


Article

Salinity Monitoring at Saline Sites with Visible–Near-Infrared Spectral Data

Wei Li ¹, Jing Liu ^{2,*} , Nisha Bao ³, Xinqi Mao ⁴, Yachun Mao ³, Yanhua Fu ⁵, Wang Cao ³, Jiaqi Huang ³ and Zhanguo Zhao ⁶

- ¹ School of Civil Engineering, University of Science and Technology Liaoning, Anshan 114051, China; ln_as_lw@163.com
 - ² Centre for Advanced Spatial Analysis, Faculty of the Built Environment, University College London, London WC1E 6BT, UK
 - ³ School of Resources and Civil Engineering, Northeastern University, Shenyang 110819, China; baonisha@mail.neu.edu.cn (N.B.); maoyachun@mail.neu.edu.cn (Y.M.); dbdxcw@163.com (W.C.); 1900994@stu.neu.edu.cn (J.H.)
 - ⁴ Software College of Northeastern University, Northeastern University, Shenyang 110819, China; maxqofficial@foxmail.com
 - ⁵ School of Architecture, Northeastern University, Shenyang 110819, China; fuyanhua@mail.neu.edu.cn
 - ⁶ China Gold Group, Beijing 100000, China; nmgjy@vip.sohu.com
- * Correspondence: ucfnruk@ucl.ac.uk

Abstract: To address the global phenomenon of the salinisation of large land areas, a quantitative inversion model of the salinity of saline soils and soil visible–near-infrared (NIR) spectral data was developed by considering saline soils in Zhenlai County, Jilin Province, China as the research object. The original spectral data were first subjected to Savitzky–Golay (SG) smoothing, multiplicative scattering correction (MSC) pre-processing, and a combined transformation technique. The pre-processed spectral data were then analysed to construct the difference index (DI), ratio index (RI), and normalised difference index (NDI), and the Spearman rank correlation coefficient (r) between these three spectral indices and the salt content in the samples was calculated, while a combined spectral index ($r > 0.8$) was eventually selected as a sensitive spectral index. Finally, a quantitative inversion model for the salinity of saline soils was developed, and the model's accuracy was evaluated based on partial least squares regression (PLSR), the random forest (RF) algorithm, and the radial basis function (RBF) neural network algorithm. The results indicated that the inversion of soil salt content using the selected combination of spectral indices based on the RBF neural network algorithm was the most effective, with the prediction model yielding an R^2 value of 0.950, a root mean square error (RMSE) of 1.014, and a relative percentage deviation (RPD) of 4.479, which suggested a good prediction effect.

Keywords: saline soils; visible–NIR spectra; spectral indices; salinity; inversion models



Citation: Li, W.; Liu, J.; Bao, N.; Mao, X.; Mao, Y.; Fu, Y.; Cao, W.; Huang, J.; Zhao, Z. Salinity Monitoring at Saline Sites with Visible–Near-Infrared Spectral Data. *Minerals* **2021**, *11*, 1086. <https://doi.org/10.3390/min11101086>

Academic Editor: Tiago Osorio Ferreira

Received: 15 September 2021
Accepted: 1 October 2021
Published: 2 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Soil is a valuable resource on which humans depend for survival. As the area of arable land continues to decrease with increasing social development and the pressure on areas due to increases in food cultivation, saline land has become an effective arable land back-up resource. In recent years, the problems of desertification and soil salinisation in arid and semi-arid regions have become increasingly serious [1–3]. According to statistics, approximately 836 million hectares (Mha) of land worldwide (3% of the total geographical area) are affected by soil salinisation, of which 48% comprises saline soils and 52% comprises sodic soils [4]. The causes of soil salinisation are divided into natural and human factors. Natural factors include climatic, hydrological, geological, and geomorphological factors and tectonic-movement-related factors [1,5], while human factors include blind land opening, excessive woodcutting, overgrazing, and unreasonable engineering

and construction [6]. The Songnen Plain in China is one of the three major sodic saline distribution areas globally, with over 3.0×10^6 hectares of saline soils at present [7], which are characterised by well-matched soil and water resources, large areas, and concentrated contiguity features [8,9]. As the salinity level affects plant growth, quantitative assessment of soil properties is essential in order to maintain and improve soil quality and enhance crop yield [10–12].

The western region of Jilin Province, China (southwestern part of the Songnun Plain) has a dry climate where the evaporation is about 2.5–4 times greater than the precipitation. Salt accumulates in spring and autumn because the evaporation is greater than the precipitation, desalinates in summer because of abundant precipitation (precipitation is mostly concentrated in June to August, accounting for 85–90%), and stops salt migration in winter because the permafrost layer hinders water–salt movement, making it easy to form saline soils because of spring droughts and summer floods. Due to the different natural and anthropogenic factors in each region of Zhenlai County, Baicheng City, Jilin Province, the distribution of saline soils has more obvious geographical variability, and the central part of Zhenlai County is mainly diffuse granite land, where waters and saline soils are interspersed, and the evaporation from the surface becomes extensive.

The traditional method of regional soil salinity status determination encompasses field surveying and analysis of soil, which are not only time-consuming and labour-intensive methods, but they also consider few measurement points, thus exhibiting a poor representativeness, and they do not achieve real-time dynamic monitoring of large areas. With the rise of remote sensing technology, the application of hyperspectral remote sensing techniques to monitor soil salinity not only saves time and effort, but also exhibits the characteristics of rapid deployment, macroscopic and dynamic features, and a superior performance that cannot be achieved by other means. In recent years, numerous scholars have studied soil salinity estimation for different soil types based on visible–near-infrared (NIR) spectroscopy [13–21]. The background, environment, measuring instruments, etc. at the time of spectral acquisition commonly impact the spectrum of a given soil sample, and spectral data pre-processing is therefore necessary [22,23]. Common pre-processing methods for spectral data include scaling (centring, normalisation, etc. [24]), baseline correction (first-order derivative, continuous wavelet transform, etc. [25]), scattering correction (multiplicative scattering correction (MSC), standard normal transform, etc. [23,26]), and smoothing (Savitzky–Golay (SG) smoothing, etc. [27,28]). The above pre-processing methods effectively remove the effects of the instrument background or drift on the signal, eliminate the effects of scattering due to the inhomogeneous distribution of particles and different particle sizes on the spectrum, remove or reduce noise, and improve the signal-to-noise ratio. As a single pre-processing method can only attenuate one specific interference aspect, a combination of different pre-processing methods can be implemented to attenuate multiple interference aspects [29–31]. Zipeng Zhang et al. compared nine pre-processing methods based on 233 soil samples retrieved from Xinjiang, Northwest China, including SG smoothing, discrete wavelet transform (DWT), first-order derivative (FD), second-order derivative (SD), MSC, standard normal variate and detrending (SNV-DT), and continuous peak removal (CR). The results revealed that SG smoothing proved to be the most effective pre-treatment method for soil organic matter (SOM) in saline soils. The pseudo-absorbance value and detrending (Abs-DT), FD, SD, SNV-DT, MSC, CR, DWT, and CMR resulted in prediction performance reductions [32]. Yiming Bi proposed a localised version of the standard normal variable (SNV) in which correction parameters were estimated from local spectral regions, and the localised approach yielded advantages in terms of spectral linearity correction, model interpretation, and prediction accuracy over the full-range approach [33]. Lihua Xu et al. employed laboratory reflectance spectra of 42 soil samples to compare four pre-treatment methods and found that FD combined with the principal component (PC)-transformed multiple linear regression (MLR) method resulted in a better total potassium (TK) prediction model [34].

In terms of the choice of the modelling method, existing modelling methods include principal component regression (PCR) [35,36], partial least squares regression (PLSR) analysis [35,37], support vector machines (SVMs) [38,39], back-propagation neural networks (BPNNs) [40], random forests (RFs) [41], and extreme learning machines (ELMs) [42]. Chengwen Chang et al. used principal component regression to predict 33 chemical, physical, and biochemical properties of 802 soil samples collected from four major land resource areas with a prediction accuracy of 0.8 [36]. Rossel R et al. compared multiple linear regression (MLR), partial least squares regression (PLSR), multivariate adaptive regression splines (MARS), support vector machines (SVMs), random forests (RFs), boosted trees (BTs), and artificial neural networks (ANNs) to estimate soil organic carbon (SOC), clay content (CC), and pH measured in water (pH) [39]. Dong Xiao et al. proposed a method to convert 1D into 2D spectral data and efficiently detect the salt content based on convolutional neural networks, gravitational search algorithms, and reservoir computation ELMs and obtained a coefficient of determination of 0.9 and a root mean square error (RMSE) value of 1.55 [43]. Gopal Ramdas Mahajan et al. compared five modelling methods—namely, linear PCR, PLSR, non-linear multiple adaptive regression splines (MARS), RF, and support vector regression (SVR)—and found that the SVR method performed the best based on 372 sites in 17 coastal areas along the west coast of India. PLSR and PCR were found to be better prediction methods than MARS and RF [14]. E.S. Mohamed et al. applied stepwise multiple linear regression (SMLR), PLSR, MARS, PCR, and an artificial neural network (ANN) to rapidly and accurately predict the salinity, organic carbon, soil moisture, and heavy metals in soils [44].

At present, there are few studies combining visible–NIR spectroscopic techniques with geostatistical methods to conduct large-scale spatial heterogeneity analyses of soil salinity in arid and semi-arid regions [45]. Therefore, in this paper, based on previous studies, original spectral data processed via four pre-processing methods were first transformed into three spectral indices—namely, the difference index (DI), ratio index (RI), and normalised index (NDI)—and Spearman's rank correlation coefficient was then calculated for these transformed spectral indices with the salt content with $r > 0.8$ (strong correlation). Final inversion of the salt content of saline soils was achieved based on PLSR, the radial basis function (RBF) neural network algorithm, and the RF algorithm.

2. Materials and Methods

2.1. Description of the Sampling Sites

Zhenlai County is located in the northwestern part of Jilin Province, China and is the convergence zone of the Songnen Plain and the Horqin Grassland. The geographic range extends from $122^{\circ}47'06.3''$ to $124^{\circ}04'33.7''$ east longitude and from $45^{\circ}28'14.3''$ to $46^{\circ}18'15.8''$ north latitude, and the area experiences a temperate continental monsoon climate. The average annual precipitation in the region is 402 mm, and evaporation is high due to the continental monsoon climate, with an average annual evaporation of 1755.9 mm, which is four times higher than the annual precipitation, resulting in drought-prone conditions in Zhenlai County. Moreover, the Songnen Plain contains accumulated clay deposits and fine sandy sediments with a poor permeability in lake facies, which are susceptible to the formation of saline land due to the accumulation of surface water and evaporation. Therefore, this area was chosen as the study area, and its geographical location is shown in Figure 1.

2.2. Soil Sampling

Field samples were collected in mid-May 2019 and soil samples were collected according to the five-point method, whereby a total of 100 samples were obtained from the surface soil layer (0–20 cm). The soil samples were transported to the laboratory, where the grass roots and stones were removed from the soil, and the samples were dried in a soil-drying oven to remove the effect of moisture on the subsequent spectral measurements, after which the samples were ground and passed through a 100-mesh sieve. The prepared

soil samples were then placed in 6-cm-diameter black round boxes for spectral acquisition and salt content determination.

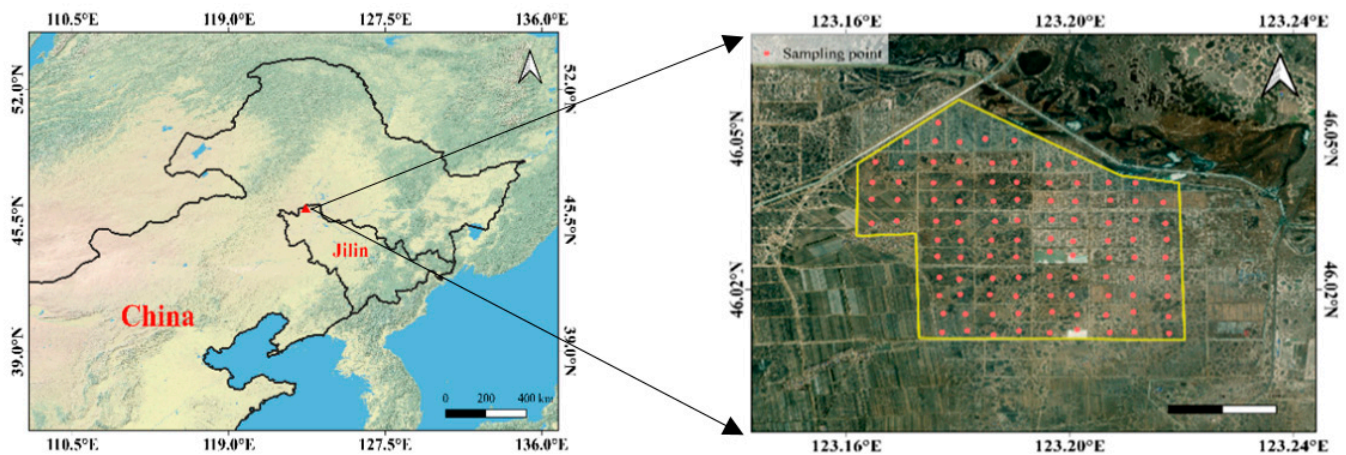


Figure 1. Geographical location of Zhenlai County and spatial distribution of the collected saline samples.

2.3. Analysis of the Soil Properties

The soil salinity of the soil samples in this paper was measured at the Chemical Analysis Center of Northeastern University, and the soil surface salinity characteristics are listed in Table 1 below.

Table 1. Statistical description of the properties of the saline soil samples.

Property	Mean	Standard Error	Minimum	Maximum	Coefficient of Variation (%)	Skewness	Kurtosis
Full dataset ($n = 100$)	5.06	0.42	0.22	17.36	83%	1.15	0.55
Calibration dataset ($n = 67$)	4.82	0.51	0.22	17.36	87%	1.35	1.20
Validation dataset ($n = 33$)	5.56	0.74	0.48	15.45	77%	0.84	−0.25

Soils were classified into non-salinized ($<1.0 \text{ g/kg}^{-1}$), mildly salinized ($1.0\text{--}2.0 \text{ g/kg}^{-1}$), moderately salinized ($2.0\text{--}4.0 \text{ g/kg}^{-1}$), severely salinized ($4.0\text{--}6.0 \text{ g/kg}^{-1}$), and saline ($>6.0 \text{ g/kg}^{-1}$) soil according to classification of soil salt content in the coastal area. In this paper, according to the focus of the research, soils were classified into non-salinized soil ($<1.0 \text{ g/kg}^{-1}$) and salt-affected soil or salinized soil ($>1.0 \text{ g kg}^{-1}$). According to the table below, we can see that the saline land type belongs to severely salinized land.

2.4. Spectroscopic Measurements

The raw spectral data of the saline soil samples were collected with an SVC HR-1024 portable geophysical spectrometer (USA) in the wavelength range from 350–2500 nm at a spectral accuracy better than $\pm 0.5 \text{ nm}$ and a minimum integration time of 1 s. To avoid any effects of the measurement background, the test samples were placed in a small circular black box with a diameter of 6 cm, and a spectrometer lens was placed perpendicularly to the sample observation surface. The experiments were carried out from 10:00 to 14:00. The sky was required to be clear and cloudless with a sun height angle of approximately 45° . To accurately determine the actual reflectance of the samples, spectroscopic experiments were conducted, after which the average of the results of five tests was calculated and adopted as the final reflectance of the saline soil samples. The procedure of the spectroscopy experiment is shown in Figure 2.



Figure 2. Spectral data collection process.

2.5. Spectral Data Pre-Processing

The visible–NIR spectral data obtained were subjected to spectral pre-processing to remove the physical variability attributed to light dispersion and eliminate any systematic variations in instrumental and environmental conditions in order to emphasise the characteristics of interest along the spectrum. The SG smoothing algorithm is widely employed to smooth and denoise spectral data and is a filtering method based on local polynomial least squares fitting in the time domain, which exhibits the advantage of smoothing and denoising of a wide range of signals regardless of sample data limitations. MSC is a pre-processing method that separates the scattering signal from the chemical absorption information contained in a spectrum, thus eliminating differences in the NIR spectrum of the same batch of samples caused by inhomogeneous sample particles during diffuse reflection. Both the SG smoothing algorithm and MSC are promising spectral pre-processing methods. However, the pre-processing effectiveness of the individual MSC and SG smoothing pre-processing methods (or in combination) varies. In regard to visible–NIR spectroscopic determination of the soil salinity, whether the above two spectral pre-treatments alone or in combination are more effective should be assessed via comparative experiments and model predictions. Little in-depth research has been performed in this area, but it is important to improve the predictive power of visible–NIR spectroscopy, especially for a complex system such as soil.

2.6. Dual-Band Spectral Index Construction

To improve the accuracy of the soil salt content inversion, three spectral indices, i.e., DI, RI, and NDI, were constructed. The established indices were applied to correlate the spectral characteristics of the experimental samples with the salt content, and the equations for each index are as follows [41]:

$$RI = \frac{R_{\lambda m}}{R_{\lambda n}} \quad (1)$$

$$DI = R_{\lambda m} - R_{\lambda n} \quad (2)$$

$$NDI = \frac{R_{\lambda m} - R_{\lambda n}}{R_{\lambda m} + R_{\lambda n}} \quad (3)$$

where:

λ_m is the position of the wavelength corresponding to point m ,

λ_n is the position of the wavelength corresponding to point n ,

$R_{\lambda m}$ is the reflectance value corresponding to the wavelength at point m , and

$R_{\lambda n}$ is the reflectance value corresponding to the wavelength at point n .

2.7. Model Calibration and Validation

The complete dataset ($n = 100$) was arranged according to the soil salinity in ascending order, and two groups were established as the calibration dataset, while one group was established as the validation dataset. Sixty-seven calibration datasets and 33 validation datasets were finally selected, and the technical flowchart is shown in Figure 3. In the prediction of the soil salt content, the collected spectral data are subject to a variety of uncertainties and exhibit a high degree of stochasticity and non-linearity, so this paper selects PLSR, the RF algorithm, and the RBF neural network algorithm as prediction models. PLSR provides a method for modelling many-to-many linear regression, especially when the sample size of the observed data is small and there are multiple correlations among the variables. PLSR modelling can provide advantages that are not offered by traditional methods, such as classical regression analysis, which can effectively simplify the data structure and thus solve the problem of a high linear correlation between multiple independent variables. The RF algorithm exhibits a clear structure, is easy to interpret, operates efficiently, has low data requirements, attains a good noise resistance, processes high-dimensional data without feature selection, is fast to train, achieves a strong generalisation capability, is relatively easy to implement under parallel computing, is less susceptible to overfitting problems, and balances errors in unbalanced data. The RBF neural network is a feed-forward neural network with only a 3-layer network structure, which attains a better function approximation ability due to its structural characteristics and effectively overcomes some of the shortcomings of the traditional BPNN.

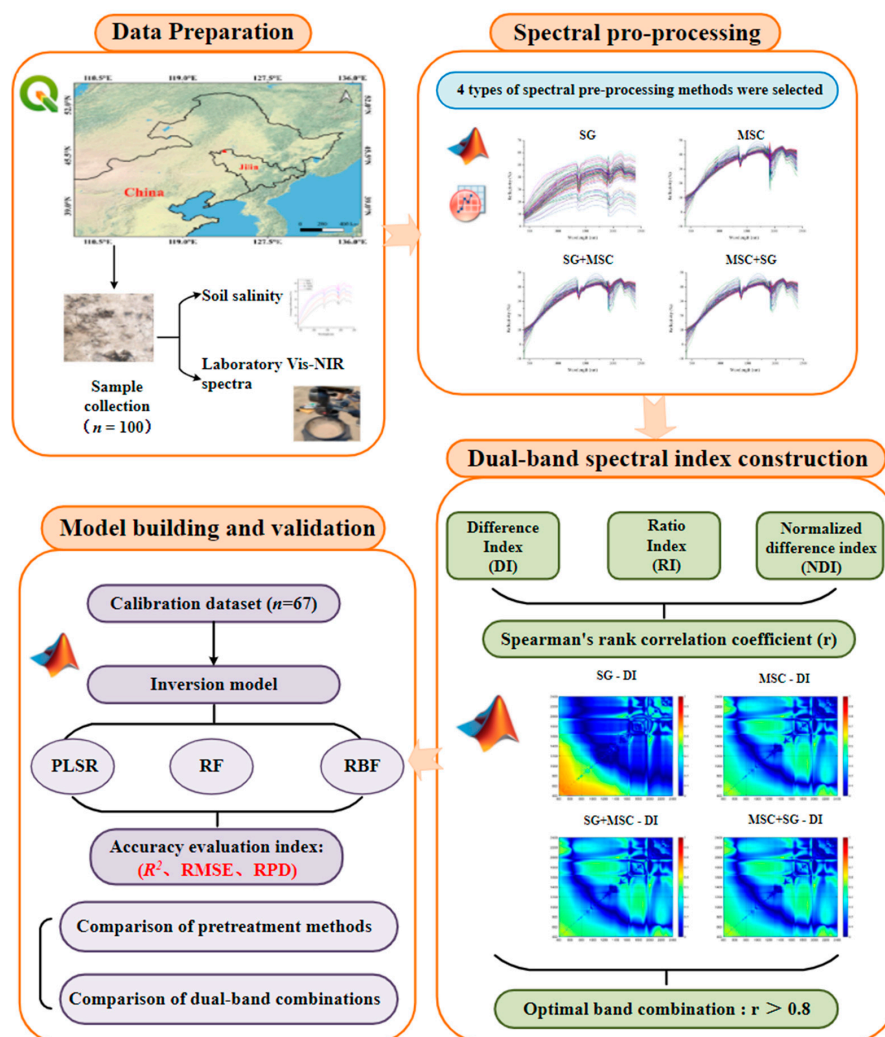


Figure 3. Flowchart of this study.

In this paper, R^2 , RMSE, and relative percentage deviation (RPD) are used to evaluate the accuracy of the model, where R^2 reflects the accuracy of modelling and validation, RMSE reflects the deviation between the salinity prediction results and the actual measured salinity values, and RPD reflects the strength of the model's prediction ability. In general, for $RPD < 1.0$, the model does not exhibit a predictive power, and for $1.0 < RPD < 1.5$, the model can only roughly estimate high and low sample values. For $1.5 < RPD < 2.0$, the model achieves a quantitative predictive power, and for $2.0 \leq RPD < 2.5$, the model attains a good predictive power, while for $2.5 \leq RPD < 3.0$, the model has a very good predictive ability. Moreover, the closer R^2 is to 1, the smaller the RMSE is, and the higher the RPD is, the better the model's accuracy will be. R^2 , RMSE, and RPD are calculated with the following equations.

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (5)$$

$$RPD = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}}{RMSEP} \quad (6)$$

where \hat{y}_i denotes the estimated value of the model, y_i denotes the observed value, and N is the number of observations of the variable to be modelled.

3. Results and Discussion

3.1. Characteristics of the Reflection Spectrum Curve of the Soil Samples

The reflectance of the 100 surface soil samples was classified according to the degree of salinisation and then averaged as spectral reflectance curves considering the different salinisation degrees. As shown in Figure 4, the four curves follow a similar trend, and the higher the soil salinity is, the higher the reflectance, with the spectral reflectance ranging from 5% to 60% and exhibiting fluctuations between 400 and 2400 nm. Between 400 and 1000 nm, the spectral reflectance increases rapidly with increasing wavelength, and between 1000 and 1400 nm, the spectral reflectance continues to increase with increasing wavelength, but the slope of the increasing curve gradually decreases. Between 1400 and 1900 nm, i.e., the two moisture absorption bands, the spectral curve still exhibits an increasing trend, but the trend is gentler. After approximately 2100 nm, the reflectance generally follows a decreasing trend. Near 1400, 1900, and 2200 nm, the moisture absorption characteristics are very obvious. The absorption characteristics at approximately 1400 nm represent water absorption combined with the vibration of the O-H bond of water. The characteristic absorption features at approximately 1900 and 2200 nm are attributed to the vibration caused by the bending and stretching of Al-OH and Mg-OH bonds [46].

3.2. Study of the Spectral Data Pre-Processing Methods

Due to the poor results of the edge band test, two bands of data from 350–399 nm and 2401–2500 nm were removed to obtain 890 bands and generate visible–NIR spectral reflectance curves of the 100 soil samples. To establish an optimised model for the analysis of the visible–NIR spectral data of the soil salinity and to thoroughly examine the influence of the spectral pre-processing method on the analytical capability of these data, the following five scenarios are discussed in detail: (1) no pre-processing (Figure 5); (2) SG smoothing pre-processing alone (Figure 6); (3) MSC pre-processing alone (Figure 7); (4) SG smoothing pre-processing followed by MSC pre-processing (Figure 8); (5) MSC pre-processing followed by SG smoothing pre-processing (Figure 9).

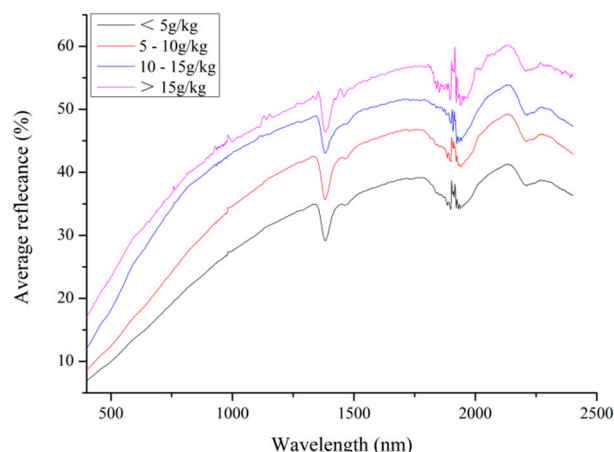


Figure 4. Raw reflectance spectra based on the soil salinity class.

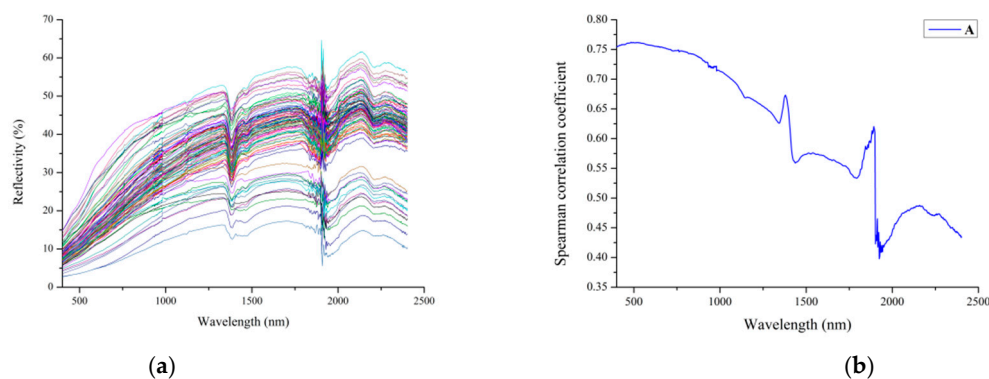


Figure 5. Original spectral curve and correlation coefficient with the soil salt content. (a): Original spectral curve (b): Correlation coefficient with the soil salt content.

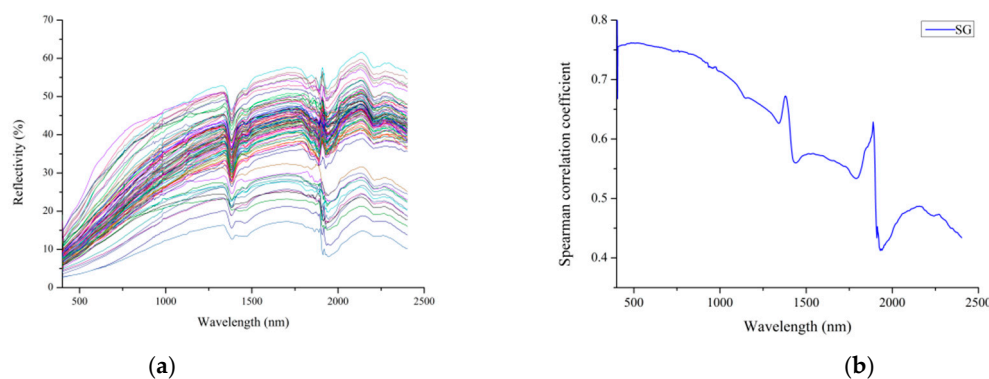


Figure 6. Spectral curve after SG smoothing and correlation coefficient with the soil salinity. (a): Spectral curve after SG smoothing (b): Correlation coefficient with the soil salinity.

Comparing the effects of the four spectral pre-treatment methods on the visible–NIR spectra and calculating the Spearman rank correlation coefficient between the pre-treated spectral data and the salt content in the soil samples, the highest positive correlation between the SG-smoothed spectral data and the salt content in the soil samples was 0.8672 at 402.4 nm, which was higher than the highest positive correlation between the original data and the salt content in the soil samples (0.7621). The lowest positive correlation between the SG-smoothed spectral data and the salt content in the soil samples reached 0.4124 at 1929.7 nm, which was higher than the lowest positive correlation between the original data and the salt content in the soil samples (0.3978). Moreover, the SG smoothing pre-processing method effectively weakened the effect of noise on the original spectra. The

highest positive correlation between the MSC-treated spectral data and the salinity of the soil samples was 0.4931 at 685.4 nm, and the highest negative correlation was -0.4897 at 402.4 nm. The highest positive correlation between the SG+MSC-treated spectral data and the salinity of the soil samples reached 0.4915 at 685.4 nm, and the highest negative correlation was -0.5689 at 405.5 nm. The highest positive correlation between the MSC+SG-treated spectral data and the salinity of the soil samples was 0.4922 at 685.4 nm, and the highest negative correlation reached -0.5684 at 405.5 nm. The combination of pre-treatment methods (SG+MSC and MSC+SG) improved the negative correlation between the treated spectral data and the salt content in the samples over the single pre-treatment method (MSC), demonstrating that the combination of pre-treatment methods combines the advantages of the two single pre-treatment methods.

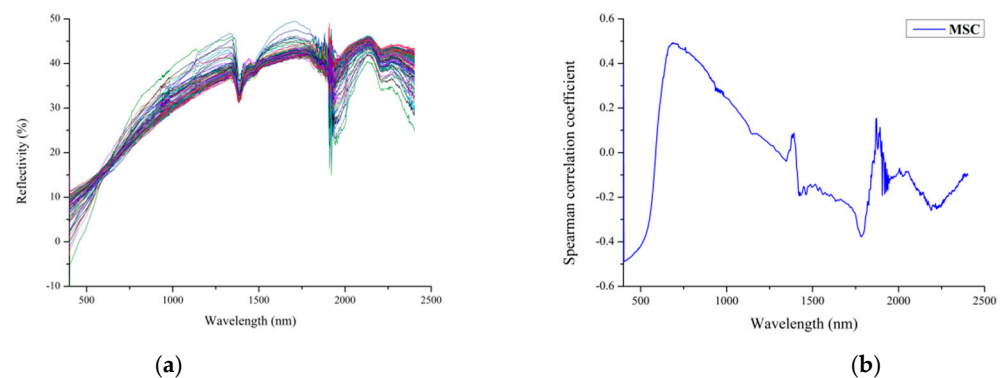


Figure 7. Spectral curve after MSC pre-processing and correlation coefficient with the soil salt content. (a): Spectral curve after MSC pre-processing (b): Correlation coefficient with the soil salt content.

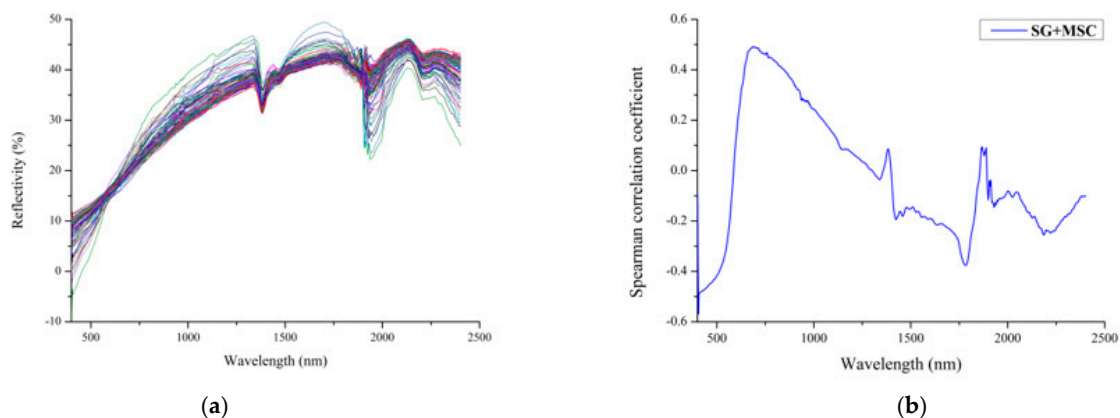


Figure 8. Spectral curve after SG + MSC pre-processing and correlation coefficient with the soil salt content. (a): Spectral curve after SG + MSC pre-processing (b): Correlation coefficient with the soil salt content.

3.3. Selection of Dual-Band Spectral Index Combinations

Three spectral indices, i.e., RI, DI, and NDI, were constructed using the spectral data processed with the above five spectral pre-processing methods, and the Spearman rank correlation coefficient between the salt content in the soil samples and these spectral indices was determined. The distribution of the correlations between DI after SG smoothing, DI after MSC pre-processing, DI after SG+MSC pre-processing, and DI after MSC+SG pre-processing and the salt content in the soil samples considering the raw spectral data is shown in Figure 10. The distribution of the correlations between RI after SG smoothing, RI after MSC pre-processing, RI after SG+MSC pre-processing, and RI after MSC+SG pre-processing and the salt content in the soil samples considering the raw spectral data of the soil salinity is shown in Figure 11. The distribution of the correlations between NDI

after SG smoothing, NDI after MSC pre-processing, NDI after SG+MSC pre-processing, and NDI after MSC+SG pre-processing and the salt content in the soil samples considering the raw soil salinity spectral data is shown in Figure 12.

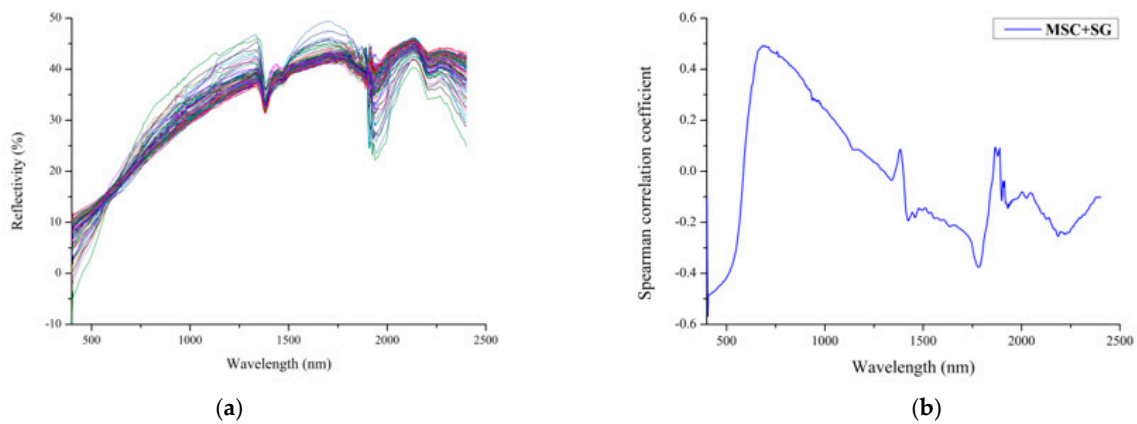


Figure 9. Spectral curve after MSC + SG pre-processing and correlation coefficient with the soil salt content. (a): Spectral curve after MSC + SG pre-processing (b): Correlation coefficient with the soil salt content.

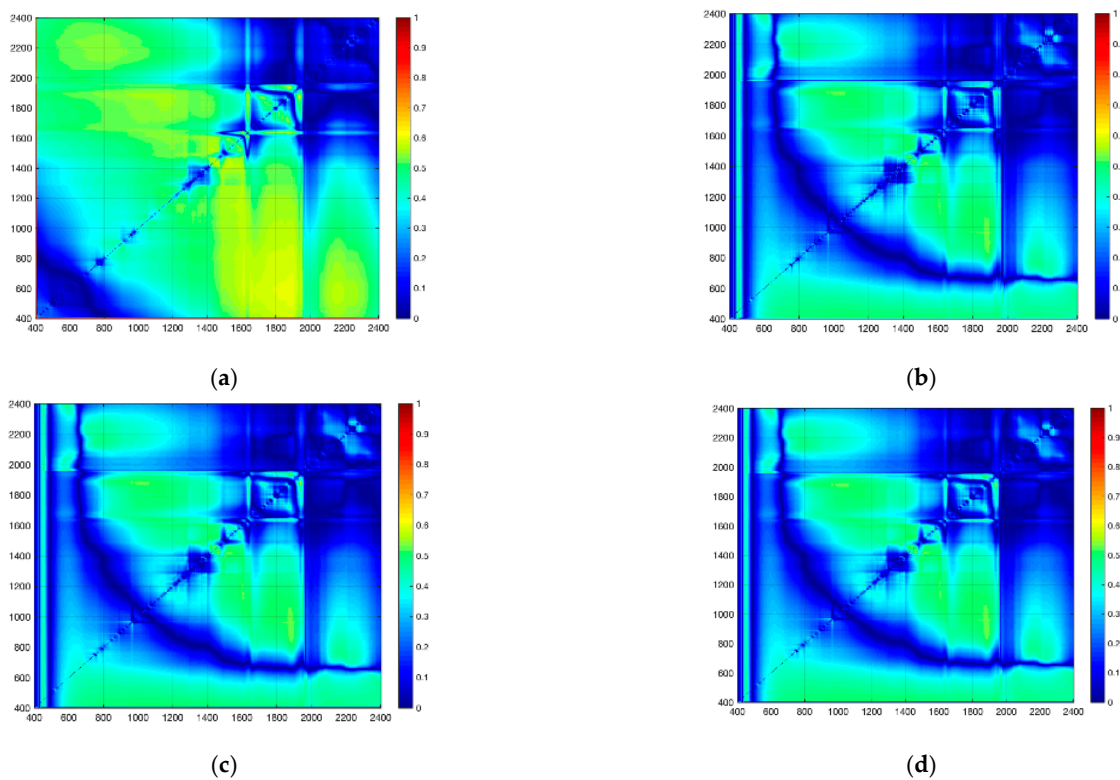


Figure 10. Correlation distribution of the four pre-treatment methods combined with the difference index and soil salinity. (a): DI after SG smoothing (b): DI after MSC pre-processing (c): DI after SG+MSC pre-processing (d): DI after MSC+SG pre-processing.

The horizontal and vertical coordinates in Figures 10–12 indicate the spectral wavelengths of the soil samples. The colours in the graphs indicate the absolute value of the correlation coefficient between the various spectral indices (RI, DI, and NDI) and the salt content in the samples at that point. The darker the red colour is, the stronger the correlation between the corresponding spectral index and the salt content. Based on the distribution plots, the effect of SG smoothing pre-processing followed by dual-band index processing is significantly better than that of the other three spectral pre-processing methods, and the correlation between RI and the salt content after SG smoothing is stronger. This approach

can effectively enhance the correlation between this spectral index and the salt content and provides an effective input dataset for subsequent modelling inversion.

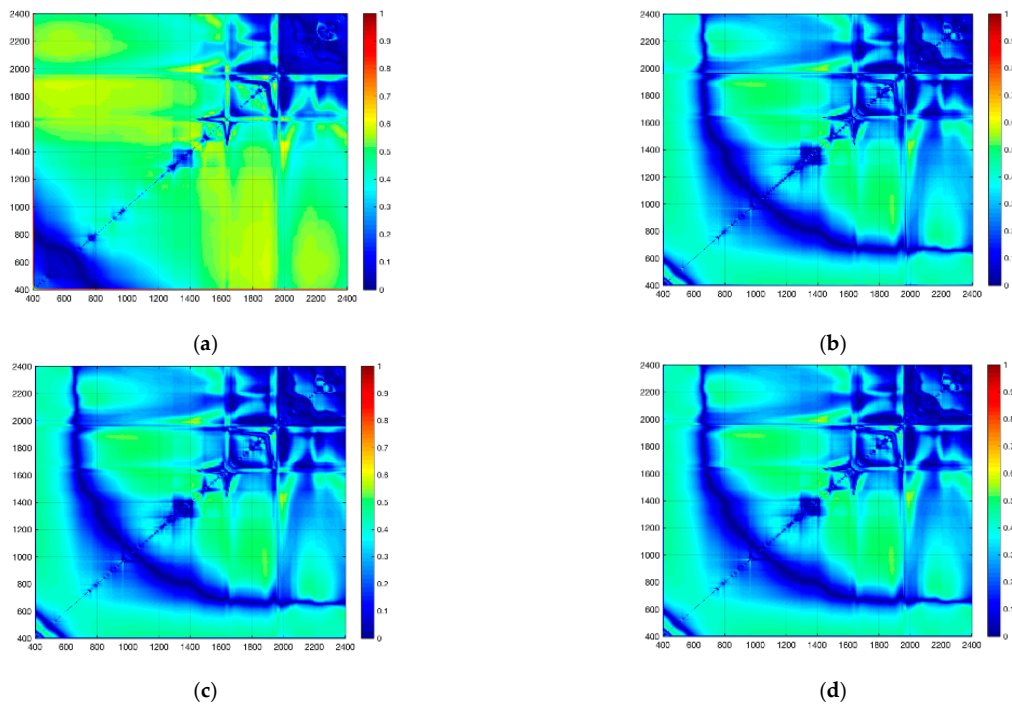


Figure 11. Correlation distribution of the four pre-treatment methods combined with the ratio index and soil salinity. (a): RI after SG smoothing (b): RI after MSC pre-processing (c): RI after SG+MSC pre-processing (d): RI after MSC+SG pre-processing.

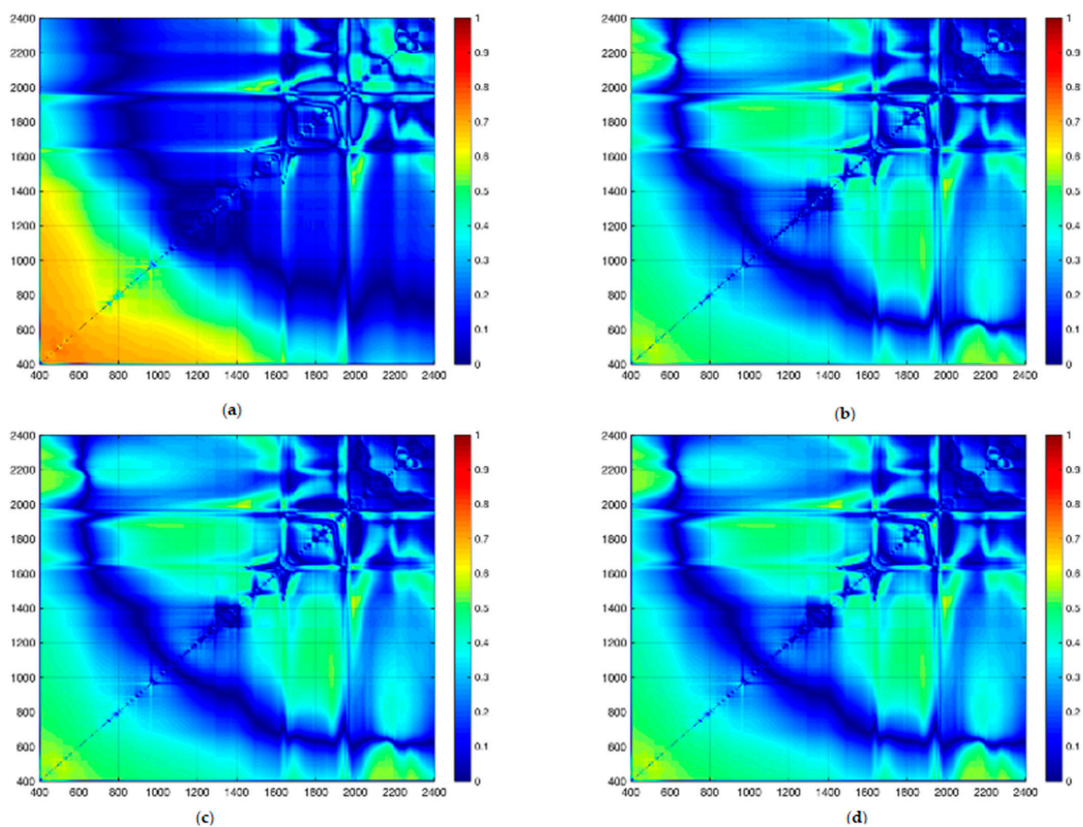


Figure 12. Correlation distribution of the four pre-treatment methods combined with the normalised index and soil salinity. (a): NDI after SG smoothing (b): NDI after MSC pre-processing (c): NDI after SG+MSC pre-processing (d): NDI after MSC+SG pre-processing.

3.4. Soil Salinity Prediction

3.4.1. Optimal Band Combination

Spearman's rank correlation coefficient, which was proposed by Spearman and is widely used in the field of statistics, is a statistical index that reflects the correlation between two groups of variables. The value of SRCC ranges from -1 to 1 , and the larger the value is, the stronger the correlation [47]. Spearman's rank correlation coefficient (r) between the above spectral indices and the salt content in the samples was calculated, and $r > 0.8$ indicates a strong correlation. Therefore, the spectral indices with $r > 0.8$ were selected to form a combination of spectral indices in this paper, and the specific selection results are listed in Table 2. According to the statistical results in the table below, the SG smoothing pre-processing algorithm is superior to raw spectral pre-processing and can effectively eliminate the effect of noise in spectral data, thus providing a solid data source for the selection of dual-band indices. Moreover, comparing the three spectral indices, RI resulted in the selection of 1334 groups, accounting for 45% of the total number of groups, and NDI resulted in the selection of 1188 groups, accounting for 40% of the total number of groups. This demonstrates that RI and NDI are more likely to be selected than DI and contribute more to the selected input dataset, which can effectively improve the modelling accuracy and effectiveness.

Table 2. Principle of the optimal selection of spectral indices of the soil salinity.

Property	Spectral Index	Pre-Processing Methods	Groups	Total
Salinity	DI	A	72	2950
		MSC+SG	42	
		SG	272	
		SG+MSC	42	
		RI	1334	
		NDI	1188	

According to the statistical results, the sensitive bands of DI selected based on a Spearman's rank correlation coefficient value of $r > 0.8$ were mainly concentrated from 400.9–700.3 nm, and those of RI and NDI were mainly concentrated from 400.9–1000 nm, 1400–2000 nm, and 2100–2400 nm, which is basically consistent with the findings of previous studies on the spectral characteristics of salinised soils.

3.4.2. Model Building and Validation

With the use of the above combined spectral index data and with the salinity data as the input data, a soil salinity inversion model was developed based on PLSR [48], the RF algorithm [49], and the RBF neural network algorithm [50], and the results of the comparison of the predicted and actual salinity values are shown in Figures 13–15. The R^2 values of the inversion comparison curves were 0.941, 0.938, and 0.950, respectively, the RMSE values were 1.035, 1.065, and 1.014, respectively, and the RPD values reached 4.327, 4.043, and 4.479, respectively.

The evaluation metrics for inversion modelling of the soil salinity based on PLSR, the RF algorithm, and the RBF neural network algorithm are summarised in the table below (Table 3). All three inversion models attain an excellent predictive capability, with R^2 reaching above 0.9 and RPD above 3.0. The RBF neural network algorithm is superior to the other two algorithms, with R^2 reaching 0.95. The reason for this finding is that the RBF network can approximate any non-linear function with an arbitrary accuracy and achieves a global approximation capability, which fundamentally solves the local optimum problem and improves the model's accuracy.

Table 3. Comparison results for the salt content inversion evaluation indices based on the random forest algorithm and partial least squares regression.

Property	Model	R ²	RMSE	RPD
Salinity	PLSR	0.941	1.035	4.327
	RF	0.938	1.065	4.043
	RBF	0.950	1.014	4.479

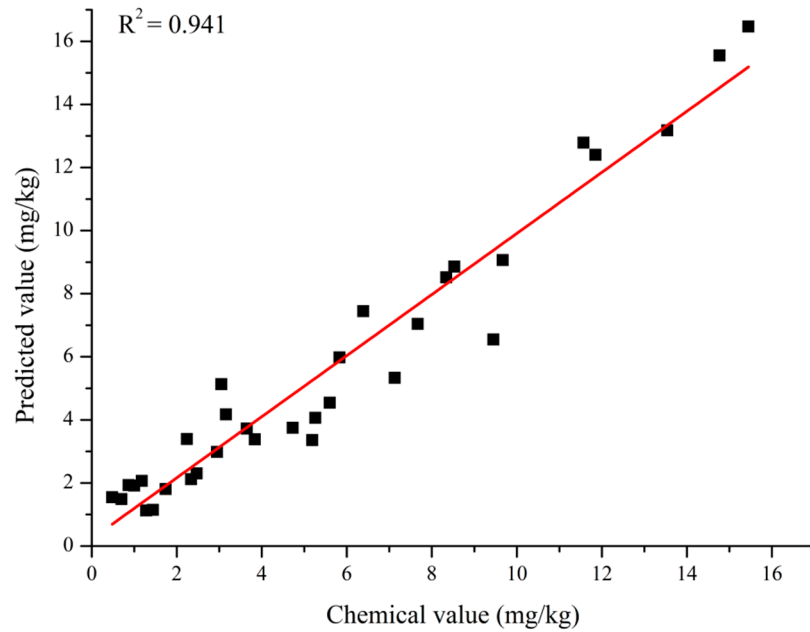


Figure 13. Comparison of the salt content predicted via partial least squares regression with the measured salt content.

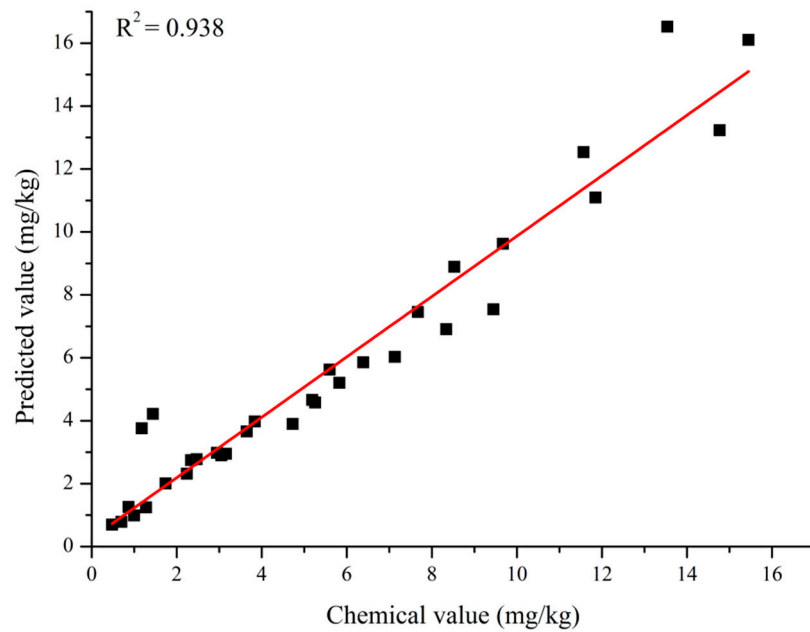


Figure 14. Comparison of the salt content predicted with the random forest algorithm with the measured salt content.

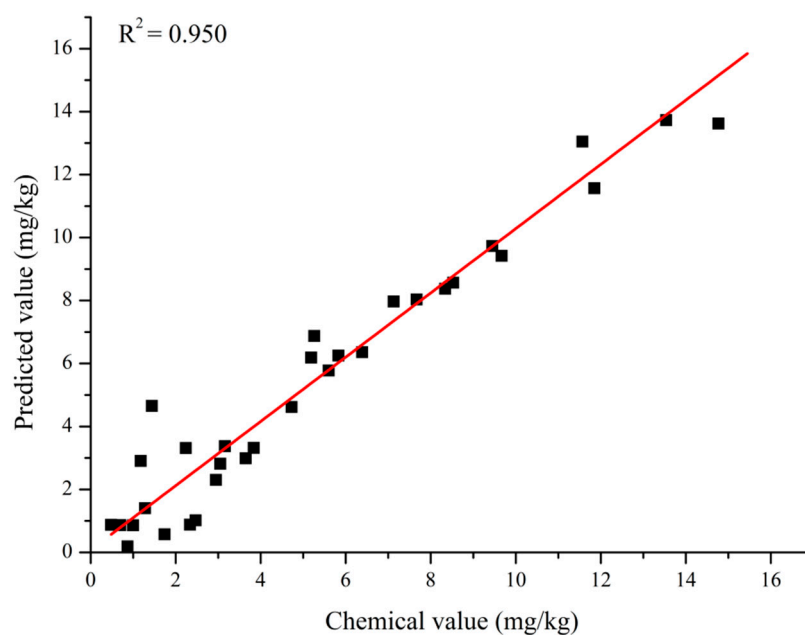


Figure 15. Comparison of the salt content predicted with the radial basis function neural network algorithm with the measured salt content.

4. Conclusions

In this paper, soil visible–NIR spectral data were used and salt content data pertaining to 100 saline soil samples retrieved from Zhenlai County, Baicheng city, Jilin Province, China were applied as the data source. The original spectral data were first subjected to Savitzky–Golay (SG) smoothing, multiplicative scattering correction (MSC) pre-processing, and a combined transformation technique. Then, we proposed a combination of soil spectral indices as the modelling input variable based on three algorithms—namely, PLSR, RF, and an RBF neural network—and compared and analysed various evaluation indices of these three modelling methods. Comparing the four pre-processing methods, SG smoothing pre-processing was able to effectively reduce the influence of environment-induced noise on the raw spectral data. Comparing the above three spectral indices, RI and NDI effectively highlighted spectral features after pre-processing, providing a solid data source for inversion modelling. Based on the statistics of the combination of the spectral indices, it was found that four spectral bands in the visible (400.9–700.3 nm) and NIR (900–1000 nm, 400.9–1000 nm, and 1400–2000 nm) wavelength regions were sensitive bands in terms of the spectral characteristics of saline soils. The superior prediction model was an RBF-neural-network-based inversion model with $R^2 = 0.950$, RMSE = 1.014, and RPD = 4.479, and it achieved excellent prediction results. This study provides technical and theoretical support for the management and improvement of saline soils in large areas.

At present, basic research on digital, intelligent, and computerised decision automation technology for global soil salinity information management (e.g., crop, moisture, salinity, soil, and climate) is limited. Remote sensing monitoring can suffer from a poor transmission timeliness and limited information coverage and is subject to various constraints. Therefore, the design and implementation of an early warning system in regard to the soil salinity is an effective and reliable way for humans worldwide to gain timely access to information in order to mitigate and prevent the damage attributed to soil salinisation.

Author Contributions: Conceptualization, W.L. and J.L.; methodology, N.B.; software, X.M.; validation, Y.M.; formal analysis, J.H.; investigation, Z.Z.; resources, Y.F.; data curation, W.C.; writing—original draft preparation, J.L.; writing—review and editing, W.L.; supervision, Y.M.; funding acquisition, Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 52074064.

Data Availability Statement: Not Applicable.

Conflicts of Interest: The authors declared that there is no conflict of interest.

References

1. He, Y.; Li, X.; Craft, C.; Ma, Z.; Sun, Y. Relationships between vegetation zonation and environmental factors in newly formed tidal marshes of the Yangtze River estuary. *Wetl. Ecol. Manag.* **2011**, *19*, 341–349. [[CrossRef](#)]
2. Bucher, E.H.; Stein, A., F. Large salt dust storms follow a 30-year rainfall cycle in the Mar Chiquita Lake (Córdoba, Argentina). *PLoS ONE* **2016**, *11*, e0156672. [[CrossRef](#)] [[PubMed](#)]
3. Riedel, F.; Denk, M.; Müller, I.; Barth, N. Prediction of soil parameters using the spectral range between 350 and 15,000nm: A case study based on the Permanent Soil Monitoring Program in Saxony, Germany. *Geoderma* **2018**, *315*, 188–198. [[CrossRef](#)]
4. Singh, K. Microbial and enzyme activities of saline and sodic soils. *Land Degrad. Dev.* **2016**, *27*, 706–718. [[CrossRef](#)]
5. Chun, X.; Su, R.; Liu, J.; Liang, W.; Yong, M.; Ulambadarkh, K. Climatic implications on variations of Qehan Lake in the arid regions of Inner Mongolia during the recent five decades. *Environ. Monit. Assess.* **2017**, *189*, 14. [[CrossRef](#)] [[PubMed](#)]
6. Legesse, D.; Vallet-Coulomb, C.; Gasse, F. Analysis of the hydrological response of a tropical terminal lake, Lake Abiyata (Main Ethiopian Rift Valley) to changes in climate and human activities. *Hydrol. Process.* **2004**, *18*, 487–504. [[CrossRef](#)]
7. Yang, F.; An, F.; Ma, H.; Wang, Z.; Zhou, X.; Liu, Z. Variations on soil salinity and sodicity and its driving factors analysis under microtopography in different hydrological conditions. *Water* **2016**, *8*, 227. [[CrossRef](#)]
8. Mahajan, G.R. Microbial and enzyme activities and carbon stock in unique coastal acid saline soils of Goa. *Proc. Natl. Acad. Sci. India Sect. B Biol. Sci.* **2015**, *86*, 961–971. [[CrossRef](#)]
9. Gao, F.; Huang, Q.; Sun, X.; Yan, Z. Study on dynamic changes of the soil salinization in the upper stream of the Tarim River based on RS and GIS. *Procedia Environ. Sci.* **2011**, *11*, 1135–1141.
10. Askari, M.S.; O'Rourke, S.M.; Holden, N. Evaluation of soil quality for agricultural production using visible–near-infrared spectroscopy. *Geoderma* **2015**, *243*, 80–91. [[CrossRef](#)]
11. Xu, S.; Zhao, Y.; Wang, M. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. *Geoderma* **2018**, *310*, 29–43. [[CrossRef](#)]
12. Farifteh, J.; Meer, F.; Meijde, M. Spectral characteristics of salt-affected soils: A laboratory experiment. *Geoderma* **2008**, *145*, 196–206. [[CrossRef](#)]
13. Jario Moreira, L.C.; Teixeira, A.; Galvao, L.S. Potential of multispectral and hyperspectral data to detect saline-exposed soils in Brazil. *Geod. Mapp. Photogramm.* **2015**, *52*, 416–436.
14. Mahajan, G.R.; Das, B.; Gaikwad, B. Monitoring properties of the salt-affected soils by multivariate analysis of the visible and near-infrared hyperspectral data. *Catena* **2020**, *198*, 105041. [[CrossRef](#)]
15. Islam, K.; Singh, B.; Mcbratney, A. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Soil Res.* **2003**, *41*, 1101–1114. [[CrossRef](#)]
16. Bilgili, A.V.; Cullu, M.A.; van Es, H.; Aydemir, A.; Aydemir, S. The use of hyperspectral visible and near infrared reflectance spectroscopy for the characterization of salt-affected soils in the Harran Plain, Turkey. *Arid Land Res. Manag.* **2011**, *25*, 19–37. [[CrossRef](#)]
17. Bilgili, A.V.; Es, H.; Akbas, F. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *J. Arid Environ.* **2010**, *74*, 229–238. [[CrossRef](#)]
18. Dlc, A.; Esa, B. Review of soil salinity assessment for agriculture across multiple scales using proximal and/or remote sensors. *Adv. Agron.* **2019**, *158*, 1–130.
19. Divya, Y.; Sanjeevi, S.; Ilamparuthi, K. A study on the hyperspectral signatures of sandy soils with varying texture and water content. *Arab. J. Geosci.* **2014**, *7*, 3537–3545. [[CrossRef](#)]
20. Gorji, T.; Yldrm, A.; Sertel, E. Remote sensing approaches and mapping methods for monitoring soil salinity under different climate regimes. *Int. J. Environ. Geoinformatics* **2019**, *6*, 33–43. [[CrossRef](#)]
21. Shahriaria, M.; Delbaria, M.; Afrasiaba, P.; Pahlavan-Rad, M.R. Predicting regional spatial distribution of soil texture in floodplains using remote sensing data: A case of southeastern Iran. *Catena* **2019**, *182*, 104149. [[CrossRef](#)]
22. Rinnan, A. Pre-processing in vibrational spectroscopy—When, why and how. *Anal. Methods* **2014**, *6*, 7124–7129. [[CrossRef](#)]
23. Rinnan, S.; Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [[CrossRef](#)]
24. Liu, Y.; Li, Y.; Peng, Y.; Yan, S.; Zhao, X.; Han, D. Non-destructive and rapid detection of the internal chemical composition of granules samples by spectral transfer. *Chemom. Intell. Lab. Syst.* **2020**, *208*, 104174. [[CrossRef](#)]
25. Tiecher, T.; Moura-Bueno, J.M.; Caner, L.; Minella, J.P.G.; Evrard, O.; Ramon, R.; Naibo, G.; Barros, C.A.P.; Silva, Y.J.A.B.; Amorim, F.F.; et al. Improving the quantification of sediment source contributions using different mathematical models and spectral preprocessing techniques for individual or combined spectra of ultraviolet–visible, near- and middle-infrared spectroscopy. *Geoderma* **2021**, *384*, 114815. [[CrossRef](#)]

26. Zhao, N.; Wu, Z.; Cheng, Y.; Shi, X.; Qiao, Y. MDL and RMSEP assessment of spectral pretreatments by adding different noises in calibration/validation datasets. *Spectrochim. Acta Part A* **2016**, *163*, 20–27. [[CrossRef](#)]
27. Schoot, M.; Kapper, C.; van Kollenburg, G.H.; Postma, G.J.; van Kessel, G.; Buydens, L.M.; Jansen, J.J. Investigating the need for preprocessing of near-infrared spectroscopic data as a function of sample size. *Chemom. Intell. Lab. Syst.* **2020**, *204*, 104105. [[CrossRef](#)]
28. Ghosh, A.K.; Das, B.S.; Reddy, N. Application of VIS-NIR spectroscopy for estimation of soil organic carbon using different spectral preprocessing techniques and multivariate methods in the middle Indo-Gangetic plains of India. *Geoderma Reg.* **2020**, *23*, e00349.
29. Bian, X.; Chen, D.; Cai, W.; Grant, E.; Shao, X. Rapid determination of metabolites in bio-fluid samples by Raman spectroscopy and optimum combinations of chemometric methods. *Chin. J. Chem.* **2011**, *29*, 2525–2532. [[CrossRef](#)]
30. Peng, L.; Jing, W.; Qian, L. Rapid identification and quantification of Panax notoginseng with its adulterants by near infrared spectroscopy combined with chemometrics. *Spectrochim. Acta Part A* **2018**, *206*, 23–30.
31. Laxalde, J.; Ruckebusch, C.; Devos, O. Characterisation of heavy oils using near-infrared spectroscopy: Optimisation of pre-processing methods and variable selection. *Anal. Chim. Acta* **2011**, *705*, 227–234. [[CrossRef](#)] [[PubMed](#)]
32. Zhang, Z.; Ding, J.; Zhu, C.; Wang, J. Combination of efficient signal pre-processing and optimal band combination algorithm to predict soil organic matter through visible and near-infrared spectra. *Spectrochim. Acta Part A* **2020**, *240*, 118553. [[CrossRef](#)]
33. Yiming, B.; Kailong, Y.; Weiqiang, X. A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation. *Anal. Chim. Acta* **2016**, *909*, 30–40.
34. Xu, L.; Xie, D.; Fan, F. Effects of pretreatment methods and bands selection on soil nutrient hyperspectral evaluation. *Procedia Environ. Sci.* **2011**, *10*, 2420–2425. [[CrossRef](#)]
35. Mouazen, A.M.; Kuang, B.; Baerdemaeker, J.D. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* **2010**, *158*, 23–31. [[CrossRef](#)]
36. Chang, C.; Laird, D.A.; Mausbach, M.J. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* **2001**, *65*, 480–490. [[CrossRef](#)]
37. Chen, X.; Dong, Z.; Liu, J. Hyperspectral characteristics and quantitative analysis of leaf chlorophyll by reflectance spectroscopy based on a genetic algorithm in combination with partial least squares regression. *Spectrochim. Acta Part A* **2020**, *243*, 118786. [[CrossRef](#)]
38. Maldonado, S.; López, J.; Vairetti, C. Time-weighted fuzzy support vector machines for classification in changing environments. *Inf. Sci.* **2021**, *559*, 97–110. [[CrossRef](#)]
39. Rossel, R.; Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **2010**, *158*, 46–54. [[CrossRef](#)]
40. Zhang, D.; Lin, J.; Peng, Q. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *J. Hydrol.* **2018**, *565*, 720–736. [[CrossRef](#)]
41. Yachun, M.; Jing, L.; Wang, C. Research on the quantitative inversion model of heavy metals in soda saline land based on visible-near-infrared spectroscopy. *Infrared Phys. Technol.* **2021**, *112*, 103602.
42. Xiao, D.; Wan, L.; Sun, X. Remote sensing retrieval of saline and alkaline land based on reflectance spectroscopy and RV-MELM in Zhenlai County. *Opt. Laser Technol.* **2021**, *139*, 106909. [[CrossRef](#)]
43. Dong, X.; Vu, Q.H.; Ba, T.L. Salt content in saline-alkali soil detection using visible-near infrared spectroscopy and a 2D deep learning. *Microchem. J.* **2021**, *165*, 106182.
44. Mohamed, E.S.; Saleh, A.M.; Belal, A.B. Application of near-infrared reflectance for quantitative assessment of soil properties. *Egypt. J. Remote. Sens. Space Sci.* **2018**, *21*, 1–14. [[CrossRef](#)]
45. Guo, X.J.; Yuan, D.H.; Jiang, J.Y. Detection of dissolved organic matter in saline-alkali soils using synchronous fluorescence spectroscopy and principal component analysis. *Spectrochim. Acta Part A* **2013**, *104*, 280–286. [[CrossRef](#)]
46. Rossel, R.; Walvoort, D.; Mcbratney, A.B. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2005**, *31*, 59–75.
47. Li, F.; Gong, H.; Chen, B. Analysis of the contribution rate of the influencing factors to land subsidence in the Eastern Beijing plain, China based on extremely randomized trees (ERT) method. *Remote Sens.* **2020**, *12*, 2963. [[CrossRef](#)]
48. Wang, N.; Xue, J.; Peng, J. Integrating remote sensing and landscape characteristics to estimate soil salinity using machine learning methods: A case study from Southern Xinjiang, China. *Remote Sens.* **2020**, *12*, 4118. [[CrossRef](#)]
49. Sačkov, I.; Barka, I.; Bucha, T. Mapping aboveground woody biomass on abandoned agricultural land based on airborne laser scanning data. *Remote Sens.* **2020**, *12*, 4189. [[CrossRef](#)]
50. Liang, L.; Guo, W.; Zhang, Y. Radial basis function neural network for prediction of medium-frequency sound absorption coefficient of composite structure open-cell aluminum foam. *Appl. Acoust.* **2021**, *184*, 108414.