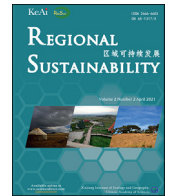




ELSEVIER

Contents lists available at ScienceDirect

## Regional Sustainability

journal homepage: [www.keaipublishing.com/en/journals/regional-sustainability](http://www.keaipublishing.com/en/journals/regional-sustainability)

## Full Length Article

## Digital mapping of soil salinization based on Sentinel-1 and Sentinel-2 data combined with machine learning algorithms

Guolin Ma<sup>a,b</sup>, Jianli Ding<sup>a,b,\*</sup>, Lijing Han<sup>a,b</sup>, Zipeng Zhang<sup>a,b</sup>, Si Ran<sup>a,b</sup><sup>a</sup> Key Laboratory of Smart City and Environment Modelling of Higher Education Institute, Xinjiang University, Urumqi, 800046, China<sup>b</sup> Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi, 830046, China

## ARTICLE INFO

## Keywords:

Salinization  
Digital soil mapping  
XGBoost  
Sentinel-1  
Sentinel-2  
Ogan-Kuqa River Oasis

## ABSTRACT

Soil salinization is one of the most important causes of land degradation and desertification, especially in arid and semi-arid areas. The dynamic monitoring of soil salinization is of great significance to land management, agricultural activities, water quality, and sustainable development. The remote sensing images taken by the synthetic aperture radar (SAR) Sentinel-1 and the multispectral satellite Sentinel-2 with high resolution and short revisit period have the potential to monitor the spatial distribution of soil attribute information on a large area; however, there are limited studies on the combination of Sentinel-1 and Sentinel-2 for digital mapping of soil salinization. Therefore, in this study, we used topography indices derived from digital elevation model (DEM), SAR indices generated by Sentinel-1, and vegetation indices generated by Sentinel-2 to map soil salinization in the Ogan-Kuqa River Oasis located in the central and northern Tarim Basin in Xinjiang of China, and evaluated the potential of multi-source sensors to predict soil salinity. Using the soil electrical conductivity (EC) values of 70 ground sampling sites as the target variable and the optimal environmental factors as the predictive variable, we constructed three soil salinity inversion models based on classification and regression tree (CART), random forest (RF), and extreme gradient boosting (XGBoost). Then, we evaluated the prediction ability of different models through the five-fold cross validation. The prediction accuracy of XGBoost model is better than those of CART and RF, and soil salinity predicted by the three models has similar spatial distribution characteristics. Compared with the combination of topography indices and vegetation indices, the addition of SAR indices effectively improves the prediction accuracy of the model. In general, the method of soil salinity prediction based on multi-source sensor combination is better than that based on a single sensor. In addition, SAR indices, vegetation indices, and topography indices are all effective variables for soil salinity prediction. Weighted Difference Vegetation Index (WDVI) is designated as the most important variable in these variables, followed by DEM. The results showed that the high-resolution radar Sentinel-1 and multispectral Sentinel-2 have the potential to develop soil salinity prediction model.

## 1. Introduction

As a global problem, soil salinization poses a serious threat to limited soil resources and ecosystem health in arid and semi-arid areas,

\* Corresponding author. Key Laboratory of Smart City and Environment Modelling of Higher Education Institute, Xinjiang University, Urumqi, 800046, China.

E-mail address: [watarid@xju.edu.cn](mailto:watarid@xju.edu.cn) (J. Ding).

<https://doi.org/10.1016/j.regsus.2021.06.001>

Received 1 December 2020; Received in revised form 12 April 2021; Accepted 17 June 2021

Available online 9 July 2021

2666-660X/© 2021 Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences. Publishing services by Elsevier B.V. on behalf of KeAi

Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

and it is one of the most important causes of land desertification and land degradation (Metternicht and Zinck, 2003). At the same time, it has also led to the continuous decline in the soil productivity, vegetation coverage, and biodiversity (Nawar et al., 2014; Zhang et al., 2020a). According to statistics, almost 3% of the world's soil resources are affected by salinization, which is expanding at a rate of  $2.00 \times 10^3 \text{ km}^2$  per year (Ramos et al., 2020). In arid and semi-arid regions, in addition to the original high water-soluble salt, limited soil resources are also threatened by secondary salinization due to scarce rainfall, high evaporation, high groundwater level, and unreasonable agricultural activities (Wang et al., 2019). Soil salinity is an important indicator to evaluate the degree of soil salinization, and there is a big difference in time and space latitude (Ding and Yu, 2014). This shows that only dynamic monitoring can fully understand the current status of soil salinization, so as to provide more quantitative information for soil restoration.

The traditional regional soil salinization evaluation method is obtained through field soil sample investigation and analysis, which is time-consuming and laborious, and cannot realize large-scale real-time dynamic monitoring. Remote sensing technology can simultaneously observe the information of the same area on a large area, which has the characteristics of macroscopic, comprehensive, dynamic, and rapid. It can quickly and accurately extract soil information, such as the degree of salinization and its spatial characteristics (Allbed et al., 2014). At present, a large amount of satellite remote sensing data with medium and high spatial and temporal resolutions have been widely used in the mapping of soil salinity and other soil attribute information, which provides new opportunities for qualitative and quantitative research on soil attributes (Zheng et al., 2009), including total phosphorus (Shen et al., 2019), organic carbon (Zhou et al., 2020), total nitrogen (Kalambukattu et al., 2018), available potassium (Dong et al., 2019), anion and cation exchange (Ivushkin et al., 2019), texture (Gholizadeh et al., 2018), bulk density (Hengl et al., 2017), and pH value (Roelofsen et al., 2015). However, the monitoring of soil attributes based on optical sensors is susceptible to weather conditions such as cloudy and rainy.

Synthetic aperture radar (SAR) remote sensing has the advantages of all-weather monitoring; it has a certain penetration ability to vegetation and the surface. Compared with optical remote sensing, it can reflect the characteristics of soil below the surface layer and make up for the lack of optical remote sensing. Soil mapping has broad application prospects (Holah et al., 2005; Baghdadi et al., 2009). SAR remote sensing images have been widely used to monitor the spatial change characteristics of soil attributes, such as soil texture (Aubert et al., 2011) and moisture (Hengl et al., 2017). In addition, some scholars have also studied the potential of SAR technology to monitor soil chemical properties, but compared with optical remote sensing images, the application of SAR images in digital soil mapping has not been well utilized. In addition to the limited availability of freely available high-resolution radar and multispectral images, the diversity and complexity of SAR images also limit its development in digital soil mapping (De Bernardis et al., 2016). The latest Sentinel series satellites launched by European Space Agency provide scientific researchers with a large number of freely available high-resolution remote sensing images, offering a broad prospect for further refined digital soil mapping (Zhou et al., 2020). These provided images have high resolution (up to 10 m) and novel spectral functions (namely, three red-side bands and two shortwave near-infrared bands). Sentinel-1 is equipped with SAR sensors with high spatial resolution ( $5 \text{ m} \times 20 \text{ m}$ ) and short revisit period (6 d) (ESA, 2015a), which provides a broad prospect for soil information monitoring. Sentinel-2 is equipped with a new generation of multi-spectrometer, with 13 different resolution bands, covering visible light, near-infrared, and short-wave near-infrared spectrum bands; its wavelength range is from 443 to 2190 nm. In addition, it has a 5-d revisit period and a width of 290 km (ESA, 2015b). The relatively high spatial and temporal resolutions of Sentinel-1 and Sentinel-2 play an important role in monitoring the spatial variability of soil salinization at the regional scale.

Sentinel-2 has been used in the research of soil salinization monitoring and mapping (Loiseau et al., 2019; Vaudour et al., 2019;

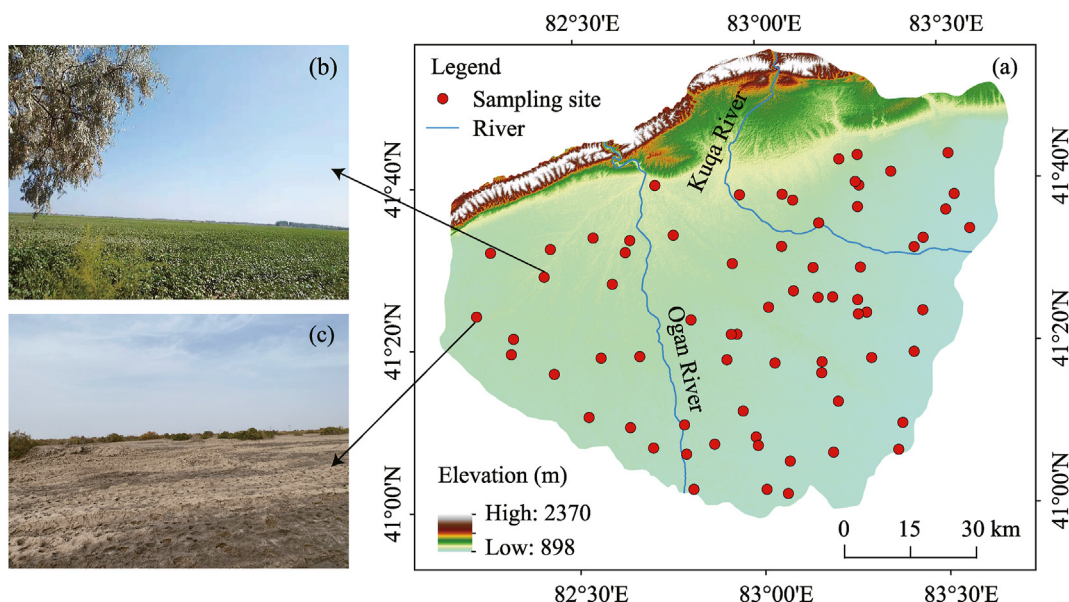


Fig. 1. Locations of the sampling sites in the Ogan-Kuqa River Oasis (a) and demonstration of the landscape of the sampling sites (b and c).

Farahmand and Sadeghi, 2020), and Sentinel-1 has also been used by many scholars in the research of other soil attribute mapping (Ma, 2018; Hoa et al., 2019; Taghadossi et al., 2019). Some scholars merged sensors with different properties in the research of attribute mapping, and found that the combination of multi-source sensors can effectively improve the accuracy of digital soil mapping (Alexakis et al., 2017; Rudiyanto et al., 2018). Although some progress has been made in the use of Sentinel-2 data in soil salinization mapping, few combinations of Sentinel-1 and Sentinel-2 data have been used to predict soil salinity.

Soil is a highly variable space-time continuum. The effect of soil-forming factors is nonlinear in the process of soil development. The nonlinear effect on soil properties is more obvious in a larger area (El Harti et al., 2016). Machine learning and data mining technology can effectively solve the nonlinear problem between soil properties and environmental factors (Allbed et al., 2014). Among these methods, tree-based models are often the most common machine learning algorithms; they are reported to have a good performance in soil salinity prediction, including random forests (RF) (Fathizad et al., 2020), classification and regression trees (CART) (Yao et al., 2013), and boosted regression trees (BRT) (Muller and van Niekerk, 2016). Other machine learning algorithms that have been used in digital soil mapping also include extreme gradient boosting (XGBoost) (Wei et al., 2019), backpropagation neural network (BPNN) (Shahabi et al., 2017), and support vector machines (SVM) (Guan et al., 2013). However, no machine learning model can show the best predictive effect under any circumstances (Hengl et al., 2017; Chen et al., 2019).

The main objectives of this study are as follows: (1) using three machine learning algorithms, i.e., CART, RF, and XGBoost, to map soil salinization in the Ogan-Kuqa River Oasis of Xinjiang, China, based on topography indices, vegetation indices, and SAR indices; (2) evaluating the accuracy of CART, RF, and XGBoost algorithms in soil salinity prediction; (3) comparing and assessing the potential of different types of sensor combinations (the combination of topography indices and vegetation indices or the combination of topography indices, vegetation indices, and SAR indices) in digital soil mapping; and (4) exploring the sensitivity of different types of sensors in soil salinity prediction and the spatial variation characteristics of soil salinity.

## 2. Materials and methods

### 2.1. Study area

The Ogan-Kuqa River Oasis is located in the central and northern Tarim Basin of Xinjiang Uygur Autonomous Region, China. It is a typical and representative alluvial fan oasis in an arid region, bred by the Ogan River and Kuqa River (Ding and Yu, 2014). It locates between 41°06′–41°58′N and 82°10′–83°50′E, and its terrain has a northwest-to-southeast tilt (Fig. 1). The terrain of this area can be divided into three parts: the fan-shaped gravel Gobi area formed by alluvial action, the Tianshan Mountains in the north, and the alluvial plain area in the south, with landscape of desert, oasis, and Gobi, in turn. The vegetation species in the study area are rare, and the vegetation coverage is low. The main species are *Phragmites australis*, *Tamarix ramosissima*, *Allhagi sparsifolia*, and so on. The oasis is mainly cultivated with cotton, corn, and winter wheat. The Ogan-Kuqa River Oasis has a typical temperate continental arid climate, with abundant light and heat resources and less precipitation. The annual total radiation is  $6.11 \times 10^5 \text{ J/cm}^2$ . The average annual precipitation is only 55.45 mm but the annual evaporation is 2356.00 mm (Ding and Yu, 2014). In recent years, due to extreme climatic conditions, population increase, and high intensity agricultural irrigation activities, the salinized soil area in the Ogan-Kuqa River Oasis has increased sharply, leading to the threat of desertification in the limited land resources, which has aroused the continuous attention of all sectors of society with regard to the sustainable development of the oasis.

### 2.2. Field sampling and analysis

We conducted the field survey from 11 July 2018 to 19 July 2018 and designed the survey route of the sampling process according to the accessibility of potential sampling sites. Based on previous field survey experience and existing soil digital maps (soil type, texture, and other characteristics), we selected a total of 70 soil sampling sites. We used the five-point sampling method to collect 5 soil samples at each sampling site (30 m × 30 m) and mixed them on site to create a representative composite sample (0–20 cm soil depth), while using a portable GPS (UniStrong G120, positioning accuracy ≤ 5 m) to record its geographic locations. We placed each soil sample in a sealed bag (to prevent soil moisture loss) and made a mark, then brought these samples back to the laboratory for further chemical analyses. All soil samples were air-dried, ground, homogenized, and sieved to ≤ 2 mm particle size. We added 100 mL of distilled water to each 20 g soil sample and mixed thoroughly (1:5 soil-water extraction solution), and then, we extracted the percolate through a multiparameter measuring device (Multi 3420 Set B, WTW GmbH, Germany) equipped with a composite electrode (TetraCon 925) at room temperature of 25 °C to measure electrical conductivity (EC). The PHS-3C was used to measure pH value (Wang et al., 2019), and the oven dry weighing method (105 °C constant temperature box, 48 h) was used to measure the soil moisture content (SMC) (Peng et al., 2019; Zhang et al., 2021).

### 2.3. Acquisition and pre-processing of remote sensing data

As an important part of the Global Monitoring for Environment and Security (GMES), the Sentinel series plays a pivotal role in land, ocean, atmosphere, climate, and global change (Drusch et al., 2012; Malenovsky et al., 2012). The Sentinel-1 launched in 2014 is equipped with a C-band SAR. It has a variety of imaging methods, which can realize multiple polarization methods, such as single polarization and dual polarization. It has a strip mode (SM), an extra wide (EW) mode, and an interferometric wide (IW) mode (Zhou et al., 2018). In this study, we obtained 2 scenes of dual-polarized Sentinel-1 images in IW mode from the European Space Agency's Copernicus Open Access Hub; the data level is Level-1, and the satellite transit date is 22 July 2018. The images were processed by the

special software SNAP for sentinel data processing developed by the European Space Agency, which sequentially performs orbit correction, thermal noise removal, radiation calibration, coherent spot filtering (window size  $7 \times 7$ ), terrain correction, and stitching and cropping, among which orbit correction and the files required for topography correction were all obtained by downloading online. The resolution of the image resampling is 30 m; we chose this resolution because it has proven to be the most suitable for soil-landscape analysis. Some scholars have pointed out that the soil information extracted by single-polarized radar images is limited, which has a certain impact on the research results (Ruecker et al., 2009). A proper combination of the polarization methods of the radar images can effectively improve the correlation between the backscatter coefficient and soil salinity (Ma, 2018). We used different polarization combinations to achieve the forecast goal, as shown in Table 1.

We downloaded the Level-1C Sentinel-2 images of the Universal Transverse Mercator/World Geodetic System (UTM/WGS) projection of the four scenes of transit on 21 July 2018 with cloud cover below 10% from the European Space Agency's Copernicus Open Access Hub. This level of product has undergone radiometric correction and geometric correction processing. We used the Sen2Cor plugin version 2.08 of SNAP to conduct atmospheric correction, in such a way that the reflectivity of the remote sensing image is changed from the reflectivity of the Top of Atmosphere (TOA) to the reflectivity of Bottom of the Atmosphere (BOA) (Clevers and Gitelson, 2013). In this step, the band with a resolution of 60 m (B1, B9, and B10) was mainly used to monitor the atmospheric characteristics, and thus, it was not included in subsequent studies. In the study of soil salinization, different degrees of soil salinization will produce stress on vegetation growth, and vegetation is sensitive to soil salinity stress. Therefore, soil salinity and its change trend can be inferred indirectly through vegetation indices. In this study, Sentinel-2 was used only to calculate vegetation indices. The bands after the atmospheric correction were resampled to 30 m by the nearest neighbor interpolation method for vegetation indices calculation. The calculation formulas of vegetation indices are shown in Table 2.

#### 2.4. Topography attribute acquisition

Topography plays an important role in the distribution of soil salinity (Gallant and Dowling, 2003). Soil salinity is high in areas with lower elevations; on the contrary, soil salinity is low in inland areas with more undulating terrain and tends to decrease with the increase of elevation (Abdel-Kader, 2011). Therefore, in the prediction of soil salinity, the topography factor is an important parameter of the predictor variables (Galin et al., 2019). In this study, digital elevation model (DEM) with a resolution of 30 m (<http://www.gscloud.cn/>) was used to calculate 9 topography indices through System for Automated Geoscientific Analyses (SAGA) GIS (Table 3).

#### 2.5. Modelling methods and evaluation indicators

CART is a non-parametric data mining technology that is used for classification or regression analysis. This technology has been continuously improved and is widely used due to its ability to suppress data noise (Breiman et al., 1984). CART includes a non-parametric regression method that adds a set of decision trees to the binomial partition algorithm. The algorithm splits predictor variables iteratively, as long as these groups are isomorphic or contain no less than the user's observations. The defined threshold, i.e., the average response value of each node, represents the predicted value of the terminal node. The regression tree can cover the missing data by substitution, which provides advantages for handling abnormal data (Vega et al., 2009). In addition to regression modelling, the hierarchical structure of classification also allows model interactions between predictor variables (Schuler et al., 2010).

RF, with the basic unit of decision tree, is an algorithm that integrates multiple trees and can be effectively used for classification and regression in multiple fields (Breiman, 2001). It operates by constructing a large number of decision trees during training and outputting the classes as the class (classification) or mean prediction (regression) mode of each tree. In essence, it makes the tree irrelevant and uncorrelated by setting the stop criteria for node splitting. The algorithm does not need to meet any assumptions about the data distribution; it can handle both grades and continuous variables at the same time and has good nonlinear data mining capabilities and generalization capabilities (Gholizadeh et al., 2018).

XGBoost is an improvement of the gradient boosting decision tree method developed by Friedman (2001). It is an open source and extensible end-to-end tree enhancement system. It has been widely used in classification and regression analysis, showing a good accuracy. Compared with ordinary gradient boosting algorithms, XGBoost can provide faster and more accurate predictions. XGBoost is a method of creating a new model that can predict the residuals of the previous model and then add them together to make the final decision; it uses the gradient descent algorithm to minimize the loss when adding a new model.

The parameters in the model need to be optimized before applying the calibration model, which are called hyperparameters and

**Table 1**  
Polarization combinations used in the study.

Polarization combination	Reference	Polarization combination	Reference
VV	Ma (2018)	VH	Ma (2018)
VV + VH	Ma (2018)	$VV^2 + VH^2$	Ma (2018)
$VV^2 + VH$	Zhang et al. (2020)	$VH^2 - VV$	Zhang et al. (2020)
$(VH^2 + VV^2)/VH$	Zhang et al. (2020)	$10 \log(VH)$	Zhang et al. (2020)
$10 \log(VV)$	Zhang et al. (2020)	$10 \log(VV) + 10 \log(VH)$	Zhang et al. (2020)

Note: VV represents the radar backscatter coefficient of vertical polarization and VH represents the radar backscatter coefficient of horizontal polarization.

**Table 2**

Calculation formulas of vegetation indices based on Sentinel-2.

Vegetation index	Index acronym	Formula	Reference
Normalized Difference Vegetation Index	NDVI	$\frac{B_8 - B_4}{B_8 + B_4}$	Tucker (1979)
Green Normalized Difference Vegetation Index	GNDVI	$\frac{B_7 - B_3}{B_7 + B_3}$	Gitelson and Merzlyak (1998)
Weighted Difference Vegetation Index	WDVI	$B_8 - 0.5 \times B_4$	Clevers (1989)
Transformed Normalized Difference Vegetation Index	TNDVI	$\sqrt{\frac{B_8 - B_4}{B_8 + B_4}} + 0.5$	Yi (2019)
Soil Adjusted Vegetation Index	SAVI	$\left( \frac{B_8 - B_4}{B_8 + B_4 + 0.5} \right) \times 1.5$	Huete (1988)
Infrared Percentage Vegetation Index	IPVI	$\frac{B_8}{B_8 + B_4}$	Crippen (1990)
Modified Chlorophyll Absorption Ratio Index	MCARI	$((B_5 - B_4) - 0.2(B_5 - B_3)) \times \frac{B_5}{B_4}$	Daughtry et al. (2000)
Red Edge In-flection Point	REIP	$700 + 40 \left( \left( \frac{B_4 + B_7}{2} \right) - B_5 \right)$	Guyot et al. (1988)
Modified Soil Adjusted Vegetation Index 2	MSAVI2	$\frac{B_6 - B_5}{2B_8 - 1 - \sqrt{(2B_8 + 1)^2 - 8}}$	Qi et al. (1994)
Difference Vegetation Index	DVI	$B_8 - B_4$	Jordan (1969)

Note:  $B$  is the reflectivity of Sentinel-2 waveband, and the subscript is the number of each band.

have a large impact on the model performance (Yin et al., 2020). Therefore, in this study, the method of grid search is used to optimize the model's hyperparameters. The prediction performance of different models was evaluated by the method of five-fold cross validation. All measured values were randomly divided into 5 groups, among which 4 groups were selected as the training set and the remaining one group as the validation set. Compared with randomly dividing the training set and the validation set, the five-fold cross validation made the model more reliable. In order to quantitatively evaluate the prediction accuracy of different models, we selected the following three indices: determination coefficient ( $R^2$ ), root mean squared error (RMSE), and ratio of performance to deviation (RPD). The larger the  $R^2$  value, the higher the model accuracy. RMSE represents the predictive ability, and its value is inversely proportional to the model accuracy. As a predictive indicator, RPD has been widely used to evaluate the model accuracy. For example, Chang et al. (2001) divided RPD into three categories: when  $RPD \geq 2.0$ , the model has excellent performance in predictive ability; when  $1.4 < RPD < 2.0$ , the model can provide multiple reasonable prediction results; and when  $RPD < 1.4$ , the model has a low feasibility and cannot predict the measured value of the sample, and thus, it is not recommended.

The above-mentioned modelling process is realized by programming in the Pycharm editor using the Python 3.7 language.

**Table 3**

Topography indices and acronym used in this study.

Topography index	Index acronym	Reference
Digital Elevation Model (m)	DEM	–
Slope	S	System for Automated Geoscientific Analyses (SAGA) GIS
Aspect	AS	SAGA GIS
Convergence Index	CI	SAGA GIS
Total catchment	TCA	SAGA GIS
Ls factor	LSF	SAGA GIS
Channel network base level	CNBL	SAGA GIS
Channel network distance	CND	SAGA GIS
Valley depth	VD	SAGA GIS
Relative slope position	RSP	SAGA GIS

**Table 4**

Descriptive statistical analysis of the soil properties.

Soil property	Maximum	Minimum	Mean	Median	SD	CV (%)
EC (dS/m)	79.70	0.08	13.77	6.25	18.24	132.49
SMC (%)	20.28	0.52	9.60	9.58	5.17	53.89
pH	10.21	7.96	8.80	8.68	0.43	–

Note: EC means electrical conductivity, SMC means soil moisture content, SD means standard deviation, and CV means coefficient of variation. –, no data available.



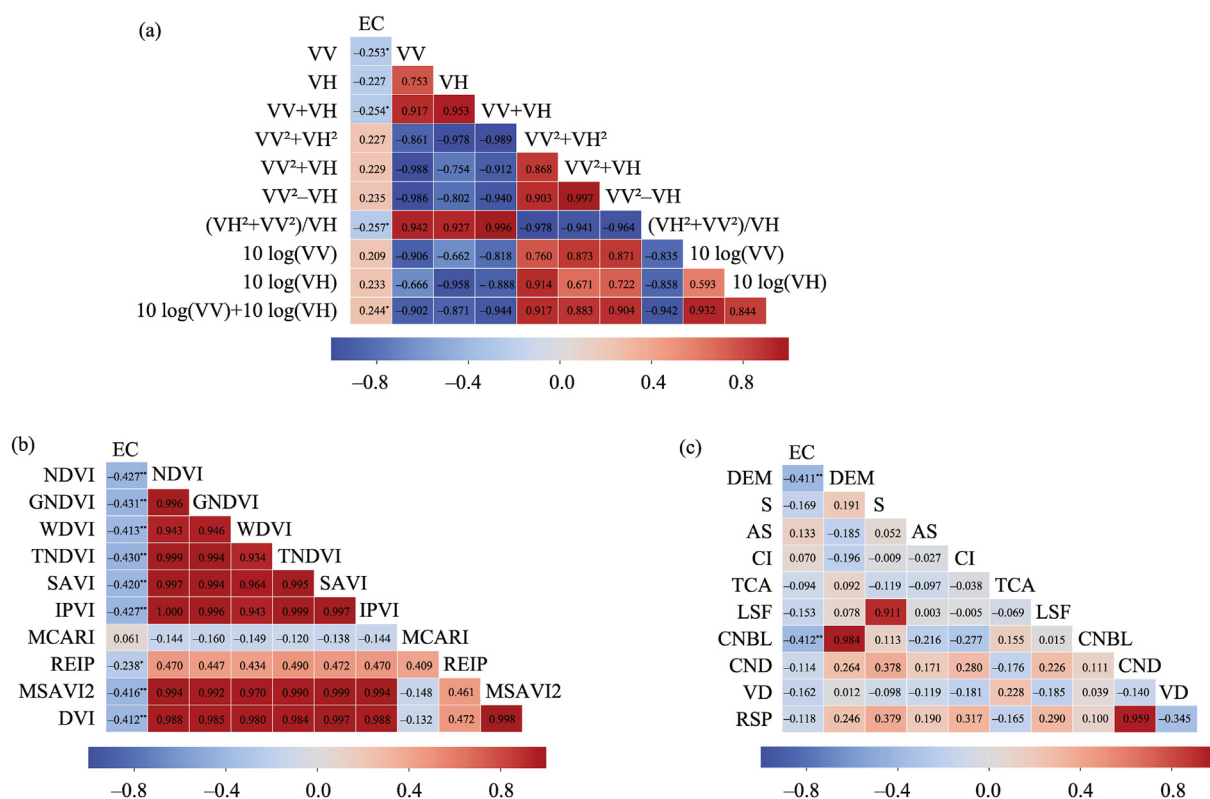
### 3. Results

#### 3.1. Descriptive statistics

EC values in the entire study area varied greatly, ranging from 0.08 to 79.70 dS/m (Table 4). The mean and median values were 13.77 and 6.25 dS/m, respectively, the standard deviation (SD) was 18.24 dS/m, and the coefficient of variation (CV) was 132.49%, which belonged to the strong variation. The SMC ranged from 0.52% to 20.28%, and the SD and CV were 5.17% and 53.89%, respectively, which indicated a moderate variation. Soil pH values in the entire study area varied from 7.96 to 10.21, with the mean and median values of 8.80 and 8.68, respectively. The soil properties in the entire study area were alkaline (Table 4).

#### 3.2. Correlation analysis between the surface parameters and measured electrical conductivity

In the machine learning methods, the choice of modelling the feature variables is critical to prediction results (Li et al., 2020). Most of the factors related to soil salinity can be obtained through band calculation, and there is a correlation between the factors. If the factors are not screened, the redundancy of variables and the accuracy of the model will be reduced (Peng et al., 2019; Wang et al., 2019). We used Pearson correlation analysis to select factors with significant correlations between the surface parameters and EC for salinization monitoring and visualization (Fig. 2). This approach can help us to choose potential predictors for the model. Among the 30 parameters (10 SAR indices, 10 vegetation indices, and 10 topography indices), the correlations between 10 parameters and EC were extremely significant ( $P < 0.01$ ), the correlations between 5 parameters and EC were significant ( $P < 0.05$ ), and the correlations between the remaining 15 parameters and EC were not significant. There were 21 parameters that were negatively correlated with EC (such as the radar backscatter coefficient of vertical polarization (VV), the radar backscatter coefficient of horizontal polarization (VH), VV + VH, and so on), of which the negative correlation between Green Normalized Difference Vegetation Index (GNDVI) and EC was the strongest, with a correlation coefficient of  $-0.431$ , while the other 9 parameters were positively correlated with EC. The positive correlation



**Fig. 2.** Correlation coefficients between measured electrical conductivity (EC) values and SAR indices (a), between measured EC values and vegetation indices (b), and between measured EC values and topography indices (c). VV, the radar backscatter coefficient of vertical polarization; VH, the radar backscatter coefficient of horizontal polarization; NDVI, Normalized Difference Vegetation Index; GNDVI, Green Normalized Difference Vegetation Index; WDV, Weighted Difference Vegetation Index; TNDVI, Transformed Normalized Difference Vegetation Index; SAVI, Soil Adjusted Vegetation Index; IPVI, Infrared Percentage Vegetation Index; MCARI, Modified Chlorophyll Absorption Ratio Index; REIP, Red Edge In-flexion Point; MSAVI2, Modified Soil Adjusted Vegetation Index 2; DVI, Difference Vegetation Index; DEM, Digital Elevation Model; S, Slope; AS, Aspect; CI, Convergence Index; TCA, Total catchment; LSF, Ls factor; CNBL, channel network base level; CND, channel network distance; VD, valley depth; RSP, relative slope position. \*\*, significance at the 0.01 probability level; \*, significance at the 0.05 probability level.

between  $10 \log(VV) + 10 \log(VH)$  and EC was the strongest, with a correlation coefficient of 0.244 (Fig. 2).

Combined with the polarization transform of SAR backscattering coefficient, the influence of noise in the image on the backscattering coefficient of ground objects is reduced effectively, and the correlation between SAR indices and EC is improved significantly. Among the 10 SAR indices, there were 4 indices significantly correlated with EC ( $P < 0.05$ ; Fig. 2a). Further, there were 8 vegetation indices that were extremely significantly correlated with EC ( $P < 0.01$ ), and Red Edge Inflection Point (REIP) was significantly correlated with EC ( $P < 0.05$ ; Fig. 2b). Among the 10 topography indices, only DEM and Channel network base level (CNBL) were significantly correlated with EC ( $P < 0.01$ ; Fig. 2c).

### 3.3. Analysis of modelling results

Table 5 shows the performance of the CART, RF, and XGBoost models in predicting EC. Among them, Model A indicates that the combination of vegetation indices and topography indices is used as a predictor variable, and Model B indicates that the combination of vegetation indices, topography indices, and SAR indices is used as a predictor variable. Through the comparative analysis of Table 5, it can be found that the model technology and the choice of predictor variables have an important influence on the prediction of soil salinity. In Model A, XGBoost model had the highest prediction accuracy ( $R^2 = 0.59$ , RMSE = 11.99 dS/m, and RPD = 1.55), and the prediction accuracy of CART and RF were similar (Table 5). The addition of SAR indices had a significant impact on the improvement of the prediction accuracy of the machine learning models. Among them, the improvement of RF prediction accuracy was the most obvious, followed by XGBoost. Specifically, for RF model,  $R^2$  increased to 0.63, RPD increased to 1.63, and RMSE decreased to 11.41 dS/m; for XGBoost model,  $R^2$  increased to 0.68, RPD increased to 1.77, and RMSE decreased to 10.56 dS/m. In general, XGBoost had a good prediction accuracy in both Model A and Model B, that is, it had the highest  $R^2$  and RPD and the lowest RMSE. Table 6 lists the main parameters of each model.

### 3.4. Relative importance of the predictors

The relative importance of the predictors was ranked, as shown in Fig. 3 (note that the importance value has been converted to percentage). The relative importance of the three types of variables in the different models was different, which revealed the differences in the main environmental characteristics in these models. For the machine learning models, the vegetation indices were the main explanatory variables for EC prediction (more than 45.00% of the total relative importance), followed by topography indices and SAR indices. Although the predictors in each model showed different hierarchical characteristics, the importance of Weighted Difference Vegetation Index (WDVI) ranked the first among all predictors, with a relative importance more than 20.00%. In addition, in CART, RF, and XGBoost models, SAR indices explained 21.00%, 17.90%, and 17.06% of EC changes, respectively, while topography indices explained 25.26%, 34.56%, and 22.20% of EC changes, respectively.

### 3.5. Spatial distribution characteristics of soil salinization

Three different machine learning algorithms were used to predict and map the spatial distribution characteristics of soil salinization (Fig. 4). According to the classification standard of the soil salinization level, we divided the soil samples in the study area into the following 5 categories:  $EC \geq 16$  dS/m is strongly saline,  $8 \text{ dS/m} \leq EC < 16$  dS/m is moderately saline,  $4 \text{ dS/m} \leq EC < 8$  dS/m is slightly saline,  $2 \text{ dS/m} \leq EC < 4$  dS/m is very slightly saline, and  $EC < 2$  dS/m means non-saline (Wang et al., 2019). In general, the three models produced similar spatial distribution characteristics of soil salinization. XGBoost model showed the subtle changes in the spatial distribution of soil salinization in detail. Moderately saline soil and strongly saline soil were mainly distributed in the marginal area of the oasis, which is caused by radiation errors of the topography. Due to the slight topographical undulations inside the oasis, coupled with the rise in the groundwater level caused by agricultural irrigation activities, non-saline soil and very slightly saline soil were mainly distributed in this area. Moderately saline soil was mainly located in the intersection of the oasis and desert. In general, soil salinity increased from the oasis to the surrounding areas. It is worth noting that CART model was affected by its own characteristics, and the predicted distribution map of soil salinity did not reflect the actual characteristics.

**Table 5**  
Evaluation of the prediction effects of the different models in predicting EC.

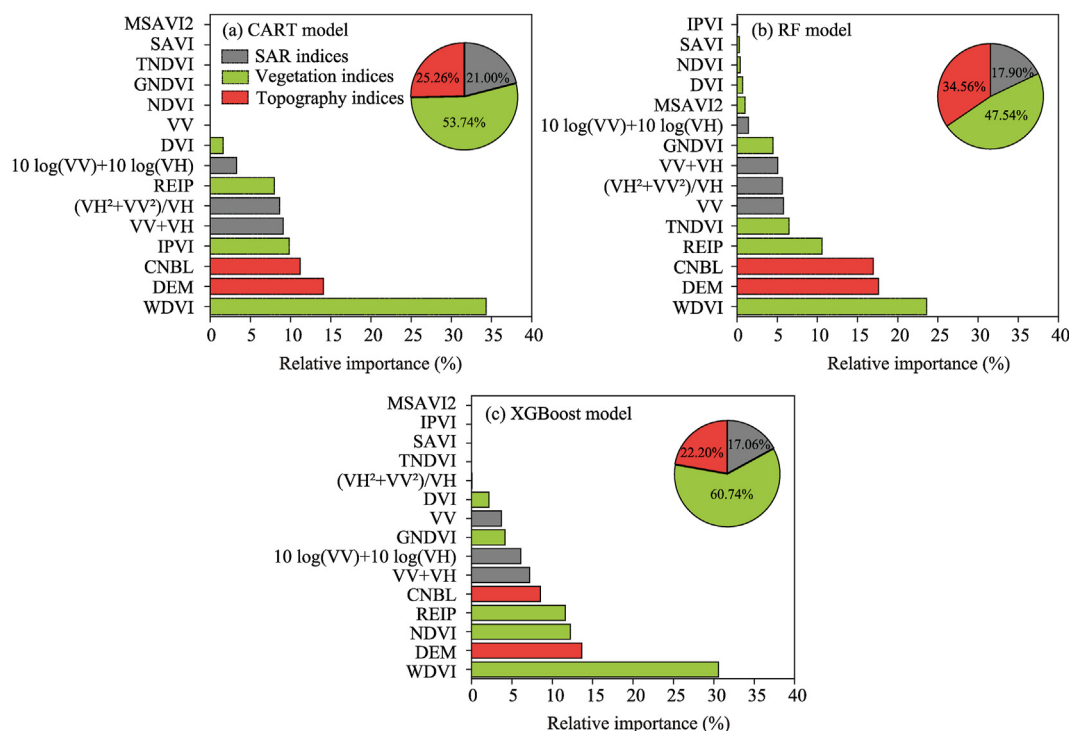
Model	Modelling technique	$R^2$	RMSE (dS/m)	RPD
Model A	CART	0.52	12.92	1.44
	RF	0.52	12.91	1.44
	XGBoost	0.59	11.99	1.55
Model B	CART	0.57	12.20	1.53
	RF	0.63	11.41	1.63
	XGBoost	0.68	10.56	1.77

Note: Model A means that the combination of vegetation indices and topography indices is used as a predictor variable and Model B indicates that the combination of vegetation indices, topography indices, and SAR indices is used as a predictor variable. CART, classification and regression tree; RF, random forest; XGBoost, extreme gradient boosting;  $R^2$ , coefficient of determination; RMSE, root mean square error; RPD, the ratio of performance to deviation. Abbreviations in the following tables are the same as Table 5.

**Table 6**

Main hyperparameters of the different models.

Model	CART	RF	XGBoost
Model A	Criterion = 'mse', Max_depth = 4, Max_features = 11, Max_leaf_nodes = 11	Criterion = 'mse', N_estimators = 7, Max_features = 8	N_estimators = 15, Learning_rate = 0.4, Max_depth = 3
Model B	Criterion = 'mse', Max_depth = 12, Max_features = 10, Max_leaf_nodes = 29	Criterion = 'mse', N_estimators = 8, Max_features = 15	N_estimators = 3, Learning_rate = 0.3, Max_depth = 5



**Fig. 3.** Relative importance of the predictor variables in the Model B for EC prediction using CART model (a), RF model (b), and XGBoost model (c). Model B means that the combination of vegetation indices, topography indices, and SAR indices is used as a predictor variable.

#### 4. Discussion

In this study, the vegetation indices were designated by the model as the most important predictor variables, followed by topography indices and SAR indices (Fig. 3). In soil salinization monitoring, vegetation is one of the key factors affecting the estimation accuracy (Yang et al., 2016). Due to the difference in vegetation coverage, variable errors will occur in the estimation of different degrees of soil salinization (Page et al., 2013). The response of these vegetation indices to soil salinization is affected by other environmental factors, and the results may be very different under various environmental conditions (Metternicht and Zinck, 2003). Although the vegetation indices have achieved good results in worldwide salinization monitoring, there is no universal vegetation indices that can achieve satisfactory results under any environmental conditions (Zhang et al., 2011; Allbed et al., 2014). In addition, although vegetation may obscure soil information, vegetation indices are more sensitive to changes in soil salinity under high vegetation cover (Peng et al., 2019). Topography, a derivative variable based on DEM, plays an important role in the formation and development of soil (Song et al., 2016); it is usually used as a key predictor in soil salinity estimation. Topography affects the flow of water and sediments, which in turn influences the spatial distribution of various soil properties such as soil salinity (Yahiaoui et al., 2015). In addition, the surface runoff caused by rainfall in the area is limited, which greatly weakens the influence of the topography factors on the redistribution of soil salinity (Akramkhanov et al., 2011; Taghizadeh-Mehrjardi et al., 2014; Peng et al., 2019). According to previous studies (e.g., Sugimori et al., 2008), some topography indices (such as DEM, CNBL, and relative slope position (RSP)) are highly correlated with EC and have significant effects on other soil characteristics. These findings are similar to our research results. DEM alters the formation and development of microclimates, which indirectly affects plant communities and soil processes (Song et al., 2016). In previous studies of soil properties (Bakr and Ali, 2019), DEM is an effective topography parameter. In this study, DEM is also identified as an important variable for soil salinity prediction, and its importance is second only to WDVI (Fig. 3).

A comparison of the prediction accuracy in this study showed that the choice of the machine learning algorithms and the combination of environmental variables can have a large impact on EC prediction (Table 5). In this study, XGBoost model performed better



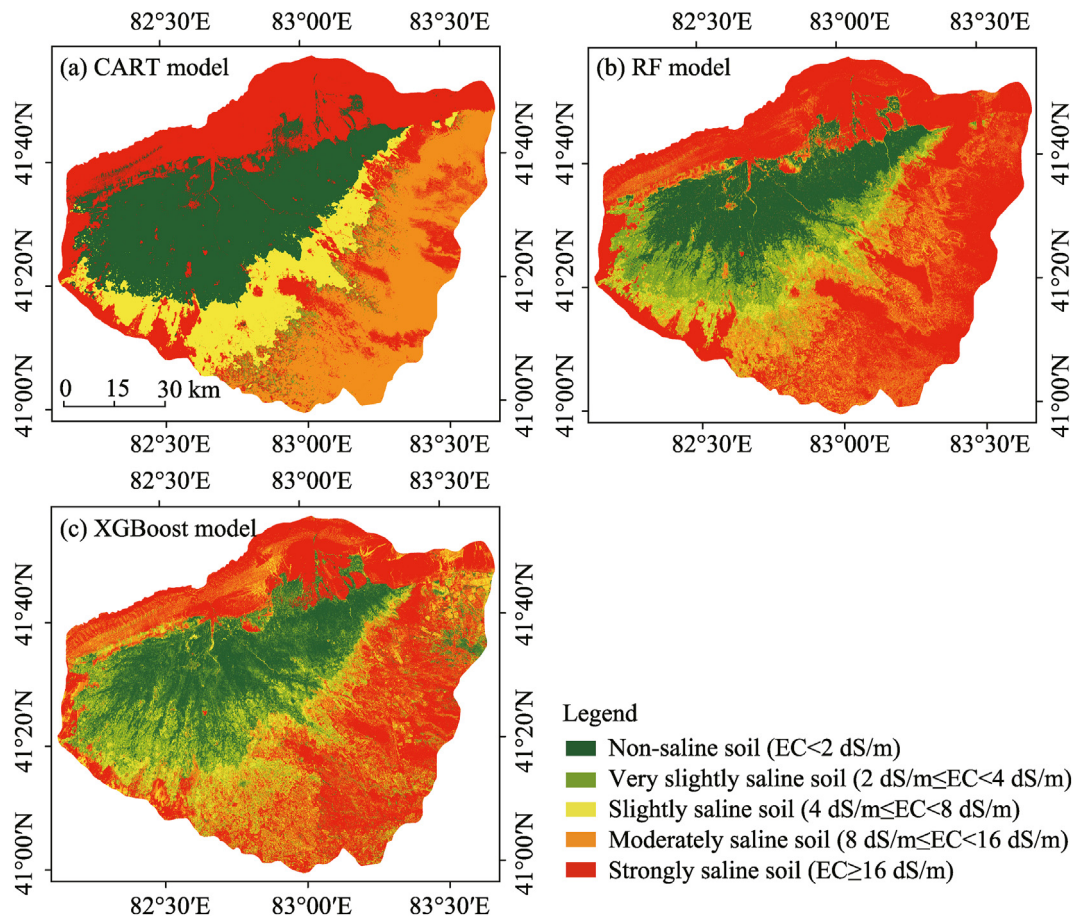


Fig. 4. Prediction maps of soil salinization based on Model B using CART model (a), RF model (b), and XGBoost model (c).

than the other two models. [Beguin et al. \(2017\)](#) reported that the selection of models and the construction of predictors have an important influence on the prediction of soil properties. [Peng et al. \(2019\)](#) used relevant covariates (based on the salt indices and vegetation indices derived from Landsat-8 and the topography indices derived from DEM) to estimate soil EC in southern Xinjiang, China. [Wang et al. \(2020\)](#) used Landsat-8 and Sentinel-2, two different types of remote sensing images, to predict soil EC in the extreme arid area of western China, and found that the types of remote sensing data, the selection of prediction models, and the differences in the study area have significant impacts on the prediction accuracy. In addition, the research results of related scholars also showed that no predictive model can perform best in all situations ([Lamichhane et al., 2019](#)). Therefore, we suggest that the prediction model should be calibrated with specific experimental data sets in different regions.

Ensemble learning is to complete the learning task by constructing multiple weak learners, which is also called a multi-classifier system ([Zhang et al., 2020b](#)). Compared with the two integrated learning algorithms of RF and CART, XGBoost inherits the advantages of feature sampling of decision tree. When training the weak learner, only the sub-data set with randomly sampled features is considered, which can increase the model diversity, avoid over-fitting, reduce the amount of model calculations, and improve the efficiency of the optimal solution to the input characteristic variables ([Zamani Joharestani, 2019](#)). Therefore, this method is suitable for the estimation of soil salinity that affected by many factors. In future studies, multiple factors should be considered. Machine learning technology integrates the advantages of different algorithms to further enhance the prediction and generalization capabilities of the model.

Our results also indicated that the optimized topography indices, SAR indices, and vegetation indices are essential for the establishment of effective EC prediction model. Some scholars emphasized that optical remote sensing images and topography factors play an important role in soil salinity prediction ([Castaldi et al., 2019](#)). However, previous studies only considered a single type of sensors, such as Landsat or MODIS ([Peng et al., 2019](#); [Wang et al., 2019](#)). In this study, results in the prediction accuracy of two different sensor combinations, Sentinel-1 and Sentinel-2, showed that the combination of radar sensors and optical sensors has a great potential in soil salinity prediction.

Remote sensing images have been widely used in digital mapping of soil salinization at different scales, but related studies are mostly limited to optical remote sensing ([Ivushkin et al., 2019](#)). The results of this study showed that both optical and SAR image data are effective predictors in the estimation of soil salinity. The contribution of SAR indices to the model mainly depends on their sensitivity to

soil moisture and surface changes (Kasischke et al., 1997). The sampling time in this study is the vegetation growing season, and the vegetation coverage is high, which weakens the radar penetration of the ground to a certain extent (Yang and Guo, 2019). In addition, SAR images are also affected by soil water content when characterizing ground surface information (Zhou et al., 2020). This may be the reason why SAR indices are not as important as the topography indices and vegetation indices in the model.

Although our research results showed that the combination of vegetation indices, SAR indices, and topography indices can effectively predict soil salinity, further improvement is still needed. In addition, our results have not yet reached a good prediction accuracy. The prediction accuracy of the optimal model XGBoost is only 0.68, which may be related to the research field, the type and quality of predictor variables, the accuracy of the prediction model, and the variability of EC (Yin et al., 2020). Owing to the dynamic changes in surface characteristics, time series Sentinel data will be considered in future studies to map soil salinity at different periods, so as to further reduce the uncertainty and improve the prediction accuracy.

## 5. Conclusions

By comparing three machine learning techniques (CART, RF, and XGBoost), we used topography indices derived from DEM, SAR indices generated by Sentinel-1, and vegetation indices generated by Sentinel-2 to study the spatial distribution of soil salinization in the Ogan-Kuqa River Oasis in Xinjiang of China. In general, XGBoost model is superior to CART and RF models in soil salinity prediction, showing the best performance. Soil salinity predicted by the three machine learning algorithms is consistent in spatial distribution. Strongly saline soil and moderately saline soil are mainly distributed in the piedmont alluvial fan and the oasis margin, while in the oasis the soil salinization degree is low. The combined method of multi-source sensors shows a good prediction performance. The freely available high-resolution remote sensing images of Sentinel-1 and Sentinel-2 provide more opportunities for dynamic monitoring of spatial changes in soil attribute information. Vegetation indices are the main explanatory variables for EC prediction, followed by topography indices and SAR indices. Further, WdVI is the most important environmental variable in soil salinity prediction.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (41771470) and the China Postdoctoral Science Foundation (2020M672776). We would like to thank the anonymous reviewers and editors for their careful review and constructive comments and suggestions, which play an important role in the further improvement of this article.

## References

- Abdel-Kader, F.H., 2011. Digital soil mapping at pilot sites in the northwest coast of Egypt: a multinomial logistic regression approach. *The Egyptian Journal of Remote Sensing and Space Science* 14 (1), 29–40.
- Akramkhanov, A., Martius, C., Park, S.J., et al., 2011. Environmental factors of spatial distribution of soil salinity on flat irrigated terrain. *Geoderma* 163 (1–2), 55–62.
- Alexakis, D.D., Mexis, F.D.K., Vozinaki, A.E.K., et al., 2017. Soil moisture content estimation based on sentinel-1 and auxiliary earth observation products. A hydrological approach. *Sensors* 17 (6), 1455.
- Albed, A., Kumar, L., Aldakheel, Y.Y., 2014. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: applications in a date palm dominated region. *Geoderma* (230–231), 1–8.
- Aubert, M., Baghdadi, N., Zribi, M., et al., 2011. Analysis of TerraSAR-X data sensitivity to bare soil moisture, roughness, composition and soil crust. *Remote Sens. Environ.* 115 (8), 1801–1810.
- Baghdadi, N., Boyer, N., Todoroff, P., et al., 2009. Potential of SAR sensors TerraSAR-X, ASAR/ENVISAT and PALSAR/ALOS for monitoring sugarcane crops on Reunion Island. *Remote Sens. Environ.* 113 (8), 1724–1738.
- Bakr, N., Ali, R.R., 2019. Statistical relationship between land surface altitude and soil salinity in the enclosed desert depressions of arid regions. *Arab. J. Geosci.* 12 (23), 715.
- Begu, J., Fugstad, G.A., Mansuy, N., et al., 2017. Predicting soil properties in the Canadian boreal forest with limited data: comparison of spatial and non-spatial statistical approaches. *Geoderma* 306, 195–205.
- Breiman, L., Friedman, J., Olshen, R., et al., 1984. Classification and regression trees. *Encyclopedia of Ecology* 40 (3), 582–588.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Castaldi, F., Hueni, A., Chabrilat, S., et al., 2019. Evaluating the capability of the Sentinel-2 data for soil organic carbon prediction in croplands. *ISPRS-J. Photogramm. Remote Sens.* 147, 267–282.
- Chang, C.W., Laird, D.A., Mausbach, M.J., et al., 2001. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* 65 (2), 480–490.
- Chen, S., Liang, Z., Webster, R., et al., 2019. A high-resolution map of soil pH in China made by hybrid modelling of sparse soil data and environmental covariates and its implications for pollution. *Sci. Total Environ.* 665, 273–283.
- Clevers, J.G.P.W., 1989. Application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. *Remote Sens. Environ.* 29 (1), 25–37.
- Clevers, J.G.P.W., Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. *Int. J. Appl. Earth Obs. Geoinf.* 23, 344–351.
- Crippen, R.E., 1990. Calculating the vegetation index faster. *Remote Sens. Environ.* 34 (1), 71–73.
- Daughtry, C.S.T., Walthall, C.L., Kim, M.S., et al., 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* 74 (2), 229–239.
- De Bernardis, C., Vicente-Guijalba, F., Martinez-Marin, T., et al., 2016. Contribution to real-time estimation of crop phenological states in a dynamical framework based on NDVI time series: data fusion with SAR and temperature. *IEEE J. Sel. Top. Appl. Earth Observ. Remote.* 9 (8), 3512–3523.

- Ding, J.L., Yu, D.L., 2014. 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* 235, 316–322.
- Dong, W., Wu, T.J., Luo, J.C., et al., 2019. Land parcel-based digital soil mapping of soil nutrient properties in an alluvial-diluvia plain agricultural area in China. *Geoderma* 340, 234–248.
- Drusch, M., Del Bello, U., Carlier, S., et al., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* 120, 25–36.
- El Harti, A., Lhissou, R., Chokmani, K., et al., 2016. Spatiotemporal monitoring of soil salinization in irrigated Tadla Plain (Morocco) using satellite spectral indices. *IEEE J. Sel. Top. Appl. Earth Observ.* 50, 64–73.
- ESA, 2015a. ESA introducing sentinel-1 [2021-06-08]. [http://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Sentinel-1/Introducing\\_Sentinel-1](http://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Introducing_Sentinel-1).
- ESA, 2015b. ESA introducing sentinel-2 [2021-06-08]. [http://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Sentinel-2/Introducing\\_Sentinel-2](http://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Introducing_Sentinel-2).
- Farahmand, N., Sadeghi, V., 2020. Estimating soil salinity in the dried lake bed of urmia lake using optical Sentinel-2 images and nonlinear regression models. *J. Indian Soc. Remote Sens.* 48 (4), 675–687.
- Fathizad, H., Hakimzadeh Ardakani, M.A., Sodaiezhadeh, H., et al., 2020. Investigation of the spatial and temporal variation of soil salinity using random forests in the central desert of Iran. *Geoderma* 365, 114233.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29 (5), 1189–1232.
- Galin, E., Guerin, E., Peytavi, A., et al., 2019. A review of digital terrain modeling. *Comput. Graph. Forum* 38 (2), 553–577.
- Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour. Res.* 39 (12), 291–297.
- Gholizadeh, A., Zizala, D., Saberioon, M., et al., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens. Environ.* 218, 89–103.
- Gitelson, A.A., Merzlyak, M.N., 1998. Remote sensing of chlorophyll concentration in higher plant leaves. *Adv. Space Res.* 22 (5), 689–692.
- Guan, X.Y., Wang, S.L., Gao, Z.Y., et al., 2013. Dynamic prediction of soil salinization in an irrigation district based on the support vector machine. *Math. Comput. Model.* 58 (3–4), 719–724.
- Guyot, G., Frederic, B., Major, D., 1988. High spectral resolution: determination of spectral shifts between the red and the near infrared. *Int. Arch. Photogramm. Remote Sens.* 11, 750–760.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., et al., 2017. SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* 12 (2), e0169748.
- Hoa, P.V., Giang, N.V., Binh, N.A., et al., 2019. Soil salinity mapping using SAR Sentinel-1 data and advanced machine learning algorithms: a case study at ben tre province of the mekong river delta (Vietnam). *Rem. Sens.* 11 (2), 128.
- Holah, N., Baghdadi, N., Zribi, M., et al., 2005. Potential of ASAR/ENVISAT for the characterization of soil surface parameters over bare agricultural fields. *Remote Sens. Environ.* 96 (1), 78–86.
- Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25 (3), 295–309.
- Ivushkin, K., Bartholomeus, H., Bregt, A.K., et al., 2019. Global mapping of soil salinity change. *Remote Sens. Environ.* 231, 111260.
- Jordan, C.F., 1969. Derivation of leaf-area index from quality of light on the forest floor. *Ecology* 50 (4), 663–666.
- Kalambukattu, J.G., Kumar, S., Arya Raj, R., 2018. Digital soil mapping in a Himalayan watershed using remote sensing and terrain parameters employing artificial neural network model. *Environ. Earth Sci.* 77 (5), 203.
- Kasischke, E.S., Melack, J.M., Dobson, M.C., 1997. The use of imaging radars for ecological applications—a review. *Remote Sens. Environ.* 59 (2), 141–156.
- Lamichhane, S., Kumar, L., Wilson, B., 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: a review. *Geoderma* 352, 395–413.
- Li, W.J., Fang, H.Y., Qin, G.X., et al., 2020. Concentration estimation of dissolved oxygen in Pearl River Basin using input variable selection and machine learning techniques. *Sci. Total Environ.* 731, 139099.
- Loiseau, T., Chen, S., Mulder, V.L., et al., 2019. Satellite data integration for soil clay content modelling at a national scale. *Int. J. Appl. Earth Obs. Geoinf.* 82, 101905.
- Ma, C., 2018. Quantitative retrieval of soil salt content based on Sentinel-1 dual polarization radar image. *Trans. Chin. Soc. Agric. Eng.* 34 (2), 153–158 (in Chinese).
- Malenovsky, Z., Rott, H., Cihlar, J., et al., 2012. Sentinels for science: potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sens. Environ.* 120, 91–101.
- Metternicht, G.I., Zinck, J.A., 2003. Remote sensing of soil salinity: potentials and constraints. *Remote Sens. Environ.* 85 (1), 1–20.
- Muller, S.J., van Niekerk, A., 2016. An evaluation of supervised classifiers for indirectly detecting salt-affected areas at irrigation scheme level. *Int. J. Appl. Earth Obs. Geoinf.* 49, 138–150.
- Nawar, S., Buddenbaum, H., Hill, J., et al., 2014. Modeling and mapping of soil salinity with reflectance spectroscopy and landsat data using two quantitative methods (PLSR and MARS). *Rem. Sens.* 6 (11), 10813–10834.
- Page, K.L., Dalal, R.C., Pringle, M.J., et al., 2013. Organic carbon stocks in cropping soils of Queensland, Australia, as affected by tillage management, climate, and soil characteristics. *Soil Res.* 51 (7–8), 584–595.
- Peng, J., Biswas, A., Jiang, Q., et al., 2019. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma* 337, 1309–1319.
- Qi, J., Chehbouni, A., Huete, A.R., et al., 1994. A modified soil adjusted vegetation index. *Remote Sens. Environ.* 48 (2), 119–126.
- Ramos, T.B., Castanheira, N., Oliveira, A.R., et al., 2020. Soil salinity assessment using vegetation indices derived from Sentinel-2 multispectral data. application to Lezíria Grande. Portugal. *Agric. Water Manage.* 241, 106387.
- Roelofs, H.D., van Bodegom, P.M., Kooistra, L., et al., 2015. An evaluation of remote sensing derived soil pH and average spring groundwater table for ecological assessments. *Int. J. Appl. Earth Obs. Geoinf.* 43, 149–159.
- Rudiyanto, Minasny, B., Setiawan, B.I., et al., 2018. Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands. *Geoderma* 313, 25–40.
- Ruecker, G.R., Agyare, A., Akramhanov, A., et al., 2009. Influence of grid cell size and flow routing algorithm on soil-landform modeling. *Journal of the Korean Geographical Society* 44 (2), 122–145.
- Schuler, U., Herrmann, L., Ingwersen, J., et al., 2010. Comparing mapping approaches at subcatchment scale in northern Thailand with emphasis on the Maximum Likelihood approach. *Catena* 81 (2), 137–171.
- Shahabi, M., Jafarzadeh, A.A., Neyshabouri, M.R., et al., 2017. Spatial modeling of soil salinity using multiple linear regression, ordinary kriging and artificial neural network methods. *Arch. Agron Soil Sci.* 63 (2), 151–160.
- Shen, Q.S., Wang, Y., Wang, X.R., et al., 2019. Comparing interpolation methods to predict soil total phosphorus in the mollisol area of northeast China. *Catena* 174, 59–72.
- Song, C.Y., Ren, H.X., Huang, C., 2016. Estimating soil salinity in the Yellow River delta, eastern China—an integrated approach using spectral and terrain indices with the generalized additive model. *Pedosphere* 26 (5), 626–635.
- Sugimori, Y., Funakawa, S., Pachikin, K.M., et al., 2008. Soil salinity dynamics in irrigated fields and its effects on paddy based rotation systems in southern Kazakhstan. *Land Degrad. Dev.* 19 (3), 305–320.
- Taghadosi, M.M., Hasanlou, M., Eftekhari, K., 2019. Soil salinity mapping using dual-polarized SAR Sentinel-1 imagery. *Int. J. Rem. Sens.* 40 (1), 237–252.
- Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., et al., 2014. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma* 213, 15–28.
- Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8 (2), 127–150.
- Vaudour, E., Gomez, C., Fouad, Y., et al., 2019. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sens. Environ.* 223, 21–33.
- Vega, F.A., Matías, J.M., Andrade, M.L., et al., 2009. Classification and regression trees (CARTs) for modelling the sorption and retention of heavy metals by soil. *J. Hazard. Mater.* 167 (1), 615–624.
- Wang, J.Z., Ding, J.L., Yu, D.L., et al., 2019. 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* 353, 172–187.

- Wang, J.Z., Ding, J.L., Yu, D.L., et al., 2020. 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.* 707, 136092.
- Wei, L.F., Yuan, Z.R., Yu, M., et al., 2019. Estimation of arsenic content in soil based on laboratory and field reflectance spectroscopy. *Sensors* 19 (18), 3904.
- Yahiaoui, I., Douaoui, A., Zhang, Q., et al., 2015. Soil salinity prediction in the Lower Chelif plain (Algeria) based on remote sensing and topographic feature analysis. *J. Arid Land* 7 (6), 794–805.
- Yang, R.M., Zhang, G.L., Liu, F., et al., 2016. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecol. Indic.* 60, 870–878.
- Yang, R.M., Guo, W.W., 2019. Using time-series Sentinel-1 data for soil prediction on invaded coastal wetlands. *Environ. Monit. Assess.* 191, 462.
- Yao, Y., Ding, J.L., Zhang, F., et al., 2013. Monitoring of soil salinization in Northern Tarim Basin, Xinjiang of China in dry and wet seasons based on remote sensing. *Chin. J. Appl. Ecol.* 24 (11), 3213–3220 (in Chinese).
- Yi, Q.X., 2019. Remote estimation of cotton LAI using Sentinel-2 multispectral data. *Trans. Chin. Soc. Agric. Eng.* 35 (16), 189–197 (in Chinese).
- Yin, L.L., Zhang, H.H., Zhou, X., et al., 2020. KAML: improving genomic prediction accuracy of complex traits using machine learning determined parameters. *Genome Biol.* 21 (1), 146.
- Zamani Joharestani, M., Cao, C.X., Ni, X.L., et al., 2019. PM<sub>2.5</sub> prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. *Atmosphere* 10 (7), 373, 2019.
- Zhang, T.T., Zeng, S.L., Gao, Y., et al., 2011. Using hyperspectral vegetation indices as a proxy to monitor soil salinity. *Ecol. Indic.* 11 (6), 1552–1562.
- Zhang, Z.P., Ding, J.L., Wang, J.Z., et al., 2020a. Prediction of soil organic matter in northwestern China using fractional-order derivative spectroscopy and modified normalized difference indices. *Catena* 185, 104257.
- Zhang, Z.P., Ding, J.L., Zhu, C.M., et al., 2020b. Combination of efficient signal pre-processing and optimal band combination algorithm to predict soil organic matter through visible and near-infrared spectra. *Spectrosc. Acta Pt. A-Molec. Biomolec. Spectr.* 240 (15), 118553.
- Zhang, Z.P., Ding, J.L., Zhu, C.L., et al., 2021. Strategies for the efficient estimation of soil organic matter in salt-affected soils through Vis-NIR spectroscopy: optimal band combination algorithm and spectral degradation. *Geoderma* 382, 114729.
- Zhang, Z.T., Du, Y.Y., Lao, C.C., et al., 2020. Inversion model of soil salt content in different depths based on radar remote sensing. *Trans. Chin. Soc. Agric. Mach.* 1–15 (in Chinese).
- Zheng, Z., Zhang, F.R., Ma, F.Y., et al., 2009. Spatiotemporal changes in soil salinity in a drip-irrigated field. *Geoderma* 149 (3–4), 243–248.
- Zhou, T., Zhao, M.F., Sun, C.L., et al., 2018. Exploring the impact of seasonality on urban land-cover mapping using multi-season Sentinel-1A and GF-1 WFV images in a subtropical monsoon-climate region. *ISPRS Int. J. Geo-Inf.* 7 (1), 3.
- Zhou, T., Geng, Y.J., Chen, J., et al., 2020. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. *Sci. Total Environ.* 729, 138244.