Ca}^{2+}]$. There were generally low correlations between soil properties (including sand content, silt content, clay content, $[\text{Na}^+]$, $[\text{Mg}^{2+}]$, $[\text{Ca}^{2+}]$, and SAR) and $R_e$ at various wavelengths, which may have also contributed to the low modeling accuracy. All of these factors were intrinsic and were classified as the first type of influencing factors.

The 27 spectral models of soil $[\text{Na}^+]$, $[\text{Mg}^{2+}]$ and $[\text{Ca}^{2+}]$ using the SMLR, PLSR, and SVM methods, were established. $R^2$ ranged from 0.269-0.689 for $[\text{Na}^+]$, from 0.233-0.645 for $[\text{Ca}^{2+}]$, and from 0.224-0.622 for $[\text{Mg}^{2+}]$. The $R^2_e$ ranged from 0.103 to 0.573. $RMSE_e$ and $RMSE_V$ values were all smaller than 1.115. The RPD values of all models were smaller than the threshold value of 1.4, indicating no model was capable of further prediction. The 9 spectral models of SAR showed differences in performance, of which 6 models, including 3 that used the SMLR method, were not effective due to RPDs < 1.4. The model using the PLSR method and the 2 models using the SVM
method had predictive ability (RPDs > 1.4). This showed that SVM improved predictive ability for SAR but PLSR and SMLR didn’t.

Model performance was also affected by the different transformation techniques of $R_e$. Among the 3 SAR models which were acceptable for predictions, the SNV transformation technique was the best (RPD=2.13), followed by the LR (RPD=1.81) and $R_e$ (RPD=1.43). By visually comparing the scatter plots of the observed and predicted SAR values, the SAR–SNV model using the SVM method, was the best model ($R^2=0.877$). Thus, outer factors including modeling methods and transformation techniques, also contributed to model performance.

Therefore, both intrinsic and outer factors contributed to the poor modeling performance. The best model of SAR could be potentially extended to a larger scale investigation in the future. It can also be applied in the field soils that contains <50% of clay and silt contents.

Acknowledgements

This study is financially supported by China National Science Foundation (51579213), China Scholarship Council for Studying Abroad (201506305014), and the China 111 project (B12007). The constructive comments of two anonymous reviewers strengthened the analyses of this paper.

References


infrared reflectance spectroscopy: comparison of partial least-square regression and support

Zhang P, Li Y. 2016. Comparisons of model establishment for soil organic matter content using two

609–618.
Table 1 Summary of descriptive statistics for the studied four soil properties of the soils samples.

<table>
<thead>
<tr>
<th>Property</th>
<th>Sub-set</th>
<th>Max</th>
<th>Min</th>
<th>Mean</th>
<th>Range</th>
<th>Standard deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Cv&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Na&lt;sup&gt;+&lt;/sup&gt;]/mmol·L&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>Total</td>
<td>28.85</td>
<td>0.18</td>
<td>4.75</td>
<td>28.66</td>
<td>6.60</td>
<td>1.99</td>
<td>3.68</td>
<td>1.39</td>
</tr>
<tr>
<td></td>
<td>Calibration</td>
<td>28.85</td>
<td>0.28</td>
<td>4.80</td>
<td>28.57</td>
<td>6.78</td>
<td>2.04</td>
<td>3.84</td>
<td>1.41</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>25.04</td>
<td>0.18</td>
<td>5.16</td>
<td>24.86</td>
<td>6.65</td>
<td>1.75</td>
<td>3.19</td>
<td>1.29</td>
</tr>
<tr>
<td>[Ca&lt;sup&gt;2+&lt;/sup&gt;]/mmol·L&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>Total</td>
<td>20.83</td>
<td>0.21</td>
<td>3.78</td>
<td>20.62</td>
<td>5.10</td>
<td>1.79</td>
<td>2.38</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>Calibration</td>
<td>20.83</td>
<td>0.21</td>
<td>4.00</td>
<td>20.62</td>
<td>5.08</td>
<td>1.50</td>
<td>1.33</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>20.53</td>
<td>0.23</td>
<td>3.90</td>
<td>20.30</td>
<td>6.13</td>
<td>1.96</td>
<td>2.70</td>
<td>1.58</td>
</tr>
<tr>
<td>[Mg&lt;sup&gt;2+&lt;/sup&gt;]/mmol·L&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>Total</td>
<td>5.47</td>
<td>0.10</td>
<td>0.82</td>
<td>5.37</td>
<td>0.94</td>
<td>2.45</td>
<td>8.08</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>Calibration</td>
<td>5.47</td>
<td>0.10</td>
<td>0.84</td>
<td>5.34</td>
<td>0.95</td>
<td>2.49</td>
<td>9.19</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>3.95</td>
<td>0.10</td>
<td>0.75</td>
<td>3.84</td>
<td>0.92</td>
<td>2.67</td>
<td>8.00</td>
<td>1.23</td>
</tr>
<tr>
<td>SAR/</td>
<td>Total</td>
<td>11.83</td>
<td>0.23</td>
<td>2.99</td>
<td>13.10</td>
<td>3.00</td>
<td>1.50</td>
<td>1.73</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Calibration</td>
<td>11.83</td>
<td>0.23</td>
<td>2.94</td>
<td>11.60</td>
<td>3.00</td>
<td>1.38</td>
<td>0.96</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>10.45</td>
<td>0.43</td>
<td>2.94</td>
<td>10.02</td>
<td>3.07</td>
<td>1.67</td>
<td>2.69</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Note: <sup>a</sup>Cv–coefficient of variability.
Table 2 Calibration and validation statistics of the established models describing cation concentration with spectral indicators.

<table>
<thead>
<tr>
<th>Cation concentration</th>
<th>Method</th>
<th>Spectral index</th>
<th>ONPLF&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Calibration</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$R_c^2$</td>
<td>$RMSE_c$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$R_v^2$</td>
<td>$RMSE_v$</td>
</tr>
<tr>
<td>Na&lt;sup&gt;+&lt;/sup&gt;</td>
<td>SMLR</td>
<td>$R_e$</td>
<td>0.471</td>
<td>0.679</td>
<td>0.442</td>
</tr>
<tr>
<td></td>
<td>SMLR</td>
<td>LR</td>
<td>0.467</td>
<td>0.690</td>
<td>0.447</td>
</tr>
<tr>
<td></td>
<td>SMLR</td>
<td>SNV</td>
<td>0.434</td>
<td>0.782</td>
<td>0.380</td>
</tr>
<tr>
<td></td>
<td>PLSR</td>
<td>$R_e$</td>
<td>0.585</td>
<td>0.624</td>
<td>0.481</td>
</tr>
<tr>
<td></td>
<td>PLSR</td>
<td>LR</td>
<td>0.553</td>
<td>0.614</td>
<td>0.432</td>
</tr>
<tr>
<td></td>
<td>PLSR</td>
<td>SNV</td>
<td>0.491</td>
<td>0.637</td>
<td>0.410</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>$R_e$</td>
<td>0.269</td>
<td>0.637</td>
<td>0.213</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>LR</td>
<td>0.301</td>
<td>0.638</td>
<td>0.242</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>SNV</td>
<td>0.689</td>
<td>0.451</td>
<td>0.573</td>
</tr>
<tr>
<td>Ca&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>SMLR</td>
<td>$R_e$</td>
<td>0.572</td>
<td>1.068</td>
<td>0.541</td>
</tr>
<tr>
<td></td>
<td>SMLR</td>
<td>LR</td>
<td>0.560</td>
<td>1.115</td>
<td>0.522</td>
</tr>
<tr>
<td></td>
<td>SMLR</td>
<td>SNV</td>
<td>0.501</td>
<td>1.033</td>
<td>0.373</td>
</tr>
<tr>
<td></td>
<td>PLSR</td>
<td>$R_e$</td>
<td>0.400</td>
<td>0.911</td>
<td>0.323</td>
</tr>
<tr>
<td></td>
<td>PLSR</td>
<td>LR</td>
<td>0.538</td>
<td>0.868</td>
<td>0.481</td>
</tr>
<tr>
<td></td>
<td>PLSR</td>
<td>SNV</td>
<td>0.468</td>
<td>0.886</td>
<td>0.372</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>$R_e$</td>
<td>0.233</td>
<td>0.914</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>LR</td>
<td>0.284</td>
<td>0.899</td>
<td>0.127</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>SNV</td>
<td>0.645</td>
<td>0.622</td>
<td>0.553</td>
</tr>
<tr>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>SMLR</td>
<td>$R_e$</td>
<td>0.594</td>
<td>0.082</td>
<td>0.510</td>
</tr>
<tr>
<td></td>
<td>SMLR</td>
<td>LR</td>
<td>0.591</td>
<td>0.081</td>
<td>0.511</td>
</tr>
<tr>
<td></td>
<td>SMLR</td>
<td>SNV</td>
<td>0.526</td>
<td>0.091</td>
<td>0.370</td>
</tr>
<tr>
<td></td>
<td>PLSR</td>
<td>$R_e$</td>
<td>0.483</td>
<td>0.094</td>
<td>0.402</td>
</tr>
<tr>
<td></td>
<td>PLSR</td>
<td>LR</td>
<td>0.505</td>
<td>0.094</td>
<td>0.400</td>
</tr>
<tr>
<td></td>
<td>PLSR</td>
<td>SNV</td>
<td>0.459</td>
<td>0.095</td>
<td>0.385</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>$R_e$</td>
<td>0.224</td>
<td>0.093</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>LR</td>
<td>0.250</td>
<td>0.092</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>SNV</td>
<td>0.622</td>
<td>0.066</td>
<td>0.527</td>
</tr>
</tbody>
</table>

Note: <sup>b</sup>ONPLF - optimum number of PLSR latent factors.
Table 3 Calibration and validation statistics of the SAR models.

<table>
<thead>
<tr>
<th>Method</th>
<th>Spectral index</th>
<th>ONPLF</th>
<th>Calibration</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$R_c^2$</td>
<td>$RMSE_c$</td>
</tr>
<tr>
<td>SMLR</td>
<td>$R_\text{e}$</td>
<td>0.497</td>
<td>2.110</td>
<td>0.473</td>
</tr>
<tr>
<td>SMLR</td>
<td>LR</td>
<td>0.491</td>
<td>2.100</td>
<td>0.480</td>
</tr>
<tr>
<td>SMLR</td>
<td>SNV</td>
<td>0.682</td>
<td>1.625</td>
<td>0.655</td>
</tr>
<tr>
<td>PLSR</td>
<td>$R_\text{e}$</td>
<td>7</td>
<td>0.728</td>
<td>1.779</td>
</tr>
<tr>
<td>PLSR</td>
<td>LR</td>
<td>6</td>
<td>0.657</td>
<td>1.794</td>
</tr>
<tr>
<td>PLSR</td>
<td>SNV</td>
<td>6</td>
<td>0.662</td>
<td>1.771</td>
</tr>
<tr>
<td>SVM</td>
<td>$R_\text{e}$</td>
<td>0.672</td>
<td>1.784</td>
<td>0.536</td>
</tr>
<tr>
<td>SVM</td>
<td>LR</td>
<td>0.791</td>
<td>1.396</td>
<td>0.712</td>
</tr>
<tr>
<td>SVM</td>
<td>SNV</td>
<td>0.878</td>
<td>1.088</td>
<td>0.814</td>
</tr>
</tbody>
</table>
Figure Captions

Figure 1 Sampling location of the study area.

Figure 2 Variations of the studied soil chemical properties and corresponding elevations.

Figure 3 Curves of reflectance for different soils. The numbers in the legends denote SAR, [Na$^{+}$], [Ca$^{2+}$], and [Mg$^{2+}$] in turn.

Figure 4 Correlations between the soil physico-chemical properties and observed reflectance at wavelength scope of 350-1810 nm.

Figure 5 Comparison of the measured and the predicted SAR values using different models. The dash line is the 1:1 line. The models are: (a) SAR~Re using the SMLR method; (b) SAR~LR using the SMLR method; (c) SAR~SNV using the SMLR method; (d) SAR~Re using the PLSR method; (e) SAR~LR using the PLSR method; (f) SAR~SNV using the PLSR method; (g) SAR~Re using the SVM method; (h) Model of SAR~LR using the SVM method; (i) SAR~SNV using the SVM method.

Figure 6 Correlations between each pair of soil particle contents and soil chemical properties.
Figure 1 Sampling location of the study area.
Figure 2 Variations of the studied soil chemical properties and corresponding elevations.
Figure 3 Curves of reflectance for different soils. The numbers in the legends denote [Na$^+$], [Mg$^{2+}$], [Ca$^{2+}$], and SAR in turn.
Figure 4 Correlations between the soil physico-chemical properties and observed reflectance at wavelength scope of 350-1810 nm.
Figure 5 Comparison of the measured and the predicted SAR values using different models. The dash line is the 1:1 line. The models are: (a) SAR~R$_e$ using the SMLR method; (b) SAR~LR using the SMLR method; (c) SAR~SNV using the SMLR method; (d) SAR~R$_e$ using the PLSR method; (e) SAR~LR using the PLSR method; (f) SAR~SNV using the PLSR method; (g) SAR~R$_e$ using the SVM method; (h) Model of SAR~LR using the SVM method; (i) SAR~SNV using the SVM method.
Figure 6 Correlations between each pair of soil particle contents and soil chemical properties.