A peer-reviewed version of this preprint was published in PeerJ on 22 January 2019.

View the peer-reviewed version (peerj.com/articles/6310), which is the preferred citable publication unless you specifically need to cite this preprint.

Wang H, Chen Y, Zhang Z, Chen H, Li X, Wang M, Chai H. 2019. Quantitatively estimating main soil water-soluble salt ions content based on Visible-near infrared wavelength selected using GC, SR and VIP. PeerJ 7:e6310 https://doi.org/10.7717/peerj.6310

Quantitatively estimating main soil water-soluble salt ions content based on Visible-near infrared wavelength selected using GC, SR and VIP

Haifeng Wang [1, 2] , **Yinwen Chen** [3] , **Zhitao Zhang** [Corresp., 1, 2] , **Haorui Chen** [4] , **Xianwen Li** [2] , **Mingxiu Wang** [5] , **Hongyang Chai** [2]

[1] Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas, Ministry of Education, Northwest A&F University, Yangling, Shaanxi, China
[2] College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi, China
[3] Department of Foreign Languages, Northwest A&F University, Yangling, Shaanxi, China
[4] Department of Irrigation and Drainage, China Institute of Water Resources and Hydropower Research, Beijing, China
[5] Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA

Corresponding Author: Zhitao Zhang
Email address: zhitaozhang@126.com

Soil salinization is the primary obstacle to the sustainable development of agriculture and eco-environment in arid regions. The accurate inversion of the major water-soluble salt ions in the soil using visible-near infrared (VIS-NIR) spectroscopy technique can enhance the effectiveness of saline soil management. However, the accuracy of spectral models of soil salt ions turns out to be affected by high dimensionality and noise information of spectral data. This study aims to improve the model accuracy by optimizing the spectral models based on the exploration of the sensitive spectral intervals of different salt ions. To this end, 120 soil samples were collected from Shahaoqu Irrigation Area in Inner Mongolia, China. After determining the raw reflectance spectrum and content of salt ions in the lab, the spectral data were pre-treated by standard normal variable (SNV). Subsequently the sensitive spectral intervals of each ion were selected using methods of gray correlation (GC), stepwise regression (SR) and variable importance in projection (VIP). Finally, the performance of both models of partial least squares regression (PLSR) and support vector regression (SVR) was investigated on the basis of the sensitive spectral intervals. The results indicated that the model accuracy based on the sensitive spectral intervals selected using different analytical methods turned out to be different: VIP was the highest, SR came next and GC was the lowest. The optimal inversion models of different ions were different. In general, both PLSR and SVR had achieved satisfactory model accuracy, but PLSR outperformed SVR in the forecasting effects. Great difference existed among the optimal inversion accuracy of different ions: the predicative accuracy of Ca^{2+}, Na^{+}, Cl^{-}, Mg^{2+} and SO_4^{2-} was very high, that of CO_3^{2-} was high and K^{+} was relatively lower, but HCO_3^{-} failed to have any predicative power. These findings provide a new approach for the optimization

of the spectral model of water-soluble salt ions and improvement of its predicative precision.

Quantitatively estimating main soil water-soluble salt ions content based on Visible-near infrared wavelength selected using GC, SR and VIP

Haifeng Wang[1,2 *], Yinwen Chen[3 *], Zhitao Zhang[1,2], Haorui Chen[4], Xianwen Li[2], Mingxiu Wang[5] and Hongyang Chai[2]

[1] Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas, Ministry of Education, Northwest A&F University, Yangling, Shaanxi, China
[2] College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi, China
[3] Department of Foreign Languages, Northwest A&F University, Yangling, Shaanxi, China
[4] Department of Irrigation and Drainage, China Institute of Water Resources and Hydropower Research, Beijing, China
[5] Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA

[*] These authors contributed equally to this work.

Corresponding Author:
Zhitao Zhang[1,2]
No.23 Weihui Road, Yangling, Shaanxi, 712100, China
Email address: zhitaozhang@126.com

ABSTRACT

Soil salinization is the primary obstacle to the sustainable development of agriculture and eco-environment in arid regions. The accurate inversion of the major water-soluble salt ions in the soil using visible-near infrared (VIS-NIR) spectroscopy technique can enhance the effectiveness of saline soil management. However, the accuracy of spectral models of soil salt ions turns out to be affected by high dimensionality and noise information of spectral data. This study aims to improve the model accuracy by optimizing the spectral models based on the exploration of the sensitive spectral intervals of different salt ions. To this end, 120 soil samples were collected from Shahaoqu Irrigation Area in Inner Mongolia, China. After determining the raw reflectance spectrum and content of salt ions in the lab, the spectral data were pre-treated by standard normal variable (SNV). Subsequently the sensitive spectral intervals of each ion were selected using methods of gray correlation (GC), stepwise regression (SR) and variable importance in projection (VIP). Finally, the performance of both models of partial least squares regression (PLSR) and support vector regression (SVR) was investigated on the basis of the sensitive spectral intervals. The results indicated that the model accuracy based on the sensitive spectral intervals selected using different analytical methods turned out to be different: VIP was the highest, SR came next and GC was the lowest. The optimal inversion models of different ions were different. In general, both PLSR and SVR had achieved satisfactory model accuracy, but PLSR outperformed SVR in the forecasting effects. Great difference existed among the optimal inversion accuracy of different ions: the predicative accuracy of Ca^{2+}, Na^{+}, Cl^{-}, Mg^{2+} and SO_4^{2-} was very high, that of CO_3^{2-} was high and K^{+} was relatively lower, but HCO_3^{-} failed to have any predicative power. These findings provide a new approach for the optimization of the spectral model of water-soluble salt ions and improvement of its predicative precision.

Introduction

Soil salinization, one of the most important causes of land desertification and deterioration, has posed serious threat to agricultural development and sustainable utilization of natural resources (Shahid & Rahman, 2011; Abbas et al. 2013). 950 million ha of saline soil worldwide has become salinized (Schofield & Kirkby, 2003). Soil salinization is eroding and degenerating the arable soil at the speed of 10 ha/min (Graciela & Alfred, 2009). Soil remediation and management are very difficult in China because of such complex natural factors as climate, terrain and geology, and human factors as unreasonable irrigation and disruption of ecological balance. The total area of saline soil in China is 36 million ha (Li et al. 2014), accounting for 4.88% of the total area available nationwide (The National Soil Survey Office, 1998). Saline soil

usually has a high concentration of salt ions with a series of effects on the plants such as physiological draught, ion toxicity and metabolic disorder, thus forming "salt damage" (Munns, 2002; Tavakkoli et al. 2011). In addition, one major cause of the inaccuracy of soil salinity spectral measurement is that pure salts seldom exist in the soil because of some trace salt ion elements are always fixed in soil crystals. Therefore, quick and accurate acquisition of the detailed information of the various salt ions content in the soil can enhance the pertinence and effectiveness of saline soil management.

The traditional quantitative estimation of soil salt contents usually includes such steps as field soil sampling in fixed points, experiments in the laboratory and comprehensive statistical analysis (Urdanoz & Aragüés, 2011). Such method is incapable of the dynamic monitoring of saline soil in a large area because of its high consumption of time and energy, small number of measuring points and poor representativeness (Ding & Yu, 2014). Compared with conventional laboratory analysis methods, remote sensing technology has been widely used due to its rich information, continuity, high precision and low cost (Ben-Dor, 2002; Viscarra Rossel et al. 2006; Viscarra Rossel & Behrens, 2010; Viscarra Rossel & Webster, 2012). The various soil constituents (contents of water, salt, organic matter and so forth) can be acquired conveniently from remote sensing data (Gomez et al. 2008; Yu et al. 2010; Periasamy & Shanmugam, 2017). Hence, with the abundant spectral reflection information within the VIS-NIR intervals of soil salinity, it is feasible to improve the accuracy of soil salinization inversion (Al-Khaier, 2003; Ben-Dor et al. 2009; Abbas et al. 2013).

The application of VIS-NIR spectral analysis technique has been proved effective in improving the accuracy of quantitative estimation and eliminating the external disturbance to some extent (Dehaan & Taylor, 2002; Metternicht & Zinck, 2003; Farifteh et al. 2008). The univariate linear regression on the basis of soil salinity index developed for CR (continuum removed) reflectance can be used as a method for soil salt content estimation (Weng et al. 2008). Due to the strong correlation between soil electrical conductivity (EC) and soil salinity, EC is also one of the important indicators for evaluating soil salinization degree. A variety of approaches have been used to acquire the EC in the field soil, including the partial least squares regression (PLSR) and multivariate adaptive regression splines (MARS) (Volkan Bilgili et al. 2010; Nawar et al. 2015), logarithmic model (Xiao et al. 2016a), Bootstrap-BP neural network model (Wang et al. 2018d) and satellite remote sensing technology (Nawar et al. 2014; Bannari et al. 2018). In addition, the differential transformation (Xia et al. 2017) and fractional derivative (Wang et al. 2017; Wang et al. 2018c) can fully utilize the potential spectral information and enhance model accuracy. The methods of spectral classification (Jin et al. 2015) and water influence elimination (Chen et al. 2016; Peng et al. 2016; Yang & Yu, 2017) work well in improving the quantitative inversion accuracy of soil salinity. Therefore, the remote sensing

technique is reliable to inverse the soil salinity quantitatively on different scales.

The quantitative analysis of VIS-NIR spectral intervals can help evaluate the content of some chemical elements (Viscarra Rossel et al. 2006; Farifteh et al. 2008; Cécillon et al. 2009; Ji et al. 2016) due to the different characteristic absorption spectrum in soil chemical elements. Besides, there exists a correlation between some principal salt ions (Na^+, Cl^-) and spectral reflectance (Jiang et al. 2017). Therefore, VIS-NIR spectroscopy technique can be used to obtain the contents of the soil salt ions to a certain extent. The spectral response characteristics of mid-infrared (MIR) spectroscopy are better than those of VIS-NIR spectroscopy in predicting soil salinity information, the latter has high predicting accuracy of the total salts content, HCO_3^-, SO_4^{2-} and Ca^{2+}, followed by Mg^{2+}, Cl^- and Na^+ (Peng et al. 2016). The spectral models have satisfactory prediction of the SAR (sodium absorption ratio) of soil salinization evaluation parameter, which is composed of the contents of Ca^{2+}, Mg^{2+} and Na^+ (Xiao et al. 2016b). Qu et al. (2009) found that the contents of the total salt, SO_4^{2-}, pH and K^++Na^+ have a higher inversion accuracy using spectral data to create PLSR model. The different pretreatment of the different ion models varies by creating and analyzing PLSR model that demonstrates relatively good predictive effects like ion contents of Ca^{2+}, Mg^{2+}, SO_4^{2-}, Cl^-, and HCO_3^- (Dai et al. 2015). Overall, PLSR is a frequently used and robust linear model for quantitative research because it has inference capabilities which are useful to model a probable linear relationship between the reflectance spectra and the salt ions content in soil. However, the non-uniform data and non-linear reflectance in spectral information of some soil chemical elements lead to the reduction in model accuracy (Viscarra Rossel & Behrens, 2010; Nawar et al. 2015). In particular, support vector regressions (SVR) based on kernel-based learning methods has the ability to handle nonlinear analysis case with high model accuracy (Vapnik, 1995; Peng et al. 2016; Hong et al. 2018b). Over the past several decades, the use of SVR for classification and regression has been extensively applied in soil VIS-NIR spectroscopy (Ben-Dor, 2002; Xiao et al. 2016b; Hong et al. 2018a). Moreover, the SVR model works well in estimating the contents of K^+, Na^+, Ca^{2+} and SO_4^{2-} in the soil (Wang et al. 2018a). Thus, the correct way of modeling helps to guarantee the model accuracy (Farifteh et al. 2007).

Many researches focused on the inversion of soil salinity using spectral information. Nevertheless, little research has explored the eight water-soluble salt ions (K^+, Ca^{2+}, Na^+, Mg^{2+}, Cl^-, SO_4^{2-}, HCO_3^- and CO_3^{2-}) using spectral information in the soil. The model fitting of ions and spectral information still needs improving (Farifteh et al. 2008; Peng et al. 2016). Apart from the suitable multivariate statistical analysis method that can partly improve the inversion effects, reduction of redundant information is another identified approach to further optimize the model (Bannari et al. 2018; Stenberg et al. 2010). Plenty of studies have demonstrated that spectral variable selection methods can not only reduce the complexity of calibration models, but also

improve the model predictive performance (Hong et al. 2018a). To select the optimal spectral variable subset, scholars have investigated varied methods such as gray correlation (GC) (Li et al. 2016; Wang et al. 2018b), stepwise regression (SR) (Zhang et al. 2018) and variable importance in projection (VIP) (Qi et al. 2017), and have achieved satisfactory effects. In addition, all the three methods have been widely applied in many studies, such as plant physiology, food engineering, mathematical statistics (Oussama et al., 2012; Maimaitiyiming et al. 2017; Liu et al. 2015). However, few studies have concentrated on the use of variable selection algorithms in the inversion of soil salt ions.

This study aims to: (1) build the optimal model of soil salt ions using VIS–NIR spectroscopy technique; (2) compare the models based on the sensitive spectral ranges selected using GC, SR and VIP methods for different soil ions; (3) compare the performance of PLSR and SVR models, and identify the optimal models for different ions.

MATERIALS AND METHODS

Study area

Hetao Irrigation District (HID), with Yin Mountains at its north, the Yellow River at its south, Ulanbuh Desert at its west and Baotou at its east, lies in Bayannur League, Inner Mongolia, China. It consists of irrigation areas of Ulan Buh, Jiefangzha, Yongji, Yichang and Urat, and it is China's largest irrigation district with a total size of 5740 km^2 (Yu et al. 2010). In addition, HID is an important production base of cereal and oil plants in China with major crops of wheat, corn and sunflower. Shahaoqu Irrigation Area (SIA), a typical region of saline soil in HID, was chosen as the study area. SIA (107°05′～107°10′E, 40°52′～41°00′N) is located in the central east of Jiefangzha Irrigation Area. SIA belongs to typical continental climate, having hot summers, chilly winters, rare precipitation and strong evaporation. Its mean annual temperature, precipitation, potential evaporation is about 7.1℃, 155 mm and 2000 mm, respectively. Physiographically, the mean elevation and slope of SIA are about 1030 m and 1/10000, respectively. According to the World Reference Base for Soil Resources (WRB), the local soil texture is mainly silty clay loam with varying degrees of saline soil. Over the years, due to its gentle terrain slope, poor groundwater runoff, intense land surface evaporation and irrational farming activities, about 60% of the land within the district has been affected by various degree of salinization, which seriously restricted the agricultural development (Wu et al. 2008; Gao et al. 2015).

Sample collection and chemical analysis

The Hetao irrigation district administration gave field permit approval to us (NO. 2017YFC0403302). To ensure the representativeness of soil samples, the samples were

randomly gathered from a total of 120 sampling units on a grid of 16 m×16 m (because the spatial resolution of GF-1 satellite imagery is 16 m) in the study area during October 12～22, 2017 (Fig. 1). In each unit, approximately 0.5 kg of topsoil (0-5 cm) was collected at four randomly selected sampling sites and then mixed thoroughly to obtain a representative sample. Overall, a total of 120 soil samples were acquired, and each sample was stored in a plastic bag, labeled and sealed. A portable global position system (GPS) was used to determine the coordinates of sampling points. Subsequently, the soil samples were transported to the lab to receive a series of such treatments as sufficient natural air-drying for two weeks and rubbing through a 2 mm sieve to exclude small stones and other impurities. Each sample was divided into two subsamples to be used for spectra collection and physiochemical analysis.

Each 50 g of soil sample was put into a respective flask, and 250 ml of distilled water (the ratio of water to soil is 5:1) were added into each flask. The water-soluble ion contents were measured in the filtrate obtained from full soaking, oscillation and filtration (Aboukila & Norton, 2017). Ca^{2+} and Mg^{2+} were measured using EDTA titration, Na^{+} and K^{+} flame photometry, CO_3^{2-} and HCO_3^{-} double indicator-neutralization titration, Cl^{-} silver nitrate titration, and SO_4^{2-} EDTA indirect complexometry (Bao, 2000). The content of CO_3^{2-} was too low (approximately 0) in some soil samples because CO_3^{2-} is liable to integrate with Ca^{2+} and Mg^{2+} as sediment in a weak alkaline solution (Table 1). Coefficient of variation (CV) reflects the degree of discreteness, and a positive correlation exists in two variables. The high CV helps to build a robust model (Dai et al. 2015). The grading of CV showed a wide range of variation among different ions, among which the ion contents of K^{+}, Na^{+} and SO_4^{2-} are over 100%, showing a strong variability, and those of CO_3^{2-}, Cl^{-}, Ca^{2+}, Mg^{2+} and HCO_3^{-} are between 10% and 100%, having a moderate variability.

Laboratory spectral measurements and pretreatments

The soil samples were put into black vessels with a diameter of 10 cm and depth of 2 cm for spectral data collection and the surfaces were smoothed with a straightedge in the laboratory. The spectral data of the soil samples were measured using ASD (Analytical Spectral Devices, Inc., Boulder, CO, USA) FieldSpec®3 spectrometer with spectral range from 350-2500 nm. This instrument is equipped with two sensors whose spectral resolutions are 1.4 nm and 2 nm, for the region of 350-1000 nm and 1000-2500 nm, respectively. The spectral data was measured in a dark room with the light sources which have halogen lamps of 50 W, 50 cm from the sample soil surfaces, and 30° incident angle to reduce the effects of external factors to the minimum. The field angle of fiber-optics probe is 5°, and it is 15 cm from the sample soil surface. The light source and spectrometer had been fully preheated, and the spectrometer had been corrected with a standardized white panel (99% reflectance) prior to each measurement to reduce measurement

error. Each sample soil was measured in four directions (3 turns, each is 90°), the spectrum was collected five times in each direction, and altogether there were 20 curves of the spectrum (Hong et al. 2018b). These curves were used as the raw spectral reflectance (R_{raw}) after having the arithmetic mean in ViewSpecPro software version 6.0. The gaps of the spectral curves near 1000 nm and 1800 nm were corrected using Splice Correction function (Xiao et al., 2016a).

The fluctuation would affect the accuracy of subsequent modeling because of such disturbance as the external environment, instrument noise and random error in spectral data collection. In general, a series of effective pretreatment, including smoothing, resampling and transformation etc., can eliminate the external noise to some degree, and then enhance the spectral characteristics (Ding et al. 2018). Therefore, it is necessary to pretreat R_{raw} in the following steps. i) The marginal wavelength (350-399 nm and 2401-2500 nm) of higher noise in each soil sample was removed, then remaining spectrum data was smoothed with filter method (window size is 5 and polynomial order is 2) using Savitzky-Golay (SG) (Savitzky & Golay, 1964) via Origin Pro software version 2017SR2. ii) The spectral data between 400 and 2400 nm was resampled with a 10 nm of samplee interval to keep the spectral features and remove redundant information (Xu et al. 2016). A new spectral curve consisting of 200 wave bands was obtained. iii) The precise $R_{raw\text{-}SNV}$ was obtained by using standard normal variable (SNV) to eliminate the effects of soil particle size, surface scattering and baseline shift on the spectrum data (Xiao et al., 2016b; Barnes et al., 1989). The spectral curves of R_{raw} and $R_{raw\text{-}SNV}$ are shown in Fig. 2A and 2B. Notably, comparison indicated that the spectral curve in Fig. 2B was much smoother than that in Fig. 2A, which made for the subsequent modeling.

Gray correlation (GC)

The GC, as one grey system theory, seeks the primary and secondary relations and analyzes the different effects of all the factors in a system (Deng, 1982; Li et al. 2016). Its calculation process is as follows: the reference sequence is $X_0 = \{x_0(t), t = 1, 2, \cdots, n\}$, the comparative sequence is $X_i = \{x_i(t), t = 1, 2, \cdots, n\}$, and the formula of the gray correlation degree (GCD) between X_0 and X_i is

$$\mathrm{GCD} = \frac{1}{n}\sum_{t=1}^{n}\gamma\left(x_0(t), x_i(t)\right) \tag{1}$$

where $\gamma\left(x_0(t), x_i(t)\right) = \dfrac{\min\limits_i \min\limits_t |x_0(t) - x_i(t)| + \rho \max\limits_i \max\limits_t |x_0(t) - x_i(t)|}{|x_0(t) - x_i(t)| + \rho \max\limits_i \max\limits_t |x_0(t) - x_i(t)|}$

ρ is the distinguishing coefficient within $[0,1]$. ρ was set as 0.1 in this paper.

The inconsistent dimension between the spectral data and the contents of different ions has some effects on the data analysis. Therefore, normalizing the spectral data preprocessing method can reduce these disadvantageous effects (Liu et al. 2015; Wang et al. 2018b). In this paper, the larger the GCD of a certain band is, the closer relation the band and the ion content has, and vice versa.

Variable importance in projection (VIP)

The VIP is a variable selection method based on PLSR (Oussama et al., 2012). The explanatory power of the independent variables to the dependent variables is achieved by calculating the VIP score. The independent variables are sequenced according to the explanatory power (Qi et al. 2017). The VIP score for the *j*-th variable is given as:

$$\mathrm{VIP}_j = \sqrt{\frac{p * \sum_{f=1}^{F} \mathrm{SSY}_f * \mathrm{W}_{jf}^2}{\mathrm{SSY}_{total} * F}} \qquad (2)$$

Where p is the number of independent variables; m is the total number of components; SSY_f is the sum of squares of explained variance for the f-th component and p the number of independent variables. SSY_{total} is the total sum of squares explained of the dependent variable, and F is the total number of components. W_{jf}^2 gives the importance of the j-th variable in each f-th component. The higher value VIP_j has, the stronger explanatory power the independent variable has over the dependent variable. The VIP scores of independent variables have been recognized as a useful measure to identify important wavelengths when the score is more than 1 (Wold et al. 2001; Maimaitiyiming et al. 2017).

Model construction and validation

Two thirds of the samples were used for modeling (n = 80) and one third for validation (n = 40) using Kennard-Stone (K-S) to calculate the Euclidian distance among different samples to ensure the statistical characteristics of modeling and the validation datasets resembled that of the whole sample set (Kennard & Stone, 1969).

The PLSR and SVR models were applied to the quantitative inversion of different water-soluble salt ion contents in the saline soil in this paper. The PLSR model is a new stoichiometric statistical model. Compared with the traditional multivariate least squares regression (MLSR), PLSR can overcome the multicollinearity among the variables, reduce the dimension, synthesize and filter the information, extract the aggregate variables with the strongest explanatory power in the system, and exclude the noise with no explanatory power (Wold et al. 2001). The optimal fitting model was built using the number of optimal principal components through full cross validation. SVR model is a new machine learning method based on the principle of structural risk minimization provided by the statistical learning theory. This model is characterized by its

ability of solving such problems as limited sample size, nonlinear data processing and spatial pattern recognition of high-dimension data (Vapnik, 1995). During the modeling in this study, the type of SVR and kernel were set as epsilon-SVR and linear function, respectively; the penalty parameter C and nuclear parameter g were acquired by a grid-searching technique and a leave-one-out cross validation procedure. The optimal values of C and g were selected when the minimum $RMSE_{CV}$ (root mean squared error of cross validation) was produced (Xiao et al. 2016b). The two models were constructed and validated using the Unscrambler software version X10.4 (CAMO AS Oslo, Norway)

Precision indices of determination coefficient of calibration (R_c^2), determination coefficient of prediction (R_p^2), root mean squared error (RMSE) and ratio of performance to deviation (RPD) were used to evaluate the performance of these models. RPD classification was adopted to facilitate the interpretation of predictive results: a model is considered as excellent when RPD ≥ 2.5, as very good when 2.0 ≤ RPD < 2.5, as good when 1.8 ≤ RPD < 2.0, and as satisfactory when 1.4 ≤ RPD < 1.8 and can only distinguish between high and low values when 1.0 ≤ RPD < 1.4 (Viscarra Rossel et al. 2007). Generally, the most robust model would be the one with the largest R_c^2, R_p^2 (approach to 1) and RPD value and the lowest RMSE value.

RESULTS

Correlation between water-soluble salt ions content and spectral reflectance

The correlation coefficients (Pearson correlation) between each soil salt ion content and $R_{raw\text{-}SNV}$ in the range of 400-2400 nm were tested with the significance level of $P < 0.01$ ($|r| = 0.234$ or above). The curves of correlation coefficients of soil salt ions were plotted in Fig. 3 and the numbers of bands passing the significance test were counted in Table 2.

The curve patterns of SO_4^{2-}, Cl^-, Ca^{2+}, Mg^{2+}, K^+ and Na^+ were similar (Fig. 3). From 400 nm to about 550 nm, the correlation coefficients rose sharply from negative to positive, moved with a gentle depression until 1400 nm, plummeted and surged up to 1560 nm (among the curves, the change of Ca^{2+} was the sharpest), and maintained a relative stable state to 1850 nm. And then from 1850 to 2400 nm, dramatic oscillating variations alternated between rise and fall. In the intervals of 400-1400 nm and 1850-2400 nm the curve pattern of CO_3^{2-} was similar to that of other ions such as SO_4^{2-}. But between 1400 nm and about 1850 nm, the curve took on a unique pattern: sustained oscillating rise. The coefficient curve of HCO_3^- displayed a smaller variation, smoothly fluctuating between -0.2 and 0.2. The complex variation of the coefficient curves of different ions revealed rich spectral information.

Selection of characteristic wavelength

Characteristic wavelength selection based on GC method

The curves of gray correlation degree for soil water-soluble salt ions content and $R_{raw\text{-}SNV}$ were

shown in Fig. 4. The correlation coefficient curves of the seven ions except CO_3^{2-} resembled those of the GCD of the $R_{raw\text{-}SNV}$. Generally, the curves exhibited patterns of "oscillatory rise, fluctuation, rapid rise and fall, and oscillatory fluctuation". The gray correlation curves of CO_3^{2-} followed a pattern of "ascending, plummeting, and smooth transition". The analysis of the GC curve amplitude showed the amplitudes of Cl^-, Mg^{2+} and Ca^{2+} were relatively large, and those of Na^+, SO_4^{2-}, K^+ and HCO_3^- were relatively small, and that of CO_3^{2-} was relatively gentle.

The order of the maximal GCD was: Cl^- (0.561) > Mg^{2+} (0.559) > Ca^{2+} (0.551) > Na^+ (0.508) > SO_4^{2-} (0.494) > K^+ (0.470) > HCO_3^- (0.465) > CO_3^{2-} (0.416). To ensure that each salt ion had sensitive bands as far as possible, the GCD threshold value was set as 0.40 to select the wavelength. The sensitive band was counted through gray correlation method (Table 3). The numbers of sensitive bands of different ions could be sequenced from the largest to the smallest as follows: Mg^{2+} (110) > HCO_3^- (105) > Cl^- (101) > Ca^{2+} (53) > Na^+ (36) > SO_4^{2-} (21) > K^+ (15) > CO_3^{2-} (14). Therefore, the orders of sensitive band numbers and maximal GCD values had great difference. Furthermore, the band intervals corresponding to the maximum GCD of different salt ions were as follows: CO_3^{2-} was near-infrared between 1740 and 1750 nm, HCO_3^- was green light between 560 and 570 nm, and the rest of six ions were near-infrared between 1650 and 1660 nm.

Characteristic wavelength selection based on SR method

Feature band intervals were selected by stepwise regression method in SPSS software version 23.0 (IBM, Chicago, USA), and the significance levels of variables acceptance and rejection were set at 0.10 and 0.15 (Zhang et al. 2018). The parameter indexes of feature band intervals selection were shown in Table 4 by stepwise regression method at maximum adjusted R^2.

Great difference existed among the optimal SR models of different ions, and the numbers of band intervals accepted by the model range from 3 to 8 (Table 4). The SR model fitted well with the adjusted R^2 greater than 0.8 when the number of selected independent variables was considered. Meanwhile, SR model of each ion was statistically significant (p<0.001). Therefore, the band intervals selected by the SR models were used as the independent variables of PLSR and SVR models.

Characteristic wavelength selection based on VIP method

Curves of VIP scores of soil water-soluble salt ions content and $R_{raw\text{-}SNV}$ were shown in Fig. 5. Max VIP scores and band intervals obtained from VIP method of soil water-soluble salt ions content and $R_{raw\text{-}SNV}$ were shown in Table 5.

The curves patterns of seven ions were similar except HCO_3^- (Fig. 5). These curves exhibited violent oscillation in the intervals of 400-800 nm and 1900-2400 nm, gentle transition between 800 nm and around 1400 nm, and fluctuant rise from 1400 to 1900 nm. In contrast, the curve of HCO_3^- showed oscillatory rise from 400 to 1400 nm, a "U" shaped motion from 1400 to 1900

nm or so, and a rapid fall and oscillation to 2400 nm. The numbers of sensitive bands based on VIP method displayed the following sequence: Cl^- (85) > Na^+ (83) > HCO_3^- (79) > SO_4^{2-} (74) > Mg^{2+} (69) = Ca^{2+} (69) = K^+ (69) > CO_3^{2-} (67). The sequence of the maximal VIP scores was HCO_3^- (2.37) > CO_3^{2-} (2.01) > Ca^{2+} (1.97) > SO_4^{2-} (1.74) > K^+ (1.73) > Na^+ (1.55) > Mg^{2+} (1.49) > Cl^- (1.42). The spectral interval of the maximal VIP scores of Cl^- was from 560 to 570 nm, Ca^{2+}, CO_3^{2-} and HCO_3^- were concentrated between 1410 and 1450 nm; and K^+, Mg^{2+}, Na^+ and SO_4^{2-} were from 1870 to1890 nm.

Construction and analysis of PLSR model

The sensitive bands were obtained using different band selection methods of GC, SR and VIP to build PLSR model. The results of PLSR model were shown in Table 6.

The models of the six ions Ca^{2+}, Cl^-, CO_3^{2-}, Mg^{2+}, Na^+ and SO_4^{2-} performed well using VIP method (R_c^2 is close to 1). The models based on the bands of Ca^{2+}, Cl^-, Mg^{2+}, Na^+ and SO_4^{2-} selected using SR method displayed good fitting effect, and those of Ca^{2+}, Mg^{2+} and Na^+ using GC method exhibited good fitting effect.

In terms of verification accuracy, VIP method had excellent prediction of Ca^{2+}, Na^+, SO_4^{2-}, SR method had excellent prediction of Ca^{2+}, Mg^{2+}, Na^+, SO_4^{2-} (the RPD of Ca^{2+} was up to 3.95), and GC method did not show strong prediction power over any ions. On the contrary, all the three models demonstrated poor forecasting power over HCO_3^-. The RPDs of SR-HCO_3^- and VIP-HCO_3^- were 0.64 and 0.93 respectively. Therefore, VIP method had the best modeling effect and SR method had the best forecasting effect, and GC method had poor modeling and forecasting effects on the salt ions inversion in the PLSR model.

Construction and analysis of SVR model

The sensitive bands were obtained by using different band selection methods of GC, SR and VIP to build SVR model. The results of SVR model were shown in Table 7.

The modeling accuracy of SVR model was similar to that of PLSR model. But the verification accuracy of ions was different between the two models. VIP method had the excellent prediction of Ca^{2+}, Cl^-, Mg^{2+}, Na^+, SR method had the excellent prediction of Ca^{2+}, Mg^{2+}, Na^+, SO_4^{2-}, and GC method did not show strong prediction power over any ions. The prediction results of Ca^{2+} were the best: the RPD of VIP and SR models were 3.93 and 3.97, respectively. Overall, in the SVR model, VIP method exhibited the best performance for modeling and predicting the salt ions content, SR method was the second, and GC method was relatively poorer.

DISCUSSION

Comparison among the results of different salt ions content in estimating

The optimal band selection method varied in some degree from the optimal modeling method (Table 6 and 7). The comparison was made between the measured value and the estimated value

of all the ions concerned under the optimal model (Fig. 6). The sequence of the forecasting power of the ions was Ca^{2+} > Na^{+} > Cl^{-} > Mg^{2+} > SO_4^{2-} > CO_3^{2-} > K^{+} > HCO_3^{-}, and it was the same as that of the modeling power.

Obviously, the verification result showed that most data points of the five ions, Ca^{2+}, Na^{+}, Cl^{-}, Mg^{2+} and SO_4^{2-}, were concentrated near line 1:1. The optimal models of these five ions had very strong predicative power with the RPD above 2.5 (Tables 6 and 7). Compared with the previous researches, model prediction effects of K^{+} and Na^{+} (Qu et al. 2009); Ca^{2+}, Na^{+} and Mg^{2+} (Viscarra Rossel & Webster, 2012); SO_4^{2-}, HCO_3^{-}, Ca^{2+}, Cl^{-}, Mg^{2+} and SO_4^{2-} (Dai et al. 2015); HCO_3^{-}, Ca^{2+} and SO_4^{2-} (Peng et al. 2016); K^{+}, Na^{+}, Ca^{2+} and SO_4^{2-} (Wang et al., 2018a) were satisfactory. Although the results of this study are not exactly the same as these previous researches, it still shows the rationality own to some extent. In addition, this result shows that band selection has realized the goal of removing the irrelevant information, and plays a major role in improving the inversion accuracy of salt ions.

In Figure 6, the data points of CO_3^{2-} and K^{+} were relatively dispersed in the verification result. The CO_3^{2-} had a relatively good predictive power (RPD = 1.80) and the K^{+} had a normal predictive power (RPD = 1.43). Notably, HCO_3^{-} had no predicative power (RPD = 0.96) because the slope were under the 1:1 line and the data points were most discrete (Figs. 6D). The predicting effect of HCO_3^{-} was different from that of Peng et al. (2016) and Dai et al. (2015), but similar to that of Wang et al. (2018a). The cause of this result needs to be further studied. Overall, it is vital to make some efforts to improve the robustness and accuracy of these ion models. Xiao et al. (2016b) failed to predict Na^{+}, Mg^{2+} and Ca^{2+}, but applied the SVR model to forecasting SAR after the SNV transformation and the performance was satisfactory (RPD = 2.13). Analogously, first derivative reflectance (FDR) index was calculated to effectively predict SAR by Xiao et al. (2016a). In addition, Viscarra Rossel & Webster (2012) forecasted the content of Na^{+} after logarithmic pretreatment with VIS-NIR spectral technique (RPD = 2.10). Thus, salt ion indexes construction and variable transformation processing are helpful approaches to improve the correlation with the spectra so as to establish satisfactory models.

A little difference existed in the applicability between PLSR and SVR models on inversing the content of ions. Both methods could produce satisfactory results in conformity with that of Peng et al. (2016). In addition, the optimal inversion models and prediction models for each ion were different: SR-PLSR model and SR-SVR model for Ca^{2+}, VIP-SVR model and SR-PLSR model for CO_3^{2-}, SR-PLSR model and VIP-PLSR model for K^{+}, VIP-PLSR model and GC-PLSR model for HCO_3^{-}, respectively. Among them, the performance of the optimal inversion model of Ca^{2+} resembled that of the prediction model. The results suggested that the ion models with poorer performance frequently demonstrated uncertainty in the inversion process (Peng et al. 2016). Generally, as the major water-soluble ion components in the two highly soluble salts of

sodium and kali, Na^+ and K^+ exhibit great difference in the spectral characterization degree (Dai et al. 2015). Therefore, the spectral characters of water-soluble salt ions are not necessarily determined by the number of dissociative ions, so more pertinent experiments and analysis should be conducted to explore the response mechanism.

Correlation analysis and inversion performance

The raw spectral reflectance curve of each soil sample presented distinct shapes (Fig. 2A). One of the prime reasons for this phenomenon is that the absorption features in these soil samples were related to soil salt crystal contents and types, as well as various chemical bonds (e.g., C-H, O-H, N-H). The results were in accordance with those in previous studies (Viscarra Rossel et al. 2006; Viscarra Rossel & Webster, 2012; Dai et al. 2015; Peng et al. 2016; Wang et al. 2018a), which demonstrated that soil VIS-NIR spectra could be used to determine part of soil salt ions contents in some degree.

Traditionally, correlation analysis helps reveal the relationships between soil salt ions content and VIS-NIR spectra, and it indicates modeling effects to some degree (Weng et al. 2008). In the current research, the number of the significant bands of different ions could be sequenced from the largest to the smallest as follows: Cl^- (96%) > Ca^{2+} (95%) > Mg^{2+} (93%) > Na^+ (90.5%) > K^+ (89%) = SO_4^{2-} (89%) > CO_3^{2-} (73%) > HCO_3^- (0.5%), the correlation coefficients of different ions ranged from the largest to the smallest as: Cl^- (-0.882) > Ca^{2+} (-0.877) > Mg^{2+} (-0.848) > Na^+ (-0.752) > SO_4^{2-} (0.749) > K^+ (0.630) > CO_3^{2-} (0.552) > HCO_3^- (0.235) (Table 2). Thereby, five ions (Cl^-, Ca^{2+}, Mg^{2+}, Na^+ and SO_4^{2-}) had more significant relationship with reflectance spectra. Although there were some differences between forecasting power ranking and correlation ranking, the optimal models of these five ions had the excellent predictive results (Fig. 6). Nevertheless, the other three ions (K^+, CO_3^{2-} and HCO_3^-) had weak correlations and unsatisfactory predictive power. In particular, HCO_3^- had only one significant band and the worst prediction effects. But in most cases, the sensitive band numbers of HCO_3^- were not the least in comparing the results of the three wavelength selection methods (Tables 3-5). Thus, we conjecture that the different calculation mechanisms cause a certain inconsistency between modeling performance and sensitivity. In addition, the optimal method of finding out their responding spectrum varies from one ion to another in the soil. In future study, it is practically significant to adopt various methods to select the optimal bands in the inversion of soil ions.

Effects of wavelength selection on estimation models

The massive complex spectra often contain a large amount of redundant information irrelevant to the ions contents. The selection of feature spectra is hence a critical step to create a robust model. From Tables 3-5, we could see the great difference exist in the number of wavelength selected with the three methods: VIP method had the largest number of wavelengths (34.5%～42.5%), SR method had the smallest number of wavelengths (1.5%～4%) and number of wavelengths

(7%～55%) varied greatly by GC method.

Our experiment with three wavelength selection methods also indicated that different methods yielded different results. Among the three methods, the VIP method produced the best results, followed by SR method, while the GC method performed least ideally. We argue that the GC method is not necessarily an inappropriate method as some results are still acceptable. However, GC method could distinguish the primary relationships among the factors in the system by calculating and comparing GCD (Deng, 1982; Liu et al. 2015). In the field of spectral analysis, the application of GC method could better identify sensitive spectral indices, select sensitive bands and optimize inversion model (Li et al. 2016). On the other hand, Wang et al (2018b) used GC method to extract the feature bands of soil organic matter content to construct the model with stronger generalization capability. Therefore, the soil compositions have a strong impact on the performance of spectral model. This conclusion is consistent with previous research results (Viscarra Rossel et al. 2006; Viscarra Rossel & Webster, 2012; Xiao et al. 2016b). The VIP values were calculated with VIP method, in the process of PLSR analysis to further evaluate the significance of each wavelength for model prediction (Wold et al. 2001; Maimaitiyiming et al. 2017; Qi et al. 2017). VIP method often produces the best results in the modeling set because it can distinguish between useful information and inevitable noises in the set. Oussama et al. (2012) adopted this method to reduce almost 75% of the total data set for a simplified model of high accuracy. Additionally, as a simplified regression linear model, SR method not only preserves significant bands but also solves multicollinearity problems effectively (Xiao et al. 2016a; Xiao et al. 2016b). It has great optimization effect on model complexity by adjusting the significance level of selected and excluded variables (Zhang et al. 2018). Compared with the selection results with VIP method, SR method could be used to extract fewer bands to establish ions (except for K^+, CO_3^{2-} and HCO_3^-) forecasting models with RPD above 1.80. Therefore, it is meaningful to make further simplification of the model while ensuring its accuracy.

Research limitations

This study clearly demonstrated that VIS-NIR spectral analysis technique is an effective method to detect salt ions content of salinity soil in the irrigated district. In terms of extracting feature wavelengths to estimate ions content, our work provides a comprehensive comparison and evaluation approaches. Such endeavor is critically and practically important to further enhance the model performance of the soil salt ions. The application of machine learning algorithms with strong applicability to solve nonlinear relationship between variables, such as Ant Colony Optimization-interval Partial Least Square (ACO-iPLS), Recursive Feature Elimination based on Support Vector Machine (RF-SVM), and Random Forest (RF) has been proved to be a useful approach to obtain the effective information of soil organic matter (Ding et al. 2018). To further improve the prediction accuracy, the more machine learning algorithms should be applied to the

analysis of sensitive spectral regions and the construction of stable models in future study. In addition, the application of multi-source remote sensing platforms such as Landsat, GaoFen-5, Hyperion and unmanned aerial vehicle (UAV) in soil salt ions estimation has not been investigated. Therefore, further research should focus on the possible combination of multiple approaches and remote sensing data at different scales to estimate soil salt ions content.

CONCLUSIONS

This study investigated the feasibility of estimating soil water-soluble salt ions content via VIS-NIR spectral model. Different methods were applied to the selection of response bands interval to construct robust inversion models. Among them, VIP method could select larger number of wavebands with the highest accuracy, SR method could select the smallest number of wavebands with good accuracy. However, the number of wavebands obtained using GC method varied greatly with poor accuracy. The PLSR and SVR models achieved good effects on the modeling and forecasting of most ions content. Moreover, the PLSR model was slightly more than the SVR model in terms of the number of ion models with good predictive effects (RPD over 2.0). The models of Ca^{2+}, Na^{+}, Cl^{-}, Mg^{2+} and SO_4^{2-} displayed the highest prediction accuracy, and the RPDs were 3.97, 3.15, 2.98 and 2.75, respectively, while those of other ions were poor. Overall, the best wavelength selection methods, models and inversion results of soil salt ions were different. In the future, the combination of band selection methods and spectral model will have a great potential for predicting some soil salt ions content in the salinization area. Such an approach can be utilized to assist decision makers toward the determination of soil salinization levels.

ACKNOWLEDGEMENTS

The authors want to thank A.P. Junying Chen for her help in language standardization of this manuscript and providing helpful suggestions. We are especially grateful to the reviewers and editors for appraising our manuscript and for offering instructive comments.

REFERENCE

Abbas A, Khan S, Hussain N, Hanjra MA, Akbar S. 2013. Characterizing soil salinity in irrigated agriculture using a remote sensing approach. *Physics and Chemistry of the Earth* **55-57**:43-52. DOI 10.1016/j.pce.2010.12.004

Aboukila EF, Norton JB. 2017. Estimation of saturated soil paste salinity from soil-water extracts. *Soil Science* **182**:107-113. DOI 10.1097/SS.0000000000000197

Al-Khaier F. 2003. *Soil salinity detection using satellite remote sensing*. Michigan Technological University.

Bannari A, El-Battay A, Bannari R, Rhinane H. 2018. Sentinel-MSI VNIR and SWIR bands

519 sensitivity analysis for soil salinity discrimination in an arid landscape. *Remote Sensing*
520 **10**:855. DOI 10.3390/rs10060855
521 **Bao S. 2000.** *Soil and agricultural chemistry analysis*. Beijing: China Agriculture Press. (in
522 Chinese).
523 **Barnes RJ, Dhanoa MS, Lister SJ. 1989.** Standard normal variate transformation and de-
524 trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy* **43**:772-777. DOI
525 10.1366/0003702894202201
526 **Ben-Dor E. 2002.** Quantitative remote sensing of soil properties. *Advances in Agronomy* **75**:173-
527 243. DOI 10.1016/S0065-2113(02)75005-0
528 **Ben-Dor E, Chabrillat S, Demattê JAM, Taylor GR, Hill J, Whiting ML, Sommer S. 2009.**
529 Using imaging spectroscopy to study soil properties. *Remote Sensing of Environment*
530 **113**:S38-S55. DOI 10.1016/j.rse.2008.09.019
531 **Cécillon L, Barthès BG, Gomez C, Ertlen D, Genot V, Hedde M, Stevens A, Brun JJ. 2009.**
532 Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy
533 (NIRS). *European Journal of Soil Science* **60**:770-784. DOI 10.1111/j.1365-
534 2389.2009.01178.x
535 **Chen H, Zhao G, Sun L, Wang R, Liu Y. 2016.** Prediction of soil salinity using near-infrared
536 reflectance spectroscopy with nonnegative matrix factorization. *Applied Spectroscopy*
537 **70**:1589-1597. DOI 10.1177/0003702816662605
538 **Dai X, Zhang Y, Peng J, Luo H, Xiang H. 2015.** Prediction and validation of water-soluble salt
539 ions content using hyperspectral data. *Transactions of the Chinese Society of Agricultural*
540 *Engineering* **31**:139-145. DOI 10.11975/j.issn.1002-6819.2015.22.019 (in Chinese).
541 **Dehaan RL, Taylor GR. 2002.** Field-derived spectra of salinized soils and vegetation as
542 indicators of irrigation-induced soil salinization. *Remote Sensing of Environment* **80**:406-
543 417. DOI 10.1016/S0034-4257(01)00321-2
544 **Deng J. 1982.** Control problems of grey systems. *Systems & Control Letters* **1**:288-294. DOI
545 10.1016/S0167-6911(82)80025-X
546 **Ding J, Yu D. 2014.** Monitoring and evaluating spatial variability of soil salinity in dry and wet
547 seasons in the Werigan–Kuqa Oasis, China, using remote sensing and electromagnetic
548 induction instruments. *Geoderma* **235-236**:316-322. DOI 10.1016/j.geoderma.2014.07.028
549 **Ding J, Yang A, Wang J, Sagan V, Yu D. 2018.** Machine-learning-based quantitative
550 estimation of soil organic carbon content by VIS/NIR spectroscopy. *PeerJ* **6**:e5714. DOI
551 10.7717/peerj.5714
552 **Farifteh J, Van der Meer F, Atzberger C, Carranza EJM. 2007.** Quantitative analysis of salt-
553 affected soil reflectance spectra: a comparison of two adaptive methods (PLSR and ANN).
554 *Remote Sensing of Environment* **110**:59-78. DOI 10.1016/j.rse.2007.02.005

555 **Farifteh J, Van der Meer F, van der Meijde M, Atzberger C. 2008.** Spectral characteristics of
556 salt-affected soils: a laboratory experiment. Geoderma **145**:196-206. DOI
557 10.1016/j.geoderma.2008.03.011
558 **Gao X, Huo Z, Bai Y, Feng S, Huang G, Shi H, Qu Z. 2015.** Soil salt and groundwater change
559 in flood irrigation field and uncultivated land: a case study based on 4-year field
560 observations. *Environmental Earth Sciences* **73**:2127-2139. DOI 10.1007/s12665-014-
561 3563-4
562 **Gomez C, Viscarra Rossel RA, McBratney AB. 2008.** Soil organic carbon prediction by
563 hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study.
564 *Geoderma* **146**:403-411. DOI 10.1016/j.geoderma.2008.06.011
565 **Graciela M, Alfred Z. 2009.** *Remote Sensing of Soil Salinization: Impact on Land Management*.
566 Boca Raton: CRC Press.
567 **Hong Y, Chen Y, Yu L, Liu Y, Liu Y, Zhang Y, Liu Y, Cheng H. 2018a.** Combining
568 fractional order derivative and spectral variable selection for organic matter estimation of
569 homogeneous soil samples by VIS–NIR spectroscopy. *Remote Sensing* **10**:479. DOI
570 10.3390/rs10030479
571 **Hong Y, Yu L, Chen Y, Liu Y, Liu Y, Liu Y, Cheng H. 2018b.** Prediction of soil organic
572 matter by VIS–NIR spectroscopy using normalized soil moisture index as a proxy of soil
573 moisture. *Remote Sensing* **10**:28. DOI 10.3390/rs10010028
574 **Ji W, Adamchuk VI, Biswas A, Dhawale NM, Sudarsan B, Zhang Y, Viscarra Rossel RA,**
575 **Shi Z. 2016.** Assessment of soil properties in situ using a prototype portable MIR
576 spectrometer in two agricultural fields. *Biosystems Engineering* **152**:14-27. DOI
577 10.1016/j.biosystemseng.2016.06.005
578 **Jiang H, Shu H, Lei L, Xu J. 2017.** Estimating soil salt components and salinity using
579 hyperspectral remote sensing data in an arid area of China. *Journal of Applied Remote*
580 *Sensing* **11**:16043. DOI 10.1117/1.JRS.11.016043
581 **Jin P, Li P, Wang Q, Pu Z. 2015.** Developing and applying novel spectral feature parameters
582 for classifying soil salt types in arid land. *Ecological Indicators* **54**:116-123. DOI
583 10.1016/j.ecolind.2015.02.028
584 **Kennard RW, Stone LA. 1969.** Computer aided design of experiments. *Technometrics* **11**:137-
585 148. DOI 10.1080/00401706.1969.10490666
586 **Li J, Pu L, Han M, Zhu M, Zhang R, Xiang Y. 2014.** Soil salinization research in China:
587 Advances and prospects. *Journal of Geographical Sciences* **24**:943-960. DOI
588 10.1007/s11442-014-1130-2
589 **Li M, Li X, Tian Y, Wu B, Zhang S. 2016.** Grey relation estimating pattern of soil organic
590 matter with residual modification based on hyper-spectral data. *The Journal of Grey System*

28:27-39.

Liu S, Yang Y, Wu L. 2015. *Grey system theory and its application*. Beijing: Science Press. (in Chinese).

Maimaitiyiming M, Ghulam A, Bozzolo A, Wilkins JL, Kwasniewski MT. 2017. Early detection of plant physiological responses to different levels of water stress using reflectance spectroscopy. *Remote Sensing* **9**:745. DOI 10.3390/rs9070745

Metternicht GI, Zinck JA. 2003. Remote sensing of soil salinity: potentials and constraints. *Remote Sensing of Environment* **85**:1-20. DOI 10.1016/S0034-4257(02)00188-8

Munns R. 2002. Comparative physiology of salt and water stress. *Plant, Cell and Environment* **25**:239-250. DOI 10.1046/j.0016-8025.2001.00808.x

Nawar S, Buddenbaum H, Hill J. 2015. Estimation of soil salinity using three quantitative methods based on visible and near-infrared reflectance spectroscopy: a case study from Egypt. *Arabian Journal of Geosciences* **8**:5127-5140. DOI 10.1007/s12517-014-1580-y

Nawar S, Buddenbaum H, Hill J, Kozak J. 2014. Modeling and mapping of soil salinity with reflectance spectroscopy and landsat data using two quantitative methods (PLSR and MARS). *Remote Sensing* **6**:10813-10834. DOI 10.3390/rs61110813

Oussama A, Elabadi F, Platikanov S, Kzaiber F, Tauler R. 2012. Detection of olive oil adulteration using FT-IR spectroscopy and PLS with variable importance of projection (VIP) scores. *Journal of the American Oil Chemists Society* **89**:1807-1812. DOI 10.1007/s11746-012-2091-1

Peng J, Ji W, Ma Z, Li S, Chen S, Zhou L, Shi Z. 2016. Predicting total dissolved salts and soluble ion concentrations in agricultural soils using portable visible near-infrared and mid-infrared spectrometers. *Biosystems Engineering* **152**:94-103. DOI 10.1016/j.biosystemseng.2016.04.015

Peng X, Xu C, Zeng W, Wu J, Huang J. 2016. Elimination of the soil moisture effect on the spectra for reflectance prediction of soil salinity using external parameter orthogonalization method. *Journal of Applied Remote Sensing* **10**:15014. DOI 10.1117/1.JRS.10.015014

Periasamy S, Shanmugam RS. 2017. Multispectral and microwave remote sensing models to survey soil moisture and salinity. *Land Degradation & Development* **28**:1412-1425. DOI 10.1002/ldr.2661

Qi H, Tarin P, Arnon K, Li S. 2017. Linear multi-task learning for predicting soil properties using field spectroscopy. *Remote Sensing* **9**:1099. DOI 10.3390/rs9111099

Qu Y, Duan X, Gao H, Chen A, An Y, Song J, Zhou H, He T. 2009. Quantitative retrieval of soil salinity using hyperspectral data in the region of Inner Mongolia Hetao Irrigation District. *Spectroscopy and Spectral Analysis* **29**:1362-1366. DOI 10.3964/j.issn.1000-0593(2009)05-1362-05

627 **Savitzky A, Golay MJE. 1964.** Smoothing and differentiation of data by simplified least squares
628 procedures. *Analytical Chemistry* **36**:1627-1639. DOI 10.1021/ac60214a047
629 **Schofield RV, Kirkby MJ. 2003.** Application of salinization indicators and initial development
630 of potential global soil salinization scenario under climatic change. *Global Biogeochemical*
631 *Cycles* **17**:1-13. DOI 10.1029/2002GB001935
632 **Shahid S, Rahman K. 2011.** *Soil salinity development, classification, assessment and*
633 *management in irrigated agriculture*. Boca Raton: CRC Press.
634 **Stenberg B, Viscarra Rossel RA, Mouazen AM, Wetterlind J. 2010.** Chapter five-visible and
635 near infrared spectroscopy in soil science. In: Donald LS, ed. *Advances in agronomy*.
636 Burlington: Academic Press, 163-215. DOI 10.1016/S0065-2113(10)07005-7
637 **Tavakkoli E, Fatehi F, Coventry S, Rengasamy P, McDonald GK. 2011.** Additive effects of
638 Na^+ and Cl^- ions on barley growth under salinity stress. *Journal of Experimental Botany*
639 **62**:2189-2203. DOI 10.1093/jxb/erq422
640 **The National Soil Survey Office. 1998.** *Soils of China*. Beijing: China Agriculture Press. (in
641 Chinese).
642 **Urdanoz V, Aragüés R. 2011.** Pre- and post-irrigation mapping of soil salinity with
643 electromagnetic induction techniques and relationships with drainage water salinity. *Soil*
644 *Science Society of America Journal* **75**:207-215. DOI 10.2136/sssaj2010.0041
645 **Vapnik VN. 1995.** *The nature of statistical learning theory*. New York: Springer-Verlag.
646 **Viscarra Rossel RA, Behrens T. 2010.** Using data mining to model and interpret soil diffuse
647 reflectance spectra. *Geoderma* **158**:46-54. DOI 10.1016/j.geoderma.2009.12.025
648 **Viscarra Rossel RA, Taylor HJ, McBratney AB. 2007.** Multivariate calibration of
649 hyperspectral γ-ray energy spectra for proximal soil sensing. *European Journal of Soil*
650 *Science* **58**:343-353. DOI 10.1111/j.1365-2389.2006.00859.x
651 **Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO. 2006.** Visible,
652 near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous
653 assessment of various soil properties. *Geoderma* **131**:59-75. DOI
654 10.1016/j.geoderma.2005.03.007
655 **Viscarra Rossel RA, Webster R. 2012.** Predicting soil properties from the Australian soil
656 visible-near infrared spectroscopic database. *European Journal of Soil Science* **63**:848-860.
657 DOI 10.1111/j.1365-2389.2012.01495.x
658 **Volkan Bilgili A, van Es HM, Akbas F, Durak A, Hively WD. 2010.** Visible-near infrared
659 reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey.
660 *Journal of Arid Environments* **74**:229-238. DOI 10.1016/j.jaridenv.2009.08.011
661 **Wang H, Jiang T, John A Y, Li Y, Tian T, Wang J. 2018a.** Hyperspectral inverse model for
662 soil salt ions based on support vector machine. *Transactions of the Chinese Society for*

Agricultural Machinery **49**:263-270. DOI 10.6041/j.issn.1000-1298.2018.05.031 (in Chinese).

Wang H, Zhang Z, Arnon K, Chen J, Han W. 2018b. Hyperspectral estimation of desert soil organic matter content based on gray correlation-ridge regression model. *Transactions of the Chinese Society of Agricultural Engineering* **34**:124-131. DOI 10.11975/j.issn.1002-6819.2018.14.016 (in Chinese).

Wang J, Ding J, Abulimiti A, Cai L. 2018c. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (VIS–NIR) spectroscopy, Ebinur Lake Wetland, Northwest China. *PeerJ* **6**:e4703. DOI 10.7717/peerj.4703

Wang J, Tiyip T, Ding J, Zhang D, Liu W, Wang F, Tashpolat N. 2017. Desert soil clay content estimation using reflectance spectroscopy preprocessed by fractional derivative. *PLOS ONE* **12**:e184836. DOI 10.1371/journal.pone.0184836

Wang X, Zhang F, Ding J, Kung H, Latif A, Johnson VC. 2018d. Estimation of soil salt content (SSC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR), Northwest China, based on a Bootstrap-BP neural network model and optimal spectral indices. *Science of The Total Environment* **615**:918-930. DOI 10.1016/j.scitotenv.2017.10.025

Weng Y, Gong P, Zhu Z. 2008. Reflectance spectroscopy for the assessment of soil salt content in soils of the Yellow River Delta of China. *International Journal of Remote Sensing* **29**:5511-5531. DOI 10.1080/01431160801930248

Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58**:109-130. DOI 10.1016/S0169-7439(01)00155-1

Wu J, Vincent B, Yang J, Bouarfa S, Vidal A. 2008. Remote sensing monitoring of changes in soil salinity: a case study in Inner Mongolia, China. *Sensors* **8**:7035-7049. DOI 10.3390/s8117035

Xia N, Tiyip T, Kelimu A, Nurmemet I, Ding J, Zhang F, Zhang D. 2017. Influence of fractional differential on correlation coefficient between EC1:5 and reflectance spectra of saline soil. *Journal of Spectroscopy* **2017**:1-11. DOI 10.1155/2017/1236329

Xiao Z, Li Y, Feng H. 2016a. Hyperspectral models and forcasting of physico-chemical properties for salinized soils in northwest China. *Spectroscopy and Spectral Analysis* **36**:1615-1622. DOI 10.3964/j.issn.1000-0593(2016)05-1615-08

Xiao Z, Li Y, Feng H. 2016b. Modeling soil cation concentration and sodium adsorption ratio using observed diffuse reflectance spectra. *Canadian Journal of Soil Science* **96**:372-385. DOI 10.1139/cjss-2016-0002

Xu C, Zeng W, Huang J, Wu J, van Leeuwen W. 2016. Prediction of soil moisture content and soil salt concentration from hyperspectral laboratory and field data. *Remote Sensing*

8:42. DOI 10.3390/rs8010042

Yang X, Yu Y. 2017. Estimating soil salinity under various moisture conditions: an experimental study. *IEEE Transactions on Geoscience and Remote Sensing* **55**:2525-2533. DOI 10.1109/TGRS.2016.2646420

Yu R, Liu T, Xu Y, Zhu C, Zhang Q, Qu Z, Liu X, Li C. 2010. Analysis of salinization dynamics by remote sensing in Hetao Irrigation District of North China. *Agricultural Water Management* **97**:1952-1960. DOI 10.1016/j.agwat.2010.03.009

Zhang Z, Wang H, Arnon K, Chen J, Han W. 2018. Inversion of soil moisture content from hyperspectra based on ridge regression. *Transactions of the Chinese Society for Agricultural Machinery* **49**:240-248. DOI 10.6041/j.issn.1000-1298.2018.05.028 (in Chinese)

Table 1(on next page)

Descriptive statistics of soil water-soluble salt ions content.

Statistical index	Minimum/（g•kg^{-1}）	Maximum/（g•kg^{-1}）	Mean/（g•kg^{-1}）	Standard deviation	Coefficient of variation/%
CO_3^{2-}	0.000	0.066	0.020	0.020	98.86
HCO_3^-	0.171	0.666	0.316	0.099	31.27
SO_4^{2-}	0.047	40.892	9.073	10.828	119.34
Cl^-	0.145	23.234	4.825	4.711	97.65
Ca^{2+}	0.08	4.111	0.697	0.669	95.95
Mg^{2+}	0.039	1.952	0.706	0.606	85.91
K^+	0.001	5.727	0.936	1.358	145.14
Na^+	0.016	23.035	5.014	5.563	110.94

1

Table 2(on next page)

Max correlation coefficient and band intervals of soil water-soluble salt ions content with standard normal variable reflectance.

Water-soluble salt ions	Number of significant bands	Maximum correlation coefficient	Maximum correlation band intervals/nm
Ca^{2+}	190	-0.877	1940～1950
Cl^{-}	192	-0.882	1990～2000
CO_3^{2-}	146	0.552	1870～1880
HCO_3^{-}	1	0.235	2200～2210
K^{+}	178	0.630	1850～1860
Mg^{2+}	186	-0.848	1990～2000
Na^{+}	181	-0.752	2010～2020
SO_4^{2-}	178	0.749	1860～1870

1

Table 3(on next page)

Max gray correlation degree and band intervals of soil water-soluble salt ions content with standard normal variable reflectance.

Water-soluble salt ions	Sensitive band numbers	Maximum gray correlation degree	Maximum gray correlation degree intervals/nm
Ca^{2+}	53	0.551	1650~1660
Cl^{-}	101	0.561	1650~1660
CO_3^{2-}	14	0.416	1740~1750
HCO_3^{-}	105	0.465	560~570
K^{+}	15	0.470	1650~1660
Mg^{2+}	110	0.559	1650~1660
Na^{+}	36	0.508	1650~1660
SO_4^{2-}	21	0.494	1650~1660

1

Table 4(on next page)

Parameter indexes of feature band intervals selection by stepwise regression method.

Water-soluble salt ions	Sensitive band numbers	Band intervals/nm	Adjusted R^2	Standard error	Sig.
Ca^{2+}	7	1040～1050，1090～1100，1900～1910，1920～1930，2200～2210，2310～2320，2370～2380	0.942	0.529	＜0.001
Cl^-	8	730～740，910～920，1890～1900，1970～1980，1990～2000，2180～2190，2200～2210，2290～2300	0.975	1.063	＜0.001
CO_3^{2-}	4	1280～1290，1360～1370，1380～1390，1420～1430	0.836	0.012	＜0.001
HCO_3^-	3	2200～2210，2260～2270，2290～2300	0.934	0.085	＜0.001
K^+	6	740～750，810～820，1160～1170，1890～1900，2210～2220，2390～2400	0.817	0.706	＜0.001
Mg^{2+}	6	1130～1140，1930～1950，1990～2000，2100～2110，2170～2180	0.973	0.152	＜0.001
Na^+	6	740～750，820～830，1860～1870，2210～2220，2260～2270，2390～2400	0.942	1.812	＜0.001
SO_4^{2-}	6	610～620，1140～1150，1960～1970，2210～2220，2290～2300，2390～2400	0.947	3.255	＜0.001

1

Table 5(on next page)

Max VIP scores and band intervals of soil water-soluble salt ions content and standard normal variable reflectance.

Water-soluble salt ions	Sensitive band numbers	Maximum VIP scores	Maximum VIP scores intervals/nm
Ca^{2+}	69	1.97	1440～1450
Cl^{-}	85	1.42	560～570
CO_3^{2-}	67	2.01	1440～1450
HCO_3^{-}	79	2.37	1410～1420
K^{+}	69	1.73	1880～1890
Mg^{2+}	69	1.49	1870～1880
Na^{+}	83	1.55	1880～1890
SO_4^{2-}	74	1.74	1880～1890

1

Table 6(on next page)

Calibration and validation results of soil water-soluble salt ions content from the PLSR inversion models using GC, SR and VIP wavelength selection methods.

Wavelength selection methods	Water-soluble salt ions	Latent variables	Calibration sets	Validation sets		
			R_c^2	R_p^2	RMSE/(g·kg^{-1})	RPD
Gray correlation	Ca^{2+}	7	0.897	0.724	0.362	1.71
	Cl^-	7	0.796	0.565	3.150	1.35
	CO_3^{2-}	5	0.660	0.649	0.012	1.21
	HCO_3^-	7	0.646	0.285	0.088	0.96
	K^+	1	0.388	0.258	1.209	0.85
	Mg^{2+}	6	0.891	0.767	0.295	1.99
	Na^+	7	0.840	0.805	2.589	1.88
	SO_4^{2-}	4	0.561	0.360	8.711	0.87
Stepwise regression	Ca^{2+}	7	0.965	0.937	0.168	3.95
	Cl^-	2	0.861	0.729	2.434	1.80
	CO_3^{2-}	4	0.685	0.742	0.010	1.80
	HCO_3^-	3	0.340	0.154	0.094	0.64
	K^+	5	0.722	0.563	0.931	1.37
	Mg^{2+}	4	0.933	0.849	0.236	2.52
	Na^+	3	0.901	0.868	2.145	2.67
	SO_4^{2-}	5	0.918	0.889	3.807	2.75
Variable importance in projection	Ca^{2+}	3	0.909	0.865	0.249	2.57
	Cl^-	4	0.930	0.862	1.725	2.48
	CO_3^{2-}	9	0.865	0.617	0.012	1.44
	HCO_3^-	9	0.704	0.263	0.090	0.93
	K^+	5	0.664	0.566	0.945	1.43
	Mg^{2+}	3	0.910	0.840	0.243	2.34
	Na^+	8	0.939	0.902	1.801	3.15
	SO_4^{2-}	8	0.919	0.872	4.038	2.75

1

Table 7 (on next page)

Calibration and validation results of soil water-soluble salt ions content from the SVR inversion models using GC, SR and VIP wavelength selection methods.

Wavelength selection methods	Water-soluble salt ions	Calibration sets	Validation sets		
		R_c^2	R_p^2	RMSE/(g·kg^{-1})	RPD
Gray correlation	Ca^{2+}	0.910	0.752	0.337	1.73
	Cl^-	0.652	0.500	3.275	1.05
	CO_3^{2-}	0.688	0.664	0.012	1.14
	HCO_3^-	0.563	0.328	0.083	0.70
	K^+	0.421	0.269	1.155	0.61
	Mg^{2+}	0.934	0.781	0.289	2.07
	Na^+	0.809	0.764	2.851	1.85
	SO_4^{2-}	0.565	0.397	9.046	0.52
Stepwise regression	Ca^{2+}	0.964	0.940	0.164	3.97
	Cl^-	0.893	0.790	2.186	2.15
	CO_3^{2-}	0.605	0.583	0.013	1.16
	HCO_3^-	0.327	0.164	0.095	0.56
	K^+	0.717	0.578	0.874	1.26
	Mg^{2+}	0.936	0.875	0.214	2.75
	Na^+	0.903	0.864	2.171	2.61
	SO_4^{2-}	0.915	0.893	3.862	2.71
variable importance in projection	Ca^{2+}	0.960	0.935	0.173	3.93
	Cl^-	0.949	0.897	1.483	2.98
	CO_3^{2-}	0.883	0.664	0.012	1.56
	HCO_3^-	0.669	0.280	0.088	0.91
	K^+	0.645	0.565	0.888	1.23
	Mg^{2+}	0.965	0.877	0.214	2.51
	Na^+	0.958	0.872	2.211	2.76
	SO_4^{2-}	0.914	0.865	4.106	2.48

1

Figure 1

Distribution of sampling sites in the study area.

(A) Location map of Shahaoqu Irrigation Area. (B) Sampling location in Shahaoqu Irrigation Area.

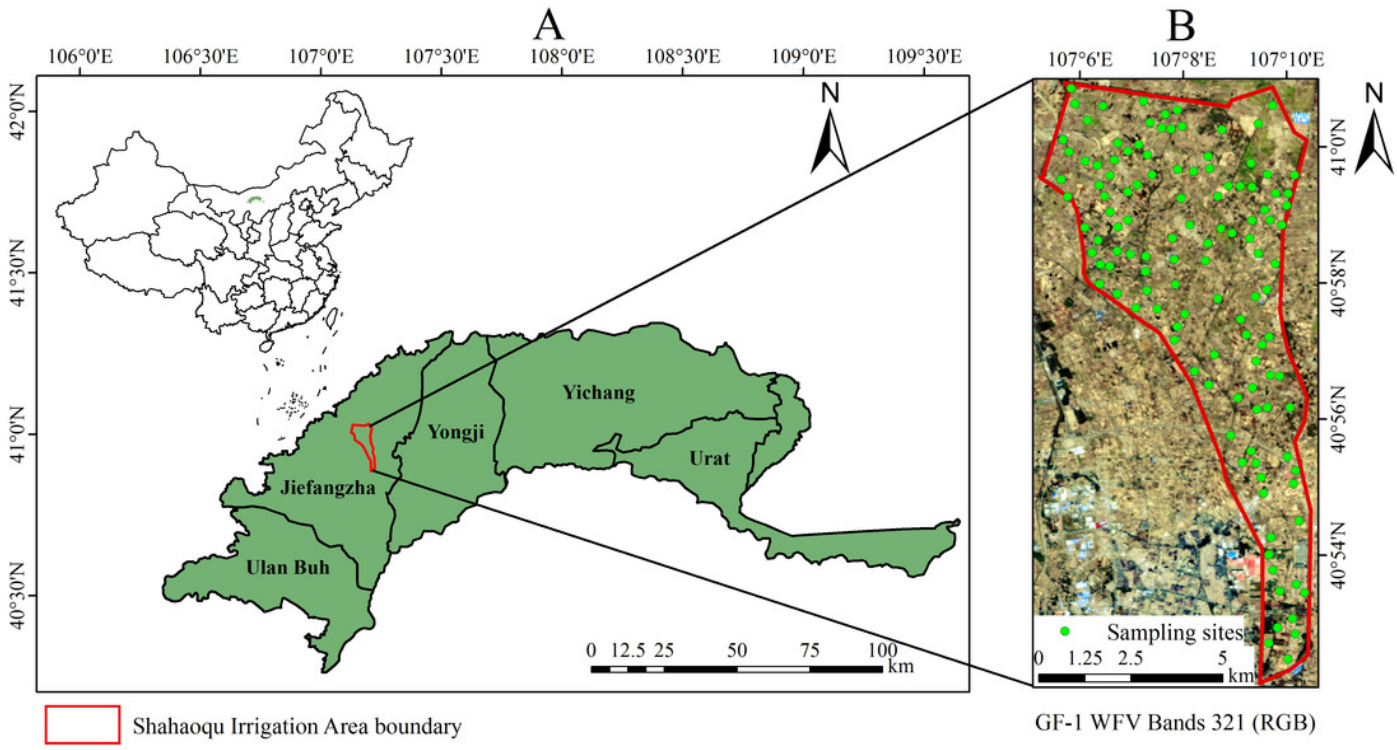

Figure 2

Spectral curves of all soil samples.

(A) Reflectance spectral curves. (B) Standard normal variable reflectance curves.

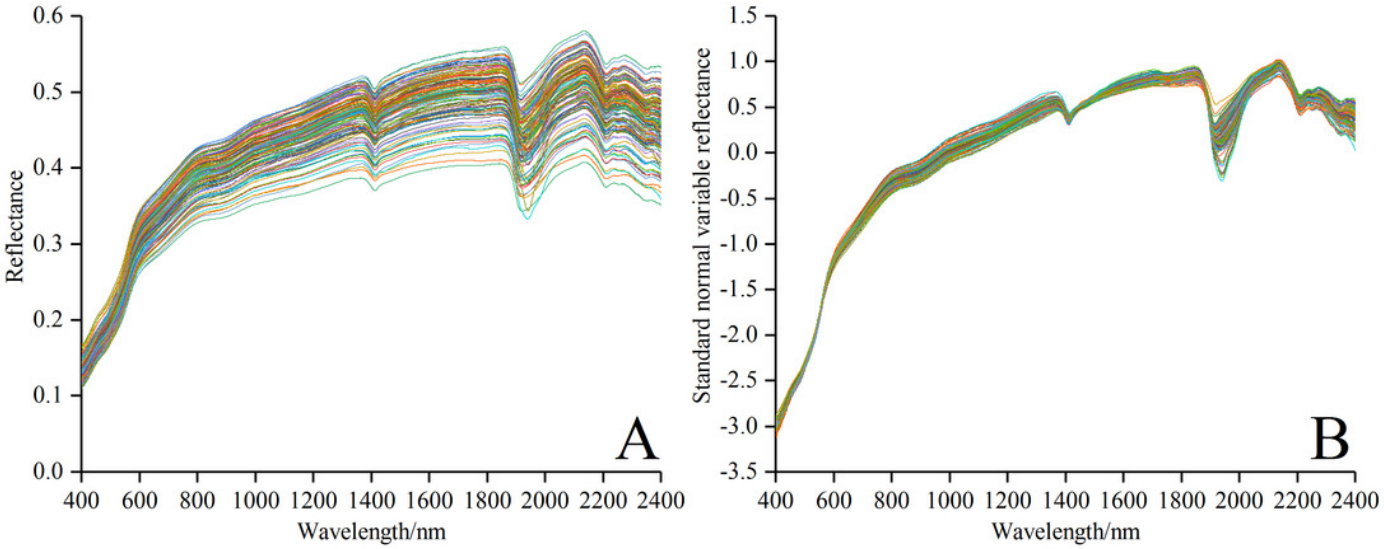

Figure 3

Correlation coefficients of soil water-soluble salt ions content with standard normal variable reflectance.

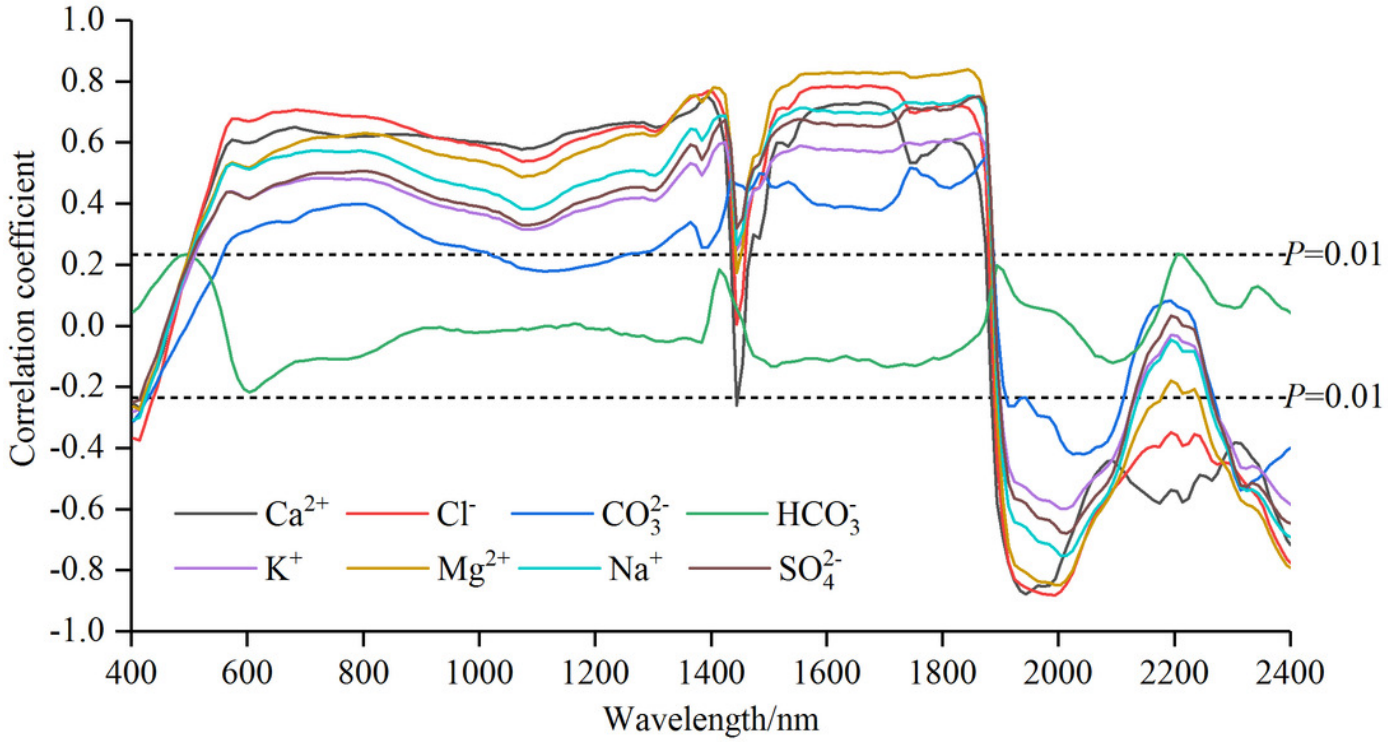

Figure 4

Gray correlation degree (GCD) for soil water-soluble salt ions content with standard normal variable reflectance.

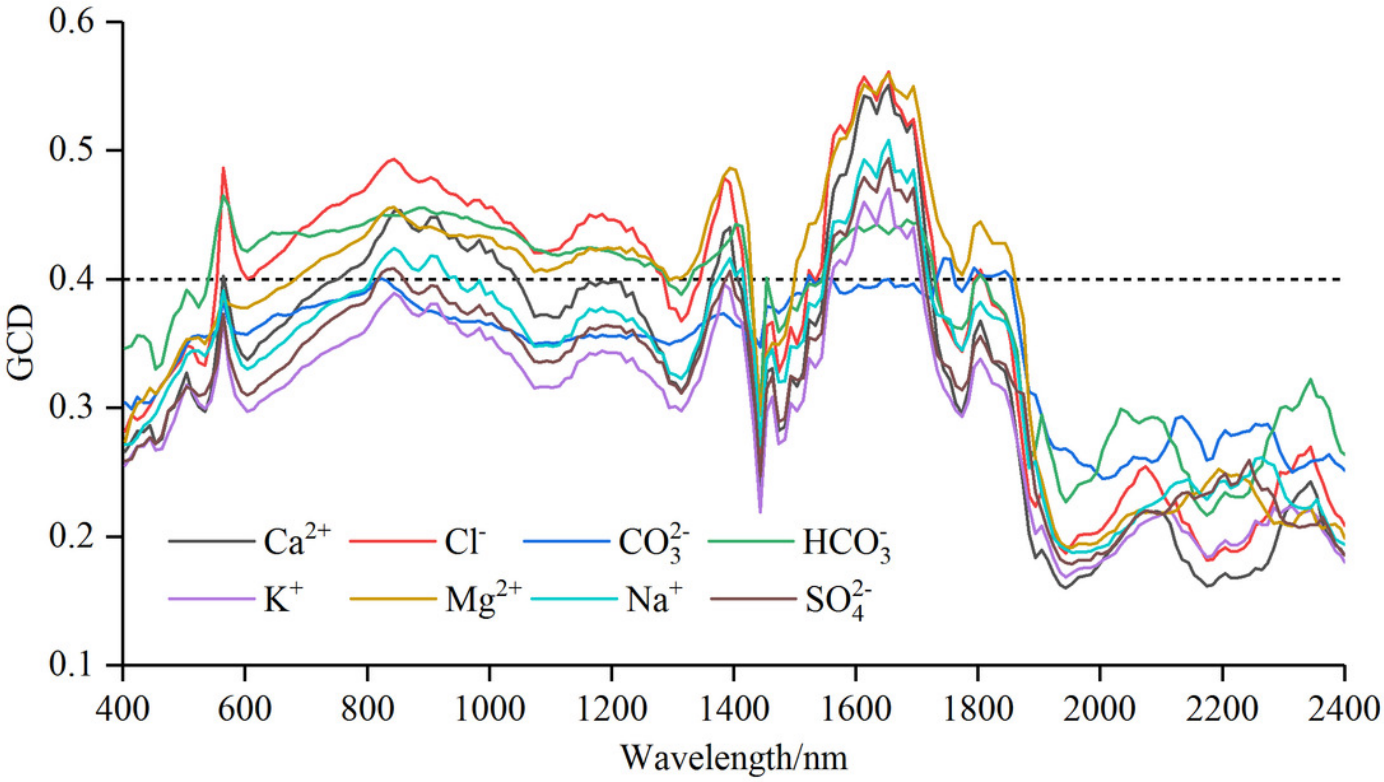

Figure 5

The Variable importance in projection (VIP) scores for soil water-soluble salt ions content with standard normal variable reflectance.

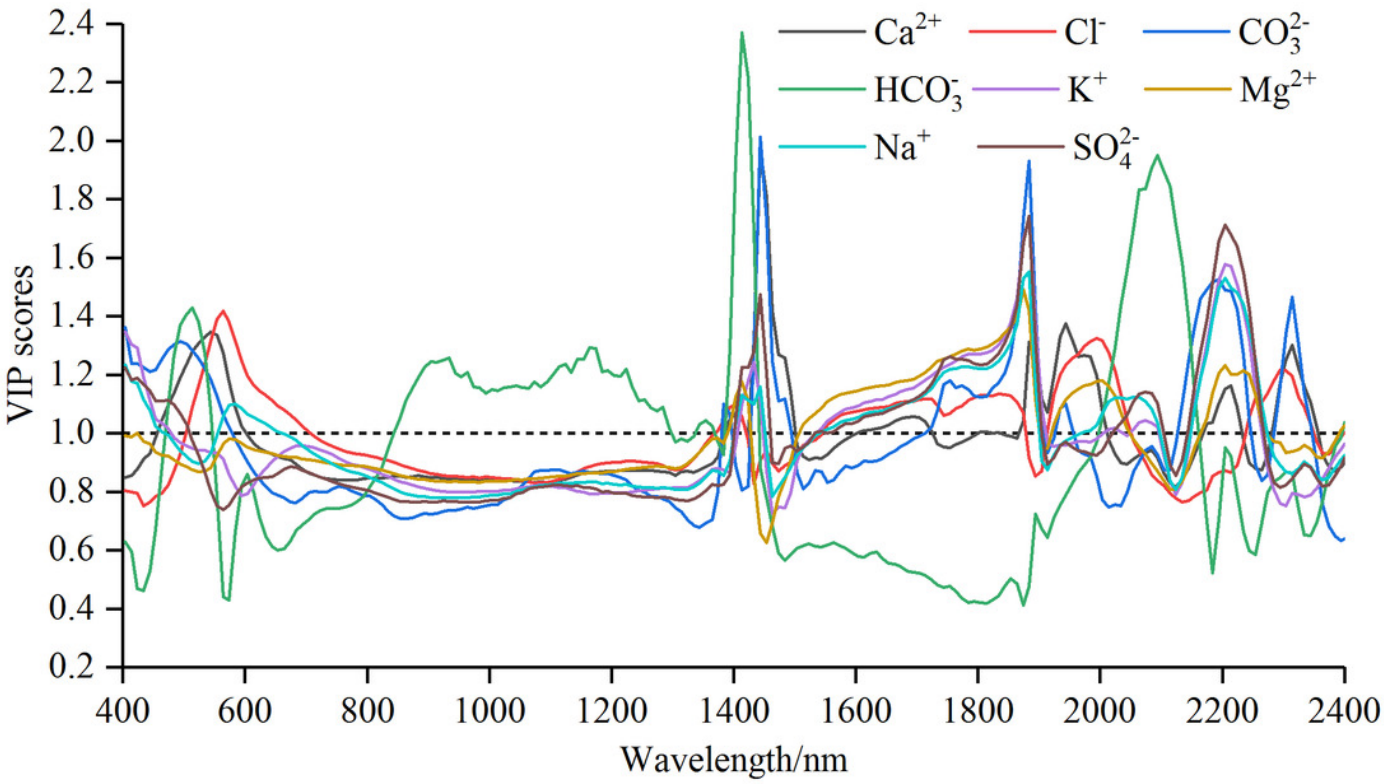

Figure 6

Validation of soil water-soluble salt ions content based on the best model.

(A) Ca^{2+} with SR-SVR model. (B) Cl^{-} with VIP-SVR model. (C) CO_3^{2-} with SR-PLSR model. (D) HCO_3^{-} with GC-PLSR model. (E) K^{+} with VIP-PLSR model. (F) Mg^{2+} with SR-SVR model. (G) Na^{+} with VIP-PLSR model. (H) SO_4^{2-} with VIP-PLSR model.