

## Article

# Soil Salinity Mapping Using Machine Learning Algorithms with the Sentinel-2 MSI in Arid Areas, China

Jiaqiang Wang <sup>1,2</sup> , Jie Peng <sup>3,\*</sup>, Hongyi Li <sup>3</sup> , Caiyun Yin <sup>4</sup>, Weiyang Liu <sup>2</sup>, Tianwei Wang <sup>1,†</sup> and Huaping Zhang <sup>4</sup>

<sup>1</sup> College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China; jiaqiangwang@webmail.hzau.edu.cn (J.W.); wangtianwei@webmail.hzau.edu.cn (T.W.)

<sup>2</sup> College of Plant Science, Tarim University, Alar 843300, China; lwyzky@taru.edu.cn

<sup>3</sup> School of Tourism and Urban Management, Jiangxi University of Finance and Economics, Nanchang 330032, China; lihongyi1981@zju.edu.cn

<sup>4</sup> Agricultural Technology Extension Station of the First Division of Xinjiang Production and Construction Corps, Alar 843300, China; 120050027@taru.edu.cn (C.Y.); 120020031@taru.edu.cn (H.Z.)

\* Correspondence: 11414049@zju.edu.cn or pjzky@163.com

† These authors have equal contribution.

**Abstract:** Accurate monitoring of soil salinization plays a key role in the ecological security and sustainable agricultural development of arid regions. As a branch of artificial intelligence, machine learning acquires new knowledge through self-learning and continuously improves its own performance. The purpose of this study is to combine Sentinel-2 Multispectral Imager (MSI) data and MSI-derived covariates with measured soil salinity data and to apply three machine learning algorithms in modeling to estimate and map the soil salinity in the study sample area. According to the convenient transportation conditions, the study area and sampling quadrat were set up, and the 5-point method was used to collect the soil mixed samples, and 160 soil mixed samples were collected. Kennard–Stone (K–S) algorithm was used for sample classification, 70% for modeling and 30% for verification. The machine learning algorithm uses Support Vector Machines (SVM), Artificial Neural Network (ANN), and Random Forest (RF). The results showed that (1) the average reflectance of each band of the MSI data ranged from 0.21–0.28. According to the spectral characteristics corresponding to different soil electrical conductivity (EC) levels (1.07–79.6 dS m<sup>−1</sup>), the spectral reflectance of salinized soil in the MSI data ranged from 0.09–0.35. (2) The correlation coefficient between the MSI data and MSI-derived covariates and soil EC was moderate, and the correlation between certain MSI data sets and soil EC was not significant. (3) The SVM soil EC estimation model established with the MSI data set attained a higher performance and accuracy ( $R^2 = 0.88$ , root mean square error (RMSE) = 4.89 dS m<sup>−1</sup>, and ratio of the performance to the interquartile range (RPIQ) = 1.96, standard error of the laboratory measurements to the standard error of the predictions (SEL/SEP) = 1.11) than those attained with the soil EC estimation models established with the RF and ANN models. (4) We applied the SVM soil EC estimation model to map the soil salinity in the study area, which showed that the farmland with higher altitudes discharged a large amount of salt to the surroundings due to long-term irrigation, and the secondary salinization of the farmland also caused a large amount of salt accumulation. This research provides a scientific basis for the simulation of soil salinization scenarios in arid areas in the future.

**Keywords:** soil salinization; Sentinel-2 MSI; remote sensing; machine learning; arid area



**Citation:** Wang, J.; Peng, J.; Li, H.; Yin, C.; Liu, W.; Wang, T.; Zhang, H. Soil Salinity Mapping Using Machine Learning Algorithms with the Sentinel-2 MSI in Arid Areas, China. *Remote Sens.* **2021**, *13*, 305. <https://doi.org/10.3390/rs13020305>

Received: 22 December 2020

Accepted: 14 January 2021

Published: 17 January 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 salinization is an important ecological and environmental problem in arid and semiarid regions globally, and it seriously affects ecological stability, regional ecology, food security, and sustainable agricultural development [1]. As a form of land degradation, soil salinization can accelerate the desertification process and cause the deterioration of the

ecological environment. Meanwhile, it also damages the functions of a series of ecological services, thus affecting human health [2]. Soil salinization directly affects soil characteristics, such as soil structure, soil microbial activity, etc., which in turn affects soil productivity and nutrient availability. At the same time, soil salinization also inhibits the absorption of water and nutrients by plants, thereby affecting physiology and biochemistry attributes of plants [3].

The timely and accurate acquisition of soil salinization information has an extremely important practical significance for the prevention and control of land degradation and ecological restoration in arid areas. Soil salinization monitoring is a basic task to reveal the occurrence, dynamics, and distribution of salinization [4]. Traditional soil salinization monitoring hardly obtains large-scale salinization distribution information, and it is difficult to monitor soil salinization dynamics on a large scale.

Currently, remote sensing data have been widely applied in soil salinization monitoring, and accurate soil salinity mapping is imperative, so research on mapping methods is particularly important.

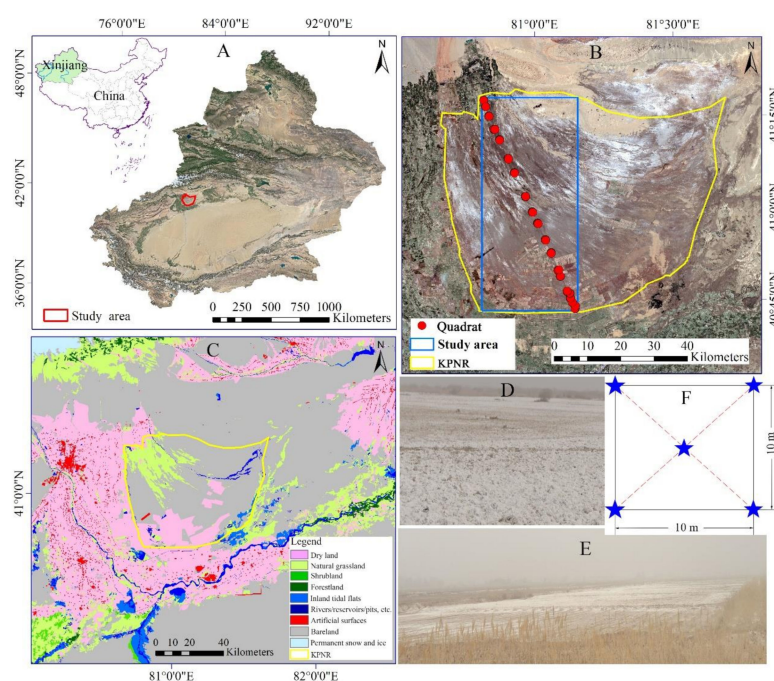
In recent years, satellite remote sensing data have played an important role in regional and even global soil salinization monitoring and mapping [5,6]. The Sentinel-2 satellite has a short revisit time, multiple wavebands, and a high spatial resolution, and it has been widely applied in resource monitoring, including soil salinization monitoring and mapping [7–9]. Davis et al. [10] compared the accuracy of the farmland soil salinity estimated with the MSI and Operational Land Imager (OLI) and found that these two sensors attain a similar salinity modeling performance, but the area of salinized land is overestimated with the OLI, and the area of salinized land covered by vegetation is underestimated; overall, due to the high spatial and temporal resolution of the MSI, it is superior to the OLI in terms of soil salinity tracking. Gorji et al. [11] used the OLI and MSI to conduct soil salinity mapping, and their results demonstrated that different salinity levels in different electrical conductivity (EC) ranges can be estimated through regression analysis of ground-measured data and satellite data. Farahmand et al. [12] evaluated the capability of various nonlinear regression models based on optical Sentinel-2 remote sensing images to estimate soil salinity. Their evaluation results confirmed that nonlinear regression models are superior to linear regression models in soil salinity estimation. It is necessary to use advanced technical methods for digital soil mapping, and there are many existing methods [13,14]. Different from statistical methods, machine learning algorithms are a branch of artificial intelligence that use learners to learn autonomously from data and then predict the results. Taghizadeh-Mehrjardi et al. used a statistical method and machine learning algorithm to predict the soil particle size fraction, and found that the ant colony optimization (ACO) had a higher accuracy [15]. Sahour et al. compared the accuracy of machine learning and statistical methods in groundwater salinity mapping, and found that extreme gradient boosting (EGB) algorithm had the best performance in the verification set [16]. Moreover, machine learning algorithms have also been applied in soil salt prediction. [17,18]. Xu et al. [19] proposed a new method for the simultaneous identification of the hyperparameters and input features of the support vector machine regression (SVR) algorithm based on an adaptive genetic algorithm for the quantitative evaluation of soil salinization. Hong et al. [20] used the Artificial Neural Network (ANN) algorithm and the SVR algorithm to estimate the soil salinity in the Yanqi Basin of Xinjiang.

Based on the advantages of machine learning algorithms that are easy to process high-dimensional data and have strong generalization capabilities, this study combines popular machine learning algorithms with Sentinel-2 MSI data and derivative parameters to evaluate and map soil salinity, which can prove that Sentinel-2 MSI and its derived variables have great application prospects in soil salinization monitoring and mapping, and can also prove that using machine learning algorithms has great potential in the prediction of soil EC. This research can provide a practical basis to achieve sustainable land use in arid areas.

## 2. Materials and Methods

### 2.1. Study Region

The Kongterik Pasture Nature Reserve (KPNR) in the Aksu Prefecture, Wensu County, is located on the northern margin of the Tarim Basin, the Xinjiang Uyghur Autonomous Region, China, between  $40^{\circ}46' \text{N}$ – $41^{\circ}15' \text{N}$  and  $80^{\circ}40' \text{E}$ – $81^{\circ}29' \text{E}$  (Figure 1A). The total area of the KPNR is  $6063.84 \text{ km}^2$ , and its altitude ranges from 922 to 1207 m above sea level, gradually decreasing from northwest to southeast. The area exhibits a sparse precipitation, intense evaporation, and extreme aridity. It has a typical continental arid climate with an average annual temperature of  $10.10^{\circ}\text{C}$ , an average annual precipitation of 65.4 mm, and an average annual evaporation of 2300 mm. Because of its relatively flat terrain, shallow groundwater burial depth, and high ratio of evaporation to precipitation, salt accumulates on the surface with water movement, resulting in soil salinization (Figure 1D). In addition, due to the influence of human activities, secondary soil salinization also occurs in the area (Figure 1E). Therefore, the main natural vegetation in the area mainly comprises halophytes such as *Tamarix chinensis* Lour., *Halocnemum strobilaceum*, *Halostachys caspica*, *Phragmites communis*, *Glycyrrhiza uralensis* Fisch, *Kareliniacaspia*, and *Kalidium foliatum*. The KPNR contains a large area of saline soil, which is a typical desert oasis transition zone and ecological degradation zone (Figure 1C). Choosing this area as the research area has a good representation value, which is of great importance to improve the ecological environment and the development of agricultural production.



**Figure 1.** Schematic diagram of the sampling points in the study area. (A) Location map of the Kongterik Pasture Nature Reserve (KPNR); (B) distribution map of sample quadrat in the study area; (C) the present land-use map of the KPNR; (D) the native soil salinization landscape in the study area; (E) the secondary soil salinization landscape in the study area; (F) a schematic diagram of the collection method of mixed soil samples.

### 2.2. Sample Collection and Analysis

A field survey and soil sample collection were performed on 14 June 2019, which coincided with the transit time of the Sentinel-2A satellite. Because there is only one highway in the entire study area, the route of the investigation process was designed based on the accessibility of potential field investigation sites. According to the local soil salt content determined in previous field investigations, local digital soil map, surface salinity characteristics, and land use/cover on remote sensing images (Figure 1B), 160 soil sampling

quadrats were established throughout the study area, and the size of each quadrat was designed to be 10 m × 10 m. With the use of the five-point sampling method, soil samples (from 0 to 20 cm) were collected at the 4 corners and the center of each plot and mixed into one mixed sample (Figure 1F). Moreover, a portable GPS instrument (Trimble JUNO, positioning accuracy ≤ 5 m) was employed to record the geographic positions. Although the positioning accuracy of the GPS instrument was insufficient, this did not affect the position alignment between the remote sensing images and sampling quadrats (since the image resolution is 10 m). All collected soil samples were transported to the laboratory to determine the moisture content and conductivity. The fresh soil samples (20.00 g) were weighed and placed in a drying cabinet at 105 °C ± 2 °C and dried to a constant weight to calculate the soil moisture content. An amount of 20.00 g from each natural air-dried soil sample was weighed to prepare a soil extract at a soil-water ratio of 1:5, and its conductivity was measured after filtration.

### 2.3. Source of the Remote Sensing Data and Their Preprocessing

Multispectral remote sensing data have been widely applied in soil salinization monitoring because of their large coverage area, easy access, and suitable spatial and spectral resolutions [21]. However, many studies have tended to use images with high spectral and spatial resolutions to obtain suitable results [22]. The launch of the Sentinel-2 satellite was the result of the joint cooperation between the European Space Agency, the European Commission, the industry, service providers, and data users [23]. Sentinel-2 data exhibit many of the technical characteristics of Landsat series data with a more frequent 5-day revisit cycle [24]. The Sentinel-2 satellite is equipped with the most advanced MSI instrument that provides high-resolution optical images. The MSI instrument of the Sentinel-2 satellite yields 4 image bands with a spatial resolution of 10 m (B2, B3, B4, and B8), 6 image bands with a spatial resolution of 20 m (B5, B6, B7, B8a, B11, and B12), and 3 image bands with a spatial resolution of 60 m (B1, B9, and B10). The relevant parameters have been described in many studies and are not provided in detail in this study [25]. In accordance with the timing of the ground survey, this study selected the MSI data of Sentinel-2B on 14 June 2019. The acquired Sentinel-2B data are reflectance data of the top of the atmosphere (TOA) at the level-1C (L1C) processing level. The L1C MSI data were converted into level-1A (L1A) MSI data with the Sen2Cor algorithm to assess the soil salinity. In particular, after atmospheric correction, the top of the atmosphere reflectance was converted into the bottom of the atmosphere or Earth surface reflectance. Four image bands with a resolution of 10 m and six image bands with a resolution of 20 m were adopted in this study. The images in the 10 m bands were resized to a 20 m pixel size, and these images were then stacked with SNAP software and clipped to obtain a subset of the study area.

### 2.4. Data Processing Method

In this study, the Sentinel-2B data were processed using three different methods to obtain various modeling factors, including 10 selected bands (after atmospheric correction), 3 bands generated after principal component analysis of the 10 selected bands, and various spectral indices constructed with these 10 bands. In addition, DEM were included as modeling factors.

#### 2.4.1. Modeling Factors

In arid regions, the spectral index is a common and effective method of soil salinity monitoring [26]. The salt spectral index was proposed based on local environmental conditions and cannot be described separately from local conditions [27]. In this study, specific satellite salinity indices were selected, and these salinity indices were screened or combined to construct a highly robust salinity index model. In addition, in this study, the original band reflectance images, the first three bands of principal component (PC) transformation, the terrain index, the tasseled cap transformation-derived wetness (TCW) [28] index, and the vegetation index (VI) were also selected (Table 1).

**Table 1.** Modeling indices used and their calculation equations.

Modeling Indices	Acronym	Equation	Reference
Resampled original band reflectance images	B2-Blue, B3-Green, B4-Red, B5-Rededge1, B6-Rededge2, B7-Rededge3, B8-NIR, B8a-Rededge4, B11-SWIR1, B12-SWIR2	The central wavelengths are 492.1 nm, 559 nm, 665 nm, 703.8 nm, 739.1 nm, 779.7 nm, 833 nm, 864 nm, 1610.4 nm, and 2185.7 nm, respectively.	
First three bands of principal component (PC) transformation	PC1, PC2, PC3	Sentinel-2B 10-m resolution and 20-m resolution images are resampled to a 20-m resolution and then subjected to principal component transformation.	
Normalized difference salinity index	NDSI	$(R - NIR)/(NIR + R)$	[29]
Salinity index	S1	$B/R$	
	S2	$(B - R)/(B + R)$	
	S3	$(G \times R)/B$	
	S5	$(B \times R)/G$	
	S6	$(R \times NIR)/G$	
	SI	$(B + R)0.5$	[30]
	SI1	$(G \times R)0.5$	
	SI2	$[(G)^2 + (R)^2 + (NIR)^2]0.5$	
	SI3	$[(R)^2 + (G)^2]0.5$	
	SI4	$(B \times R)0.5$	
Intensity index 1	Int1	$(G + R)/2$	[31]
Intensity index 2	Int2	$(G + R + NIR)/2$	
Vegetation Index	NDVI	$(NIR - R)/(NIR + R)$	
	EVI	$2.5 \times [(NIR - R)/(NIR + 6 \times R - 7.5 \times B + 1)]$	
	CRSI	$[(R \times NIR) - (B \times G)]/[(R \times NIR) + (B \times G)]$	
	RVI	$NIR/R$	
	SAVI	$(1 + L)[(NIR - R)/(NIR + R + L)]$	
	GDVI	$(NIR^n - R^n)/(NIR^n + R^n), n = 2$	
Tasseled cap wetness	TCW	$0.1509 \times B + 0.1973 \times G + 0.3272 \times R + 0.3406 \times NIR - 0.7112 \times SWIR1 - 0.4573 \times SWIR2$	[32]

Note: B = B2 (492.1 nm), G = B3 (559 nm), R = B4 (665 nm), NIR = B8 (833 nm), SWIR1 = B11 (1610.4 nm), SWIR2 = B12 (2185.7 nm).

#### 2.4.2. Modeling Methods and Accuracy Verification

In this study, the total data set ( $n = 160$ ) was divided into a modeling set (112 soil samples, 70% of the total soil samples) and a verification set (48 soil samples, 30% of the total soil samples) by Kennard–Stone (K–S) algorithm. In the total data set, according to the sampling order, one sample was selected every four samples as a verification sample. Three modeling methods were applied to evaluate the soil salinity in the study area, namely, Support Vector Machines (SVM), Random Forest (RF) algorithm, and ANN algorithm. When establishing the soil EC estimation model, according to the principle of minimum mean square error of cross-validation ( $RMSE_{CV}$ ), the kernel function selected by SVM was a polynomial, the penalty parameter  $C$  was 6, the regression accuracy  $\epsilon$  was 0.1, and the  $\gamma$  value was 2.0. The ANN model selected a multi-layer perceptron (MLP), set a hidden layer, and 30 hidden layer nodes. The number of decision trees  $N$  of the RF model was 100, the feature variable  $K$  selected each time was 34, the maximum tree depth  $D$  was 10, and the minimum child node size was 5. The K–S algorithm and three modeling algorithms were implemented in matlab2016a.



Four basic parameters were considered to evaluate the model: The determination coefficient ( $R^2$ ), root mean square error (RMSE), ratio of the performance to the interquartile range (RPIQ), and ratio of the standard error of the laboratory measurements to the standard error of the predictions (SEL/SEP).

### 3. Results

#### 3.1. Statistics of the Soil Sample EC Values and Sentinel-2B Reflectance Data

##### 3.1.1. Descriptive Statistics of the Soil Samples Electrical Conductivity (EC) Values

The total salinity data set was divided into two parts: One was the modeling set, accounting for 70% of the total data set, and the other was the verification set, accounting for 30% of the total data set (Table 2).

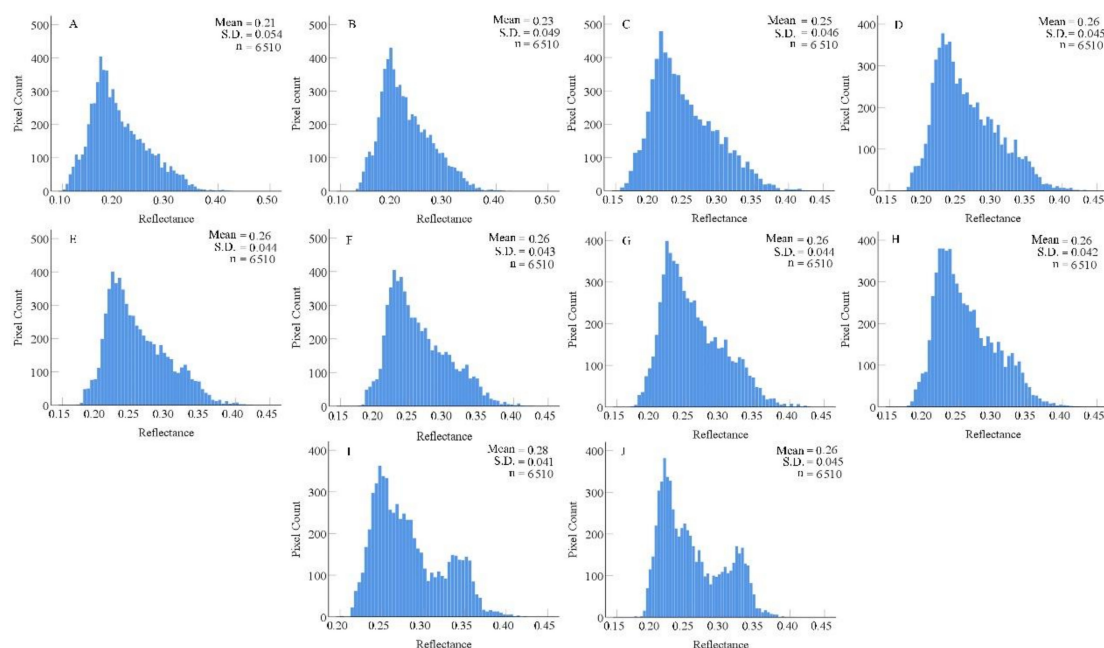
**Table 2.** Descriptive statistics of the total, modeling, and verification soil EC data sets.

Data Set	<i>n</i>	Mean	Min.	Max.	S.D.	C.V.
Total data set	160	24.03	1.07	79.6	10.70	44.53
Modeling data set	112	23.86	1.07	79.6	10.65	44.64
Verification data set	48	24.72	6.32	64.65	10.64	43.04

Note: Table 2 provides the data range of the three data sets, and the standard deviation and coefficient of variation are consistent. In the total data set, the soil EC ranges from 1.07 to 79.6  $\text{dS}\cdot\text{m}^{-1}$ , the standard deviation is 10.70  $\text{dS}\cdot\text{m}^{-1}$ , and the coefficient of variation is 44.53%, which is a moderate variation. The data range of the modeling set is consistent with that of the total data set, while the data range of the validation set is a subset of their ranges, i.e., included in their ranges, which is 6.32–64.65  $\text{dS}\cdot\text{m}^{-1}$ . The standard deviation and coefficient of variation of the three data sets are not significantly different. The standard deviations of the modeling set and the validation set are 10.65  $\text{dS}\cdot\text{m}^{-1}$  and 10.64  $\text{dS}\cdot\text{m}^{-1}$ , respectively, and the coefficients of variation are 44.64% and 43.04%, respectively. The above statistics demonstrate that the division of data sets meets the modeling conditions.

##### 3.1.2. Descriptive Statistics of the Soil Sample Sentinel-2B Reflectance Data

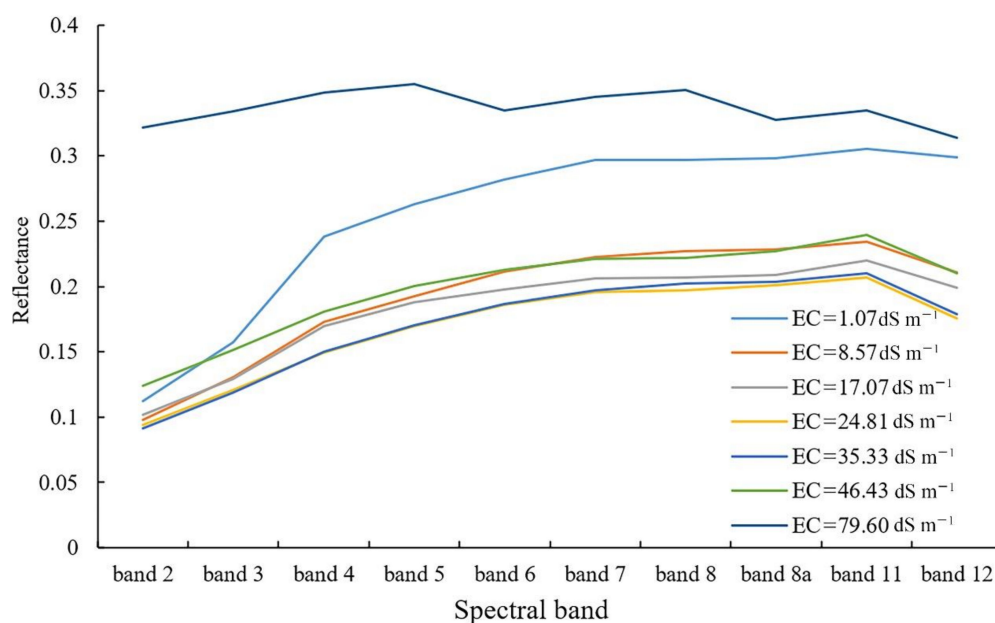
To detect the characteristics of the spectral bands of the Sentinel-2 images, 6510 random pixels (not including vegetation and water mask pixels) were selected in each band, and a statistical analysis of the pixel distribution characteristics was performed (Figure 2).



**Figure 2.** Pixel count statistics of 6510 random soil pixels in the Sentinel-2 MSI spectral band. (A) Band 2-Blue; (B) band 3-Green; (C) band 4-Red; (D) band 5-Rededge1; (E) band 6-Rededge2; (F) band 7-Rededge3; (G) band 8-NIR; (H) band 8a-red edge4; (I) band 11-SWIR; (J) band 12-SWIR.

Figure 2 shows that the reflectance values in all bands range from 0 to 1, and the values in each band are relatively similar. The mean reflectance is between 0.21 and 0.28, and the standard deviation is between 0.041 and 0.054.

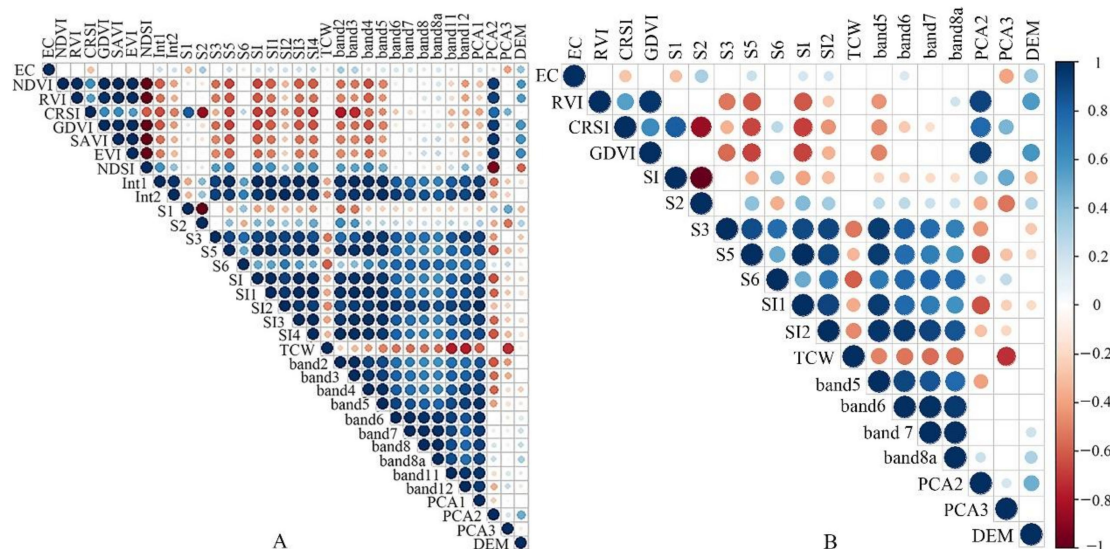
Seven representative soil samples were selected with different salinity levels to analyze the corresponding reflectance characteristics in the Sentinel-2 band (Figure 3). The reflectance curves of these soil samples were similar in shape. The soil samples with an EC of  $79.60 \text{ dS}\cdot\text{m}^{-1}$  and the soil samples with an EC of  $1.07 \text{ dS}\cdot\text{m}^{-1}$  attained a higher reflectance than the other soil samples. Among them, the reflectance of the soil sample with an EC of  $79.60 \text{ dS}\cdot\text{m}^{-1}$  was between 0.32 and 0.36, which was the highest value. The reflectance of the soil sample with an EC of  $1.07 \text{ dS}\cdot\text{m}^{-1}$  was between 0.12 and 0.31. The reflectance curve of the soil samples with an EC ranging from  $8.57 \text{ dS}\cdot\text{m}^{-1}$  to  $79.60 \text{ dS}\cdot\text{m}^{-1}$  was more concentrated, and the reflectance in the 10 wavebands was low. In this study, the soil samples with the highest and lowest EC values did not correspond to the highest- and lowest-reflectance curves, respectively, which should be closely related to the soil moisture content and salt composition [33,34].



**Figure 3.** Reflectance of the soil samples with different salinity levels in the Sentinel-2 spectral bands.

To examine the sensitivity of the Sentinel-2B MSI-derived covariates (spectral bands, PC image, vegetation index, TCW, DEM, and satellite salinity indices) to the soil EC, Pearson correlation analysis was performed, and a correlogram was established (Figure 4). As shown in Figure 4A, there was a significant statistical correlation between the 35 covariates generated from the Sentinel-2 MSI data and soil EC. Seven spectral indices, namely, NDVI, RVI, GDVI, SAVI, EVI, NDSI, and B12, failed the significance test ( $p < 0.05$ ). In this study, PCA2 and PCA3 attained the highest correlation with the soil EC, while SSM exhibited the strongest relationship with S3 and B12, rather than with the soil EC. We found that although there was a statistically significant correlation between the measured soil salinity and TCW, the correlation coefficient was not the highest. In particular, the correlation between the soil salinity and the surface soil moisture index was low in this study area. In addition, good correlations between the soil EC and nine spectral bands were observed. In general, most of the MSI-derived covariates exhibited significant correlations with the soil EC in the study area (Figure 4A). Figure 4A showed that the correlation coefficients between many factors are very high, which indicates that there is multicollinearity among factors, and multicollinearity will increase the variance of the regression coefficients and make the established model unstable. Therefore, by calculating the variance inflation factor (VIF), we screened the variables and selected the variables with  $1 < \text{VIF} < 10$ , thereby

reducing the multicollinearity among the factors (Figure 4B). Finally, 18 variables were selected to establish the soil salinity estimation model for improving the accuracy and stability of the model.



**Figure 4.** Correlation coefficients between the laboratory-measured soil electrical conductivity (EC) values and Sentinel-2B MSI-derived covariates based on 160 samples. (A) Correlogram of all factors; (B) correlogram of selected factors.

### 3.2. Construction of the Optimal Soil EC Estimation Models

Based on Figure 4, the original Sentinel-2B MSI images, their derived features (e.g., satellite salinity indices, vegetation index, principal component factors, and TCW) and the DEM were adopted as RS data sources (covariates) to estimate the soil EC. With the use of 18 spectral parameters as the independent variables required by the soil EC prediction model and with the soil EC data as the dependent variables, three machine learning algorithm estimation models were constructed with ANN, RF, and SVM (Table 3).

**Table 3.** Accuracy statistics of the Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Machines (SVM) soil EC estimation models. RMSE: Root mean square error; RPIQ: Ratio of the performance to the interquartile range; SEL/SEP: Standard error of the laboratory measurements to the standard error of the predictions.

Modeling Method	$R_m^2$	Modeling			$R_v^2$	Verification		
		RMSE	RPIQ	SEL/SEP		RMSE	RPIQ	SEL/SEP
SVM	0.71	5.78	1.75	1.26	0.88	4.89	1.96	1.11
RF	0.81	4.67	1.85	1.42	0.27	10.61	0.65	1.79
ANN	0.80	4.53	2.06	1.27	0.57	8.15	1.26	1.34

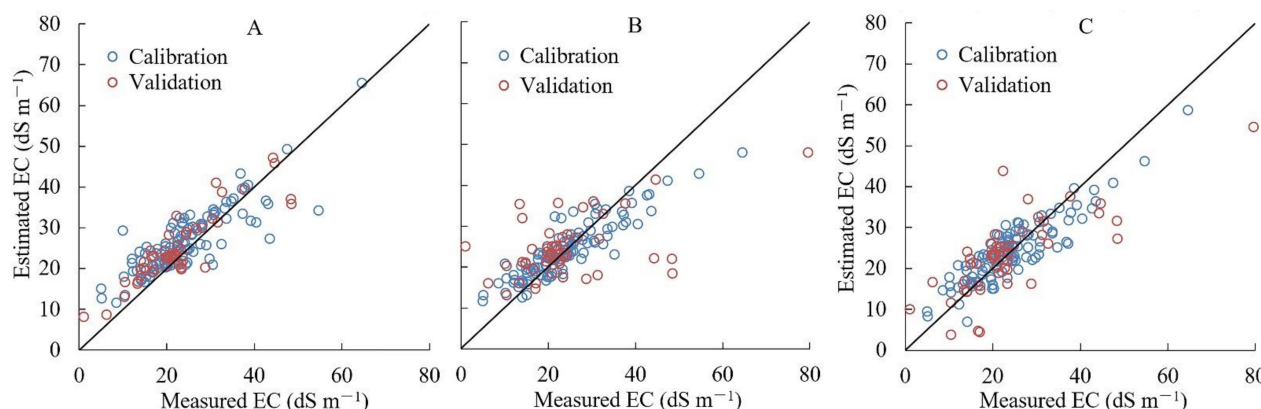
18 spectral variables were selected to establish soil EC models. To evaluate the modeling effect and accuracy, the predicted soil EC based on SVM, RF, and ANN was validated against the measured soil EC. Four parameters, namely,  $R^2$ , RMSE, RPIQ, and SEL/SEP, were considered for evaluation in this study. Among them, the  $R^2$  value is directly proportional to the model accuracy. The closer the  $R^2$  value is to 1, the higher the model fitting accuracy is. The RMSE value is inversely proportional to the accuracy of the model. The closer this value is to 0, the lower the deviation between the measured value and the model-predicted value is, and the stronger the prediction ability is. The RPIQ is the ratio of the interquartile range to the RMSE. The interquartile range is the difference between the 75% and 25% sample values. It is generally accepted that  $RPIQ < 1.7$  indicates



a low model reliability,  $1.7 \leq \text{RPIQ} < 2.2$  indicates that the model exhibits a relatively balanced predictive ability, and  $\text{RPIQ} \geq 2.2$  indicates that the model achieves an excellent predictive effect [35]. The ideal value of SEP/SEL is 1, which indicates that the variability in the predicted values is equal to the variability in the measured values, and the farther the SEP/SEL value is from 1, the higher the variability between the predicted and measured values is [36].

The statistical results (Table 3) of the model parameters showed that among the modeling data sets obtained with these three models, in regard to the RF model, the  $R^2$  value was the highest, and the RMSE was relatively low, at 0.81 and 4.67, respectively, while in the SVM model, the  $R^2$  value was the lowest, at 0.71, the RPIQ was 1.75, and the SEL/SEP was the closest to 1. Neither the RF model nor the ANN model satisfied the modeling requirements. Among the validation sets of the three models, in regard to the SVM model, the  $R^2$  value was 0.88, which was the highest value, and the RMSE was the lowest, at 4.89. Moreover, the RPIQ was between 1.7 and 2.2, and SEL/SEP was also the closest to 1. Therefore, Table 3 indicates that the SVM model is the most robust model among the three models.

In regard to the SVM model, RF model, and ANN model, there were obvious outliers in the estimated values of the soil samples based on EC ( $20\text{--}50 \text{ dS m}^{-1}$ ). They occurred on both sides of the 1:1 line, and these points were relatively discrete (Figure 5). The estimated data points obtained with the SVM model were more concentrated than those obtained with the RF model and the ANN model (Figure 5A).



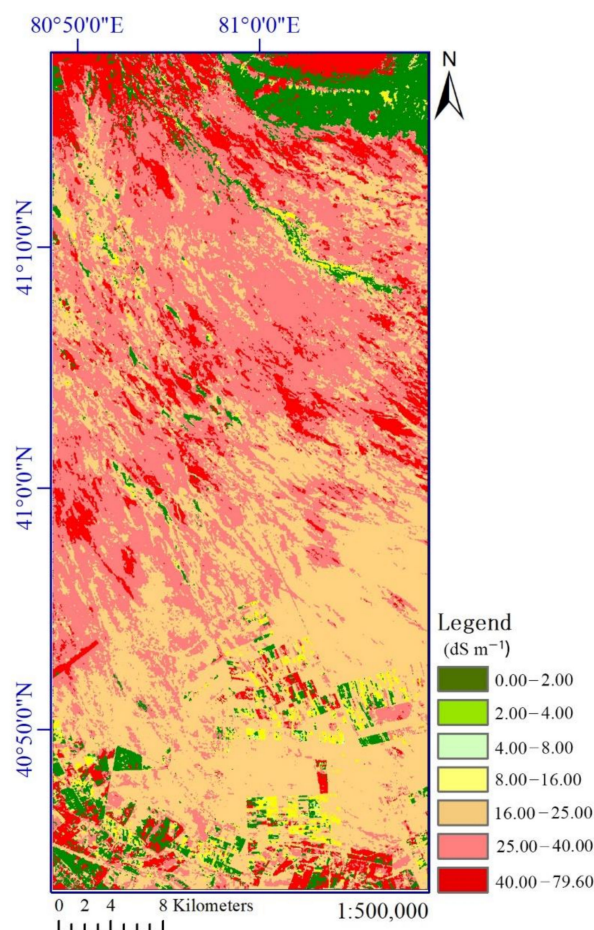
**Figure 5.** Scatter plots of the measured and estimated soil EC values derived from the SVM, RF, and ANN regression models using the Sentinel-2 MSI data. (A) SVM model, (B) RF model, (C) ANN model. The black solid line represents the line where the ratio of the measured values to the estimated values is 1:1.

### 3.3. Soil EC Mapping Based on the Optimal Estimation Models

Based on the RS data sets (Sentinel-2B MSI) and corresponding SVM models, we generated a soil EC distribution map of the KPNR (Figure 6).

The soil EC distribution map (Figure 6) highlights those areas with a continuous distribution of saline soils. For further analysis and visualization, a commonly used soil salinity classification scheme was adopted to classify the soil salinity levels of the predicted images (Schoeneberger et al., 2002): Non-saline ( $0 \text{ dS} \cdot \text{m}^{-1} < \text{EC} \leq 2 \text{ dS} \cdot \text{m}^{-1}$ ), very slightly saline ( $2 \text{ dS} \cdot \text{m}^{-1} < \text{EC} \leq 4 \text{ dS} \cdot \text{m}^{-1}$ ), slightly saline ( $4 \text{ dS} \cdot \text{m}^{-1} < \text{EC} \leq 8 \text{ dS} \cdot \text{m}^{-1}$ ), moderately saline ( $8 \text{ dS} \cdot \text{m}^{-1} < \text{EC} \leq 16 \text{ dS} \cdot \text{m}^{-1}$ ), and strongly saline ( $>16 \text{ dS} \cdot \text{m}^{-1}$ ). Figure 6 shows that the area with a strong salinity ( $>16 \text{ dS} \cdot \text{m}^{-1}$ ) occupied the majority of the study sample area. The areas with the highest salinity occurred in the northwest area with the highest altitude and the southern area with the lowest altitude in the study area. Most of these landscapes are flat. The northwestern area with high elevations is located in the upper and middle parts of the alluvial fan. As a large amount of salt is discharged from the cultivated

land areas in the upper and middle parts of the alluvial fan, the discharged salt flows to the surrounding area with a low altitude. Hence, high soil salinization occurs in the surrounding area of cultivated land. In the southern low-elevation area, land cultivation leads to a shallow groundwater depth, which causes serious secondary soil salinization when the land is left uncultivated. Soils with an EC of  $0\text{--}8\text{ dS}\cdot\text{m}^{-1}$  (non-saline, very slightly saline, and slightly saline soils) were mainly distributed in the cultivated farmlands and areas with relatively large topographic changes, such as parts of the northeast and south of the study area. Soils with an EC ranging from  $8\text{ dS}\cdot\text{m}^{-1}\text{--}16\text{ dS}\cdot\text{m}^{-1}$  (moderately saline soils) largely occurred in some abandoned farmlands in the southern part of the study area.



**Figure 6.** Soil EC distribution map derived from the SVM models based on the Sentinel-2B MSI data.

#### 4. Discussion

##### 4.1. Soil Salinity Detection Based on the Sentinel-2 MSI Data

The Sentinel 2 multispectral sensor is similar to other multispectral sensors in that it uses the spectral information reflected by ground objects to detect useful geographic information [37]. Soils with different salinities have different spectral characteristics, which is the basis of the remote sensing monitoring of soil salinization. The area covered by a white salt crust has a high salt content. However, in each band of the Sentinel-2 MSI data, the spectral reflectance of the soil samples did not necessarily increase with increasing soil salinity (Figure 3). This makes it difficult to directly use multispectral bands and their derived spectral indices to monitor and map the soil surface salinity. According to previous studies, the salinity index and vegetation index were used to estimate the soil salinity [38,39]. Due to differences in geographic location, topography, and vegetation types, the soil salinity under vegetation cover varied greatly, ranging from non-saline soil to heavily saline soil [40,41]. However, in many previous studies, regions with

vegetation coverage were directly identified as non-salinized regions or slightly salinized regions [42,43]. MSI data with a high spatial resolution contain few mixed pixels, which reduces the impact of the above issue (Figure 5). Therefore, in this study, we did not mask the vegetation coverage area before modeling, and we also collected samples in vegetated areas to use their spectral parameters to model and estimate the soil salinity and obtain the true soil salinity in the vegetation coverage area. The vegetation cover and the soil index are indispensable environmental variables for soil salinization monitoring, and these variables change with the environmental conditions. Therefore, environmental information reflecting changes in soil properties such as vegetation cover, phenology, and plant growth should be carefully considered.

#### *4.2. Accuracy of the Soil Salt Estimation Model Based on the Spectral Variables*

The key to the successfully inversion of the soil salt content using spectral variables is to choose an effective mathematical regression model. Algorithms such as MLR, PLSR, and BP neural networks have been widely applied in the inversion and modeling of soil component contents [44,45]. Machine learning has the ability of autonomous learning and can solve the problem of complex nonlinear function approximation in soil salinization monitoring. Wang et al. [24] compared the accuracy of the OLI and MSI in soil salinity mapping. The  $R^2$  value of the MSI-based soil EC estimation model reached 0.912, while the  $R^2$  value of the estimated model in this study was only 0.783, which mainly occurred due to the difference in the number of samples. The former study had only 64 samples, while in this study, 160 samples were used for modeling. Therefore,  $R^2$  is low, but the soil salinity mapping in this study should be more realistic and objective. The performance of the SVM soil salinity estimation model is better than that of the ANN model and the RF model, which may be due to their own algorithm characteristics. SVM is a small sample learning method with solid theoretical foundation. It is based on the principle of structural risk minimization, which ensures that the learning machine has good generalization ability. By introducing kernel function, the global optimality of the algorithm is guaranteed, and the empirical component in the neural network is avoided. ANN is a learning method based on statistics. Its performance depends on the number of samples in the model training process, and in most cases, the number of samples is limited. A large amount of sample data with different value ranges will influence the RF model. If the value range is small, the variance will be small and the offset will be large, making the model precision on the training set much higher than that on the test set. In this study, there are 34 variables and 160 soil samples. In terms of the number of samples, the SVM model has more advantages than ANN model. Due to the large number of variables and the small value range of some variables (such as 10 bands of MSI), the accuracy of the RF model is also greatly affected. Therefore, the SVM soil salinity estimation model has the best performance among the three models.

Based on 18 variables and 3 machine learning algorithms, 3 soil salinity estimation models were established in this study. It was found that only the SVM model meets the accuracy requirements and can be used for the quantitative inversion of the soil EC. Xing et al. [46] proposed a data-driven model based on the support vector machine to predict the daily soil temperature in different climates at the continental scale with a relatively high accuracy. Zhang et al. [47] used a combination of partial least squares (PLS), multiple linear regression (MLR), and support vector machine (SVM) to establish a prediction model for the soil organic matter, total nitrogen, total phosphorus, and total potassium contents. Their results revealed that the SPA-SVM model attains the best applicability for all soil nutrient contents. Jiang et al. [20] compared the performance of soil electrical conductivity (EC) estimation models established by support vector machines and artificial neural networks. Their results showed that the support vector machine regression algorithm is superior to the artificial neural network algorithm in soil salinity monitoring. The SVM is a nonlinear model estimation method, and its accurate estimation effect has been verified [48,49]. On this basis, other methods, such as deep learning and gene

expression programming, can also be applied, or other factors related to soil salt transport, such as the temperature vegetation dryness index (TVDI) and surface temperature ( $T_s$ ), can be included to further improve the accuracy of soil salinity estimation.

#### 4.3. Uncertainty Analysis of Soil Salinity Mapping Based on the Sentinel-2 MSI Data

Uncertainty is an important problem in soil property mapping. In this study, there are two main aspects of the uncertainty: One is the uncertainty of the model, and the other is the uncertainty of the relationship between the soil salinity data and MSI data. In this study, mixed soil samples from 0 cm–20 cm below the surface were collected according to the usual sampling principles [26,50]. However, were the spectral variables indicating salinity characteristics obtained from the MSI data suitable to reflect the EC value of the 0–20 cm mixed soil samples? The data could be more suitable to reflect the EC value of 0–5 cm or 0–10 cm mixed soil samples. These spectral variables (salinity index, vegetation index, etc.) are affected by many environmental factors, such as soil organic matter, soil moisture, soil surface roughness, and soil metal mineral content. Moreover, even if the MSI data were subjected to geometric correction and atmospheric correction, the images would still be affected by the terrain conditions and shadows. The sample size is not large enough, which may also lead to potential uncertainties. In future research, we will increase the number of samples and sampling points, and choose more sampling depths to reduce the uncertainty of soil salt prediction.

It should be pointed out that the inversion capability of a single satellite image is always limited. We could apply multiple satellites, scales, and spectral dimensions to map soil properties to achieve more accurate prediction results [50,51]. Finally, combining the classic theory of soil science and remote sensing with data mining algorithms used for big data analysis is essential for better soil salinity mapping.

## 5. Conclusions

In this study, we analyzed the spectral characteristics of MSI images, established SVM, RF, and ANN soil EC estimation models, and verified the performance of each model. Moreover, we conducted soil EC mapping in the study area. The main conclusions are as follows:

1. The average reflectance of each band of the MSI data ranges from 0.21–0.28. According to the spectral characteristics corresponding to the different soil EC levels, the spectral reflectance of salinized soil in the MSI data ranges from 0.09–0.35.
2. In general, the correlation coefficient between the MSI data and MSI-derived covariates and soil EC was moderate, and the correlation between certain MSI data sets and soil EC was not significant.
3. The SVM soil EC estimation model established with the MSI data set attained a better performance and accuracy than those attained with the soil EC estimation models established with the RF and ANN models.
4. We applied the SVM soil EC estimation model to map the soil salinity in the study area, which provides a scientific basis for the simulation of soil salinization scenarios in arid areas in the future.

**Author Contributions:** Conceptualization, J.W., J.P., and T.W.; investigation, C.Y., W.L. and H.Z.; methodology, J.W., H.L. and T.W.; formal analysis, J.W.; resources, C.Y.; writing—original draft, J.W.; writing—review and editing, J.P.; funding acquisition, J.P. and H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by grants from the Young and middle-aged Innovative Leaders Program of Xinjiang Production and Construction Corps (Grant Nos. 2020CB032), the National Key Research and Development Program of China (Grant Nos. 2018YFE0107000), and the National Science Foundation of China (Grant Nos. 42071068 and 31860172).

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

**Informed Consent Statement:** Not applicable.



**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

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

## References

1. Xiao, Y.; Zhao, G.; Li, T.; Zhou, X.; Li, J. Soil salinization of cultivated land in Shandong Province, China—Dynamics during the past 40 years. *Land Degrad. Dev.* **2019**, *30*, 426–436. [\[CrossRef\]](#)
2. Litalien, A.; Zeeb, B. Curing the earth: A review of anthropogenic soil salinization and plant-based strategies for sustainable mitigation. *Sci. Total Environ.* **2020**, *698*, 134235. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Hafez, E.M.; Omara, A.E.D.; Alhumaydhi, F.A.; El-Esawi, M.A. Minimizing hazard impacts of soil salinity and water stress on wheat plants by soil application of vermicompost and biochar. *Physiol. Plant.* **2020**. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Yang, J.; Zhao, J.; Zhu, G.; Wang, Y.; Ma, X.; Wang, J.; Guo, H.; Zhang, Y. Soil salinization in the oasis areas of downstream inland rivers—Case Study: Minqin oasis. *Quat. Int.* **2020**, *537*, 69–78. [\[CrossRef\]](#)
5. Wang, X.; Zhang, F.; Ding, J.; 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. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Guo, B.; Yang, F.; Han, B.; Fan, Y.; Jiang, L. A model for the rapid monitoring of soil salinization in the yellow river delta using landsat 8 OLI imagery based on VI-SI feature space. *Remote Sens. Lett.* **2019**, *10*, 796–805. [\[CrossRef\]](#)
7. Gašparović, M.; Jogun, T. The effect of fusing sentinel-2 bands on land-cover classification. *Int. J. Remote Sens.* **2018**, *39*, 822–841. [\[CrossRef\]](#)
8. Wang, J.; Ding, J.; Yu, D.; Ma, X.; 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. [\[CrossRef\]](#)
9. Filho, M.G.; Kuplich, T.M.; Quadros, F.L.F.D. Estimating natural grassland biomass by vegetation indices using sentinel 2 remote sensing data. *Int. J. Remote Sens.* **2020**, *41*, 2861–2876. [\[CrossRef\]](#)
10. Davis, E.; Wang, C.; Dow, K. Comparing Sentinel-2 MSI and Landsat 8 OLI in soil salinity detection: A case study of agricultural lands in coastal North Carolina. *Int. J. Remote Sens.* **2019**, *40*, 6134–6153. [\[CrossRef\]](#)
11. 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. [\[CrossRef\]](#)
12. Farahmand, N.; Sadeghi, V.; Farahmand, S. Estimating soil salinity in the dried lake bed of Urmia Lake using optical Sentinel-2b images and multivariate linear regression models. *J. Indian Soc. Remote Sens.* **2020**, *48*, 675–687. [\[CrossRef\]](#)
13. Mulder, V.L.; Bruin, S.D.; Schaepman, M.E.; Mayr, T.R. The use of remote sensing in soil and terrain mapping—A review. *Geoderma* **2011**, *162*, 1–19. [\[CrossRef\]](#)
14. Van Zijl, G.; Van Tol, J.; Bouwer, D.; Lorentz, S.; Roux, P.L. Combining historical remote sensing, digital soil mapping and hydrological modelling to produce solutions for infrastructure damage in Cosmo City, South Africa. *Remote Sens.* **2020**, *12*, 433. [\[CrossRef\]](#)
15. Taghizadeh-Mehrjardi, R.; Toomanian, N.; Khavaninzadeh, A.R.; Jafari, A.; Triantafyllis, J. Predicting and mapping of soil particle-size fractions with adaptive neuro-fuzzy inference and ant colony optimization in central Iran. *Eur. J. Soil Sci.* **2016**, *67*, 707–725. [\[CrossRef\]](#)
16. Sahour, H.; Gholami, V.; Vazifedan, M. A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer. *J. Hydrol.* **2020**, *591*, 125321. [\[CrossRef\]](#)
17. Zeng, W.; Zhang, D.; Fang, Y.; Wu, J.; Huang, J. Comparison of partial least square regression, support vector machine, and deep-learning techniques for estimating soil salinity from hyperspectral data. *J. Appl. Remote Sens.* **2018**, *12*, 022204. [\[CrossRef\]](#)
18. Cao, X.; Ding, J.; Ge, X.; Wang, J. Estimation of soil electrical conductivity based on spectral index and machine learning algorithm. *Acta Pedol. Sin.* **2020**, *57*, 867–877. (In Chinese) [\[CrossRef\]](#)
19. Xu, H.; Chen, C.; Zheng, H.; Luo, G.; Yang, L.; Wang, W.; Wu, S.; Ding, J. AGA-SVR-based selection of feature subsets and optimization of parameter in regional soil salinization monitoring. *Int. J. Remote Sens.* **2020**, *41*, 4470–4495. [\[CrossRef\]](#)
20. Jiang, H.; Rusuli, Y.; Amuti, T.; He, Q. Quantitative assessment of soil salinity using multi-source remote sensing data based on the support vector machine and artificial neural network. *Int. J. Remote Sens.* **2019**, *40*, 284–306. [\[CrossRef\]](#)
21. Setia, R.; Lewis, M.; Marschner, P.; Segaran, R.R.; Chittleborough, D. Severity of salinity accurately detected and classified on a paddock scale with high resolution multispectral satellite imagery. *Land Degrad. Dev.* **2013**, *24*, 375–384. [\[CrossRef\]](#)
22. Ramos, T.B.; Castanheira, N.; Oliveira, A.R.; Paz, A.M.; Darouich, H.; Simionesei, L.; Farzaman, M.; Gonçalves, M.C. Soil salinity assessment using vegetation indices derived from sentinel-2 multispectral data. Application to Lezíria Grande, Portugal. *Agric. Water Manag.* **2020**, *241*, 106387. [\[CrossRef\]](#)
23. Abderrazak, B.; Ali, E.B.; Rachid, B.; Hassan, R. Sentinel-MSI VNIR and SWIR bands sensitivity analysis for soil salinity discrimination in an arid landscape. *Remote Sens.* **2018**, *10*, 855. [\[CrossRef\]](#)
24. 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. [\[CrossRef\]](#) [\[PubMed\]](#)

25. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [\[CrossRef\]](#)
26. Peng, J.; Biswas, A.; Jiang, Q.S.; Zhao, R.Y.; Hu, J.; Hu, B.F.; Shi, Z. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma* **2018**, *337*, 1309–1319. [\[CrossRef\]](#)
27. Verstraete, M.M.; Pinty, B. Designing optimal spectral indexes for remote sensing applications. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 1254–1265. [\[CrossRef\]](#)
28. Han, L.; Liu, D.; Cheng, G.; Zhang, G.; Wang, L. Spatial distribution and genesis of salt on the saline playa at Qehan Lake, Inner Mongolia, China. *Catena* **2019**, *177*, 22–30. [\[CrossRef\]](#)
29. 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. [\[CrossRef\]](#)
30. 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. [\[CrossRef\]](#)
31. 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. [\[CrossRef\]](#)
32. Baig, M.H.A.; Zhang, L.; Shuai, T.; Tong, Q. Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sens. Lett.* **2014**, *5*, 423–431. [\[CrossRef\]](#)
33. Peng, J.; Liu, H.; Shi, Z.; Xiang, H.; Chi, C. Regional heterogeneity of hyperspectral characteristics of salt-affected soil and salinity inversion. *Trans. Chin. Soc. Agric. Eng.* **2014**, *30*, 167–174.
34. Xu, C.; Zeng, W.; Huang, J.; Wu, J.; Van Leeuwen, W. Prediction of soil moisture content and soil salt concentration from hyperspectral laboratory and field data. *Remote Sens.* **2016**, *8*, 42. [\[CrossRef\]](#)
35. Hong, Y.; Shen, R.; Cheng, H.; Chen, Y.; Zhang, Y.; Liu, Y.; Zhou, M.; Yu, L.; Liu, Y.; Liu, Y. Estimating lead and zinc concentrations in peri-urban agricultural soils through reflectance spectroscopy: Effects of fractional-order derivative and random forest. *Sci. Total Environ.* **2018**, *651*, 1969–1982. [\[CrossRef\]](#)
36. Wang, J.Q.; Wu, W.M.; Wang, T.W.; Cai, C.F. Estimation of leaf chlorophyll content and density in *Populus euphratica* based on hyperspectral characteristic variables. *Spectrosc. Lett.* **2018**, *51*, 485–495. [\[CrossRef\]](#)
37. Zeraatpisheh, M.; Ayoubi, S.; Jafari, A.; Tajik, S.; Finke, P. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* **2019**, *338*, 445–452. [\[CrossRef\]](#)
38. Allbed, A.; Kumar, L. Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: A review. *Adv. Remote Sens.* **2013**, *2*, 373–385. [\[CrossRef\]](#)
39. Celleri, C.; Zapperi, G.; Trilla, G.G.; Pratolongo, P. Assessing the capability of broadband indices derived from landsat 8 operational land imager to monitor above ground biomass and salinity in semiarid saline environments of the bahía blanca estuary, argentina. *Int. J. Remote Sens.* **2019**, *40*, 4817–4838. [\[CrossRef\]](#)
40. Metternicht, G.I.; Zinck, J.A. Remote sensing of soil salinity: Potentials and constraints. *Remote Sens. Environ.* **2003**, *85*, 1–20. [\[CrossRef\]](#)
41. Bui, E.N. Soil salinity: A neglected factor in plant ecology and biogeography. *J. Arid Environ.* **2013**, *92*, 14–25. [\[CrossRef\]](#)
42. Pakparvar, M.; Gabriels, D.; Aarabi, K.; Edraki, M.; Raes, D.; Cornelis, W. Incorporating legacy soil data to minimize errors in salinity change detection: A case study of Darab Plain, Iran. *Int. J. Remote Sens.* **2012**, *33*, 6215–6238. [\[CrossRef\]](#)
43. Ding, J.; Yu, D. 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*, 316–322. [\[CrossRef\]](#)
44. Yang, J.; Wang, X.; Wang, R.; Wang, H. Combination of convolutional neural networks and recurrent neural networks for predicting soil properties using Vis-Nir spectroscopy. *Geoderma* **2020**, *380*, 114616. [\[CrossRef\]](#)
45. Gholami, V.; Sahour, H.; Hadian, M.A. Soil erosion modeling using erosion pins and artificial neural networks. *Catena* **2021**, *196*, 104902. [\[CrossRef\]](#)
46. Xing, L.; Li, L.; Gong, J.; Ren, C.; Liu, J.; Chen, H. Daily soil temperatures predictions for various climates in united states using data-driven model. *Energy* **2018**, *160*, 430–440. [\[CrossRef\]](#)
47. Zhang, C.; Liu, Y.M.; Sun, Y.N.; Liu, J.H. Hyperspectral prediction model of soil nutrient content in the loess hilly-gully region, China. *J. Appl. Ecol.* **2018**, *29*, 2835–2842.
48. Banerjee, K.; Krishnan, P.; Mridha, N. Application of thermal imaging of wheat crop canopy to estimate leaf area index under different moisture stress conditions. *Biosyst. Eng.* **2017**, *166*, 13–27. [\[CrossRef\]](#)
49. Deiss, L.; Margenot, A.J.; Culman, S.W.; Demyan, M.S. Tuning support vector machines regression models improves prediction accuracy of soil properties in mir spectroscopy. *Geoderma* **2020**, *365*, 114227. [\[CrossRef\]](#)
50. Wang, Z.; Zhang, X.L.; Zhang, F.; Chan, N.; Kung, H.; Liu, S.H.; Deng, L.F. Estimation of soil salt content using machine learning techniques based on remote-sensing fractional derivatives, a case study in the Ebinur Lake Wetland National Nature Reserve, Northwest China. *Ecol. Indic.* **2020**, *119*, 106869. [\[CrossRef\]](#)
51. Zhang, K.; Chao, L.J.; Wang, Q.Q.; Huang, Y.C.; Liu, R.H.; Hong, Y. Using multi-satellite microwave remote sensing observations for retrieval of daily surface soil moisture across china. *Water Sci. Eng.* **2019**, *12*, 85–97. [\[CrossRef\]](#)