

Determination of in-situ salinized soil moisture content from visible-near infrared (VIS–NIR) spectroscopy by fractional order derivative and spectral variable selection algorithms

Congcong Lao^{1,2}, Zhitao Zhang^{1,2*}, Junying Chen^{1,2*}, Haorui Chen⁴, Zhihua Yao^{1,2},
Zheng Xing^{1,2}, Xiang Tai^{1,2}, Jifeng Ning⁵, Yinwen Chen³

(1. Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas, Ministry of Education, Northwest A&F University, Yangling, 712100, China;

2. College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, 712100, China;

3. Department of Foreign Languages, Northwest A&F University, Yangling, 712100, China;

4. Department of Irrigation and Drainage, China Institute of Water Resources and Hydropower Research, Beijing 100048, China;

5. College of Information Engineering, Northwest A&F University, Yangling, 712100, China)

Abstract: The measurement of salinized soil moisture content (SSMC) is essential to precise irrigation and avoidance of secondary salinization. Visible and near infrared (VIS–NIR) spectroscopy has been effectively used to estimate soil moisture content (SMC) but not for SSMC. The direct application of in-situ VIS–NIR spectroscopy to the estimation of SSMC can help save a large amount of time and labor, but the in-situ VIS–NIR was interfered by many factors, such as soil texture, soil surface debris and environmental temperature. Spectral derivatives can be used to eliminate unnecessary interference for optimal spectral information, but traditional integer derivatives (i.e. first and second derivatives) often ignored some spectral information due to different integer order differential curves were obviously different. In addition, the full spectrum usually contains redundant spectral variables. These variables would affect the accuracy and estimation velocity of the model. Different combinations of fractional order derivative (FOD) and spectral variable selection algorithms (i.e. variable importance projection (VIP), competitive adaptive weighted sampling (CARS) and random frog algorithm (RFA)) may offer some alternative solutions to these problems. In order to test the effects of these combinations on VIS–NIR spectral model optimization, we measured the in-situ soil spectra of 163 sites in Shahaoqu Irrigation Area, Inner Mongolia, China. Meanwhile, we collected soil samples and measured their SSMC and soil salt content (SSC). Then the Extreme Learning Machine (ELM) model was applied to the SSMC estimation. The results showed that SSC and SSMC had obvious effects on in-situ spectra. With the increase of differential order, the spectral resolution increased gradually, but the spectral intensity decreased at the same time. So, the spectral information may not increase. However, FOD can balance the contradiction between spectral resolution and spectral intensity. The estimation of ELM models based on 0.75 order derivatives that is the most accurate among the full spectrum ELM models. The coefficient of determination (R^2) was 0.83 and ratio of the performance to deviation (RPD) was 2.44. In all the models (twenty-seven different combinations of FOD and variable selection algorithms), the best model was based on the combination of 0.75 derivative spectrum and random frog algorithm ($R^2 = 0.94$, RPD = 3.80). The results of this study also confirmed that the combination of RFA and FOD could effectively improve the accuracy of the in-situ spectral estimation of SSMC. However, VIP was chosen as an alternative due to computational efficiency.

Keywords: salinized soil moisture content, in-situ visible and near-infrared spectroscopy, fractional order derivative, random frog algorithm, extreme learning machine

DOI: 10.33440/j.ijpaa.20200303.98

Citation: Lao C C, Zhang Z T, Chen J Y, Chen H R, Yao Z H, Xing Z, Tai X, Ning J F, Chen Y W. Determination of in-situ salinized soil moisture content from visible-near infrared (VIS–NIR) spectroscopy by fractional order derivative and spectral variable selection algorithms. Int J Precis Agric Aviat, 2020; 3(3): 21–34.

1 Introduction

Soil moisture content (SMC), as an important component of

soil, is of great significance to heat balance, soil temperature, plant growth and so on^[1]. It has become a common sense that salt

Received date: 2020-07-06 **Accepted date:** 2020-09-06

Biographies: **Congcong Lao**, Master Student, research interests: Application of remote sensing technology in water-saving irrigation and water resources, Email: laocc365@163.com; **Haorui Chen**, Senior Engineer, research interests: Irrigation and drainage, Email: chenhr@iwhr.com; **Zhihua Yao**, Master Student, research interests: Application of remote sensing technology in water-saving irrigation and water resources, Email: yaozhihua@nwfau.edu.cn; **Zheng Xing**, Master Student, research interests: Application of remote sensing technology in water-saving irrigation and water resources, Email: yizhi199708@163.com; **Xiang Tai**, Master Student, research interests: Application of remote sensing technology in water-saving irrigation and water resources, Email: taixiang23@

163.com; **Jifeng Ning**, PhD, Professor, research interests: Image processing algorithm, Email: njf@nwsuaf.edu.cn; **Yinwen Chen**, lecturer, research interests: Language polishing, Email: martinchen@nwfau.edu.cn.

* **Corresponding author: Zhitao Zhang**, PhD, Associate Professor, research interests: Application of remote sensing technology in water-saving irrigation and water resources, Northwest A&F University, Xinong Road, Yangling District, Xianyang City, 712100, China. Email: zhitaozhang@126.com. **Junying Chen**, PhD, Professor, research interests: Water conservancy engineering and agricultural water and soil engineering research, Northwest A&F University, Xinong Road, Yangling District, Xianyang City, 712100, China. Email: cjoyrose@126.com.

flows with water. Soil salinity is closely related to the moisture content. In the arid area of northern China, flood irrigation in farmland is extremely water-wasting, and prone to secondary salinization^[2]. Therefore, agricultural scholars have endeavored to develop precision irrigation, the key to which is the accurate, quick and low-cost monitoring of saline soil moisture content.

Accurate ex-situ measurements of soil moisture content (SMC) can be conducted in the laboratory (drying method)^[3]. However, the main problem of this traditional method is water evaporation of soil samples during the transportation and storage of samples, so it can be classified as discontinuous measurement of SMC^[3]. The SMC measured by this method cannot represent the real soil moisture in the field. Compared to the ex-situ, in-situ SMC measurement is a more accurate and reliable alternative. There are many methods for in-situ SMC measurement, such as time domain reflectometry, capacitive probe, and electrical resistivity tomography^[4-7]. However, these methods have drawbacks in the measurement cycle and soil nondestructive testing. Because these measurement methods all invade the soil, causing damage to the soil structure. Moreover, the accuracy of many traditional measurement instruments would be affected by water soluble salt ions, and it is impossible to accurately measure the saline soil moisture content (SSMC) in salinized areas. Visible-near infrared (VIS-NIR) spectroscopy is a promising method for measuring in-situ SMC, and it characterizes objects based on their reflectance in the wavelength range of 350 to 2500 nm^[7]. VIS-NIR spectroscopy has turned out to be an effective substitute for SSMC prediction in the laboratory. According to Wang et al., soil salinity had some negative effect on the quantitative SSMC prediction, so they used the salt-resistant spectral index for such prediction^[8]. Although Wang et al used the in-situ soil dataset to verify that the highest precision coefficient of determination (R^2) = 0.72, it was still difficult to accurately quantify the SSMC^[8]. Therefore, it is necessary to further determine the salt-resistant spectral band, and further improve the accuracy of VIS-NIR spectrum to detect the in-situ SSMC.

In order to improve the estimation accuracy of the in-situ SSMC, it is often necessary to preprocess the spectral data^[9]. The common preprocessing methods include data transformation, spectral smoothing, scattering correction, spectral derivative and so on. The preprocessing of the original reflection spectrum can improve the signal-to-noise ratio, facilitate the spectral feature and optimal variable extraction for subsequent analysis and modeling^[10]. Spectral derivative is a highly applicable pretreatment method, which is widely used in the fields of soil, vegetation and chemical spectrums^[11-13]. However, greatly different in morphology from the original spectral reflectance curve, the traditional integer-order derivative curve (first derivative and second derivative) may neglect useful information on such monitored property as the band information sensitive to the SSMC. The Fractional Order Derivative (FOD) algorithm, a mathematical method for reflectivity analysis, allows interpolation between integer-order derivatives to extract finer details from the spectrum^[14]. In recent years, FOD have been widely used in soil near-infrared spectroscopy, especially in monitoring soil organic matter and soil salinity^[15-17]. So far, few studies have applied the FOD algorithm to SSMC monitoring.

Soil VIS-NIR spectrum is characterized by multi-collinearity, non-specificity and extensiveness. In addition, soil is a complex mixture of substances, and the interaction between different

substances may affect the spectrum. Consequently, the soil properties studied will be obscured, and the general linear regression model will not perform well^[18,19]. It is worth noting that the relationship between spectral data and soil properties (such as soil organic matter content, soil salt content) is largely nonlinear, and especially when the soil is salinized, the linear relationship between moisture and spectrum may be weakened. Such defects of the linear model lead to the development of many nonlinear models. Extreme learning machine (ELM) is one of them, which can solve the nonlinear problems and usually has relatively ideal prediction accuracy^[20-22].

Spectral variable selection is an important method to further optimize the regression model. Many studies have shown that this selection algorithm can reduce the complexity of the calibration model and improve the prediction performance of the model to some extent^[23,24]. Variable selection is used to identify the most informative subset of a large number of variables in order to minimize the number of variables and exclude unreliable variables^[25]. Many of these methods have been well applied in soil spectral modeling, including competitive adaptive reweighted sampling (CARS), variable importance for projection (VIP), successive projections algorithm (SPA), and gray correlation (GC), genetic algorithm (GA), uninformative variable elimination (UVE), etc^[25-30]. For example, Xu et al found that CARS combined with support vector machine can better predict rice root density^[31]; Hong et al. discovered that CARS combined with FOD had the best performance in predicting wet soil organic matter^[9]. Similarly, VIP has a good advantage in variable selection and has been applied in many research fields^[32,33]. Compared with the previous variable selection algorithms, the random frog algorithm (RFA) is a relatively new algorithm based on Reversible Jump Markov Chain Monte Carlo (RJCMC) method^[34]. This algorithm was first proposed for cancer-related gene selection^[35]. The advantage of this algorithm is that it requires neither any harsh mathematical formulas nor some previous distributions to be specified as in the formal RJCMC method, which makes it easier to implement. But this algorithm is rarely used in the field of soil spectroscopy, so it is necessary to test the applicability of RFA to soil spectral variable selection^[36]. Different studies have shown that none of these three methods (i.e., VIP, CARS, RFA) is optimal across different data sets, so we need to systematically test them in the same dataset. In addition, few studies have used variable selection algorithms combined with FOD to estimate in-situ SSMC. Therefore, FOD combined with appropriate variable selection is expected to be used to establish a better prediction model of in-situ SSMC using VIS-NIR spectroscopy.

This study aimed to explore: (1) the effect of SSMC and SSC on soil spectrum; (2) the effect of FOD on the improvement of SSMC prediction accuracy; (3) the applicability of random frog algorithm in soil spectral variable selection; (4) the modeling performance of FOD combined with various variable selection algorithms, and (5) the estimated potential and spatial mapping effect on in-situ SSMC. It is hoped that this study will contribute to the SSMC estimation.

2 Materials & Methods

2.1 Study Area

The area chosen for SMC estimation is called Shahaouq Irrigation Area (SIA), located in Hetao Irrigation District (HID), Inner Mongolia, China (Figure 1). HID is one of the most

important grain and economic crop production bases and the third largest irrigation area in China^[37]. The geographical environment of SIA lies to the north of the Yellow River and the south of Yinshan Mountain. The climate of SIA is typically continental, with severe cold and little snow in winter and high temperature and little rain in summer. The average annual precipitation is 155 mm, which is concentrated in June-August. However, the annual

average evaporation of SIA (20 cm evaporating dish) is 2000 mm. It is a typical area where agriculture fully relies on irrigation. Long-term unreasonable high irrigation and low drainage has led to a high groundwater level in the irrigation area. This problem, coupled with such other factors as climate conditions, soil geology and topography, has made secondary salinization in SIA very severe^[38].

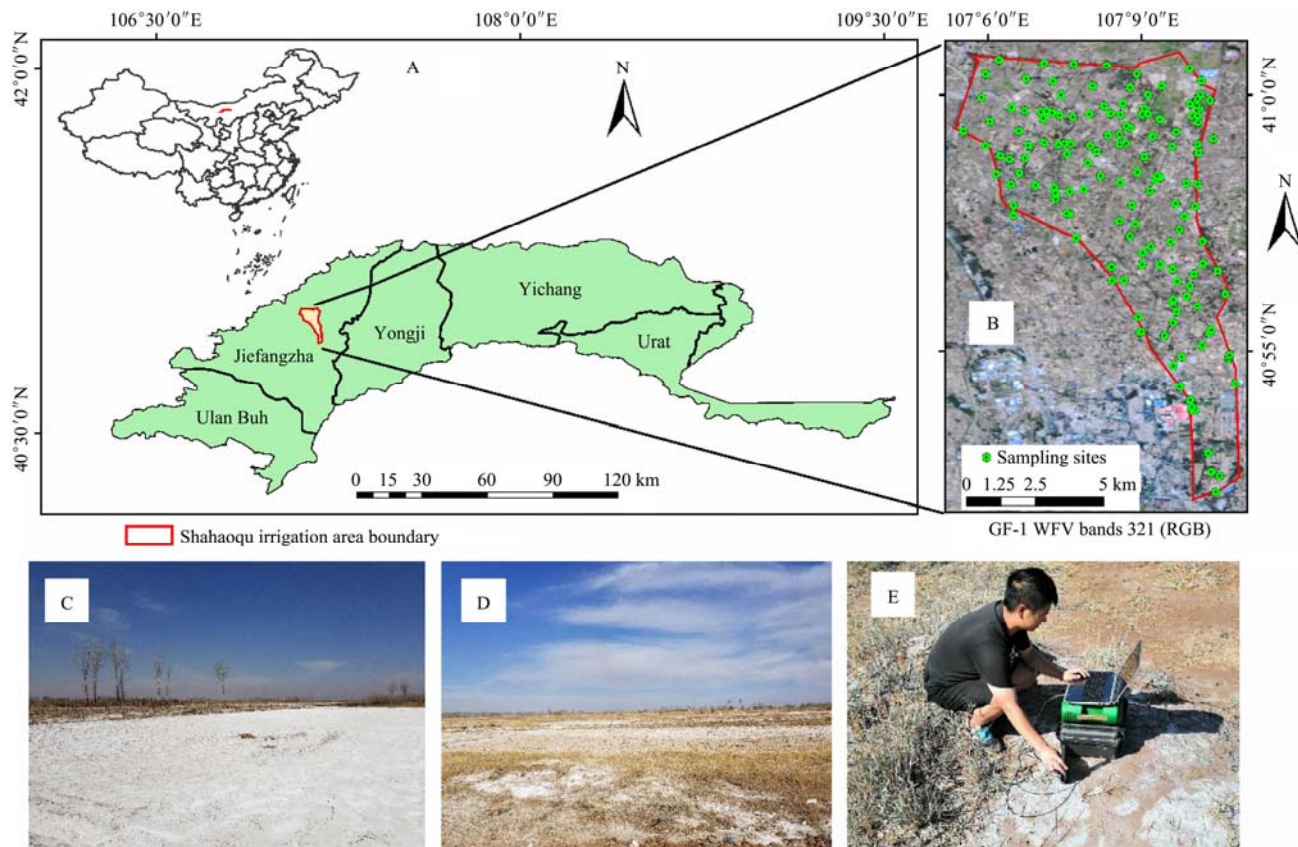


Figure 1 Location maps of the study area (A inner mongolia, B. shahaoqu irrigation area). Photographs of the landscape (C saline soil, D severe salinization) E in-situ field experiment using the Analytical Spectral Device (ASD) FieldSpec®3 spectrometer

2.2 Field Sampling and In-Situ Spectral Acquisition

From April 10 through 20, 2019, based on the typical landscape features of SIA (107°05'00"-107°10'00"E, 40°52'00"-41°00'00"N), 162 sampling units were selected on a grid of 16 m × 16 m, and the sample units were distributed in the cultivated area of the study area (Figure 1). The soil samples and in-situ soil spectral data were collected at the same time. The position of each sampling point was determined and recorded by GPS. The specific operation is as follows.

In this study, the in-situ spectral reflectance of the soil samples were acquired by the spectrometer FieldSpec®3 (350-2500 nm), which is manufactured by Analytical Spectral Devices Company (ASD, <https://mtri.org/asd.html>). The sampling intervals of the spectrometer were 350~1000 nm to 1.4 nm, 1000~2500 nm to 2 nm, and the resampling interval was 1 nm. Soils were scanned with a high intensity ASD contact probe connected to the spectrometer with an optical fiber cable. This probe was used to prevent the disturbance from stray light during the acquisition, which has a viewing area of 2 cm in diameter and built-in halogen light source^[39]. The contact probe touched the soil surface vertically to acquire the relative reflectance (Figure 1E).

In the field measurement process, we would choose a relatively flat and clean soil area to make the probe fully fit the soil surface to avoid the interference of environmental light. To obtain accurate

relative reflectance, a white spectralon panel was used to calibrate the spectrometer before each spectral acquisition. In order to reduce random noise, five repetitive acquisitions were averaged for each soil sample. The effect of dark current was considered. Before the spectrum measurement, we used the dark current removal function of the RS3 software to reduce the impact on the spectrum measurement. During the acquisition of in-situ spectra in the field, it is necessary to avoid hard stones so as to avoid irreversible damage to the probe.

After the in-situ VIS-NIR acquisition, each soil sample was immediately put in two aluminum boxes which were used for the determination of SSMC and SSC respectively. And sealed with one sampling bag to avoid evaporation of soil moisture, labeled and transported to the lab (up to a depth <10 cm).

2.3 SSMC and SSC (Soil Salt Content) Data Acquisition and Pre-Processing

We recorded the number and weight of an aluminum box for each soil sample. Then we put the uncovered aluminum boxes into the drying box and measured the SSMC by the drying method under the constant temperature of 105°C in a continuous period of 24 hours. The formula of SSMC is

$$\text{SSMC} = \frac{M_1 - M_2}{M_2 - M_3} \times 100\% \quad (1)$$

where, M_1 , M_2 and M_3 stand for is the weights of wet soil sample (containing aluminum box), dry soil sample (containing aluminum box) and empty aluminum box, respectively.

In order to acquire SSC data, another group of the same soil samples were air-dried and the impurities were removed. The conductivity of soil solution ($EC_{1.5}$, dS/m) was determined with a conductivity meter (DDS-307A, Shanghai Youke Instrument Company) equipped with the composite electrode (DJS-1C and T-818-B-6) in a 1:5 soil-water extraction solution at room temperature (25°C). Then the SSC (%) was calculated by the following empirical formula^[40]:

$$SSC = (0.2882EC_{1.5} + 0.0183) \quad (2)$$

where, $EC_{1.5}$ is the conductivity of 1:5 soil-water extraction solution.

The edge wavelengths (350-399 nm and 2401-2500 nm) of higher noise in the original spectra were first removed, and the original spectral data were narrowed to 400-2400 nm. Then the smoothing Savitzky-Golay (5 filter widths and second-order polynomial) was used to reduce such effects as the instrument noise and external environment in spectral data collection. The spectrum is resampled at 10 nm intervals to reduce the complexity of the spectral matrix. Therefore, a total of 201 wavebands (400-2400 nm) were acquired for the subsequent analysis.

2.4 Fractional Order Derivative (FOD)

FOD is a branch of mathematical theory that extends the classical integer-order differential to any order differential. There are three main types of FOD algorithm: Grünwald–Letnikov (G-L), Riemann–Liouville and Caputo. G-L is a discrete form of definition, which is convenient for numerical calculations and highly computationally efficient, so this formula was adopted in this study. The formula of G-L is

$$d^q f(x) = \lim_{h \rightarrow 0} \frac{1}{h^q} \sum_{m=0}^{t-a} (-1)^m \frac{\Gamma(q+1)}{m! \Gamma(q-m+1)} f(x-mh) \quad (3)$$

where, q is the differential order; h is the differential step size; t and a are the upper and lower limits of the differential, respectively. The formula contains the gamma function:

$$\Gamma(\beta) = \int_0^{\infty} e^{-t} t^{\beta-1} dt = (\beta-1)! \quad (4)$$

If $f(\lambda)$ is a one-dimensional spectrum and $[a, t]$ is its wavelength interval ($\lambda \in [a, t]$), the wavelength interval is equally divided into n portions by the unit h . Given $h=1$, $n = \frac{(t-a)}{h} = t-a$. Then the formula of the FOD is derived from equation (3) as follows:

$$\begin{aligned} \frac{d^v f(\lambda)}{d\lambda^v} \approx & f(\lambda) + (-v)f(\lambda-1) + \frac{(-v)(-v+1)}{2} f(\lambda-2) + \dots \\ & + \frac{\Gamma(-v+1)}{n! \Gamma(-v+n+1)} f(\lambda-n) \end{aligned} \quad (5)$$

where, v is the differential order; When $v=1$ or 2 , equation (5) is the same as the first and second order differential equation. But when $v=0$, the spectral data are not processed. In this paper, equation (5) was used to calculate the 0~2 differential of soil spectrum (with an interval of 0.25 order). This process was realized by MATLAB R2018a (The Math Works Inc.: Natick, MA, USA).

2.5 Spectral Variable Selection Algorithms

The effect of the model was the accuracy of the prediction. Overly complex models may lead to overfitting. Overfitting models often perform poorly on the validation set. So we needed to make the variables a little less, make the model simpler, and

sacrificed the accuracy of some models on the calibration set. On the contrary, it may improve the performance of the model on the validation set. So, three spectral variable selection algorithms were used to extract sensitive informative spectral variables, including VIP, CARS and RFA.

VIP is a variable selection method based on partial least squares regression (PLSR). VIP selects variables by integrating the importance, which is reflected by the weight of each component in the PLSR model^[32]. This method can obviously reflect the importance of each band in the interpretation of dependent variables. The formula is as follows:

$$VIP_j = \sqrt{\frac{p \sum_{h=1}^m [SS_h (W_{hj} / \|W_h\|)^2]}{\sum_{h=1}^m SS_h}} \quad (6)$$

where, VIP_j is a measure of the contribution of the j_{th} independent variable; p is the number of independent variables; m is the number of ingredients extracted on the basis of PLSR; SS_h represents the sum of the squares explained by the h_{th} component ($h=1,2,3,\dots,m$); $(W_{hj}/\|W_h\|)^2$ indicates the importance of the j_{th} variable. Its threshold value is 1, and the higher the VIP value, the stronger the effect on the interpretation of dependent variables. The sorting of independent variables was realized according to the explanatory power.

CARS is a promising variable selection method proposed by Li et al^[41]. This method mimics the “survival of the fittest” principle in Darwin's biological evolution theory, and selects the effective band variables by adaptive reweighted sampling (ARS) and exponentially decreasing function (EDF). In the actual variable selection, we can obtain multiple valid variable subsets by selecting variables with relatively large absolute values of PLSR coefficients and eliminating variables with relatively small absolute values of regression coefficients. Then cross-validation method is used to model, calculate and compare the root-mean-square error of cross-validation ($RMSE_{CV}$). The subset of variables with the smallest $RMSE_{CV}$ is the subset of optimal variables. This method, capable of selecting wavelength variables that are sensitive to soil properties, and overcoming the combinatorial explosion when variables are being selected, is suitable for high-dimensional data. The specific details of the algorithm were given by Vohland et al and LI et al^[23,41].

RFA is a relatively new variable selection algorithm proposed by Li et al^[35]. This method is a Reversible Jump Markov Chain Monte Carlo (RJMC) variable selection method based on Bayesian inference^[34]. Its feature is that it can use a small number of variables for iterative modeling, which is a very effective method for selecting high-dimensional variables. The main working principle of this algorithm is: (1) construct a Monte Carlo chain in the model space with each model as a node; (2) obtain the posterior distribution of the model from the convergence of the chain; and (3) calculate the variable selection probability by sampling this distribution, which can guide the variable selection. Specific details of this algorithm can be found in Li et al and Hu et al^[35,36]. The key parameters of this algorithm in our study are described below: N (the number of iterations) = 10000; Q (initial number of variable to sample) = 2; A (The maximal number of latent variables for cross-validation) = 10.) The procedure of VIP, CARS, RFA were carried out in MATLAB R2018a with lib PLS toolbox^[42].

2.6 Model Calibration and Validation

We divided the samples by the concentration gradient. First, 162 soil samples were sorted in ascending order according to the SSMC. Then all the samples were divided into 53 levels, and each level had about 3 samples. Finally, one sample was selected in each level as the validation dataset (with a total of 53 soil samples, the second samples of each of the 53 levels were selected). The remaining samples were used as calibration dataset.

The ELM is a model that can be easily used for classification and regression. This model proposed by Huang et al is based on the single hidden layer feedforward neural network (SLFN)^[43]. Compared with the traditional neural network learning algorithm (such as backpropagation neural network), the training process of the ELM has the advantages of high training speed and good generalization ability^[21]. In addition, when ELM is used to train the data set, the algorithm can generate an obvious optimal solution without adjusting the input weight of the network and the deviation of the hidden nodes. We only need to choose the appropriate activation function and determine the number of neurons in the hidden layer. In this study, the activation function used in the ELM model is the tan-sigmoid function. The method of determining the number of the optimal hidden layer nodes was regulated by varying the nodes from 2 to 100 with an increment of 2 at each step^[44]. We used the ELM package to conduct the algorithm in R3.5.2 software.

The model accuracy is evaluated by R^2 , RMSE, and ratio of the performance to deviation (RPD). The model is divided into six levels based on the value of the RPD: $RPD < 1.0$ (unsuccessful prediction/model); $1 \leq RPD < 1.4$ (poor prediction/model); $1.4 \leq$

$RPD < 1.8$ (fair prediction/model); $1.8 \leq RPD < 2$ (good prediction/model); $2 \leq RPD \leq 2.5$ (very good, quantitative prediction/model); $2.5 \leq RPD$ (excellent prediction/model)^[45]; In summary, the best model has the largest R^2 (close to 1) and RPD, as well as the smallest RMSE value.

3 Results

3.1 SSMC, SSC and VIS-NIR Spectra

Statistical distribution of measured SSMC for the whole, calibration and validation datasets is shown in Figure 2a. The SSMC of the study area varied widely between 0.92% and 37.13%, with an average SSMC of 22.66%, standard deviation of 6.29%, and a high coefficient of variation of 360.25%. Compared with the range of the SSMC (0.92%-37.19%) for the calibration dataset, the validation dataset had a similar range of 3%-37.08% with mean and standard deviation of 22.55% and 6.21%, respectively. During the sampling period, the SSMC in the SIA was concentrated around 22%, the SSC in SIA was mainly concentrated in 0-0.75%. The distribution of the SSMC and SSC in each point was shown in Figure 2b, and points were mainly concentrated in the area when $0\% < SSC < 1\%$ and $15\% < SSMC < 30\%$ in Figure 2b. The distribution of SSMC and SSC in different conditions might have some effect on the SSMC prediction. The results showed that the distribution of the SSMC of all the datasets was basically the standardized normal distribution. The statistical results of SSMC in both calibration and validation dataset were similar to those of the whole dataset. Therefore, the SSMC of both datasets can adequately represent the entire dataset.

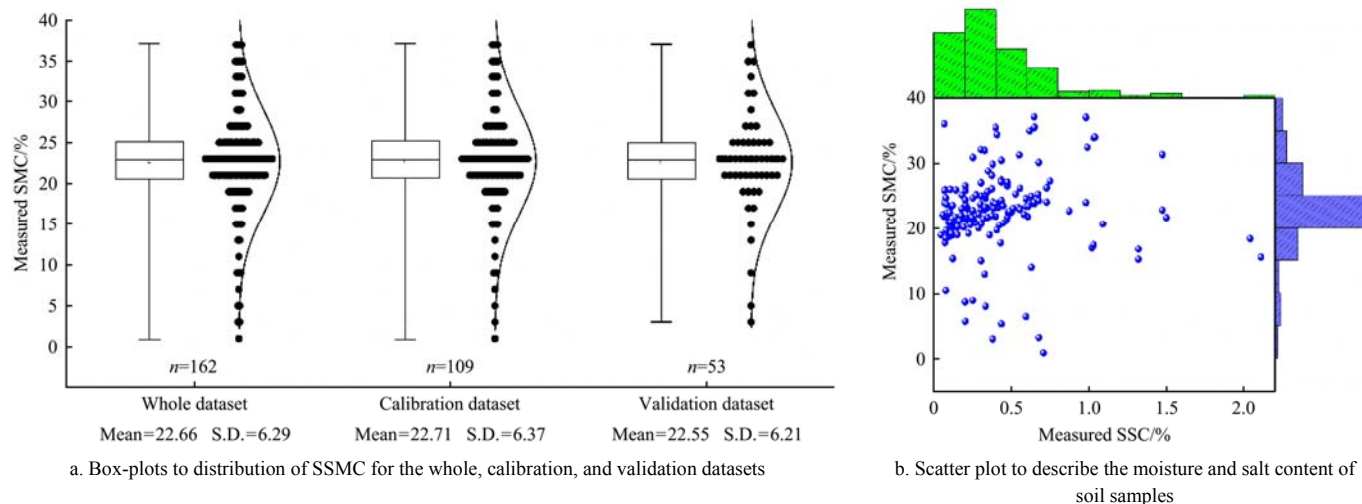


Figure 2 Box-plots, scatter plot and distribution statistics of SSMC and SSC (S.D.: standard deviation)

For convenient analysis of the effects of different factors on the in-situ soil spectrum, we selected representative spectra from the measured spectra. Figure 4 shows the original spectral reflectance of different SSMC under the same SSC and the original spectral reflectance of different SSC under the same SSMC. In the vicinity of 1450 nm, 1950 nm and 2200 nm bands, different SSC spectra and different SSMC spectra showed obvious absorption characteristics (Figure 3). With the same SSMC, the spectral reflectance increased with the increase of SSC. However, the bands near 1950 nm were mainly affected by the SSMC, and the spectral reflectance change was not obvious (Figure 3b). When the SSC was the same, the spectral reflectance increased with the increase of SSMC, but the soil spectral reflectance

decreased after the SSMC reached 22.3% (Figure 3a). It can be seen from the above description that the salinized soil spectrum is affected by both SSMC and SSC. Therefore, when the spectrum is used to predict the SSMC, the difference of SSC will inevitably affect the prediction effect.

3.2 FOD Spectra

In this study, the FOD (order=0~2; interval=0.25 steps) of the in-situ soil spectra were calculated using formula (5). The pretreated spectra are shown in Figure 4. When the order increased from 0 to 1, the negative peaks at 425 nm, 498 nm and 2340 nm increased, which might be affected by SSC (Figure 4a-e). After the fractional order derivative treatment, the vibration of water molecules became gradually prominent, and the absorption valleys

at 1400 nm, 1900 nm and 2200 nm became more obvious, which proved that the overlapping peaks were eliminated^[9]. It is generally considered that the absorption band is a composition of hydroxyl (-OH) band, H₂O band, hydroxyl stretching and bending vibrations of Al-OH and Mg-OH^[46]. As the order increased from 1 to 2, the positive peaks at 2250 nm and 2300 nm were gradually highlighted, while the absorption valleys at 2200 nm and 2340 nm

were attenuated (Figure 4f-i). The 2340 nm is mainly related to soil salinity, and 2250 nm may be related to other properties of the soil. As the 0-2 order FOD changes, most of the reflectance values gradually approached zero, demonstrating that the baseline drift was eliminated. In general, the FOD highlights not only the curve features but also other interference characteristics, so the appropriate band selection algorithm is especially important.

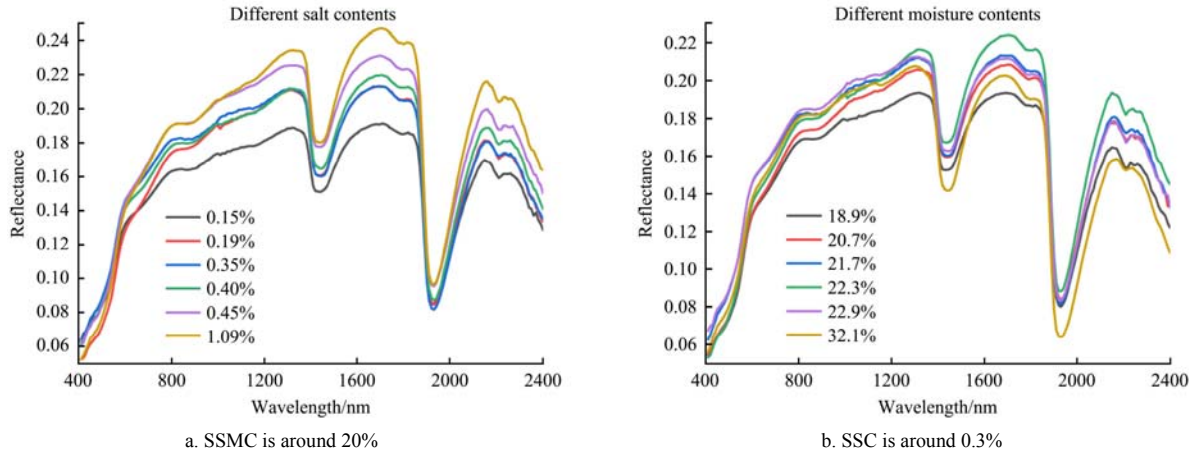


Figure 3 Spectral reflectance curves under different conditions

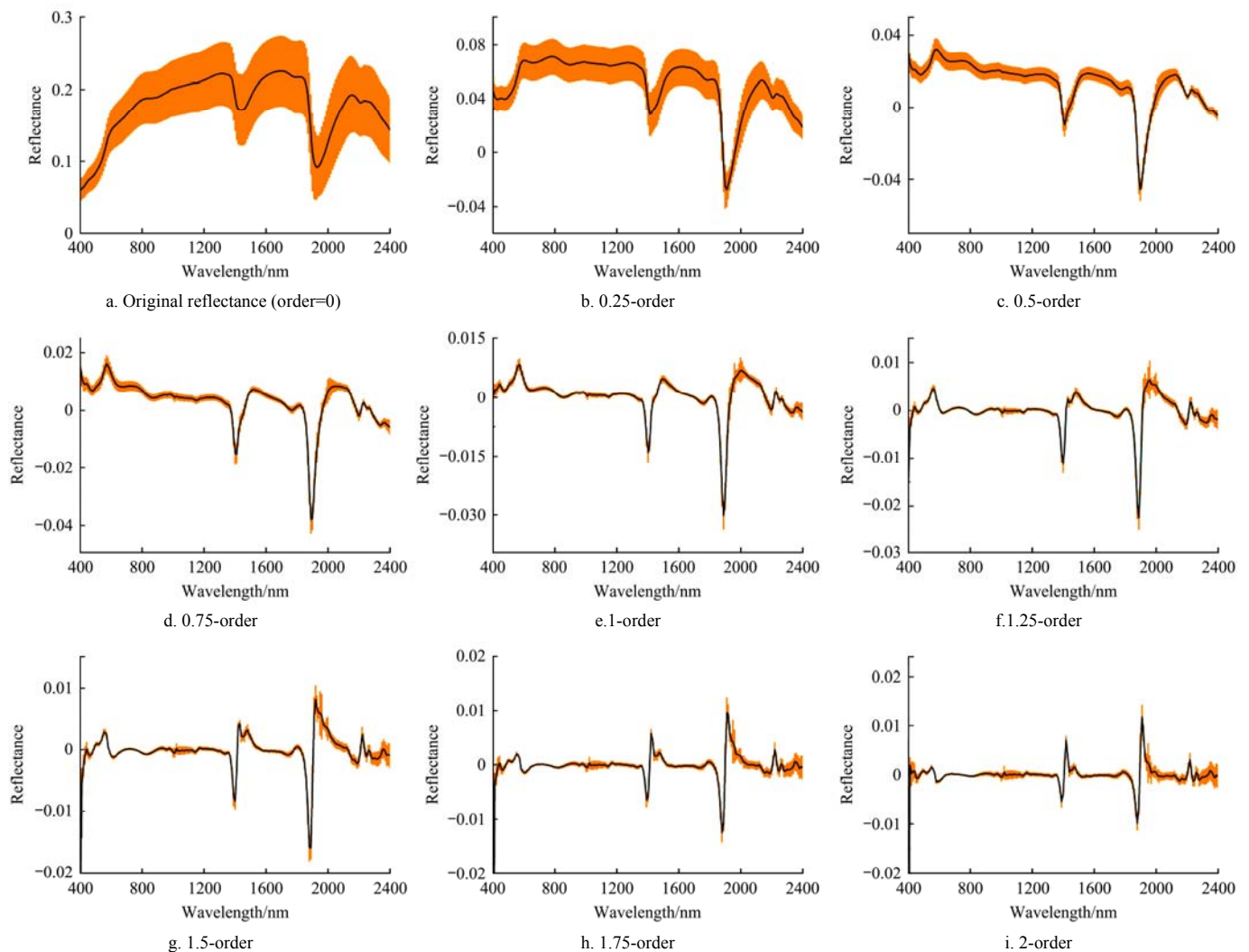


Figure 4 Mean FOD spectra of soil samples ($n = 162$). The yellow shaded areas represent the standard deviations of the spectra

3.3 Spectral Variables Selection by VIP, CARS and RFA

VIP analysis of the SSMC and different FOD spectra was conducted, and the VIP score curves are shown in Figure 5. As is

indicated, as the order increased, the VIP scores in the band around 1400, 1900 nm gradually increased, but decreased slightly after 1.25 order. In order to achieve the band selection, we judged the

explanatory power of the independent variable over the dependent variable according to the “VIP>1” criterion, and recorded the VIP selection results of SSMC and FOD (Figure 6). When the derivative order was below 1, the number of selected spectral variables was relatively large, but when the derivative order was over 1, the number of selected spectral variables was relatively

small. The selected variables were mainly concentrated near the bands that are more sensitive to SSMC (1400, 1900, 2200 nm). In all FOD transformations, some related wavelengths were always selected, including 1390, 1420, and 1860, 1870, 1880, 1890, 1900, 1910, 1920, 1950 nm.

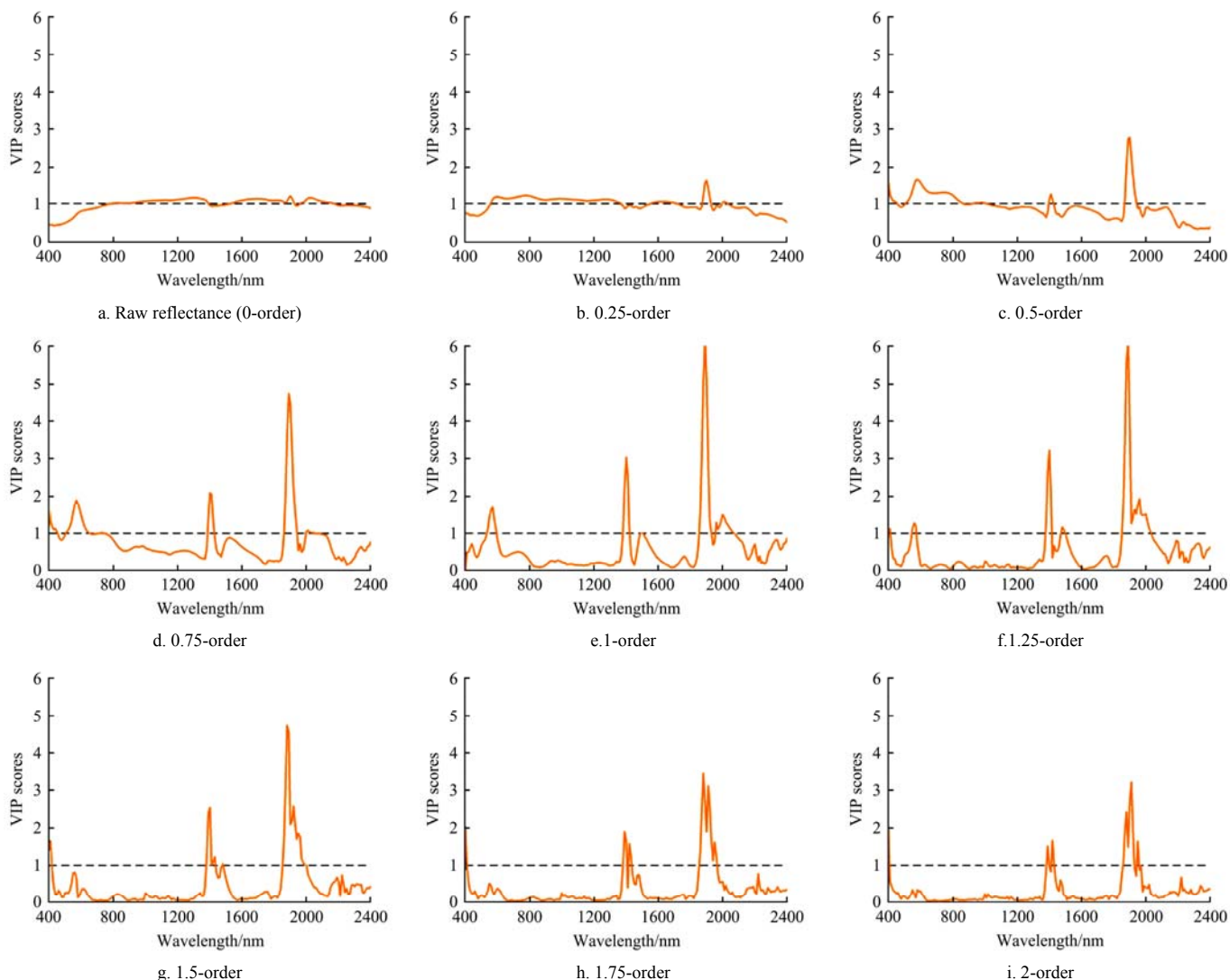


Figure 5 Plots of VIP Scores

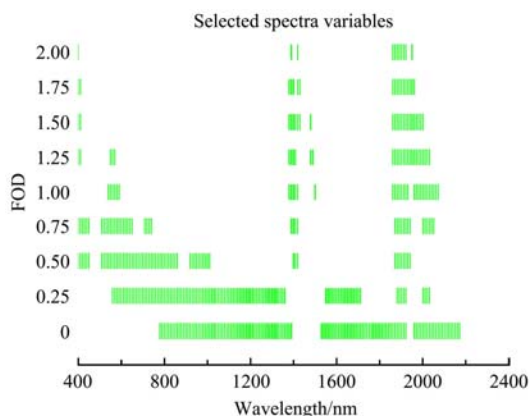


Figure 6 The distributions of spectral variables selected by VIP algorithm at different derivative orders

The following is the illustration of the CARS selection by an example of the 0.75-order FOD spectrum (Figure 7). As is shown in Figure 7, as the number of sampling runs increased, the number

of sampling spectral variables decreased, displaying a trend of fast-to-slow decrease (Figure 7a). Figure 8b shows the change of the RMSE_{cv} value with the increase of the number of sampling runs. When the number of sampling runs was 38, the RMSE_{cv} value was the smallest. When the number of sampling runs is less than 38 or more than 38, it means that the variables that fail to provide information and the variables that have lost their information are both removed. Figure 7c shows the variation of the regression coefficient path for each variable during the sampling process. As can be seen from both Figures 7b and 7c, the optimal subset of the CARS method selection corresponded to the region with the lowest RMSE_{cv} value, which was marked with a vertical asterisk line. We retained the six selected spectral variables as the subset with the most information, and the number of these variables was only 2.99% of the full spectrum. The spectral variables ultimately selected for all FOD transformations are shown in Figure 8. The selected variables were basically concentrated around 1400 and 1900 nm, and the number of variables was very small.

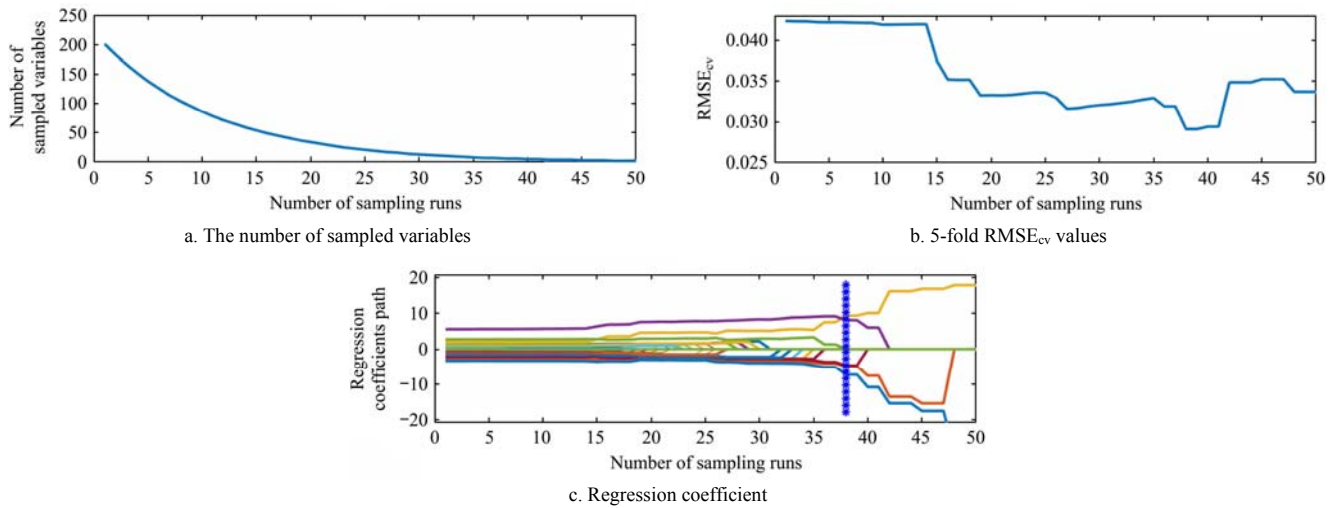


Figure 7 CARS variable selection of 0.75-order FOD spectra

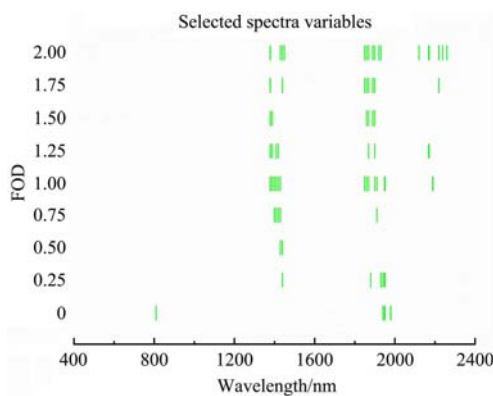


Figure 8 The distribution of spectral variables selected by CARS algorithm at different derivative orders

For the RFA, the number of optimal iterations was selected according to the $RMSE_{cv}$ value. Then the optimal number of

spectral variables was determined by the number of optimal iterations. Finally, the selected bands were determined according to the order of the variable selection probability. Take the 0.75-order derivative spectrum as an example (Figure 9). As Figure 9a shows, as the number of iterations increased, the number of selected spectral variables changed. The curve showed an oscillation with no obvious regularity. Figure 9b shows the change in the $RMSE_{cv}$ (root mean square error during RFA operation) value as the number of iterations increased. It can be seen from Figure 9b that the $RMSE_{cv}$ reached a minimum when the number of iterations was 2062. Figure 9a shows that when $RMSE_{cv}$ reached a minimum, the number of selected optimal variables was 51. As is shown in Figures 9a and 9c, the sequence of the band selection probabilities ultimately determined the selected variables. The selected spectral variables accounted for 25.37% of the full spectrum. The spectral variables ultimately selected for all FOD transformations are shown in Figure 10.

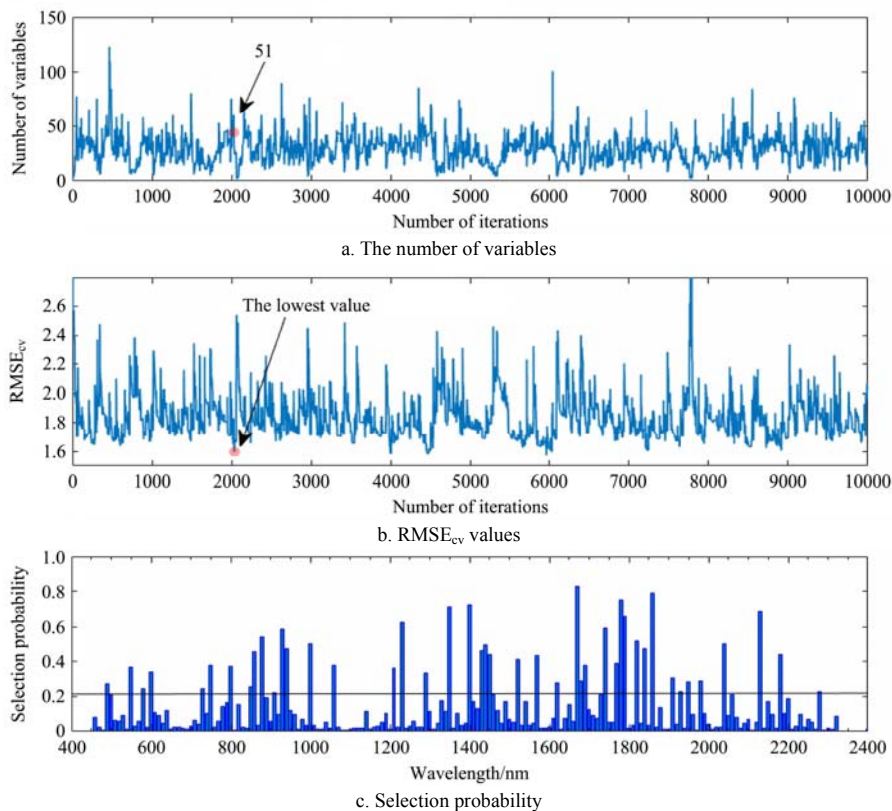


Figure 9 Random forg algorithm (RFA) variable selection of 0.75-order derivative spectra

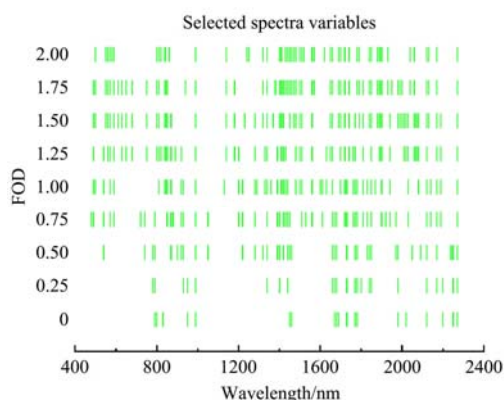


Figure 10 The distribution of spectral variables selected by RFA at different derivative orders

3.4 Comparison of Predictions Using Different FOD Transformations and Variable Selection Algorithms

SSMC estimation model based on ELM was established for different FOD, and a series of calculation results are shown in Table 1. These nine ELM models demonstrated different modeling results. Except the three models (order=1.5, 1.75, and 2), all other FOD transformations have improved the accuracy of the models, compared to the original reflectance (order=0). This may be due to the ability of these methods to reduce the adverse effects (such as baseline effects) so as to improve model performance. However, when the order = 1.5, 1.75, and 2, the FOD would amplify the interference information, resulting in a decrease in the model accuracy and robustness. The most accurate model is the ELM based on the 0.75-order FOD spectrum ($R_v^2=0.83$, $RMSE_v=2.87\%$, $RPD=2.44$), followed by the 0.5-order FOD spectrum. For the first and the second derivative spectrums, their RPD values were reduced by 0.28 and 1.07, compared with the 0.75-order FOD spectrum. It can be seen that most of the FOD can improve the model performance, compared with the original spectrum (order=0) and the integer derivative spectrum (i.e. the first and second order derivative).

Table 1 Modeling results of full-spectrum ELM models for SSMC with different FOD preprocessing techniques

Order	N ^a	Calibration dataset (n=109)		Validation dataset (n=53)		
		R_c^2	$RMSE_c/\%$	R_v^2	$RMSE_v/\%$	RPD
0	201	0.69	3.52	0.71	3.36	1.84
0.25	201	0.81	2.76	0.73	3.19	1.95
0.5	201	0.84	2.55	0.82	2.61	2.38
0.75	201	0.88	2.22	0.83	2.54	2.44
1	201	0.84	2.54	0.78	2.87	2.16
1.25	201	0.78	2.96	0.75	3.10	2.00
1.5	201	0.82	2.69	0.69	3.44	1.81
1.75	201	0.73	3.29	0.57	4.11	1.51
2	201	0.72	3.36	0.46	4.54	1.37

Note: ^a Number of the spectral variables.

In order to study the influence of FOD and variable selection algorithms on SSMC estimation, the ELM method was used to establish spectral models on the same calibration and verification datasets (but different spectral wavelengths were selected). The model accuracy was evaluated by R^2 , RMSE, and RPD. The descriptive regression statistics is shown in Table 2. The cross-validation results of the model indicated that in the spectral variables selection, the simplified model showed better performance and higher R_c^2 values (from 0.74 to 0.96) and lower

$RMSE_c$ values (from 1.22% to 3.18%) than the corresponding full-spectrum ELM model (Table 1). In addition, the number of selected spectral variables was greatly reduced (variable ratios range from 1.49% to 62.18%), and the selected spectral variables contained the most useful information related to the SSMC. In general, RFA selects more spectral variables than VIP and CARS. These results showed that the variable selection could help simplify the model structure and improve its prediction accuracy.

Table 2 Modeling results of VIP-ELM, CARS-ELM and RFA-ELM for SSMC with different derivative order preprocessing techniques

Selection	Order	T ^a	N ^b	Calibration dataset (n=109)		Validation dataset (n=53)		
				R_c^2	$RMSE_c/\%$	R_v^2	$RMSE_v/\%$	RPD
VIP	0	0.0058	125	0.85	2.47	0.82	2.61	2.37
	0.25	0.0057	108	0.87	2.29	0.84	2.45	2.53
	0.5	0.0061	65	0.91	1.91	0.88	2.17	2.85
	0.75	0.0063	45	0.90	2.05	0.88	2.17	2.86
	1	0.0065	33	0.87	2.27	0.86	2.28	2.72
	1.25	0.0076	30	0.86	2.35	0.81	2.67	2.53
	1.5	0.0065	25	0.84	2.51	0.81	2.68	2.32
	1.75	0.0053	19	0.83	2.57	0.81	2.73	2.27
	2	0.0054	12	0.82	2.64	0.77	2.92	2.12
CARS	0	0.33	5	0.78	2.96	0.74	3.12	1.99
	0.25	0.27	6	0.81	2.75	0.77	2.89	2.14
	0.5	0.28	3	0.86	2.32	0.83	2.51	2.47
	0.75	0.28	6	0.88	2.15	0.88	2.18	2.85
	1	0.29	14	0.89	2.12	0.89	2.01	3.09
	1.25	0.29	8	0.85	2.41	0.86	2.29	2.71
	1.5	0.30	7	0.82	2.68	0.81	2.69	2.31
	1.75	0.31	9	0.81	2.74	0.80	2.71	2.29
	2	0.31	17	0.78	2.94	0.77	2.95	2.10
RFA	0	59.35	20	0.87	2.29	0.83	2.49	2.49
	0.25	57.04	24	0.89	2.09	0.87	2.27	2.74
	0.5	74.82	39	0.89	2.05	0.89	2.06	3.02
	0.75	84.07	51	0.96	1.22	0.94	1.63	3.80
	1	110.91	53	0.94	1.52	0.91	1.82	3.42
	1.25	122.36	56	0.91	1.89	0.86	2.28	2.72
	1.5	125.12	63	0.85	2.48	0.86	2.34	2.65
	1.75	106.45	61	0.82	2.71	0.76	3.00	2.07
	2	94.81	52	0.81	2.77	0.75	3.06	2.03

Note: ^a Computation time (seconds) in the processes of variable selection; ^b Number of the selected spectral variables.

In the VIP method, the ELM model based on the 0.75-order derivative spectrum has the highest prediction accuracy for SSMC ($R_v^2=0.88$, $RMSE_v=2.17\%$, $RPD=2.86$). The simulation results showed the 0.75-VIP-ELM model had a good prediction performance (Figure 13). Similarly, the ELM models based on the first derivative spectrum and the 0.75 order derivative spectrum had the most accurate estimation of SSMC in the CARS and the RFA methods, respectively. It is worth noting that the VIP, CARS, and RFA methods have a positive effect on building a more accurate and concise ELM model than the full-spectrum ELM model (Table 1). Figure 11 shows the scatter plots of measured SSMC and predicted SSMC by different variable selection algorithms. We can see from Figure 11d that the 0.75-order FOD spectrum combined with RFA could make the scatter of the

measured and the predicted SSMC values closer to the 1:1 line. However, the points of the measured and the predicted SSMC values based on the 0.75-order FOD full spectrum were significantly more scattered and the fit line was more deviant from the 1:1 line (Figure 11a). The comparison of the RPD average conversion of different FOD (Figure 12a) demonstrated that the

ELM models based on the 0.75-order FOD spectrum had the best performance in all the FOD transformations. Their average RPD is 2.65, displaying a good pretreatment effect in SSMC prediction. The mean RPD of each variable selection algorithm is shown in Figure 12b. The results showed that RFA had the best performance, and VIP and CARS followed.

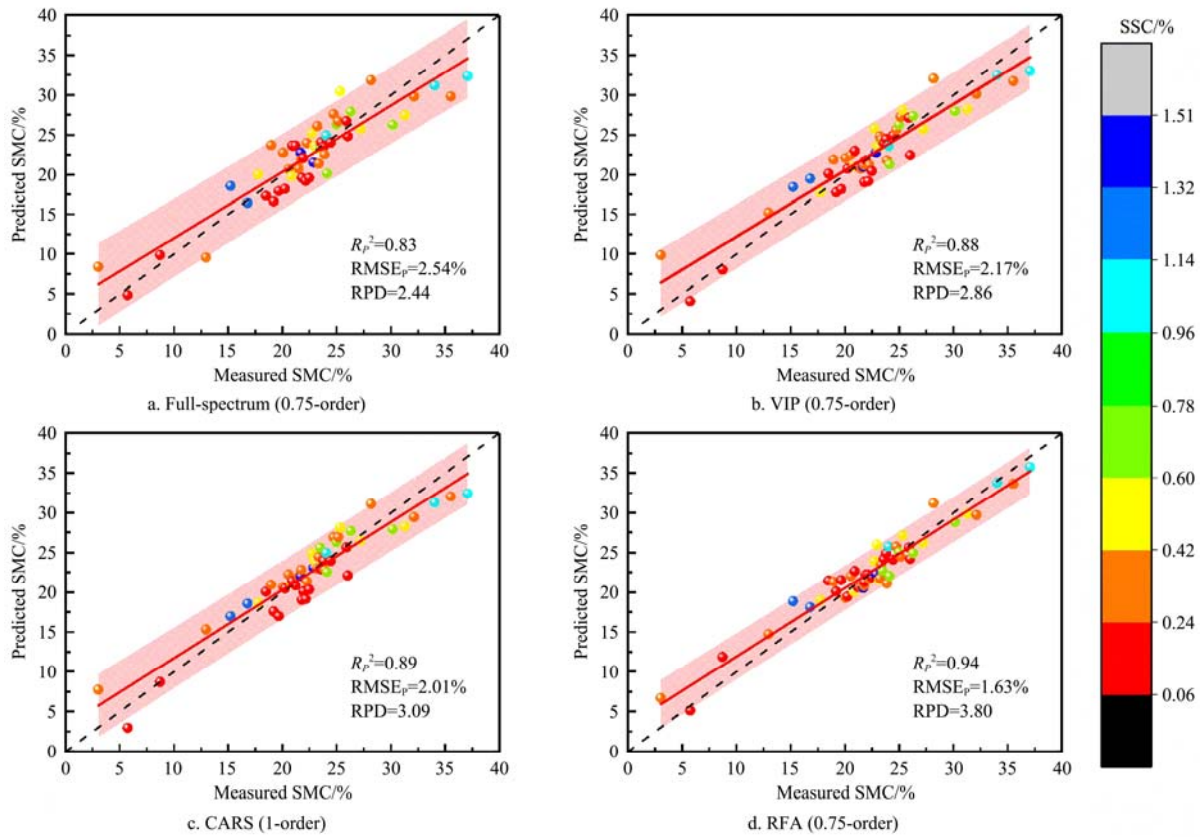


Figure 11 Scatter plots of measured SSMC and predicted SSMC by different variable selection algorithms. The color of the scatter points indicates the degree of salinization. The black lines denote as the 1:1 line

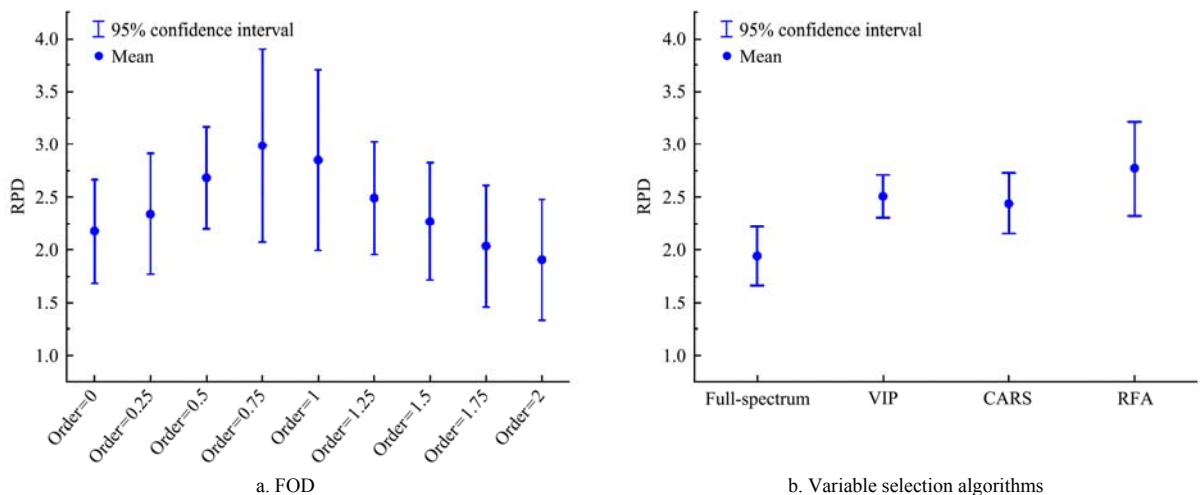


Figure 12 The mean and 95% confidence interval of the RPD values of SSMC prediction

The optimal model is usually a model with the best predictive power and the highest accuracy. However, the choice of the variable selection algorithm is a compromise between predictive performance and computation time. The time of each variable selection algorithm was calculated in MATLAB R2018a and these algorithms were further compared (Table 2). All the processing was conducted in the same computer (Intel Core i7 2.50 GHz Processor, 16GB of RAM, Windows 10). In all FOD

transformations, the VIP was completed in 0.01 second, and the CARS in 1 second, but the RFA always over 50 seconds. In the case of considering more efficiency, VIP was the most suitable method, although its predictive accuracy was lower than RFA, but similar to CARS (Figure 13). In the case of considering more precision, RFA was the most suitable method, although its effectiveness was lower than VIP and CARS, but its predictive accuracy was best (Figure 13).

3.5 Digital mapping

We applied the models using the full-band and different variable selection methods to all sampling points in the study area, and predicted the moisture content of each sample point. The distribution diagrams of the measured, predicted and residual SSMC were plotted using the inverse distance weighted method of ARCGIS 10.5, as is shown in Figure 13. The spatial distribution of the predicted SSMC showed that the north was higher than the south and the west was higher than the east, which was consistent across all tested models (Figure 13b1-b4). All the tested models showed a similar pattern of residual spatial distribution. Specifically, all the models underestimated the SSMC in the northwest, and the degree of underestimation was $0.75\text{-FULL-ELM} > 1\text{-CARS-ELM} \approx 0.75\text{-VIP-ELM} > 0.75\text{-RFA-ELM}$.

For other areas, the water content was overestimated to varying degrees, and the degree of overestimation was $0.75\text{-FULL-ELM} > 1\text{-CARS-ELM} \approx 0.75\text{-VIP-ELM} > 0.75\text{-RFA-ELM}$. The degree of similarity between the predicted map and the measured map was $0.75\text{-RFA-ELM} > 1\text{-CARS-ELM} \approx 0.75\text{-VIP-ELM} > 0.75\text{-FULL-ELM}$. In summary, the ranking of model performance in spatial distribution was $0.75\text{-RFA-ELM} > 1\text{-CARS-ELM} \approx 0.75\text{-VIP-ELM} > 0.75\text{-FULL-ELM}$. These results confirmed that the 0.75-RFA-ELM model displayed a good performance in space simulation. According to the results in this part and that in Section 3.5, we recommend 0.75-VIP-ELM or 0.75-RFA-ELM with appropriate reductions in the number of iterations (i.e., reduced program run time) be used in the SSMC estimation.

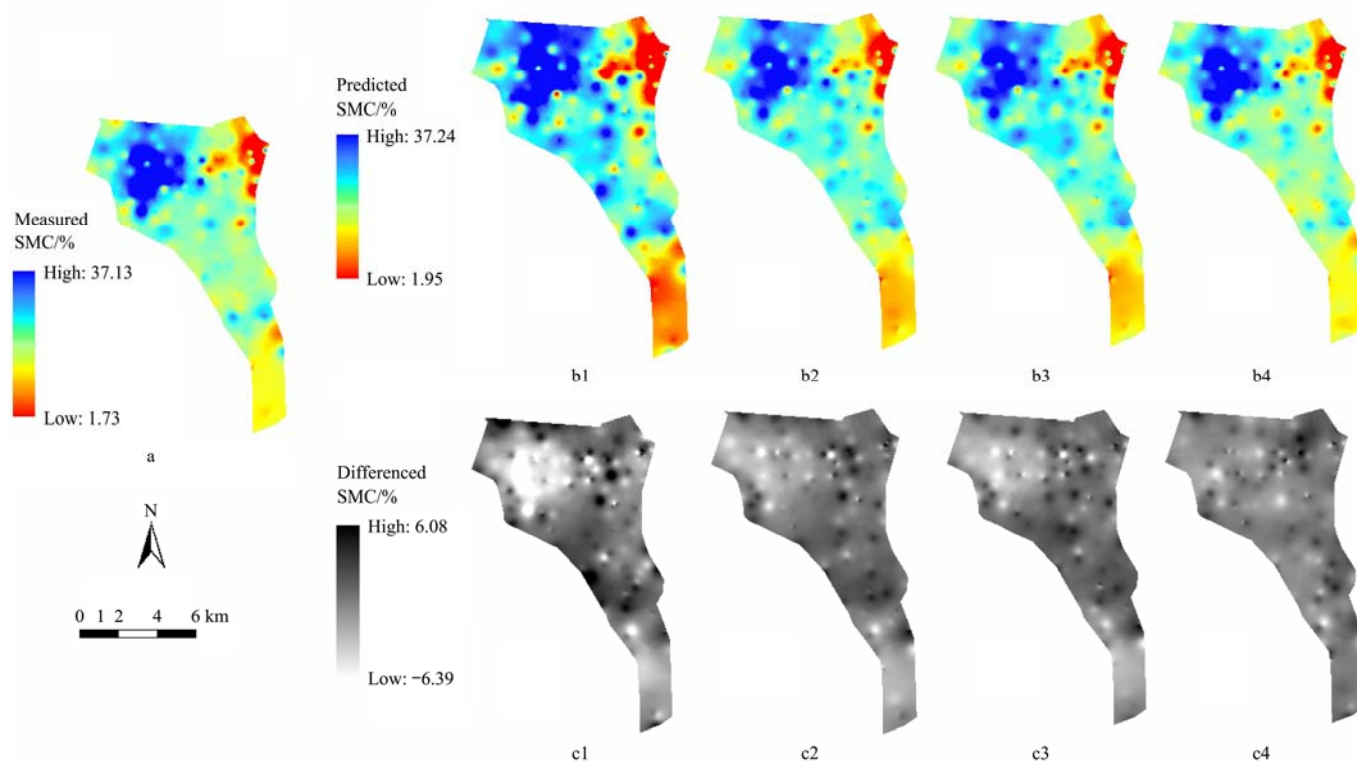


Figure 13 Spatial distribution diagrams of (a) the measured SSMC; The predicted SSMC: (b1) 0.75 order derivative full-spectrum ELM (0.75-FULL-ELM), (b2) 0.75 order derivative VIP ELM (0.75-VIP-ELM); (b3) 1 order derivative CARS ELM (1-CARS-ELM);(b4) 0.75 order derivative RFA ELM (0.75-RFA-ELM); Residuals: (c1) 0.75-FULL -ELM, (c2) 0.75-VIP-ELM; (c3)1-CARS-ELM; (c4)0.75-RFA-ELM (White and black are the overestimation and underestimation of SSMC, respectively)

4 Discussion

Although the VIS-NIR spectroscopy for SSMC prediction in laboratory has been relatively mature, there is very rare in-situ research. If in-situ soil spectroscopy can be applied to SSMC estimation, much time and labor will be saved. Our results indicated that both SSC and SSMC affected the soil spectrum, so the VIS-NIR moisture detection in saline soils is somewhat different from non-saline soils (Figure 3). This is consistent with other studies: the spectral reflectance of most of the bands rises as the soil salinity increases; the soil salinity even changes the effect of soil moisture on the spectrum. This is mainly because when the soil salinity is high, on the surface of the soil will form a salt shell, which is mainly white (Figure 1C, 1D, 1E). The white substance has strong reflection performance, which can enhance the reflection ability of the soil, thereby greatly increasing the reflectivity of some bands^[8]. In addition, when acquiring in-situ spectra, such external factors as soil surface conditions, soil texture,

and so on also affect the spectral characteristics^[10,47]. Although the SSC and the external factors have some negative effect on the in-situ soil spectroscopy, the combination of different variable selection algorithms in this study with the ELM algorithm displayed excellent predictive performance (best validation $R^2=0.94$, RPD=3.80). Overall, our results indicated that the in-situ spectroscopy had great potential for SSMC estimation, and the highest model accuracy is slightly higher than the research reports. Oltra-Carrió et al. used the spectral indices to invert the SSMC under different soil clay content, and its R^2 was up to 0.91^[48]; Wang et al. applied salt-resistant indices to modeling, and used field data to verify the accuracy (R^2 was up to 0.72)^[8]; The study by Cai et al. in north central Tarim Basin in Xinjiang showed that ELM could achieve the best performance in SSMC prediction, and verified that R^2 could reach up to 0.9354^[22].

In the spectral analysis, the first-order differential can eliminate the effect of the partially linear or near-linear noise spectrum and the background on the target spectrum, and improve

the overlapping spectral resolution; the second-order differential can eliminate baseline drift and background signal, and improve the analysis accuracy^[49]. The FOD is an extension of the integer differential, extending the order from an integer to a non-integer. Usually, the first-order differential of the spectral curve is defined as the slope of the spectral line, and the second-order differential is defined as the curvature of the spectral line, corresponding to the inflection point and the extreme value, respectively. However, the fractional order can be considered as the sensitivity to the slope and curvature of the line. That is, when the fractional order increases from 0 to 1, the sensitivity of the fractional result to the spectral reflectance decreases, and the sensitivity to the slope of the line is increased; When the order increases from 1 to 2, the sensitivity of the fractional result to the slope of the line decreases, and the sensitivity to the curvature of the line increases. Thus, similar to the integer order derivative, FOD can also eliminate oblique baselines, but its ability to eliminate baselines of different curvatures proves to be superior to integer order derivatives^[14,15,17,50]. In all full-spectrum ELM models (Table 1), the ELM model based on the 0.75-order derivative spectrum has the highest accuracy (RPD=2.44). Compared with the first derivative spectrum, the second derivative spectrum and the original spectral reflectance, the RPD is increased by 0.28, 1.07 and 0.60, respectively.

The VIS-NIR spectrum often contains redundant wavelength information, increasing the complexity of the model^[23,25-28]. It can be seen that the SSMC model based on spectral variable selection, regardless of the FOD transformation, has better performance than the full-spectrum ELM model (Table 2). The RPD of the optimal model in the variable selection is 3.80, while for the full-spectrum ELM model based on the 0-order derivative spectrum (i.e. no differential processing), the RPD is only 1.52. The optimization model can eliminate unnecessary spectral variables and reduce over-fitting or under-fitting, which means that variable selection can simplify the model structure and improve the model robustness^[51].

Among the three spectral variable selection methods used in this study, RFA showed the best prediction performance in all FOD transformations (Figure 13). These models, whose RPD values ranged from 2.03 to 3.80, were classified as models with very good predictions to excellent predictions. Compared with RFA, the results of VIP were relatively poor, and those of CARS were the worst (Figure 12b, considering only the mean). Nevertheless, RFA took the longest computation time (average = 92.77s), and the complete iteration of possible combinations of variables was also complicated^[35,36]. In order to improve the calculation accuracy and efficiency, VIP method is suggested to replace RFA method or to reduce the number of iterations of RFA when the quality of variable selection is guaranteed. At present, many studies have proved that VIP is a useful tool for spectral variables selection. For example, Wang et al. found that the spectral variables selected by VIP could be used to establish a more accurate model of soil water-soluble salt ions^[30]. Guezenc et al discovered that the combination of VIP and coefficient maps was beneficial to obtaining quality variables of soil potassium^[52]; Jia et al. found that combination of VIP with recursive partial least squares regression could predict soil organic carbon and nitrogen steadily^[53]. In this study, the application of CARS did not produce satisfying results (Table 2 and Figure 12). Although the number of variables was greatly reduced and the validation model was simplified, the average effect of CARS was not as good as VIP (Figure 12).

However, many studies have shown that CARS is an excellent variable selection method. For example, Vohland et al. found that the CARS method could effectively select the key informative spectral variables in the spectral inversion of soil organic matter and other properties^[54]; Xu et al. found that CARS in combination with support vector machine regression could predict rice root density relatively accurately^[31]. The relatively poor performance of CARS in this study is probably because the variables selected by CARS may also include some unstable or disturbance variables. The direct evidence is that, among the three variable selection methods, the number of variables selected by CARS (no more than 17, accounting for less than 8.45% of the total band) was the smallest (Table 2).

Compared with the other two variable selection algorithms, RFA was more successful in the selection of soil spectral variables. From our research as well as Yao's, we found that RFA could be used to select more effective spectral variables to predict the soil properties^[55]. However, the number of iterations in RFA had a great effect on the calculation time. Although a larger number of iteration means more effective variables obtained, it is necessary to balance the contradiction between the number of iterations and the calculation time. In all the FODs, the bands selected by RFA were relatively dispersed compared with that by VIP and CARS, but, with the three algorithms, bands could always be selected near 1400 and 1900 nm. This is consistent with previous conclusions, because these two bands are mainly for water absorption^[56]. The relative dispersion of the bands selected by RFA may be related to the effect of soil salinity^[35,36,55,57]. Namely, the soil contains high moisture salts, such as MgCl₂ and CaCl₂, which absorb some water^[58]. Consequently, some spectral information of SSMC exists in SSC-sensitive spectral bands. However, we should pay attention to the fact that the result of variable selection method varies according to the material and sample characteristics to be studied.

ELM algorithm has been widely used to solve complex regression or non-linear problems^[21,43,44]. In recent years, Khosravi et al compared the multivariate prediction methods and studied their effects on the prediction of heavy metals in soils^[21]. The results showed that ELM had the best performance in estimating such heavy metals as Zn and Pb among the four commonly used multivariate regressions. Our research further supports their findings. Previous studies have shown that soil moisture has a non-linear effect on reflectance spectra^[20,21,43,44,59]. Variable selection will inevitably reduce the spectral variables involved in modeling, but these processes can't change the non-linear relationship between the SSMC and the reflectance spectra^[9]. Therefore, the application of ELM model in this study can well solve these complex non-linear regression problems.

The cultivated land soil in this study area is mostly clay loam^[37]. Therefore, the measured soil spectra have similar spectral characteristics, which provide a basis for the establishment of SSMC prediction model. Although the prediction of SSMC in this study is more accurate, all the estimation models are based on in-situ point spectra, which can't be directly applied to surface spectral data (i.e., spectral images). However, the combination of the spectral pretreatment and variable selection models reported in this paper is a promising start. We intend to validate these methods using different data sets including different soil types, soil texture and salinized areas in future studies in order to provide a reference for precise and rapid determination of SSMC by satellite or UAV.

5 Conclusions

This study investigated the effectiveness of FOD and spectral variable selection methods in SSMC estimation. The main conclusions are:

(1) The spectrum of salinized soil is affected by SSMC and SSC, but the effect of SSC can be weakened by the selection of effective bands.

(2) In some cases, FOD (e.g. 0.5 and 0.75 orders) can have better estimation than the integer order derivatives (i.e., 1st and 2nd orders) and the original reflection spectrum.

(3) The ELM model based on the 0.75 order derivative spectrum and the random frog algorithm provides an optimal model prediction. Our study confirmed the potential of in-situ VIS-NIR spectroscopy for SSMC estimation.

(4) The good applicability of the random frog algorithm in soil spectroscopy was further confirmed in this study.

(5) Variable selection algorithms (i.e., VIP, CARS, and RFA) are capable of selecting useful spectral variables, simplifying the model, and improving the prediction accuracy. In general, RFA has the best prediction but a long calculation time. An alternative approach is the application of the VIP method due to its less required time to process the algorithm with a relatively high precision.

Acknowledgements

The researchers thank the National Key Research and Development Program of China (2017YFC0403302, 2016YFD0200700) and the Fundamental Research Funds for the Central Universities under Grant (2452019180) for their support. The study was also supported by Natural Science Foundation of Shanxi (2019JM-066), Science and Technology Plan Project of Yangling (2018GY-03) and Humanities and Social Science Program of Northwest A&F University (Z109021405).

Author Contributions

Congcong Lao, Junying Chen and Zhitao Zhang designed experiment. Congcong Lao did the experiments, analyzed the data and wrote the paper. Zhihua Yao, Zheng Xing and Xiang Tai did part of the experiments. Haorui Chen and Jifeng Ning provided some help during the experiment. Congcong Lao and Yinwen Chen were involved in drafting and revising the paper.

Conflicts of Interest

The authors declare no conflict of interest.

[References]

- [1] Kevin Z M, Nashon K R M, Moses M, et al. The Role of Moisture in the Successful Rehabilitation of Denuded Patches of a Semi-Arid Environment in Kenya. *Journal of Environmental Science and Technology*, 2010; 3(4): 195–207. doi: 10.3923/jest.2010.195.207.
- [2] Chen L, Feng Q, Li F, et al. A bidirectional model for simulating soil water flow and salt transport under mulched drip irrigation with saline water. *Agricultural Water Management*, 2014; 146: 24–33. doi: 10.1016/j.agwat.2014.07.021.
- [3] Susha Lekshmi S U, Singh D N, Shojaei Baghini M. A critical review of soil moisture measurement. *Measurement*, 2014; 54: 92–105. doi: 10.1016/j.measurement.2014.04.007.
- [4] Robinet J, von Hebel C, Govers G, et al. Spatial variability of soil water content and soil electrical conductivity across scales derived from Electromagnetic Induction and Time Domain Reflectometry. *Geoderma*, 2018; 314(1): 60–74. doi: 10.1016/j.geoderma.2017.10.045.
- [5] Eller H, Denoth H. A capacitive soil moisture sensor. *Journal of Hydrology*, 1996; 185(1): 37–46. doi: 10.1016/0022-1694(95)03003-4.
- [6] Xu D, Sun R, Yeh T J, et al. Mapping soil layers using electrical resistivity tomography and validation: Sandbox experiments. *Journal of Hydrology*, 2019; 575(5): 23–36. doi: 10.1016/j.jhydrol.2019.05.036.
- [7] Alamry A S, Meijde M, Noomen M, et al. Spatial and temporal monitoring of soil moisture using surface electrical resistivity tomography in Mediterranean soils. *Catena*, 2017; 157(3): 88–96. doi: 10.1016/j.catena.2017.06.001.
- [8] Wang Q, Li P, Pu Z, et al. Calibration and validation of salt-resistant hyperspectral indices for estimating soil moisture in arid land. *Journal of Hydrology*, 2011; 408(2): 76–85. doi: 10.1016/j.jhydrol.2011.08.012.
- [9] Hong Y, Chen Y, Yu L, et al. Combining Fractional Order Derivative and Spectral Variable Selection for Organic Matter Estimation of Homogeneous Soil Samples by VIS–NIR Spectroscopy. *Remote Sens.-Basel*, 2018; 10: 479. doi: 10.3390/rs10030479.
- [10] Nawar S, Buddenbaum H, Hill J, et al. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil and Tillage Research*, 2016; 155(5): 10–22. doi: 10.1016/j.still.2015.07.021.
- [11] Yue J, Tian Q, Tang S, et al. A dynamic soil endmember spectrum selection approach for soil and crop residue linear spectral unmixing analysis. *International Journal of Applied Earth Observation and Geoinformation*, 2019; 78(3): 06–17. doi:10.1016/j.jag.2019.02.001.
- [12] Wang X, Zhang F, Kung H, et al. Spectral response characteristics and identification of typical plant species in Ebinur lake wetland national nature reserve (ELWNNR) under a water and salinity gradient. *Ecological Indicators*, 2017; 81(2): 22–34. doi: 10.1016/j.ecolind.2017.05.071.
- [13] Ji R, Zhao Z, Yu X, et al. Determination of rhodamine B in capsicol using the first derivative absorption spectrum. *Optik*, 2019; 181: 796–801. doi: 10.1016/j.ijleo.2018.12.141.
- [14] Zhang D, Tiyyip T, Ding J, et al. Quantitative Estimating Salt Content of Saline Soil Using Laboratory Hyperspectral Data Treated by Fractional Derivative. *Journal of Spectroscopy*, 2016; 2016: 1–11. doi: 10.1155/2016/1081674.
- [15] Wang J, Ding J, Abulimiti A, et al. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (VIS–NIR) spectroscopy, Ebinur Lake Wetland, Northwest China. *PeerJ*, 2018; 6: e4703. doi: 10.7717/peerj.4703.
- [16] Wang X, Zhang F, Kung H, et al. New methods for improving the remote sensing estimation of soil organic matter content (SOMC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR) in northwest China. *Remote Sensing of Environment*, 2018; 218(1): 04–18. doi: 10.1016/j.rse.2018.09.020.
- [17] Hong Y, Liu Y, Chen Y, et al. Application of fractional-order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy. *Geoderma*, 2019; 337(7): 58–69. doi: 10.1016/j.geoderma.2018.10.025.
- [18] Vohland M, Ludwig M, Thiele-Bruhn S, et al. Quantification of Soil Properties with Hyperspectral Data: Selecting Spectral Variables with Different Methods to Improve Accuracies and Analyze Prediction Mechanisms. *Remote Sens.-Basel*, 2017; 9: 1103. doi: 10.3390/rs9111103.
- [19] Rossel R A V, Behrens T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 2010; 158: 46–54. doi: 10.1016/j.geoderma.2009.12.025.
- [20] Liu X, Xu L. The universal consistency of extreme learning machine. *Neurocomputing*, 2018; 311: 176–82. doi: 10.1016/j.neucom.2018.05.066.
- [21] Khosravi V, Doulati Ardejani F, Yousefi S, et al. Monitoring soil lead and zinc contents via combination of spectroscopy with extreme learning machine and other data mining methods. *Geoderma*, 2018; 318: 29–41. doi: 10.1016/j.geoderma.2017.12.025.
- [22] Cai L, Ding J. Prediction for Soil Water Content Based on Variable Preferred and Extreme Learning. *Spectroscopy and Spectral Analysis*, 2018; 38(7): 2209–2214. doi: 10.3964/j.issn.1000-0593(2018)07-2209-06. (in Chinese)
- [23] Vohland M, Ludwig M, Thiele-Bruhn S, et al. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma*, 2014; 223(225): 88–96. doi: 10.1016/j.geoderma.2014.01.013.
- [24] Sarathjith M C, Das B S, Wani S P, et al. Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. *Geoderma*, 2016; 267: 1–9. doi: 10.1016/j.geoderma.2015.12.031.
- [25] Zou X, Zhao J, Malcolm J W, et al. Variables selection methods in

- near-infrared spectroscopy. *Analytica Chimica Acta*, 2010; 667(1-2): 14–32. doi: 10.1016/j.aca.2010.03.048.
- [26] Bin J, Ai F, Fan W, et al. An efficient variable selection method based on variable permutation and model population analysis for multivariate calibration of NIR spectra. *Chemometrics & Intelligent Laboratory Systems*, 2016; 158: 1–13. doi: 10.1016/j.chemolab.2016.08.006.
- [27] Chong I, Jun, C. Performance of some variable selection methods when multicollinearity is present. *Chemometrics & Intelligent Laboratory Systems*, 2005; 78(1): 03–12. doi: 10.1016/j.chemolab.2004.12.011.
- [28] Abrahamsson C, Johansson J, Sparén A, et al. Comparison of different variable selection methods conducted on NIR transmission measurements on intact tablets. *Chemometrics & Intelligent Laboratory Systems*, 2003; 69: 3–12. doi: 10.1016/S0169-7439(03)00064-9.
- [29] Araújo M C U, Saldanha T C B, Galvão R K H, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometr. Intell. Lab. 2001*; 57(2): 65–73. doi: 10.1016/S0169-7439(01)00119-8.
- [30] Wang H, Chen Y, Zhang Z, et al. Quantitatively estimating main soil water-soluble salt ions content based on Visible-near infrared wavelength selected using GC, SR and VIP. *Peer J*, 2019, 7: e6310. doi: 10.7717/peerj.6310.
- [31] Xu S, Zhao Y, Wang M, et al. Determination of rice root density from Vis-NIR spectroscopy by support vector machine regression and spectral variable selection techniques. *Catena*, 2017; 157: 12–23. doi: 10.1016/j.catena.2017.05.008.
- [32] Oussama A, Elabadi F, Platikanov S, et al. 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*, 2012; 89(18): 07–12. doi: 10.1007/s11746-012-2091-1.
- [33] Rahman A, Faqeerzada M A, Joshi R, et al. Quality Analysis of Stored Bell Peppers Using Near-Infrared Hyperspectral Imaging. *Transactions of the Asabe*, 2018; 61(1): 199–207. doi: 10.13031/trans.12482.
- [34] GREEN B P J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 1995; 4(7): 11–32. doi: 10.1093/biomet/82.4.711.
- [35] Li H, Xu Q, Liang Y. Random frog: An efficient reversible jump Markov Chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification. *Analytica Chimica Acta*, 2012; 740(31): 20–6. doi: 10.1016/j.aca.2012.06.031.
- [36] Hu M, Dong Q, Liu B, et al. Estimating blueberry mechanical properties based on random frog selected hyperspectral data. *Postharvest Biology and Technology*, 2015; 106(3): 1–10. doi: 10.1016/j.postharvbio.2015.03.014.
- [37] Yu R, Liu T, Xu Y, et al. Analysis of salinization dynamics by remote sensing in Hetao Irrigation District of North China. *Agricultural Water Management*, 2010; 97(19): 52–60. doi: 10.1016/j.agwat.2010.03.009.
- [38] Gao X, Huo Z, Bai Y, et al. Soil salt and groundwater change in flood irrigation field and uncultivated land: a case study based on 4-year field observations. *Environmental Earth Sciences*, 2015; 73(21): 27–39. doi: 10.1007/s12665-014-3563-4.
- [39] Xu C, Zeng W, Huang J, et al. Prediction of Soil Moisture Content and Soil Salt Concentration from Hyperspectral Laboratory and Field Data. *Remote Sensing*, 2016; 8(42): 1–20. doi: 10.3390/rs8010042.
- [40] Huang Q, Xu X, Lü L, et al. Soil salinity distribution based on remote sensing and its effect on crop growth in Hetao Irrigation District. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2018; 34(1): 102–109. doi: 10.11975/j.issn.1002-6819.2018.01.014. (in Chinese)
- [41] Li H, Liang Y, Xu Q, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal Chim Acta*, 2009; 648(1): 77–84. doi: 10.1016/j.aca.2009.06.046.
- [42] Li H, Xu Q, Liang Y. libPLS: An integrated library for partial least squares regression and linear discriminant analysis. *Chemometr Intell Lab*, 2018; 176(15): 34–43. doi: 10.1016/j.chemolab.2018.03.003.
- [43] Huang G, Zhu Q, Siew C. Extreme learning machine: Theory and applications. *Neurocomputing*, 2006; (70): 489–501. doi: 10.1016/j.neucom.2005.12.126.
- [44] Hong Y, Chen S, Zhang Y, et al. Rapid identification of soil organic matter level via visible and near-infrared spectroscopy: Effects of two-dimensional correlation coefficient and extreme learning machine. *Sci Total Environ*, 2018; 644(12): 32–43. doi: 10.1016/j.scitotenv.2018.06.319.
- [45] Viscarra Rossel R A, Taylor H J, McBratney A B. Multivariate calibration of hyperspectral ray energy spectra for proximal soil sensing. *Eur J Soil Sci*, 2007; 58(3): 43–53. doi: 10.1111/j.1365-2389.2006.00859.x.
- [46] Guo Y, Ni Y, Kokot S. Evaluation of chemical components and properties of the jujube fruit using near infrared spectroscopy and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2016; (153): 79–86. doi: 10.1016/j.saa.2015.08.006.
- [47] Morellos A, Pantazi X, Moshou D, Alexandridis T, Whetton R, Tziotziou G, Wiebensohn J, Bill R, Mouazen A M. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosyst. Eng*, 2016; 152(1): 04–16. doi: 10.1016/j.biosystemseng.2016.04.018.
- [48] Oltra-Carrió R, Baup F, Fabre S, et al. Improvement of Soil Moisture Retrieval from Hyperspectral VNIR-SWIR Data Using Clay Content Information: From Laboratory to Field Experiments. *Remote Sens.-Basel*, 2015; 7(3): 184–205. doi: 10.3390/rs70303184.
- [49] Tian G, Yuan H, Chu X, et al. Near Infrared Spectra (NIR) Analysis of Octane Number by Wavelet Denoising-Derivative Method. *Spectroscopy and Spectral Analysis*, 2005; 25(04): 516–520. doi: 10.1016/j.saa.2004.06.052.
- [50] Hong Y, Chen S, Liu Y, et al. Combination of fractional order derivative and memory-based learning algorithm to improve the estimation accuracy of soil organic matter by visible and near-infrared spectroscopy. *Catena*, 2019; 174(1): 04–16. doi: 10.1016/j.catena.2018.10.051.
- [51] Raj A, Chakraborty S, Duda B M, et al. Soil mapping via diffuse reflectance spectroscopy based on variable indicators: An ordered predictor selection approach. *Geoderma*, 2018; (314): 146–59. doi: 10.1016/j.geoderma.2017.10.043.
- [52] Guezennoc J, Bassel L, Gallet-Budynek A, et al. Variables selection: A critical issue for quantitative laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2017; (134): 6–10. doi: 10.1016/j.sab.2017.05.009.
- [53] Jia S, Li H, Wang Y, et al. Recursive variable selection to update near-infrared spectroscopy model for the determination of soil nitrogen and organic carbon. *Geoderma*, 2016; 268(9): 2–9. doi: 10.1016/j.geoderma.2016.01.018.
- [54] Vohland M, Ludwig M, Thiele-Bruhn S, et al. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma*, 2014; 223(225): 88–96. doi: 10.1016/j.geoderma.2014.01.013.
- [55] Yao X, Yang W, Li M, et al. Prediction of Total Nitrogen in Soil Based on Random Frog Leaping Wavelet Neural Network. *IFAC-PapersOnLine*, 2018; 51(6): 1–5. doi: 10.1016/j.ifacol.2018.08.121.
- [56] Wijewardane N K, Ge Y, Morgan, et al. Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. *Geoderma*, 2016; 26(7): 92–101. doi: 10.1016/j.geoderma.2015.12.014.
- [57] Yun Y, Li H, Wood L R, et al. An efficient method of wavelength interval selection based on random frog for multivariate spectral calibration. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2013; 111(3): 1–6. doi: 10.1016/j.saa.2013.03.083.
- [58] Yu L, Zhu Y, Hong Y, et al. Determination of soil moisture content by hyperspectral technology with CARS algorithm. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2016; 32(22): 138–145. doi: 10.11975/j.issn.1002-6819.2016.22.019. (in Chinese)
- [59] Zhang K, Luo M. Outlier-robust extreme learning machine for regression problems. *Neurocomputing*, 2015; 151(15): 19–27. doi: 10.1016/j.neucom.2014.09.022.