International Journal of Computational Intelligence Systems and Applications. Volume 10, Issue 1, (2023) 16-22
http://www.cayley-nielson.com/ijcisa/

Remote Sensing Inversion of Soil Salinity Considering Environmental Factors

Chao Hui[1], Gao Yunbao[2*], Zhang Bin[3]
[1]North China University of Science and Technology, Tangshan, China
[2]Rizhao Natural Resources and Planning Bureau, Rizhao, China
[3]Huarun Electric Power North China Region (Tangshan Company), Tangshan, China
Email: *[1]hchao@ncst.edu.cn,[2]996518122@qq.com,[3]1034878268@qq.com*

Abstract

Soil desalinization is one of the most important environmental problems facing the world at present, which poses a serious threat to global agricultural production and seriously affects local agricultural development and ecological construction. The phenomenon of soil desalinization in coastal zone is becoming more and more serious. Understanding the spatial distribution information of soil salinity in coastal zone can provide decision-making basis for land management, ecological restoration and sustainable development of agriculture in this area. Based on Sentinel-1/2 image data and The height data of ASTERGDEM with 30m resolution, combined with the measured data obtained synchronously from the field and the satellite, adopt Stepwise Regression (SR), Random Forest (RF), Support Vector Machine (SVM), SVM) and Partial Least Squares Regression (PLSR) are used to study the inversion of soil salt content, and the spatial distribution map of surface soil salt in the study area is drawn with the optimal feature combination and method.

Keywords: Remote sensing inversion, land desalinization, remote sensing monitoring, normalization processing.

1. Introduction

Soil desalinization refers to the process of gradually accumulating soluble salt in soil and finally forming high-salinity soil due to the comprehensive influence of human activities and environmental factors, in which human activities are mainly unreasonable irrigation and fertilization, while environmental factors include local topography, parent materials, climate and hydrology. Soil desalinization, as a common phenomenon of soil degradation, will not only affect the growth environment of crops, and then affect the output of crops, limit the sustainable development of local agriculture, but also destroy the local ecological balance, thus leading to the deterioration of the local ecological environment and the possibility of ecological crisis, which has become an environmental problem that needs to be solved urgently.

Microwave remote sensing is a common means of obtaining surface information, which has good penetrability and all-weather detection ability, makes up for the deficiency of optical remote sensing, and has certain advantages in soil composition prediction. TAGHADOSI Support Vector Regression (SVR) analysis based on texture is applied to Sentinel-1 SAR data to draw salinity, and the method of directly correlating radar intensity with soil salinity is studied, and good inversion results of soil salinity are obtained. By using the back scattering coefficient extracted from Radarsat-2 image and soil salinity measurement data.

2. Research methods and technical route

The inversion model of surface soil salt content in Theta tank farm in Inner Mongolia was established. LASNE et al. research shows that in micro. The wave frequency is in the range of 1~7GHz (the center frequency of Sentinel-1 image data is 5.404GHz), and the imaginary part. It is sensitive to soil salinity, while the solid part has a greater relationship with water content. Therefore, the effective use of parametric SAR data can provide a new way to obtain soil salinity information in the study area in a wide range, and can provide technical support for agricultural production practice in the study area dynamically and timely.

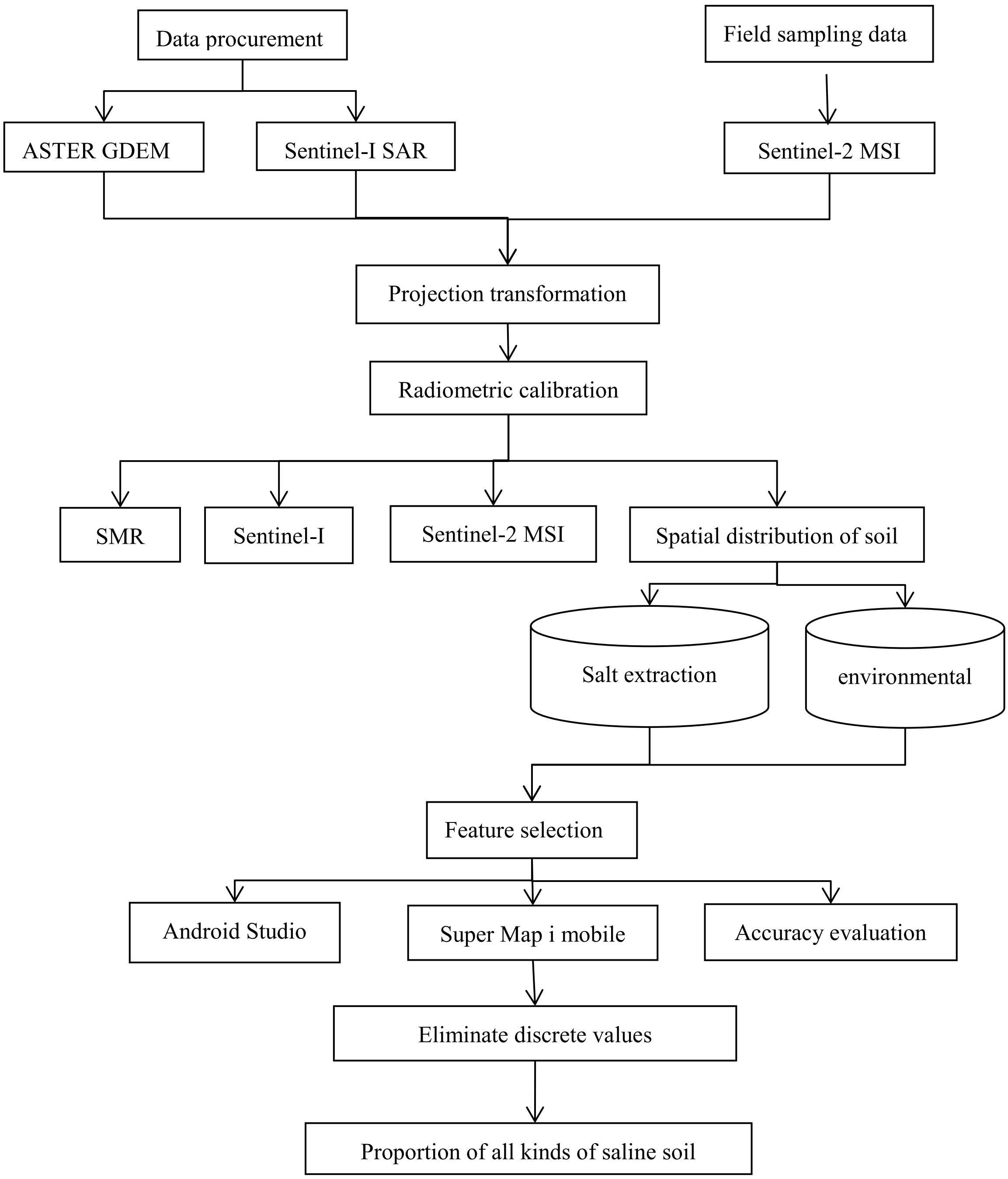

Fig. 1 Research technology road map

Based on the Sentinel-1 SAR radar image, Sentinel-2 MSI optical image and 30m DEM elevation data, this paper extracts the radar image characteristics, salinity vegetation index and topographic factors, and establishes the soil salinity characteristic data set in the study area by combining with the field soil salinity measurement data of the same period. Firstly, the discrete values in the data set are put forward. According to the method of cross-validation, this data set is divided into training set and validation set, and then the driving factors of soil salinity in the study area are sorted and screened by the method of feature importance of random forest model, and the features that contribute the most to soil salinity in the study area are extracted. Then, on the basis of the optimal features selected, the relative accuracy of soil salinity inversion in the study area based on different data source features and their combinations is further analyzed and compared, and the most suitable data source feature combination for soil salinity inversion in the study area is found.

Based on the best feature combination of multi-source data, this paper analyzes and compares the model performance and prediction effect based on random forest model, support vector regression, multiple stepwise regression and partial least squares regression, and uses the best model to predict the soil salinity in the study area, so as to estimate the soil salinity in a large area in this area and count the proportion of saline soil in various degrees. In this paper, the following data processing and methods are advanced. Its main stages and research technical route are shown in Figure1.

3. Project implementation

In order to analyze and model the measured data of soil salt content and the selected characteristics, this paper adopts the method of machine learning. At present, the commonly used methods for monitoring soil salt content are stepwise multiple regression. Stepwise multiple regression (SMR), Support vector regression (SVR), Partial least squares regression (PLSR) and Random forest regression (RFR). In this paper, the above four methods are used to model and analyze the soil salt content in this area.

(1) Stepwise multiple regression

Stepwise multiple regression (SMR) is a multiple linear regression model with screening variables, which can well eliminate unimportant parameters in variables and solve the problem of high collinearity in variables. Its operation process is to import the independent variables into the model one by one, and each imported independent variable has to pass the F test, and the selected variables are tested one by one.

Because the imported new variables make the old variables no longer significant, the old variables will be automatically eliminated, so that the optimal variable set model will be obtained. The common methods of stepwise regression are forward method and backward method, and the forward method is used here, that is, variables are added one by one from less to more until there are no variables to add.

(2) Random forest

Random forest regression (RFR) is an integrated learning algorithm which was put together by BREIMAN in 2001. This algorithm has the advantages of nonlinear mining ability, good anti-noise ability, data distribution cannot meet any assumptions, strong adaptability to data sets and fast training speed.

Random forest randomly samples the original training set into n training subsets by bootstrap aggregation algorithm, and randomly selects k features ($K<n$) from each training set. According to these k features, M sub-decision trees are established repeatedly and their prediction results are obtained. Finally, the classification model is voted, and the model with the highest number of votes is selected as the final decision. The specific process is as follows.

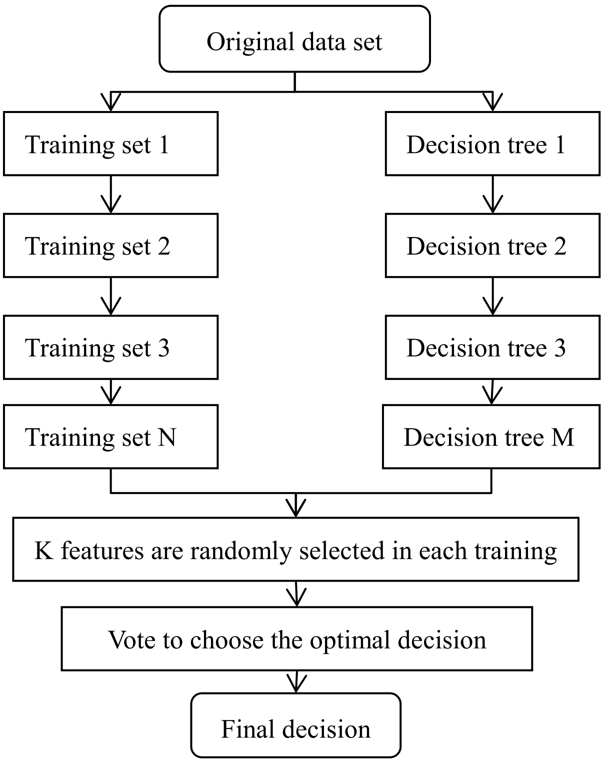

Fig. 2 Random forest flow chart

4. Data source introduction

The salt content of soil is closely related to the formation of soil. According to the theory of soil formation, the formation of soil is influenced by parent material, climate, organism, relief and time. The specific selected parameters are shown in Table 1.

Table 1. Characteristic variables derived from Sentinel-1/2 and DEM

Feature type Parameter quantity	Candidate variable Number	Candidate variable Number
H/A/a polarization decomposition parameter	17	Alpha,Entropy,Anisotropy, delta,lambda, Alphal, Alpha2
Back scattering coefficient	2	Soil-regulated vegetation index SAVI. Vegetation soil salinity index VSSI, normalization
Salinity vegetation index	14	Canopy salinity responds to vegetation index CRSI, brightness index BI and salinity index
DEM derivative factor	18	Set RSP, elevation DEM, slope AS, slope s, fluctuation QFD, roughness，Degree RS, plane curvature SC

The comprehensive influence of five soil-forming factors. Therefore, considering the factors that may affect the soil salt content will greatly improve the accuracy of the experimental model. The back scattering coefficient of radar data has a certain relationship with dielectric constant, and the change of dielectric constant will affect the change of soil salt content.

On the basis of retaining the advantages of radar data, this paper will combine various environmental factors to retrieve the soil salinity information in this area. Terrain factor is the most commonly used environmental factor, and the terrain factor data in this paper is extracted from DEM data with 30M resolution provided by geospatial data cloud platform. Biological factors such as normalized vegetation index (NDVI).

Table 2. Importance Ranking of Characteristic Variables

Importance ranking	parameter	Importance ranking	Band
1	CRSI	0.572	3
2	SI6	0.524	5
3	SI5	0.238	8
4	AH	0.226	4
5	VV	0.152	6
6	VD	0.053	1

The vegetation indices are extracted from Sentinel-2 image data, and the parent material and soil elements are also based on. Salinity indexes were calculated from Sentinel-2 image data. Therefore, a total of 17 H/A/a polarization decomposition parameters, 2 back scattering coefficients, 13 salinity vegetation indices and 18 topographic indices were selected as the characteristic variables for soil salinity prediction. The specific selected parameters are shown in Table 2.

The higher the importance (VI) of the variable, the closer the relationship between this characteristic and soil EC. Therefore, you can root. According to VI value, some characteristics closely related to EC value are selected.

The visual images of the gray values of these features are displayed in ARCGIS environment, and the comparative value of pixels can reveal the response of soil salt characteristics, as shown in Figure 3. Among these 10 features in this study area, polarization decomposition parameter Entropy, back scattering coefficient VV and optical index. CRSI, SI5 and SI6 have more spatial and texture information and richer visual information, while the rest of the terrain. The higher the importance (VI) of factor visual belief variable, the closer the relationship between this feature and soil EC. Therefore, some features closely related to EC value can be selected according to VI value. The specific selected parameters are shown in Fig. 3.

In this chapter, four machine learning algorithms are used to model, analyze and compare the characteristic variables extracted from different data sources, and in order to improve the accuracy of model prediction, all factors are comprehensively considered in the modeling and prediction of soil salt content in the study area. The results show that, The overall performance of the model that comprehensively considers the remote sensing features extracted from remote sensing image data and terrain factors extracted from DEM elevation data to predict the soil salt content in the study area is better than that predicted by remote sensing features or terrain factors respectively, and the performance of the model predicted by remote sensing features is better than that predicted by terrain factors. In the process of predicting soil salt content using all parameters, the best model is PLSR algorithm, followed by PLSR algorithm.

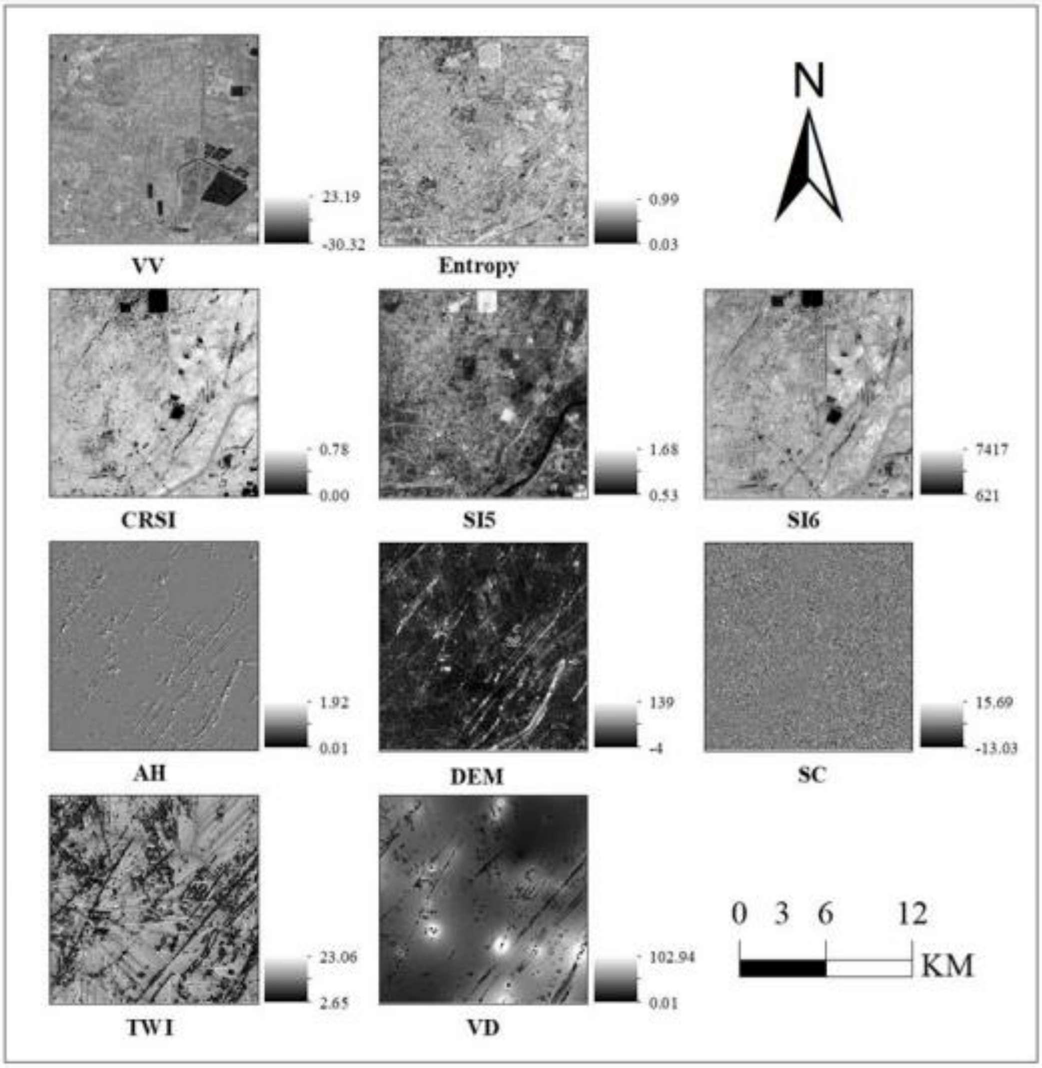

Fig. 3 Visualization of feature

RFR algorithm and SMR algorithm, SVR algorithm has the weakest prediction effect on soil salt content in this process. Finally, according to the best combination of characteristic parameters and the best performance model, the spatial distribution map of soil salt in the study area is generated, and the distribution law of soil salt is summarized and analyzed in detail.

References

[1] Loozen Y, Rebel K T, De Jong S M, et al. Mapping canopy nitrogen in European forests using remote sensing and environmental variables with the random forests method. Remote Sensing of Environment, 2020, 247(7):11.

[2] HOA P V, GIANG N V, BINH N A, et al. 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). Remote Sensing, 2019, 11(2):21.

[3] Friedman J H, Meulman J J. Multiple additive regression trees with application in epidemiology Stat Med, 2003, 22(9):1365-1381.

[4] Zeng W Z, Zhang D Y, Fang Y H, et al. 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(2):4-6.

[5] Fan X W, Liu Y B, Tao J M, et al. Soil Salinity Retrieval from Advanced Multi-Spectral Sensor with Partial Least Square Regression. Remote Sensing, 2015, 7(1):488-511.

[6] Rodriguez-Febereiro M, Dafonte J, Fandino M, et al. Evaluation of Spectroscopy and Methodological Pre-Treatments to Estimate Soil Nutrients in the Vineyard. Remote Sensing, 2022, 14(6):16-26.

[7] Wang K, Qi Y B, Guo W J, et al. Retrieval and Mapping of Soil Organic Carbon UsingSentinel-2A Spectral Images from Bare Cropland in Autumn. Remote Sensing, 2021, 13(6):72-91.

[8] Huang X, Shi Z H, Zhu H D, et al. Soil moisture dynamics within soil profiles and associated environmental controls. Catena, 2016, 136(6):189-196.

[9] Shen J Q, Shuai Y M, Li P X, et al. Extraction and Spatio-Temporal Analysis of Impervious Surfaces over Dongying Based on Landsat Data. Remote Sensing, 2021, 13(18):3666-3688.