



## Article

# Combination of Hyperspectral and Machine Learning to Invert Soil Electrical Conductivity

Pingping Jia <sup>1,2,†</sup>, Junhua Zhang <sup>3,4,†</sup> , Wei He <sup>1</sup>, Yi Hu <sup>1</sup>, Rong Zeng <sup>1</sup>, Kazem Zamanian <sup>1,5</sup> , Keli Jia <sup>2</sup> and Xiaoning Zhao <sup>1,\*</sup>

<sup>1</sup> School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China; 20211156001@nuist.edu.cn (P.J.); 20201211002@nuist.edu.cn (W.H.); 003053@nuist.edu.cn (Y.H.); rzeng@nuist.edu.cn (R.Z.); kzamani@gwdg.de (K.Z.)

<sup>2</sup> School of Geography and Planning, Ningxia University, Yinchuan 750021, China; jiakl@nxu.edu.cn

<sup>3</sup> School of Ecology and Environment, Ningxia University, Yinchuan 750021, China; jhzhzhang@nxu.edu.cn

<sup>4</sup> Breeding Base for State Key Laboratory of Land Degradation and Ecological Restoration in Northwestern China, Ningxia University, Yinchuan 750021, China

<sup>5</sup> Institute of Soil Science, Leibniz University of Hannover, 30419 Hannover, Germany

\* Correspondence: jasminezxsx@msn.com

† These authors contributed equally to this work.

**Abstract:** An accurate estimation of soil electrical conductivity (EC) using hyperspectral techniques is of great significance for understanding the spatial distribution of solutes and soil salinization. Although spectral transformation has been widely used in data pre-processing, the performance of different pre-processing techniques (or combination methods) on different models of the same data set is still ambiguous. Moreover, extremely randomized trees (ERT) and light gradient boosting machine (LightGBM) models are new learning algorithms with good generalization performance (soil moisture and above-ground biomass), but are less studied in estimating soil salinity in the visible and near-infrared spectra. In this study, 130 soil EC data, soil measured hyperspectral data, topographic factors, conventional salinity indices such as Salinity Index 1, and two-band (2D) salinity indices such as ratio indices, were introduced. The five spectral pre-processing methods of standard normal variate (SNV), standard normal variate and detrend (SNV-DT), inverse (1/OR) (OR is original spectrum), inverse-log (Log(1/OR) and fractional order derivative (FOD) (range 0–2, with intervals of 0.25) were performed. A gradient boosting machine (GBM) was used to select sensitive spectral parameters. Models (extreme gradient boosting (XGBoost), LightGBM, random forest (RF), ERT, classification and regression tree (CART), and ridge regression (RR)) were used for inversion soil EC and model validation. The results reveal that the two-dimensional correlation coefficient highlighted EC more effectively than the one-dimensional. Under SNV and the second order derivative, the two-dimensional correlation coefficient increased by 0.286 and 0.258 compared to the one-dimension, respectively. The 13 characteristic factors of slope, NDI, SI-T, RI, profile curvature, DOA, plane curvature, SI (conventional), elevation, Int2, aspect, S1 and TWI provided 90% of the cumulative importance for EC using GBM. Among the six machine models, the ERT model performed the best for simulation ( $R^2 = 0.98$ ) and validation ( $R^2 = 0.96$ ). The ERT model showed the best performance among the EC estimation models from the reference data. The kriging map based on the ERT simulation showed a close relationship with the measured data. Our study selected the effective pre-processing methods (SNV and the 2 order derivative) using one- and two-dimensional correlation, 13 important factors and the ERT model for EC hyperspectral inversion. This provides a theoretical support for the quantitative monitoring of soil salinization on a larger scale using remote sensing techniques.

**Keywords:** soil EC inversion; hyperspectral reflectance; extremely randomized trees; light gradient boosting machine; northwestern China



**Citation:** Jia, P.; Zhang, J.; He, W.; Hu, Y.; Zeng, R.; Zamanian, K.; Jia, K.; Zhao, X. Combination of Hyperspectral and Machine Learning to Invert Soil Electrical Conductivity. *Remote Sens.* **2022**, *14*, 2602. <https://doi.org/10.3390/rs14112602>

Academic Editor: Fabio Castaldi

Received: 6 April 2022

Accepted: 25 May 2022

Published: 28 May 2022

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



**Copyright:** © 2022 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

Salinization is a major soil degradation process that threatens the ecosystem, agricultural production and sustainability of the ecological environment in arid and semiarid regions worldwide [1–4]. There are nearly  $10 \times 10^8$  hm<sup>2</sup> of area harmed by salinization worldwide, and more than 33 million hm<sup>2</sup> in China [5,6]. This has a negative impact on crop yields and agricultural productivity, seriously threatening ecosystem health and economic sustainability. Therefore, timely and accurate monitoring of soil salt content is important to combat soil salinity and improve agricultural productivity in the face of climate change and human activities.

In agricultural production, soil salinization monitoring plays a very important guiding role in crop management [7,8]. Soil electrical conductivity (EC) is widely used in the study of saline soil [9], which can directly reflect soil salinity [10]. EC is an important indicator in measuring the degree of soil salinization and evaluating crop yield at regional scales [11,12].

Hyperspectral remote sensing techniques can fully exploit spectral information in order to realize real-time and non-contact monitoring of soil salinization, which has become a major detection method for soil salinization monitoring at present [13,14]. Due to the complex causes of soil salinization, its spectral characteristics are significantly affected by soil texture, organic matter content, soil moisture, soil salt content and other factors. Hyperspectral resolution is high. The fine spectral resolution reflects the fine characteristics of the ground object spectrum, and inhibits the influence of other interference factors to a large extent. Several scholars have explored models of the function between measured soil EC and spectral reflectance, and have successfully predicted salt content in soils using reflectance spectroscopy [15,16]. The pre-processing of hyperspectral data is the key to improving the accuracy of the inversion. At present, in terms of data pre-processing, the common methods include spectral transformation and the Savitzky-Golay filtering method, among the others [17,18]. However, no single pre-processing technique (or combination method proposed) is suitable for different data sets. Moreover, these studies mainly considered the sensitivity of the spectrum without deeply studying the interaction between the spectral bands. The optimal band combination algorithm overcame this problem by calculating the spectral index (two-band (2D) salinity indices) and reducing the interference of irrelevant wavelengths [19]. It enhances the relationship between soil attributes and spectral features, and minimizes the effects of irrelevant wavelengths [20,21]. Thus, the optimal band combination algorithm has been widely used locally [10,22]. However, this method is not currently used in the study of hyperspectral inversion of soil salinization.

The causes of soil salinization and the composition of soil salinity are complex. They differ according to different regions in the selection of sensitive bands, salinity index, vegetation index, topographic factors and other environmental variables in the remote sensing and monitoring of soil salinity [23–26]. Although most of these variables can be obtained by band operation, there are different degrees of information redundancy. Therefore, band selection strategies need to be developed [27,28], such as Pearson's correlation coefficient (PCC), gray relational analysis (GRA), and variable importance in projection (VIP). This feature filtering method reduces information redundancy, but it is difficult to obtain the optimal inversion parameter subset. Compared with the above variable optimization method, gradient boosting machines (GBM) can effectively construct and run the enhanced tree, perform parallel computation, and effectively process sparse data [29]. However, it is rarely used in the optimization of characteristic variables in soil salinization modeling.

Soil is a spatio-temporal continuum with high variability, and the non-linear effects of soil-forming factors on soil development lead to obvious varied properties in larger areas [20,30,31]. At present, remote sensing technology is the most effective means of soil surface monitoring, but the lack of data mining has become a bottleneck to high efficiency and high precision monitoring. The research shows that the machine learning (ML) inversion model has strong nonlinear fitting ability and excellent data mining ability, which will increase the use of spectral reflection information [32,33]. Back propagation neural networks (BPNN), support vector machines (SVM), multiple adaptive regression

splines (MARS), etc., have all been used to invert soil salt content [34–36]. However, the effectiveness of different modelling methods varies. Ensemble learning methods have the advantages of high flexibility and generalization. The new developed extremely randomized trees (ERT) [37] and light gradient boosting machine (LightGBM) [38] are simple and fast learning algorithms. They have shown good results for adsorption energy of metal ions [39], soil moisture [40] and above-ground biomass [41], but are currently less well studied for applications in estimating soil salinity in visible-near-infrared (Vis-NIR) spectra.

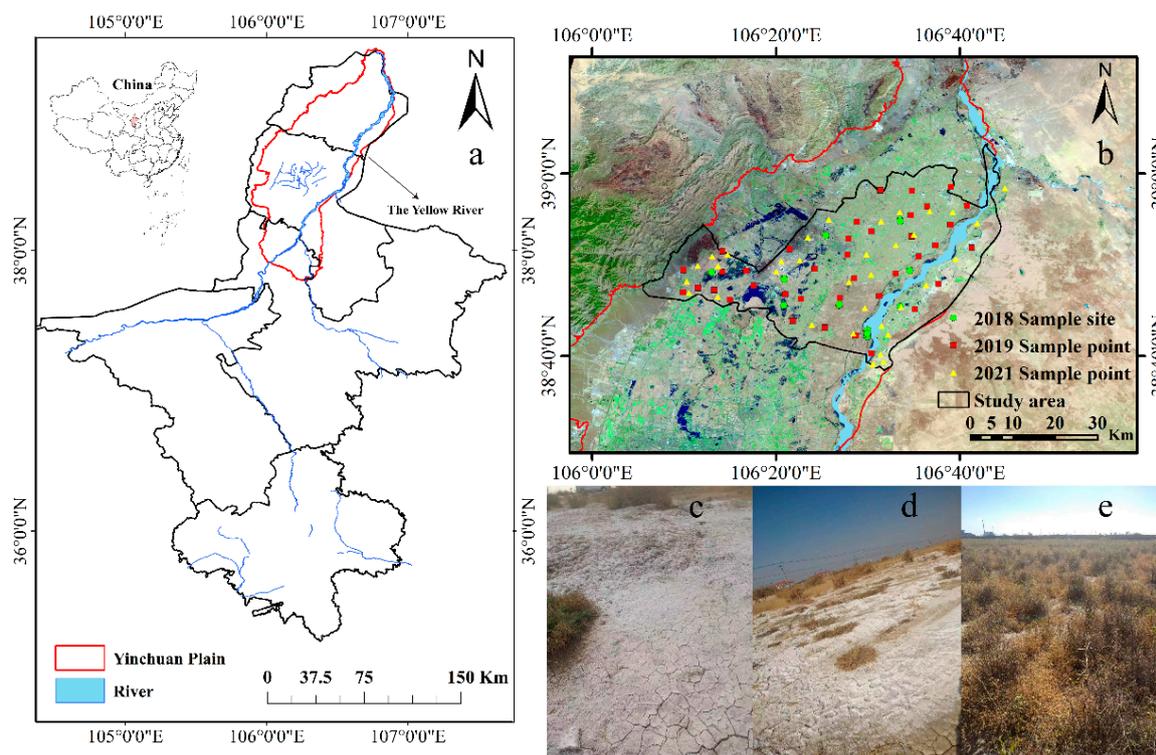
Yinchuan Plain irrigation area is located in the upper reaches of the Yellow River, where the salinized soil area is about 2406 km<sup>2</sup> (where alkaline represents takyrs solonchaks). Among salinized soils, the area with high salt content is mainly located in Pingluo County in the north, a typical salinized land in Yinchuan Plain, where light, moderate and heavily salinized soils account for 25.2%, 39.8% and 2.7% of the county area, respectively [42]. In addition, the prediction and inversion results of EC by previous pre-treatment technologies and multivariate methods are different due to the region, soil type and spectral range. Few studies have simultaneously explored multiple forms of pre-processing and modeling methods in the same database. This study aims to provide a new train of thought for assessing soil salinization using hyperspectral analysis.

For this purpose, the saline-alkali soil samples were collected and hyperspectral data were acquired in the study area from 2018, 2019 and 2021. Spectral pre-processing methods and machine learning methods were used for data simulation, best model selection and validation. The main objectives of our research were as follows: (1) explore the response of saline properties to sensitive spectral wavelengths; (2) compare the optimal spectral parameters under one- and two-dimensional correlation coefficients, and point out the influencing factors for the EC model; (3) acquire the best-performing model for predicting soil salinity, and map soil salinity in the study region; (4) apply and provide technical support for saline soil evaluation.

## 2. Materials and Methods

### 2.1. Study Area

The study area is located in Pingluo county (38°26′60″–39°14′09″N, 105°57′40″–106°52′52″E), northern Yinchuan Plain, Ningxia Province, China (Figure 1), covering an area about 2060 km<sup>2</sup>. The area is located in the irrigated middle and upper reaches of the Yellow River and lies between the diluvial fan and plain at the eastern foot of Helan Mountain. The study area experiences a warm temperate monsoon climate, with an annual mean temperature of 9 °C, low precipitation (annual mean: 150–203 mm), and strong evaporation (annual mean > 1825 mm). The research area is one of the most serious areas of soil salinization in Ningxia Province, due to the low-lying terrain, poor drainage conditions, shallow groundwater depth, strong evaporation, water salinity pooling, the terrain, and unreasonable irrigation. The major types of land use and land cover include water bodies, deserts, wastelands and basic farmland. The major soil types are lime calcite, saline, alkaline, and irrigated silt (Calcite Solonchack, Petrosalic Solonchack, Sodic Solonchack, Haplic Cambisol Salic according to the World Reference Base for Soil Resources (WRB)). The parent materials are mainly carbonate. The natural vegetation is dominated by salt tolerant vegetation (such as *Nitraria tangutorum* and *Phragmites australis*) [43].



**Figure 1.** Locations of the Yinchuan Plain, China (a), distribution of sampling (b) sampling sites in 2018 (green circle), 2019 (red square) and 2021 (yellow triangle), and typical landscape photographs (c–e).

## 2.2. Data Collection

### 2.2.1. Soil Sampling and Laboratory Analysis

Based on soil surface features, pH conditions, and land use patterns, nine sampling sites (57 samples) were selected throughout Pingluo County from northern Yinchuan Plain in Ningxia Province in October 2018 (Figure 1). The sampling sites included basic farmland (non-alkaline or slightly saline-alkaline soil), medium- and low-yielding farmland (moderate-strongly saline-alkaline soil), and abandoned land (strongly saline-alkaline or alkaline soil) with varying levels of alkalinity. At each site, a soil auger was used to collect intact soil cores (0–20 cm in length) at intervals of 30, 60, 100, 200, and 300 m, after conducting the hyperspectral measurements. The sampling method was the same in March 2019 and May 2021. A total of 130 soil samples were collected, including 57 in 2018, 41 in 2019 and 32 in 2021. Sampling was carried out in a 5 km × 5 km grid of sample points (Figure 1). The collected soil samples (0–20 cm, non-mixed soil samples) were stored in sealed bags until use. The latitude and longitude of the sampling sites were recorded by a handheld global positioning system (GPS). The information including the surface salinity accumulation, land use patterns, vegetation types and cover were documented. After the samples were brought back to the laboratory, the soil moisture content (water content by weight) was determined using the drying method, and the soil EC was determined using the electrical conductivity method [44]. According to the definition of Brady and Weil [45], the soil in the study area was partitioned into five levels of salinity: non saline (0–0.4 dS m<sup>-1</sup>), very slightly saline (0.4–0.8 dS m<sup>-1</sup>) slightly saline (0.8–1.6 dS m<sup>-1</sup>), moderately alkaline (1.6–2.4 dS m<sup>-1</sup>) and strongly alkaline (>2.4 dS m<sup>-1</sup>).

### 2.2.2. Hyperspectral Measurement and Data Processing

Field spectra were acquired at the time of soil sampling in each year. Soil spectroscopy was conducted at each sampling site using an SR-3500 spectrometer (Spectral Evolution, Esses, MA, USA), at wavelengths of 350 to 2500 nm. The spectral resolution was set

at 3.5 nm from 350 to 1000 nm, 10 nm from 1000 to 1500 nm, and 7 nm from 1500 to 2100 nm. Measurements were carried out between 10:00–14:00 on a sunny day. During the hyperspectral measurements, the spectrometer was vertically downwards with the probe at about 80 cm (waist height) above the surface. Before each measurement, the reference panel on the spectrometer was initialized, then five measurements per sampling site were obtained and averaged to minimize instrument noise.

### 2.3. Extraction of Salinization Related Factors

#### 2.3.1. Spectral Reflectance Transformation and Selection of Spectral Indices

In order to eliminate the instrument noise and environmental background interference, the edge bands (350–399 nm and 2401–2500 nm) with excessive noise were removed. The spectral curves consisting of 201 band numbers were obtained by resampling the 400–2400 nm spectral data at 10 nm intervals original spectrum (OR), taking into account the smoothing and features of the spectral curves. Five types of spectrum pre-processing methods, including standard normal variate (SNV), standard normal variate and detrend (SNV-DT), inverse (1/OR), inverse-log (Log(1/OR) and fractional order derivative (FOD) (range 0–2, with intervals of 0.25, 0 order means OR), were implemented on the OR.

Spectral index is a linear or non-linear combination of reflectance in different bands. Spectral index was used to establish the correlation between spectral data and specific targets, and to provide a scientific basis for soil salinity research [46]. This research mainly applies the spectral characteristic indices including Deviation of arch (DOA), Salinity index (Table 1) and Two-band (2D) index (Table 2):

- (1) Deviation of arch (DOA) [47]

$$\text{DOA} = R_{600} - (R_{550} - R_{650})/2 \quad (1)$$

- (2) Salinity index (conventional)

**Table 1.** Reference overview of studies of spectral salinity indices and formula.

Acronym	Spectral Index	Formula	Reference
SI-T	Salinity Index	$R/\text{NIR} \times 100$	[48]
SI	Salinity Index	$(B \times R)^{1/2}$	[48]
SI1	Salinity Index 1	$(G \times R)^{1/2}$	[49]
SI2	Salinity Index 2	$(G^2 + R^2 + \text{NIR}^2)^{1/2}$	[49]
SI3	Salinity Index 3	$(G^2 + R^2)^{1/2}$	[49]
S1	Salinity Index I	$B/R$	[50]
S2	Salinity Index II	$(B - R)/(B + R)$	[50]
S3	Salinity Index III	$G \times R/B$	[50]
Int1	Intensity Index 1	$(G + R)/2$	[51]
Int2	Intensity Index 2	$(G + R + \text{NIR})/2$	[51]
NDSI	Normalized Difference Salinity Index	$(R - \text{NIR})/(R + \text{NIR})$	[52]

B, G, R, and NIR correspond to reflectance in blue (455–492 nm), green (492–577 nm), red (622–770 nm), and near-infrared (770–1050 nm) after conversion by the best way, respectively.

- (3) Two-band (2D) index

**Table 2.** Reference overview of studies of spectral indices and formula.

Acronym	Spectral Indices	Formula	Reference
DI	Difference Index	$R_i - R_j$	[53]
RI	Ratio Index	$R_i/R_j$	[53]
NDI	Normalized Index	$(R_i - R_j)/(R_i + R_j)$	[54]
PI	Product Index	$R_i \times R_j$	[54]

Table 2. Cont.

Acronym	Spectral Indices	Formula	Reference
SI	Sum Index	$R_i + R_j$	[54]
RDVI	Renormalized Difference Vegetation Index	$(R_i - R_j)/(R_i + R_j)^{1/2}$	[55]
NPDI	Nitrogen Planar Domain Index	$(R_i + R_j) \times R_j$	[56]

Note:  $R_i$  and  $R_j$  in the formula belong to any two wavelengths in 400–2400 nm, and  $R_i \neq R_j$ . All thirteen spectral transformations were involved in the calculation of the seven spectral indices mentioned above. For each spectral index, the wavelength combination with the largest correlation with soil EC was extracted and deemed to be the optimal band combination.

### 2.3.2. Topographical Factors

Topography is the main factor of soil formation and development in arid and semi-arid regions, affecting surface material energy and redistribution. Digital Elevation Model (DEM) data were downloaded from the website of Geospatial Data Cloud (<http://www.gscloud.cn/>, 24 May 2022) at a spatial resolution of 30 m. The DEM of each sampled point was extracted using the Extract Multi Values to Points tool in Spatial Analyst Tools in ArcGIS 10.4, along with slope, aspect, plane curvature, profile curvature and topographic wetness index (TWI) as input variables to the model.

## 2.4. Machine Learning Algorithms

### 2.4.1. Feature Selection Based on Gradient Boosting Machine

Twenty-four variables (eleven conventional soil salinity indices, seven 2D indices, and six terrain parameters) were selected as feature descriptors. In consideration of the possible over-fitting risk, GBM was introduced to screen out the most important features from the 24 feature descriptors for participation in the subsequent construction of the soil EC model.

### 2.4.2. Modeling Strategies and Accuracy Assessment

In order to achieve EC predictions and to ensure the generalization and robustness of the models, we divided the dataset into two disjoint sets, and the training and validation sets were assigned by the 5-fold cross validation method [57]. XGBoost, LightGBM, RF, ERT, CART, and RR were used to build an EC inversion model based on the factors selected by GBM. In the toolkit Scikit-Learn (<http://scikit-learn.org>, 24 May 2022), ML models were first trained with training sets and then the model was used to predict the EC of the validation set. The main parameters were grid searched [58], and the default values of other parameters were Scikit-Learn. The optimal hyperparameters of the model are shown in Table 3. The determination coefficient ( $R^2$ ), mean squared error (MSE), correlation coefficient, standard deviation and root mean square error (RMSE) between the predicted and true values were calculated to evaluate the predictability. A scoring mechanism was developed to pre-evaluate the six ML models. The model with the largest  $R^2$  and lowest MSE values was considered the most robust.

Table 3. Optimal hyperparameters of six machine learning methods.

Category	Method	Optimal Hyperparameters
Boosting	Extreme gradient boosting (XGBoost)	n_estimators = 5, max_depth = 4, min_child_weight = 2, learning_rate = 0.32
	Light gradient boosting machine (LightGBM)	n_estimators = 16, objective = regression, num_leaves = 31, learning_rate = 0.32

Table 3. Cont.

Category	Method	Optimal Hyperparameters
Bagging	Random forest (RF)	n_estimators = 27, max_depth = 10, max_features = 4, random_state = 1
	Extremely randomized trees (ERT)	n_estimators = 21, max_depth = 14, random_state = 1
	Classification and Regression Trees (CART)	max_depth = 6, max_features = 4, max_leaf_nodes = 12, random_state = 1
Linear	Ridge regression (RR)	alpha = 0.1

### 2.5. Kriging

Kriging interpolation is a spatial local interpolation method [42], which makes use of the original data and the structure of the semi-variance function in order to get the unbiased best estimate of the unsampled regional variables. It mainly analyzes the structural and random characteristics of the regional variables, and then obtains their spatial distribution characteristics. The soil EC model with the highest inversion accuracy was selected. The kriging interpolation method was used to invert the spatial distribution of the soil EC. The inverted soil EC values were then compared with the interpolation results of 320 measured data by our research group, from 0–20 cm depths of soils from the whole Yinchuan Plain in 2019 and 2021, in order to verify the adaptability of the model on a large scale.

## 3. Results

### 3.1. The Spectral Characteristic of the Soil Samples

All hyperspectral characteristic curves of soil were analyzed (Figure 2). The soil spectral reflectance increased with an increase in the wavelength and with a certain volatility. The pattern of spectra curves was similar across different saline levels, with absorption valleys at 1400 nm and 1900 nm. The spectral curves of salinized soil between 400 and 1400 nm show regular changes with the increase of salinization, that is, the soil reflectance increases with the increase of salinization. Although this rule is not obvious after 1400 nm, this regularity can distinguish different degrees of salinization; based on this we can accurately distinguish different salinization soils through certain treatment.

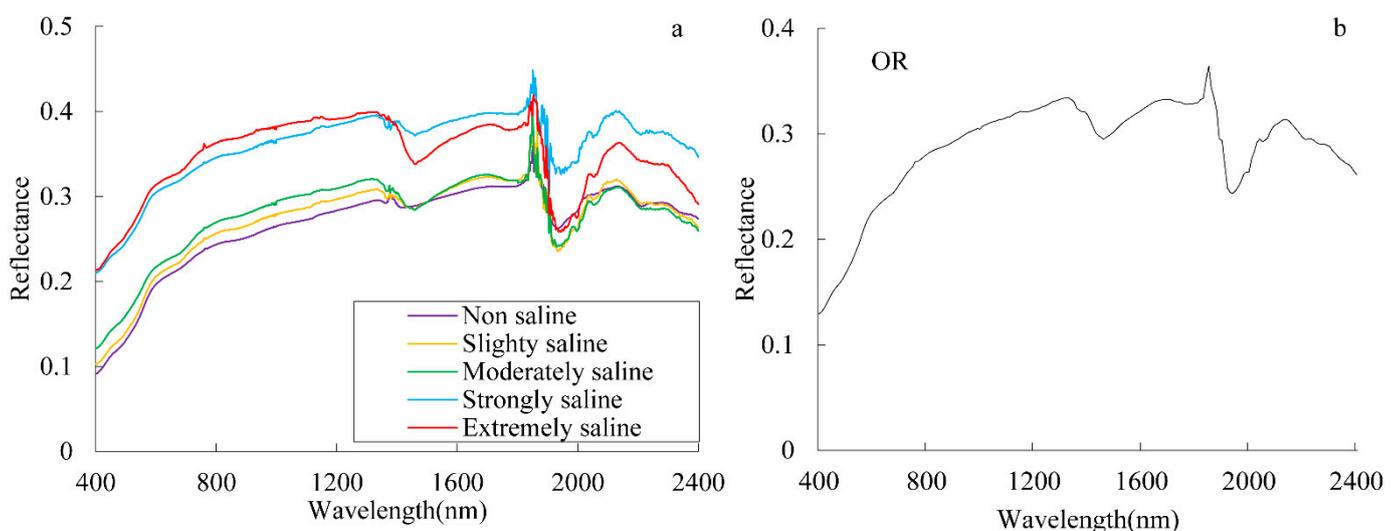
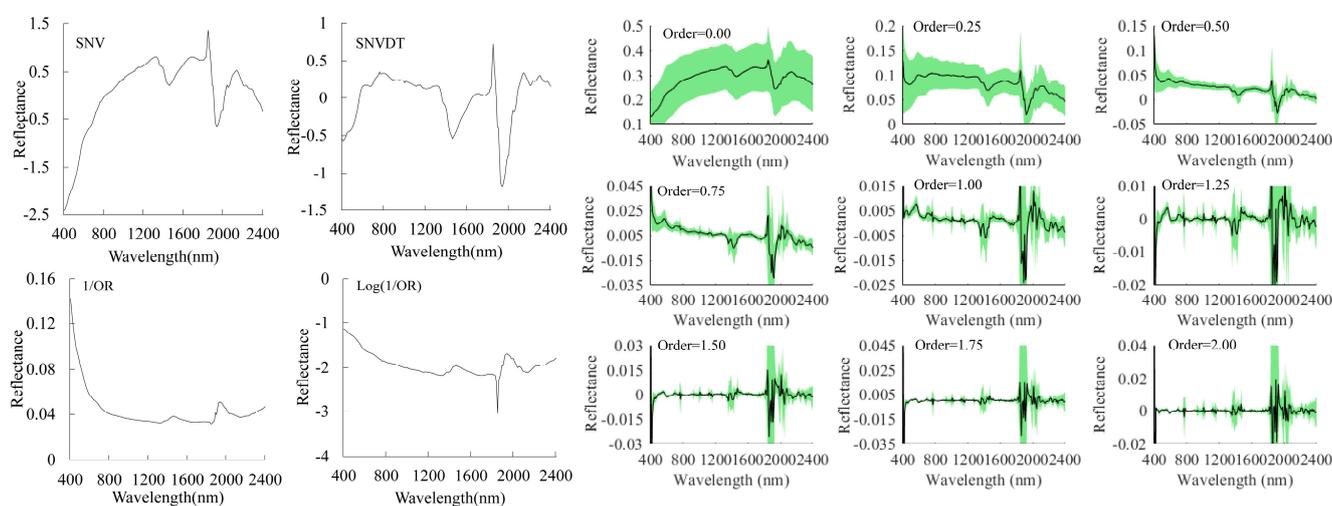


Figure 2. Hyperspectral reflectance of the soil measured on the ground under different saline levels (a) and mean original spectral reflectance ( $n = 130$ ) (b).

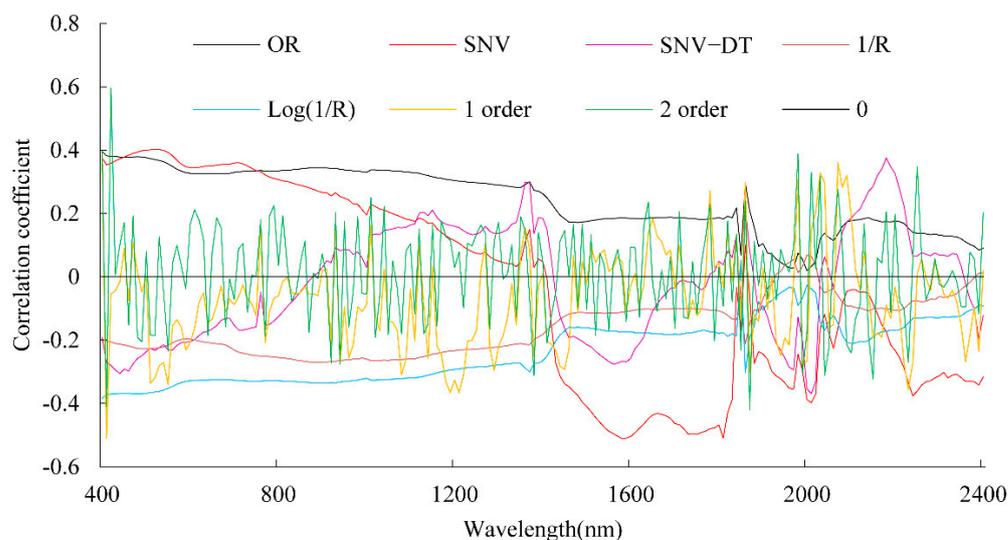
The pattern of SNV transformation was similar to that of the OR reflectance curve, but the characteristics of the absorption valley and reflection peak of the curve were obviously enhanced (Figure 3). After SNV-DT transformation, the spectral absorption and reflection characteristics of the curve were enhanced, and several new characteristics lacking in the OR were present, including two reflection peaks in the visible region (near 700 and 800 nm) and the absorption valley (near 2200 nm). After 1/OR and Log (1/OR) transformation, the reflectance showed a downward trend, compared with the OR, the absorption valley features were weakened, and at 1400 and 1900 nm, contrary to the OR features. As the FOD increases from 0 to 2, the intensity of spectral signals weakened, but the spectral detail increased. The absorption valleys at 1400 nm and 1900 nm became more and more obvious, and two small absorption valleys and a reflection peak were observed at 1400 nm. Meanwhile, the absorption valley peaked at 1900 nm gradually with the increase of overall absorption characteristics of the visible region (FOD 0–2). When the order increased from 1 to 2, most of the reflectance values approached zero, one positive peaked at 580 nm, and two negative peaks at 480 nm and 660 nm. Moreover, the 1 order derivative better identified both positive and negative peaks.



**Figure 3.** Pre-processing of mean spectral curves including standard normal variate (SNV), standard normal variate and detrend (SNV-DT), inverse (1/OR) (OR is original spectrum), inverse-log (Log(1/OR) and fractional order derivative (FOD) (range 0–2, with intervals of 0.25) (green areas represent the standard deviations of the spectra) collected from soil measured on the ground.

### 3.2. Correlation Analysis of Spectral Reflectance and EC

The correlation coefficients between EC and the reflectance processed by SNV, SNV-DT, 1/OR, Log (1/OR), 1 order and 2 order derivatives in the range of 400~2400 nm were computed and plotted, respectively (Figure 4). In the OR, the correlation coefficient showed a steady downward trend with the increase of wavelength. The SNV correlation was slightly higher than the OR correlation in the range of 400~700 nm, and then the correlation coefficient decreased and became negative from 1400 nm. The SNV-DT correlation trends increased in the 400~1100 nm range, crossing with the SNV correlation near 1100 nm, and then the correlation gradually became negative. In 1/OR and Log (1/OR), EC was negatively correlated with the spectral reflectance over the whole wavelength range with smooth change and stronger correlation in the visible and near-infrared parts. The correlation coefficients in the 1 order and 2 order derivatives were alternately positive and negative in wavelength, which increased the correlation of several specific bands. The maximum absolute correlation coefficient (MACC) between EC was 0.596 at 420 nm in 2 order derivation reflectance, which was only 0.396 at 400 nm in OR.



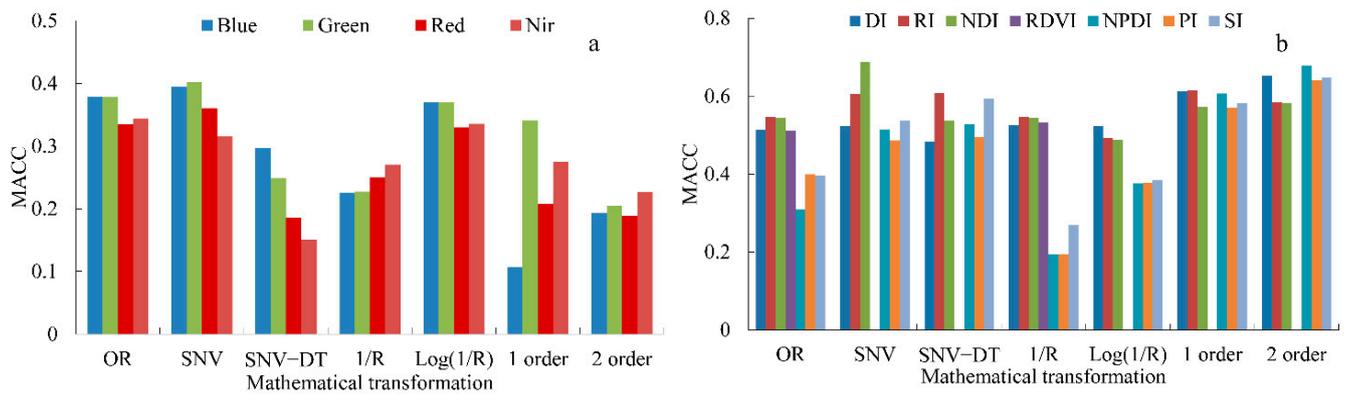
**Figure 4.** One-dimensional correlation coefficients between EC and partial conversion transformation (OR, SNV, SNV-DT, 1/OR, Log (1/OR), 1 and 2 order derivative) in the range of 400~2400 nm.

In each spectral transformation, the band with the largest correlation coefficient was extracted (Table 4). The correlation coefficient was best in all four bands under SNV transformation, because the salinity index was calculated according to required optimum wavelengths varying from 450 to 1050 nm, i.e., blue: 455~492 nm; green: 492~577 nm; red: 622~770 nm; and near-infrared: 770~1050 nm (Figure 5a). Therefore, the corresponding best reflectance under SNV transformation was selected for salinity index calculation.

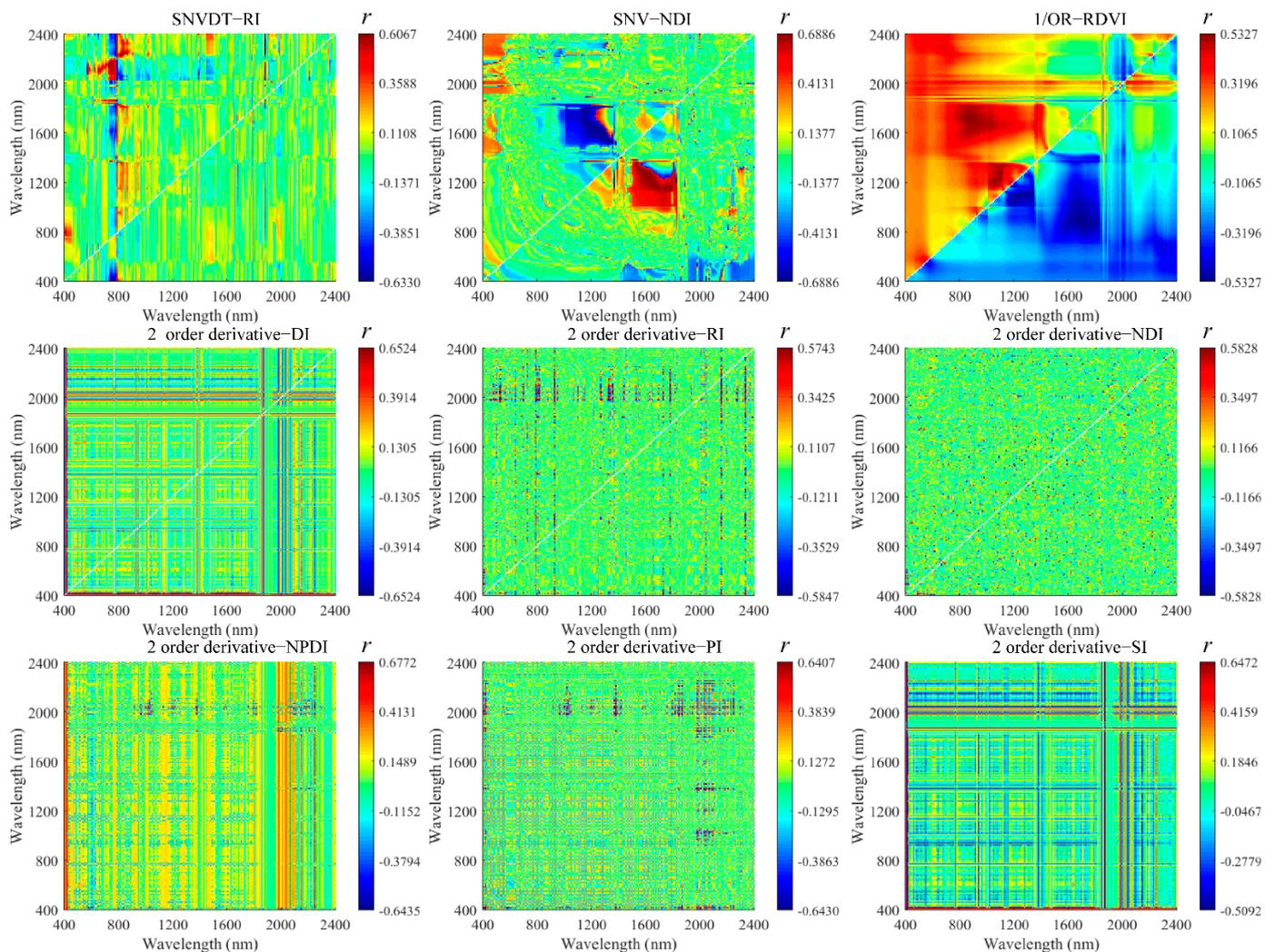
**Table 4.** Most sensitive spectral parameters by one- and two-dimensional correlation coefficients.

Dimensionality	Spectral Parameters	MACC
One-dimensional	OR <sub>400</sub> , SNV <sub>1580</sub> , SNVDT <sub>2180</sub> , 1/OR <sub>880</sub> , Log(1/OR) <sub>400</sub> , FOD(0.25 <sub>400</sub> , 0.51 <sub>860</sub> , 0.75 <sub>400</sub> , 1.410, 1.25 <sub>410</sub> , 1.5 <sub>410</sub> , 1.75 <sub>490</sub> , 2.420)	0.396, 0.488, 0.376, 0.270, 0.384, (0.395, 0.412, 0.396, 0.509, 0.443, 0.421, 0.437, 0.596)
One-dimensional (using SNV)	Blue <sub>480</sub> , Green <sub>520</sub> , Red <sub>710</sub> , NIR <sub>780</sub>	0.394, 0.403, 0.360, 0.315
Two-dimensional	DI (2 order derivative <sub>930</sub> , 2 order derivative <sub>430</sub> ), RI (SNVDT <sub>2410</sub> , SNVDT <sub>760</sub> ), NDI (SNV <sub>1220</sub> , SNV <sub>1290</sub> ), SI (2 order derivative <sub>1790</sub> , 2 order derivative <sub>430</sub> ) RDVI (1/OR <sub>1120</sub> , 1/OR <sub>1290</sub> ), NPDI (2 order derivative <sub>1830</sub> , 2 order derivative <sub>430</sub> ), PI (2 order derivative <sub>430</sub> , 2 order derivative <sub>420</sub> )	0.652, 0.633, 0.689, 0.647, 0.533, 0.677, 0.643
Two-dimensional (using 2 order derivative)	DI (930, 430), RI (2050, 510), NDI (520, 1200), NPDI (1830,430), PI (430, 420), SI (1790, 430)	0.652, 0.585, 0.583, 0.677, 0.643, 0.647

The bands with high correlation of SNVDT-RI, SNV-NDI, 1/OR-RDVI, 2 order derivative-DI, 2 order derivative-NDI, 2 order derivative-NPDI and 2 order derivative-PI to EC, respectively (Figure 6). The best bands of SNVDT-RI, SNV-NDI, and 1/OR-RDVI were more concentrated, while the best bands under 2 order derivative were more dispersed, mostly in the form of grids and dots. The MACC between SNV-NDI and EC was 0.69, and the high bands were mainly concentrated around 1200 nm and 1600 nm with the explicit expression  $[(R_i - R_j)/(R_i + R_j)]$ . Overall, all 2D indices under 2 order derivative were subsequently selected to participate in the EC modeling, because the best spectral indices under 2 order derivative transformation were more numerous (Figure 5b).



**Figure 5.** Maximum absolute correlation coefficient (MACC) of visible-near-infrared (Vis-NIR) (a) and two-band index (b) under different reflectance conversion modes.



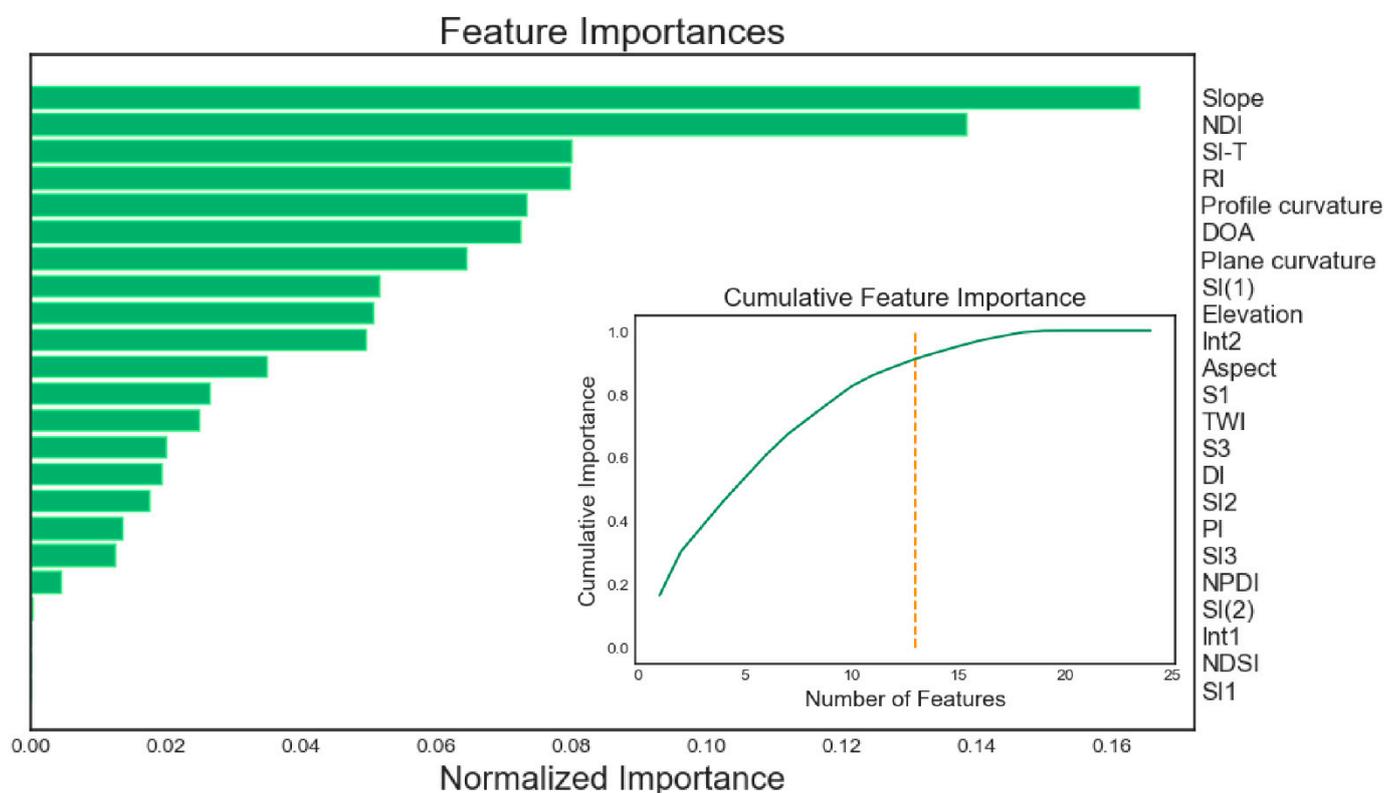
**Figure 6.** Two-dimensional correlation coefficients between EC and optimal spectral index under different transformation reflectance and two derivative orders (The x and y axis represent the wavelength 400–2400 nm, the right-side color bar indicates the color of the PCC values. The colors dark red and dark blue represent a relatively high PCC (red for positive and blue for negative) between the measured EC and the band combinations).

In general, two-dimensional correlation coefficients show higher correlation values compared to one-dimensional. In the case of NDI at SNV, the best correlation coefficient between spectral reflectance and EC in one dimension was 0.488, which increased by 0.201 at a two-dimensional spectral index (Table 4).

### 3.3. Model Inversions and Comparisons

#### 3.3.1. Feature Selection and Importance Analysis

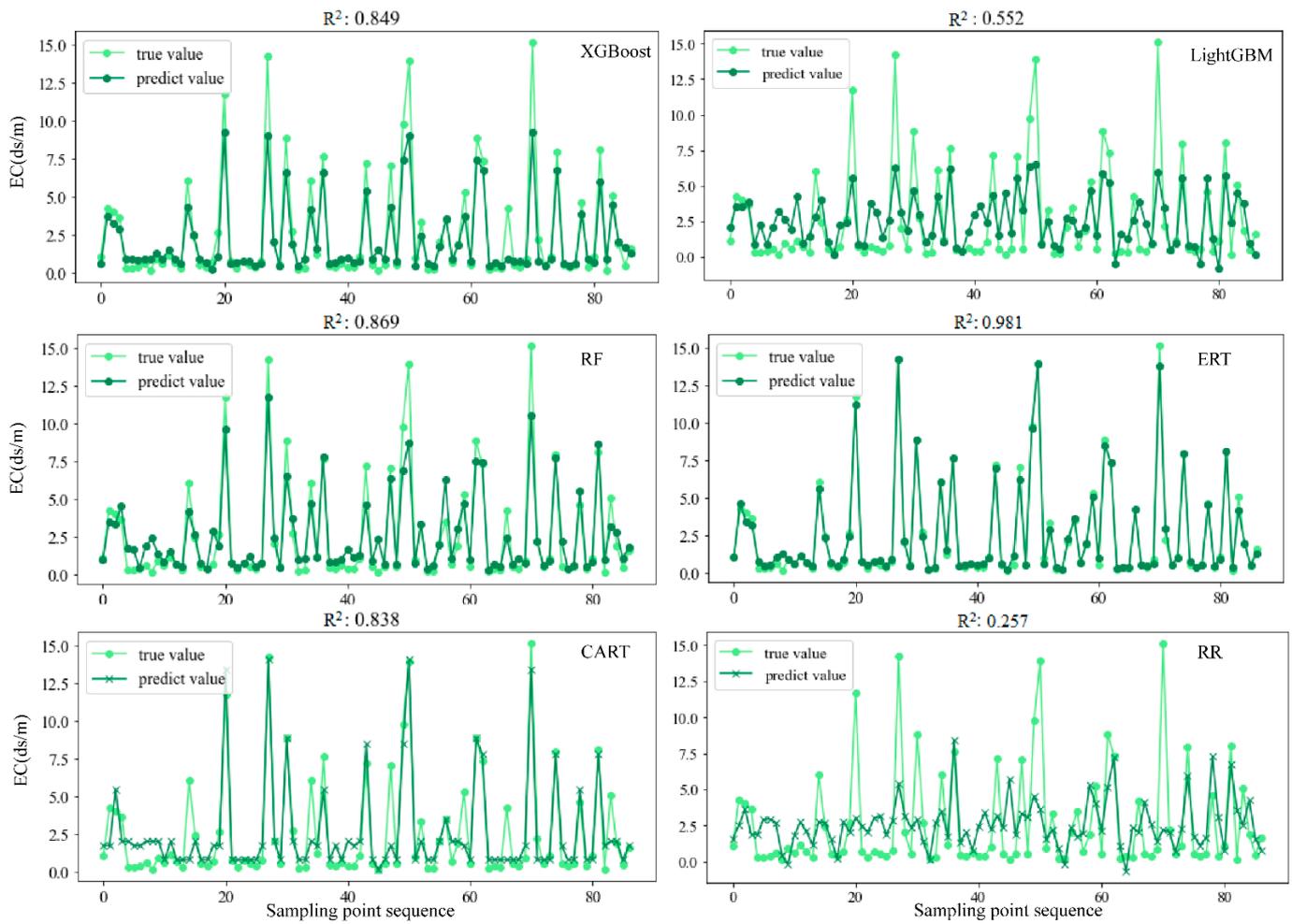
Based on GBM results, the top 13-ranked features could provide more than 90% of the cumulative importance for the model (Figure 7) (the salinity index SI was denoted as SI (1), and the 2D index SI as SI (2)). Therefore, the top 13 feature variables were selected as the independent variables for the subsequent model.



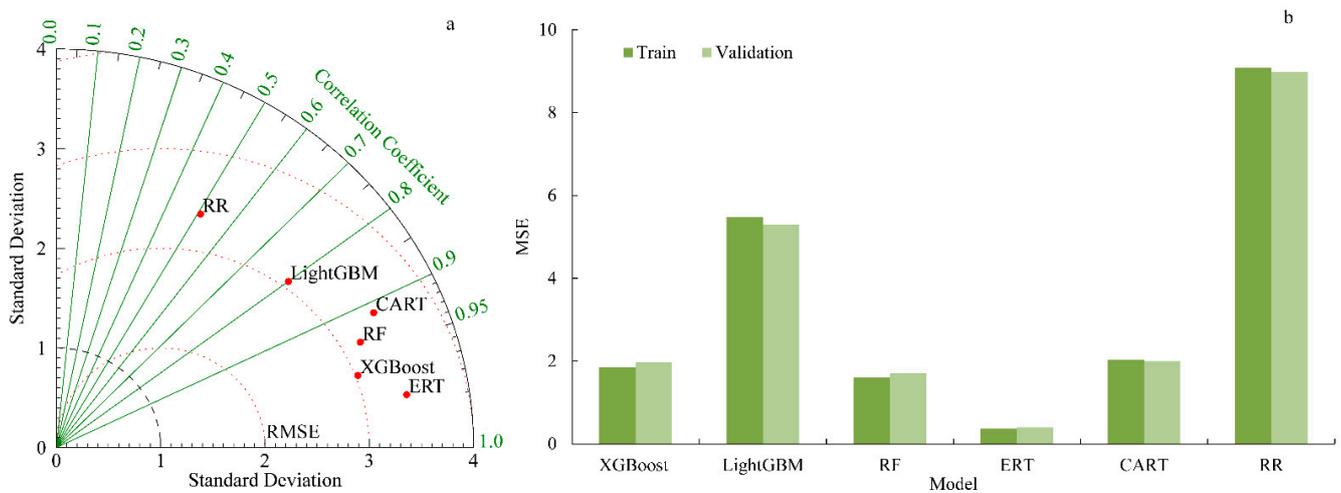
**Figure 7.** Feature importance of spectral index and topographic factors ranking using gradient boosting machine (GBM). Dotted vertical line representing cumulative feature importance reached 0.9 (90%).

#### 3.3.2. Establishment and Verification of Soil EC Inversion Models

The selected spectral parameters and topographic factors by GBM and EC contents were applied as input and output datasets, respectively, to construct an EC model via the XGBoost, LightGBM, RF, ERT, CART and RR methods. The prediction effect of different models on the training set is shown in Figure 8. The ERT model had excellent results with the most reliable estimation ( $R^2 = 0.98$ ), which was followed by the RF method ( $R^2 = 0.87$ ), XGBoost methods ( $R^2 = 0.85$ ), the CART model ( $R^2 = 0.84$ ), LightGBM methods ( $R^2 = 0.55$ ) and the RR model ( $R^2 = 0.26$ ). The standard deviation of the tree model does not differ significantly based on the Taylor diagram (Figure 9a). The MSE (Train and Validation) order was ERT (0.37, 0.40) < RF (1.60, 1.72) < XGBoost (1.85, 1.96) < CART (2.02, 2.00) < LightGBM (5.47, 5.3) < RR (9.08, 9.00) (Figure 9b), and the correlation coefficient order was ERT (0.98) > XGBoost (0.96) > RF (0.94) > CART (0.91) > LightGBM (0.80) > RR (0.51) (Figure 9a). Overall, the ERT model performs the best. ERT, XGBoost, RF and CART better predicted soil EC within the MSE range of 0.37 to 2.02 (Figure 9b).



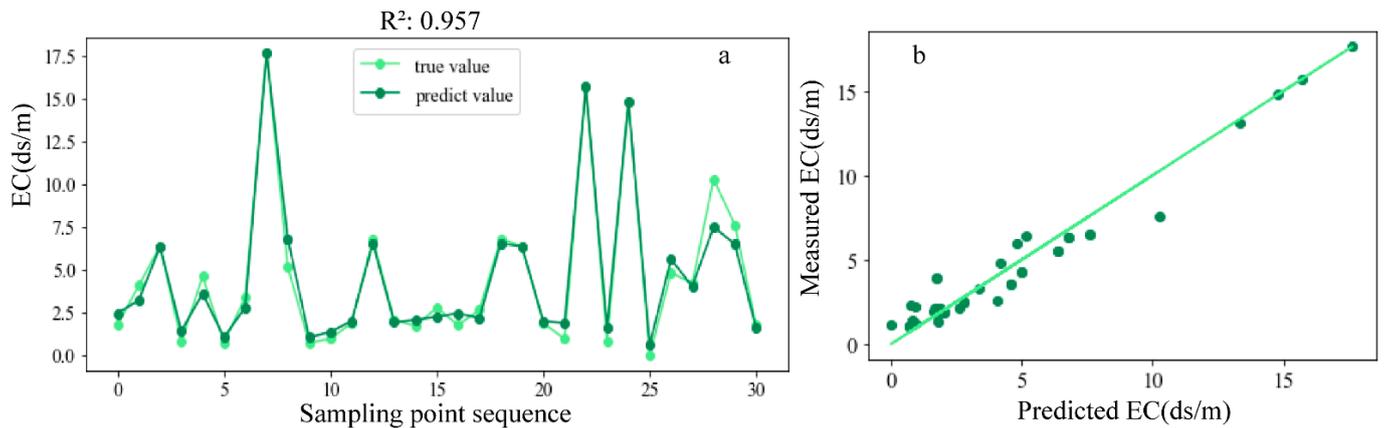
**Figure 8.** The training set fitting effect diagrams and scores of six machine learning methods: extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), random forest (RF), extremely randomized trees (ERT), classification and regression tree (CART), and ridge regression (RR).



**Figure 9.** The normalized Taylor diagrams of the predicted and measured EC data (a) and the model accuracy comparison and mean squared error (MSE) of the six methods (b).

### 3.3.3. Testing of Predictive Models

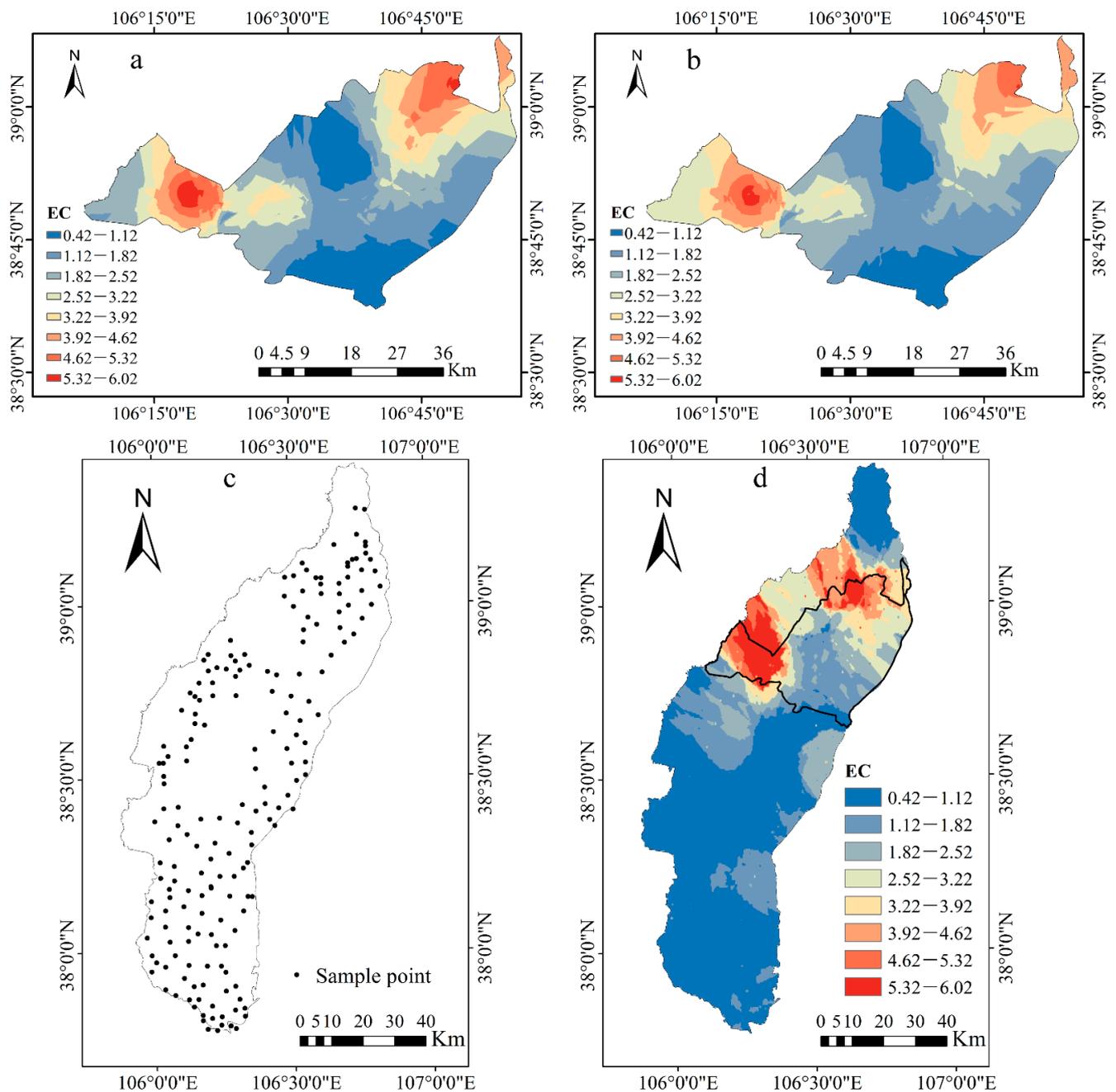
This study employed the trained model for the July 2018 data ( $n = 42$ ) to examine the ERT model's prediction performance for soil EC inversion. All of the procedures for calculating salinity indices and modeling were the same as those stated previously. The model showed significant correlation with the measured (true) data ( $R^2 = 0.96$ ) (Figure 10a). The model had a strong linear relationship between the measured and simulated (true) values ( $y = ax + b$ ,  $R^2 = 0.96$ ) (Figure 10b). Therefore, the ERT model was the best model for EC simulation.



**Figure 10.** Fitting effect graph of ERT for EC (a) and model using validation correlation diagram of ERT for EC (b) in July 2018.

### 3.4. Digital Soil Maps of EC

The kriging interpolation method of inversion of the spatial distribution of soil EC was used (Figure 11). The ERT model inversion effect and the measured value were very close. In general, overall trends in soil EC in the north-east and midwest are relatively serious, with salinity conditions from north to south continuously reducing. The results were consistent with the field investigation and demonstrated the validity of the model to some extent. To verify the application of the model on a large scale, 320 data collected in 2019 and 2021 from Yinchuan Plain were used. Topographically, the areas with the lowest degree of salinization were mainly located in the mountainous areas in the southern part of the plain. The highest salinity in the region was found in the Xidatan area in the northwest. In addition, it can be seen that the EC values in Pingluo county in Yinchuan Plain are relatively consistent with the validation effect of the model, which indicates that the model can be used for large-scale inversion.



**Figure 11.** Spatial distribution of soil EC (a) measured value, (b) ERT of study area, (c) sample point in Yinchuan Plain, (d) measured value in Yinchuan Plain.

## 4. Discussion

### 4.1. Hyperspectral Pre-Processing

Spectral data pre-processing can effectively remove background noise, baseline effect and multiplier interference, which is an important step in extracting useful spectral information and optimizing quantitative effects of models [59]. Selecting the most suitable pre-processing method to process all the datasets is difficult and infeasible [13]. In this study, five kinds of mathematical transformations (SNV, SNV-DT, 1/OR, Log (1/OR), and FOD (range 0–2, with intervals of 0.25)) were carried out on the basis of the original characteristic spectrum, pre-processing the spectral data with different degrees of change (Figure 3). Through pre-processing, the reflection and absorption peak valley were better identified in FOD transformation. Although spectral intensity gradually decreased, spectral

details increased. Many tiny peaks began to appear and grew with the increase of the derivative order, which achieved the purpose of refining the variation trend and reducing the information defect [60].

According to the analysis of the correlation between the reflectance of spectral transformation and EC, the MACC of the two-dimensional correlation coefficient was much higher than that of the one-dimensional (Table 4). In one-dimensional, the MACC of SNV and of FOD were higher than that of OR. These results confirmed that partial spectral pre-processing eliminates or reduces unwanted side effects in reflection spectra. The two-dimensional correlation coefficient was much larger than the one-dimensional, indicating that the two-band spectral index fully considers the interaction between spectral bands and effectively eliminates the overlapping absorption of soil components [21]. The two-dimensional correlation coefficient further reveals the potential use of the optimal band algorithm in determining sensitive spectral variables related to EC content. It can be seen that SNV has the best transformation effect in the Vis-NIR band, which may be because SNV improves the signal-to-noise ratio of the original absorbance spectrum and enhances the spectral absorption information related to component content [61]. The best transformation band correlated with EC in the 2 order derivative ( $PCC_{420nm} = 0.596$ ), because the spectral derivative effectively deals with nonlinear problems, enhances the difference between similar spectra, and eliminates background noise. SNV and 2 order derivative have the largest improvement in two dimensions compared with one dimension, which were 0.286 and 0.258, respectively. Peon et al. [62] successfully transferred the spectral reflectance of the laboratory to different satellite sensors by means of different spectral response functions, or directly extracted sensitive spectral parameters from satellite sensors. Therefore, we can use SNV and 2 order derivative to calculate the optimal spectral index in order to simplify the input variables and develop increasingly efficient and specific EC estimation models.

#### 4.2. Feature and Model Selection

Studies show that the robustness of the model improves by removing potentially irrelevant environmental variables [19]. In this study, the importance of variables was ranked based on GBM, and it was found that DEM, slope, aspect, plane curvature, profile curvature and TWI all participate in EC modeling. Topographic variables determine the movement direction of run-off water, thus changing the accumulation mode and location of salt in soil [25]. Taghizadeh-mehrjardi et al. [63] proved that DEM and its derivatives were significantly correlated with soil salinity and have great potential in monitoring soil salinity. Peng et al. [64] established an EC inversion model of the Aksu River Basin in Xinjiang using terrain attributes and Landsat 8 OLI index, with  $R^2$  reaching 0.92. It is possible to apply this hyperspectral feature and model for remote sensing.

Prediction accuracy is the most important factor for soil properties inversion. At present, scholars have conducted relevant studies on the inversion of soil EC, but the models have mixed results [65–69]. In this study, ERT and RF models perform better than XGBoost, LightGBM, CART, and RR (Figure 8). This may be due to introducing random attribute selection during the training of the RF model, and due to the extraction of data based on randomness and difference [70], which greatly improved the accuracy of decision making [33,71]. Studies have shown that ensemble learning and variable selection can improve the consistency of models with predicted values and true values when variables are complex [72]. Compared with bagging, boost's prediction of EC was slightly inferior, and the comparison of the two boosting methods shows that the XGBoost algorithm has a better prediction effect. XGBoost avoids the over-fitting problem to a large extent by introducing regularization terms, thus improving the generalization ability of the model [39]. Therefore, XGBoost can be used as an effective method to build a simulation estimation model of salt content in a certain region. LightGBM is an efficient implementation of a gradient boosting decision tree (GBDT). As a distributed gradient lifting framework, it is mainly optimized in the training speed and memory of the model. Due to the unique leaf-wise strategy of LightGBM, it is easier to control the model complexity, however, it is difficult to give full

play to its advantages in the case of small data sets [38]. This may be the reason for the poor performance of LightGBM in the problem background of this study. It should be pointed out that the best single model cannot guarantee the highest accuracy under changing input data or future conditions [73].

#### 4.3. Spatial Distribution of Soil Salinity

Interpolation through the model between the predicted values and measured values show that soil EC values in the midwest and northeast of Pingluo County are more serious. From north to south, the salinity condition continuously decreases, where most soils with low salinity are distributed. In general, both in the region landform, groundwater, drainage and other related factors, distribution of Sha lakes, the west lake, and Mingshui lake in midwest Pingluo county, in the west Xidatan town was more serious. In the north of the plain, due to shallow groundwater, poor drainage and a high evaporation ratio, salt crusts commonly form. In the south, the terrain elevation is higher and the drainage is smooth. Therefore, the salinization degree is low, and most soils are slightly saline or non-saline.

Yinchuan Plain is an important ecological barrier of China's western region and reserved cultivated land resources, with 30% arable land. Yinchuan Plain provides 50% of the food produced in Ningxia Province. However, soil salinization in the Yinchuan Plain has become a particular inhibitor of crop growth and of the healthy development of agricultural and ecological systems [74]. Thus, various approaches, including leaching with low EC water and adding gypsum, as well as cultivating salt-tolerant plants, have already been attempted in order to alleviate salinization effects [75]. Nevertheless, the success of such approaches needs continuous monitoring of changes in soil EC, which can be time consuming or infeasible for large areas. An accurate remote sensing approach to monitor soil salinization, as suggested in this study, will overcome these problems and will enable the avoidance of future salinization, especially when the cultivated land area and grain yield should be increased to guarantee national food security.

#### 4.4. Uncertainty Analysis of Soil EC Inversion Model

The accuracy of hyperspectral based monitoring of soil salinization is susceptible to data pre-processing, and factors such as soil texture, moisture content, organic matter content, etc. that affect the soil spectrum and the selection of salt indices and vegetation indices. Moreover, the division of modeling validation sets also affects the model accuracy [76,77]. In the inversion study of soil electrical conductivity, the correlations were mainly between the reflectance and the degree of salinity. Although the properties of saline soils are comparatively changeable and so may affect the reflectance, such changes, or the intensity of changes and consequences for the reflectance, are also due to or dependent on the salinity levels. Our results show that the 2 order derivative is the best spectral processing method and that the terrain factor plays a dominant role in the model, which makes it superior to spectral variables and provides a basis for the simulation of soil salinization at larger scales. Nevertheless, no universal spectral index or variable can give a single satisfactory result under any environmental condition. Therefore, the selection of optimal modeling parameters should be based on the regional and environmental conditions, rather than on fixed parameters.

Simple near-ground remote sensing methods cannot reflect the characteristics of salinization. Therefore, comprehensive multisource data is needed as a new way to study the complex salinization monitoring problem, by obtaining more soil and training information [78]. The soil samples used in this paper are mainly concentrated in arid areas: whether the established optimal model can be applied to other areas remains to be further discussed and verified. The applicability of different ensemble learning algorithms needs to be explored for arid oasis and for coastal saline soils, and other types of models should be selected for comparison and optimization. Our approach can also be applied to predict other soil properties such as texture and clay mineralogy, or to determine different soil layers or soil types based on different data sets and verifications. Nevertheless, in the

future, the study range and soil sample size need further expansion to complete the soil hyperspectral database, and the environmental variables should be appropriately considered in order to improve the accuracy of the model. It is necessary to study how soil properties affect the response mechanism of the hyperspectral inversion of the EC content, and to combine models and images in order to conduct large-scale mapping and monitoring of soil salinization [10].

## 5. Conclusions

To find the best hyperspectral inversion of soil electrical conductivity, this study was conducted at northern Yinchuan Plain, China, where serious soil salinization is faced. The performance of different pre-processing techniques on different models of the same data set were applied. The extremely randomized trees (ERT) and light gradient boosting machine (LightGBM) models were studied. We used soil measured hyperspectral and salinity data to explore the feasibility of identifying EC via Vis–NIR spectral model outputs. The correlation between reflectance and EC was analyzed using different hyperspectral pre-processing methods, and the spectral indices were calculated. Under different spectral transformation and salinity indices, the correlation between 2 order derivative and SVN-NDI and EC were the largest, with 0.596 and 0.689, respectively. The SNV and 2 order derivative pre-processing techniques exerted a strong influence on improving the correlations. The salinity features selected based on GBM include slope, NDI, SI-T, RI, profile curvature, DOA, plane curvature, SI (conventional), elevation, Int2, aspect, S1 and TWI. Among the six EC inversion models (XGBoost, LightGBM, RF, ERT, CART and RR), the ERT model performs best. The optimal parameter combination after the grid search was as follows: the  $n\_estimator$  was 21, the  $max\_depth$  was 14, and  $R^2$  and MSE were 0.98 and 0.37, respectively. Based on the validation of the prediction model, the machine learning model (ERT) can be applied to the prediction of EC. The ERT model provides a new method for the accurate inversion of soil EC. Our study provides a reference for the inversion of soil EC, and a basis for soil salinization simulation on a larger scale, which can support the sustainable development of local agriculture and protect the ecological environment in arid areas.

**Author Contributions:** Writing-original draft, visualization, P.J.; Data curation, P.J., K.J. and J.Z.; Validation, P.J., K.J., R.Z. and X.Z.; Formal analysis, W.H., Y.H. and X.Z.; Supervision, J.Z. and X.Z.; Review, editing, and revision, K.J., K.Z. and X.Z.; Funding acquisition, K.J., J.Z., K.Z. and X.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (Grant numbers 41877109; 42050410320; 42067003; 42061047); the Jiangsu Specially-Appointed Professor Project, China (Grant number R2020T29); the Key R&D Project of Ningxia, China (Grant number 2021BEG03002); the National Key R&D Program of China (Grant number 2021YFD1900602); and the Thousand Young Talents Program, China (Grant number Y772121).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Metternicht, G.I.; Zinck, J.A. Remote sensing of soil salinity: Potentials and constraints. *Remote Sens. Environ.* **2003**, *85*, 1–20. [[CrossRef](#)]
2. Zhang, T.T.; Qi, J.G.; Gao, Y.; Ouyang, Z.T.; Zeng, S.L.; Zhao, B. Detecting soil salinity with MODIS time series VI data. *Ecol. Indic.* **2015**, *52*, 480–489. [[CrossRef](#)]
3. Tóth, G.; Hermann, T.; da Silva, M.R.; Montanarella, L. Monitoring soil for sustainable development and land degradation neutrality. *Environ. Monit. Assess.* **2018**, *190*, 57. [[CrossRef](#)] [[PubMed](#)]
4. Huang, F.; Ding, X.X.; Li, W.W.; Jia, H.T.; Wei, X.R.; Zhao, X.N. The effect of temperature on decomposition of the different parts of maize residues in a Solonchak. *Catena* **2021**, *201*, 105207. [[CrossRef](#)]
5. Nachshon, U. Cropland soil salinization and associated hydrology: Trends, processes and examples. *Water* **2018**, *10*, 1030.

6. Zaman, M.; Shahid, S.A.; Heng, L. *Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques*; Springer: Cham, Switzerland, 2018.
7. Chen, Z.X.; Ren, J.Q.; Tang, H.J.; Shi, Y.; Leng, P.; Liu, J.; Wang, L.M.; Wu, W.B.; Yao, Y.M.; Hasiyuyua. Progress and perspectives on agricultural remote sensing research and applications in China. *J. Remote Sens.* **2016**, *20*, 748–767.
8. Tripathi, A.; Tiwari, R.K. A Simplified Sub-Surface Soil Salinity estimation using Synergy of Sentinel-1 SAR and Sentinel-2 multispectral satellite data, for early stages of wheat crop growth in Rupnagar, Punjab, India. *Land Degrad. Dev.* **2021**, *32*, 3905–3919. [[CrossRef](#)]
9. Peng, J.; Wang, J.Q.; Xiang, H.Y.; Teng, H.F.; Liu, W.Y.; Chi, C.M.; Niu, J.L.; Guo, Y.; Shi, Z. Comparative study on hyperspectral inversion accuracy of soil salt content and electrical conductivity. *Spectrosc. Spectr. Anal.* **2014**, *34*, 510–514.
10. Wang, J.Z.; Ding, J.L.; Yu, D.L.; Ma, X.K.; Zhang, Z.P.; Ge, X.Y.; Teng, D.X.; Li, X.H.; Liang, J.; Lizaga, I.; et al. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region. Xinjiang, China. *Geoderma* **2019**, *353*, 172–187. [[CrossRef](#)]
11. Sun, X.; Gao, Y.; Wang, D.E.; Chen, J.H.; Zhang, F.G.; Zhou, J.B.; Li, Y.H. Stoichiometric variation of halophytes in response to changes in soil salinity. *Plant Biol.* **2017**, *19*, 360–367. [[CrossRef](#)]
12. Gorji, T.; Sertel, E.; Tanika, A. Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey. *Ecol. Indic.* **2017**, *74*, 384–391. [[CrossRef](#)]
13. Viscarra Rossel, R.A.; Behrens, T.; Ben-Dor, E.; Brown, D.J.; Demattê, J.A.M.; Shepherd, K.D.; Shi, Z.; Stenberg, B.; Stevens, A.; Adamchuk, V.; et al. A global spectral library to characterize the world's soil. *Earth Sci. Rev.* **2016**, *155*, 198–230. [[CrossRef](#)]
14. Douglas, R.K.; Nawar, S.; Alamar, M.C.; Mouazen, A.M.; Coulon, F. Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using Vis-NIR spectroscopy and regression techniques. *Sci. Total Environ.* **2018**, *616*, 147–155. [[CrossRef](#)] [[PubMed](#)]
15. Ding, J.L.; Yu, D.L. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan–Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments. *Geoderma* **2014**, *235–236*, 316–322. [[CrossRef](#)]
16. Nawar, S.; Buddenbaum, H.; Hill, J. Estimation of soil salinity using three quantitative methods based on visible and near-infrared reflectance spectroscopy: A case study from Egypt. *Arab. J. Geosci.* **2015**, *8*, 5127–5140. [[CrossRef](#)]
17. Zhang, Z.P.; Ding, J.L.; Zhu, C.M.; Wang, J.Z. Combination of efficient signal pre-processing and optimal band combination algorithm to predict soil organic matter through visible and near-infrared spectra. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2020**, *240*, 118553. [[CrossRef](#)] [[PubMed](#)]
18. Kooistra, L.; Wanders, J.; Epema, G.F. The potential of field spectroscopy for the assessment of sediment properties in river floodplains. *Anal. Chim. Acta* **2003**, *484*, 189–200. [[CrossRef](#)]
19. Wang, H.F.; Chen, Y.W.; Zhang, Z.T.; Chen, H.R.; Chai, H.Y. Quantitatively estimating main soil water-soluble salt ions content based on Visible-near infrared wavelength selected using GC, SR and VIP. *PeerJ* **2019**, *7*, e6310. [[CrossRef](#)]
20. Yu, X.; Liu, Q.; Wang, Y.B.; Liu, X.Y.; Liu, X. Evaluation of MLSR and PLSR for estimating soil element contents using visible/near-infrared spectroscopy in apple orchards on the Jiaodong peninsula. *Catena* **2016**, *137*, 340–349. [[CrossRef](#)]
21. Hong, Y.S.; Liu, Y.L.; Chen, Y.Y.; Liu, Y.F.; Yu, L.; Liu, Y.; Cheng, H. Application of fractional-order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy. *Geoderma* **2019**, *337*, 758–769. [[CrossRef](#)]
22. Hong, Y.S.; Chen, S.C.; Zhang, Y.; Chen, Y.Y.; Lei, Y.; Liu, Y.F.; Liu, Y.L.; Cheng, H.; Liu, Y. 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*, 1232–1243. [[CrossRef](#)] [[PubMed](#)]
23. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [[CrossRef](#)]
24. Zhao, X.N.; Othmanli, H.; Schiller, T.; Zhao, C.Y.; Sheng, Y.; Zia, S.; Müller, J.; Stahr, K. Water Use Efficiency in Saline Soils under Cotton Cultivation in the Tarim River Basin. *Water* **2015**, *7*, 3103–3122. [[CrossRef](#)]
25. Vermeulen, D.; Van Niekerk, A. Machine learning performance for predicting soil salinity using different combinations of geomorphometric covariates. *Geoderma* **2017**, *299*, 1–12. [[CrossRef](#)]
26. Wang, J.Z.; Ding, J.L.; Yu, D.L.; Teng, D.X.; He, B.; Chen, X.Y.; Ge, X.Y.; Zhang, Z.P.; Wang, Y.; Yang, X.D.; et al. Machine learning-based detection of soil salinity in an arid desert region, Northwest China: A comparison between Landsat-8 OLI and Sentinel-2 MSI. *Sci. Total Environ.* **2020**, *707*, 136092. [[CrossRef](#)] [[PubMed](#)]
27. Genuer, R.; Poggi, J.M.; Tuleau-Malot, C. Variable selection using random forests. *Pattern Recognit. Lett.* **2010**, *31*, 2225–2236. [[CrossRef](#)]
28. Zou, X.; Zhao, J.; Povey, M.J.W.; Mel, H.; Mao, H. Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* **2010**, *667*, 14–32.
29. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
30. Dotto, A.C.; Dalmolin, R.S.D.; ten Caten, A.; Grunwald, S. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. *Geoderma* **2018**, *314*, 262–274. [[CrossRef](#)]
31. Douglas, R.K.; Nawar, S.; Cipullo, S.; Alamar, M.C.; Coulon, F.; Mouazen, A.M. Evaluation of Vis-NIR reflectance spectroscopy sensitivity to weathering for enhanced assessment of oil contaminated soils. *Sci. Total Environ.* **2018**, *626*, 1108–1120. [[CrossRef](#)]

32. Wang, X.P.; Zhang, F.; Ding, J.L.; Kung, H.T.; Latif, A.; Johnson, V.C. Estimation of soil salt content (SSC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR), Northwest China, based on a Bootstrap-BP neural network model and optimal spectral indices. *Sci. Total Environ.* **2018**, *615*, 918–930. [[CrossRef](#)] [[PubMed](#)]
33. Wang, F.; Shi, Z.; Biswas, A.; Yang, S.T.; Ding, J.L. Multi-algorithm comparison for predicting soil salinity. *Geoderma* **2020**, *365*, 114211. [[CrossRef](#)]
34. Triki Fourati, H.; Bouaziz, M.; Benzina, M.; Bouaziz, S. Modeling of soil salinity within a semi-arid region using spectral analysis. *Arab. J. Geosci.* **2015**, *8*, 11175–11182. [[CrossRef](#)]
35. Xu, S.X.; Zhao, Y.C.; Wang, M.Y.; Shi, X.Z. 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](#)]
36. Das, B.; Manohara, K.K.; Mahajan, G.R.; Sahoo, R.N. Spectroscopy based novel spectral indices, PCA- and PLSR-coupled machine learning models for salinity stress phenotyping of rice. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2019**, *229*, 117983. [[CrossRef](#)]
37. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [[CrossRef](#)]
38. Ke, G.L.; Meng, Q.; Finley, T.; Wang, T.F.; Chen, W.; Ma, W.D.; Ye, Q.W.; Liu, T.Y. Light GBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2017; pp. 3147–3155.
39. Zhang, R.H.; Wei, X.; Lu, Z.H.; Ai, Y.J. Training a model for predicting adsorption energy of metal ions based on machine learning. *J. Inorg. Mater.* **2021**, *36*, 1178–1184. [[CrossRef](#)]
40. Cheng, Y.; Li, Y.X.; Li, F.; He, L. Soil moisture retrieval using extremely randomized trees over the Shandian river basin. *Natl. Remote Sens. Bull.* **2021**, *25*, 941–951.
41. Zhang, Y.Z.; Ma, J.; Liang, S.L.; Li, X.S.; Li, M.Y. An evaluation of eight machine learning regression algorithms for forest aboveground biomass estimation from multiple satellite data products. *Remote Sens.* **2020**, *12*, 4015. [[CrossRef](#)]
42. Chen, R.H.; Shang, T.H.; Zhang, J.H.; Wang, Y.J.; Jia, K.L. Effect of different spectra types on the accuracy and correction of soil salt content inversion in Yinchuan Plain, China. *J. Appl. Ecol.* **2021**, *33*, 922–930. [[CrossRef](#)]
43. Jia, P.P.; Shang, T.H.; Zhang, J.H.; Sun, Y. Inversion of soil pH during the dry and wet seasons in the Yinbei region of Ningxia, China, based on multi-source remote sensing data. *Geoderma Reg.* **2021**, *25*, e00399. [[CrossRef](#)]
44. Lu, R.K. *Soil Agrochemistry Analysis Protocols*; China Agriculture Science Press: Beijing, China, 1999.
45. Brady, N.C.; Weil, R.R. *The Nature and Properties of Soils*, 14th ed.; Li, B.G.; Xu, J.M., Translators; Science Press: Beijing, China, 2019.
46. Ge, X.Y.; Wang, J.Z.; Ding, J.L.; Cao, X.Y.; Zhang, Z.P.; Liu, J.; Li, X.H. Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring. *PeerJ.* **2019**, *7*, e6926. [[CrossRef](#)] [[PubMed](#)]
47. Jiao, C.X.; Zheng, G.H.; Xie, X.L.; Cui, X.F.; Shang, G. Prediction of soil organic matter using Visible-Short Near-Infrared imaging spectroscopy. *Spectrosc. Spectr. Anal.* **2020**, *40*, 3277–3281.
48. Allbed, A.; Kumer, L.; Aldakheel, Y.Y. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region. *Geoderma* **2014**, *230–231*, 1–8. [[CrossRef](#)]
49. Douaoui, A.E.K.; Nicolas, H.; Walter, C. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma* **2006**, *134*, 217–230. [[CrossRef](#)]
50. Abbas, A.; Khan, S. Using remote sensing techniques for appraisal of irrigated soil salinity. In *Proceedings of the Advances and Applications for Management and Decision Making Land, Water and Environmental Management: Integrated Systems for Sustainability MODSIM07*; Modelling and Simulation Society of Australia and New Zealand: Canberra, Australia, 2007; pp. 2632–2638.
51. Cao, L.; Ding, J.L.; Umot, H.; Su, W.; Ning, J.; Miu, C.; Li, H. Extraction and modeling of regional soil salinization based on data from GF-1 satellite. *Acta Pedol. Sin.* **2016**, *53*, 1399–1409.
52. Khan, N.M.; Rastoskuev, V.V.; Sato, Y.; Shiozawa, S. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agric. Water Manag.* **2005**, *77*, 96–109. [[CrossRef](#)]
53. Jin, X.L.; Song, K.S.; Du, J.; Liu, H.J.; Wen, Z.D. Comparison of different satellite bands and vegetation indices for estimation of soil organic matter based on simulated spectral configuration. *Agric. Meteorol.* **2017**, *244–245*, 57–71. [[CrossRef](#)]
54. Hong, Y.S.; Shen, R.L.; Cheng, H.; Chen, S.C.; Chen, Y.Y.; Guo, L.; He, J.H.; Liu, Y.L.; Yu, L.; Liu, Y. Cadmium concentration estimation in peri-urban agricultural soils: Using reflectance spectroscopy, soil auxiliary information, or a combination of both? *Geoderma* **2019**, *354*, 113875. [[CrossRef](#)]
55. Ihuoma, S.O.; Madramootoo, C.A. Narrow-band reflectance indices for mapping the combined effects of water and nitrogen stress in field grown tomato crops. *Biosyst. Eng.* **2020**, *192*, 133–143. [[CrossRef](#)]
56. Rukeya, S.; Nijat, K.; Abdugheni, A.; Hu, L.; Ahunaji, Y.; Balati, M.; Shi, Q. Possibility of optimized indices for the assessment of heavy metal contents in soil around an open pit coal mine area. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 14–25.
57. Martens, H.A.; Dardenne, P. Validation and verification of regression in small data sets. *Chemometr. Intell. Lab. Syst.* **1998**, *44*, 99–121. [[CrossRef](#)]
58. Liashchynskiy, P.; Liashchynskiy, P. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. *arXiv* **2019**, arXiv:1912.06059.
59. Rinnan, A.; Berg, V.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](#)]
60. Zhao, Q.D.; Ge, X.Y.; Ding, J.L.; Wang, J.Z.; Zhang, Z.H.; Tian, M.L. Combination of Fractional order differential and machine learning algorithm for spectral estimation of soil organic carbon content. *Laser Optoelectron. Prog.* **2020**, *57*, 253–261.
61. Shi, Z. *Principle and Method of Soil Surface Hyperspectral Remote Sensing*; Science Press: Beijing, China, 2014; pp. 30–31.

62. Peon, J.; Recondo, C.; Fernandez, S.; Calleja, J.F.; De Miguel, E.; Carretero, L. Prediction of topsoil organic carbon using airborne and satellite hyperspectral imagery. *Remote Sens.* **2017**, *9*, 1211. [[CrossRef](#)]
63. Taghizadeh-Mehrjardi, R.; Minasny, B.; Sarmadian, F.; Malone, B.P. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma* **2014**, *213*, 15–28. [[CrossRef](#)]
64. Peng, J.; Biswas, A.; Jiang, Q.; Zhao, R.; Hu, J.; Hu, B.; Shi, Z. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma* **2018**, *337*, 1309–1319. [[CrossRef](#)]
65. Gorji, T.; Yildirim, A.; Hamzehpour, N.; Tanik, A.; Sertel, E. Soil salinity analysis of Urmia Lake Basin using Landsat-8 OLI and Sentinel-2A based spectral indices and electrical conductivity measurements. *Ecol. Indic.* **2020**, *112*, 106173. [[CrossRef](#)]
66. Paz, A.M.; Castanheira, N.; Farzamian, M.; Paz, M.C.; Conceição Gonçalves, M.; Monteiro Santos, F.A.; Triantafylis, J. Prediction of soil salinity and sodicity using electromagnetic conductivity imaging. *Geoderma* **2020**, *361*, 114086. [[CrossRef](#)]
67. Habibi, V.; Ahmadi, H.; Jafari, M.; Moeini, A. Mapping soil salinity using a combined spectral and topographical indices with artificial neural network. *PLoS ONE* **2021**, *16*, e0228494. [[CrossRef](#)] [[PubMed](#)]
68. Melendez-Pastor, I.; Navarro-Pedreño, J.; Gómez Lucas, I.; Koch, M. Identifying Optimal Spectral Bands to Assess Soil Properties with VNIR Radiometry in Semi-Arid Soils. *Geoderma* **2008**, *147*, 126–132. [[CrossRef](#)]
69. Pang, G.J.; Wang, T.; Liao, J.; Li, S. Quantitative Model Based on Field-Derived Spectral Characteristics to Estimate Soil Salinity in Minqin County, China. *Soil Sci. Soc. Am. J.* **2014**, *78*, 546. [[CrossRef](#)]
70. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
71. Hengl, T.; Mendes, D.J.J.; Heuvelink, G.B.; Ruiperez, G.M.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauermarschallinger, B. Soil Grids 250 m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e0169748. [[CrossRef](#)]
72. Heung, B.; Ho, H.C.; Zhang, J.; Knudby, A.; Bulmer, C.E.; Schmidt, M.G. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* **2016**, *265*, 62–77. [[CrossRef](#)]
73. Shafizadeh-Moghadam, H.; Valavi, R.; Shahabi, H.; Chapi, K.; Shirzadi, A. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *J. Environ. Manag.* **2018**, *217*, 1–11. [[CrossRef](#)]
74. Xu, Y.Z.; Guan, C.; Ding, H.; Ci, D.W.; Qin, F.F.; Zhang, Z.M.; Dai, L.X. Effects of salt and drought stresses on rhizosphere soil bacterial community structure and peanut yield. *J. Appl. Ecol.* **2020**, *31*, 1305–1313.
75. Tang, X.; Shang, H.; Liu, G.M.; Yao, Y.T.; Zhang, F.H.; Yang, J.S.; Zhou, L.X.; Chu, R. Effects of Combined Amendment on Improvement of Salinized Soil and Plant Growth. *Soils* **2021**, *53*, 1033–1039.
76. Liu, Y.; Pan, X.Z.; Wang, C.K.; Li, Y.L.; Shi, R.J. Predicting soil salinity with Vis-NIR spectra after removing the effects of soil moisture using external parameter orthogonalization. *PLoS ONE* **2015**, *10*, e0140688. [[CrossRef](#)]
77. Wang, F.; Ding, J.L.; Wei, Y.; Zhou, Q.Q.; Yang, X.D.; Wang, Q.F. Sensitivity analysis of soil salinity and vegetation indices to detect soil salinity variation by using Landsat series images: Applications in different oases in Xinjiang, China. *Acta Ecol. Sin.* **2017**, *37*, 5007–5022.
78. O'Rourke, S.M.; Stockmann, U.; Holden, N.M.; McBratney, A.B.; Minasny, B. An assessment of model averaging to improve predictive power of portable vis-NIR and XRF for the determination of agronomic soil properties. *Geoderma* **2016**, *279*, 31–44. [[CrossRef](#)]