Postprint of Soil Salinity Prediction Based on Random Forest Algorithm

Authors: Xiaoyu Li, Jia Keli, Wei Huimin, Ruihua Chen, Wang Yijing

Date: 2023-08-26T00:00:00+00:00

Abstract

Rapid monitoring of regional soil salinization information is of great significance for salinization management and ecological environment protection. This study employs Sentinel-2A and Landsat 8 OLI remote sensing imagery as data sources, takes the Yinchuan Plain as the study area, and utilizes the Google Earth Engine (GEE) platform to estimate soil salt content based on the random forest algorithm by establishing the relationship between spectral index features and ground-measured soil salt content. The results demonstrate that GEE can provide reliable data support for soil salt content prediction; the random forest model established using Sentinel-2A as the data source achieves superior prediction accuracy (R2=0.789, RMSE=1.487) compared to Landsat 8 OLI, and can be applied to high-resolution remote sensing estimation of soil salt content, thereby providing theoretical support for large-scale soil salt content monitoring efforts.

Full Text

Abstract

Rapid monitoring of regional soil salinization information is of great significance for salinization control and ecological environment protection. This paper uses remote sensing imagery as the data source and the Yinchuan Plain as the study area, employing the Google Earth Engine (GEE) platform and Landsat-8 OLI and Sentinel-2A satellite data to estimate soil salt content based on the random forest algorithm. The relationship between spectral index characteristics and ground-measured soil salt content was established for soil salinity estimation. The results demonstrate that Google Earth Engine can provide reliable data support for soil salt content prediction. The random forest model established with Sentinel-2A as the data source exhibits better prediction accuracy (R^2 = 0.789, RMSE = 1.487) and can be used for high-resolution remote sensing estimation of soil salt content, providing theoretical support for large-scale soil salinity monitoring.

Keywords: soil salt content; Google Earth Engine; random forest; prediction; Yinchuan Plain

1. Introduction

Soil salinization represents a critical ecological and environmental issue facing global arid and semi-arid regions. Worldwide, the area affected by salinization exceeds 10^8 hectares, with China' s saline soil area surpassing 10^7 hectares and increasing at approximately 1.5×10^{6} hectares annually. Influenced by global climate warming and human activities, soil salinization has become a key factor limiting agricultural development and sustainable resource utilization, representing a primary cause of land degradation that severely impacts regional ecological stability, food security, and sustainable agricultural development. Consequently, rapid, accurate, and large-scale monitoring of soil salinization is essential for agricultural production and ecological environmental protection.

Traditional soil salinity monitoring methods are time-consuming and labor-intensive. Remote sensing technology offers significant advantages for large-scale monitoring compared to conventional approaches and plays a crucial role in global and regional soil salinization monitoring. With the development of remote sensing technology, data from various sources provide support for soil salinization monitoring, and establishing effective methods for extracting salinization information is key to rapidly obtaining such data. In recent years, the random forest algorithm has been gradually applied to soil salinity estimation due to its superiority, efficiency, robustness, and accuracy. By establishing complex nonlinear relationships between measured soil salinity and spectral information, the algorithm fully utilizes sensor-derived spectral data to improve estimation accuracy.

Regarding platform applications, Google Earth Engine (GEE) integrates data acquisition, storage, processing, and analysis, enabling efficient preprocessing and computational analysis of massive image datasets. Scholars have utilized this platform for research on land use and crop classification algorithms, crop yield estimation, biomass estimation, and dynamic monitoring of water bodies and forest changes. These studies demonstrate that the GEE platform can efficiently process large volumes of image data, though few have attempted to apply it to soil salinity monitoring. This study employs JavaScript programming in GEE' s online editor, using the Yinchuan Plain in Ningxia as an example, to estimate soil salt content by constructing a random forest model based on spectral indices derived from Landsat-8 OLI and Sentinel-2A satellite imagery. The objective is to enrich the application of random forest algorithms and the GEE platform in soil salinization monitoring, providing technical support for accurate prediction and monitoring of soil salinization information in irrigation districts.

1.1 Study Area Overview

The Yinchuan Plain (37.83°-39.38°N, 104.28°-107.65°E) is located in northern Ningxia Hui Autonomous Region [Figure 1: see original paper]. The plain extends approximately 165 km from north to south and 10–50 km from east to west, covering an area of 7,615 km^2. Situated in the middle and upper reaches of the Yellow River irrigation region, between the alluvial fan at the eastern foothills of the Helan Mountains and the plain itself, it serves as a pilot zone for ecological protection and high-quality development in the Yellow River basin and represents a typical area of soil salinization in northwestern arid and semi-arid regions. The study area features a temperate arid climate with warm temperate monsoon characteristics, an average annual temperature of 8–9°C, annual precipitation of 150–203 mm, and annual evaporation of 2,500–3,100 hours. Characterized by poor drainage, shallow groundwater depth, strong evaporation, concentrated water-salt conditions, high terrain, and irrational irrigation, it is one of the most severely salinized areas in Ningxia. The main crops include rice, wheat, and corn.

1.2 Soil Salinity Data Acquisition

Soil samples were collected in May 2021. A 5 km × 5 km grid was established across the study area for sampling point placement. At each sampling point, a plum-blossom sampling method was employed within a 30 m × 30 m range. Surface soil (0–20 cm) was mixed, and approximately 500 g was retained using the quartering method, sealed in bags, and brought back to the laboratory. Sampling point numbers, coordinate information, land use, and vegetation growth conditions were recorded simultaneously. A total of 120 soil samples were collected from 105 sampling points. After removing weeds and gravel, all soil samples were air-dried, ground, and passed through a 1 mm sieve. Salt content was measured using the electrical conductivity method with a 5:1 water-to-soil ratio. After removing outliers, 105 surface soil samples (0–20 cm) were obtained for analysis.

1.3 Data Processing

To eliminate dimensional effects among different indicators, data were normalized using the normalization function to scale all input data between 0 and 1. The normalization formula is: $x' = (x_i - x_{min})/(x_{max} - x_{min})$, where x_i is the measured sample value, x_{max} is the maximum sample value, x_{min} is the minimum sample value, and x' is the normalized data.

1.4 Remote Sensing Image Acquisition and Processing

Landsat-8 OLI and Sentinel-2A Level-2A surface reflectance data were obtained from the GEE platform (https://code.earthengine.google.com). The data datum is WGS84 with a map projection of UTM. The imaging time range was set from April 15 to May 15, 2021, corresponding to the field sampling period. Since the Yinchuan Plain is composed primarily of alluvial-pluvial plains with low-lying terrain and poor drainage, image mosaicking on the GEE platform can cause overlap and discontinuity issues. Therefore, based on the set time range, the median value of the image collection was calculated using the median function, and the image with the smallest difference from this median image was selected for mosaicking. Cloud masking was performed using the QA band integrated in the data for image quality assessment. After mosaicking, six commonly used bands (Green, Red, NIR, SWIR1, SWIR2) were selected for band composition, and the study area image data were obtained by clipping with the study area vector boundary. All operations were performed on the GEE platform using JavaScript API programming.

2. Methods

2.1 Spectral Index Selection

In arid and semi-arid regions, spectral indices are effective methods for monitoring soil salinization. Previous studies indicate that scholars commonly use vegetation indices and salinity indices when selecting spectral parameters. Since vegetation indices alone cannot accurately reflect spectral information in sparsely vegetated areas, numerous studies have combined vegetation and salinity indices to invert soil salinization. Considering the study area' s location in an arid/semi-arid region, this study integrated both vegetation and salinity indices. Salinity indices include SI, SI_1, SI_2, SI_3, and SRI, while vegetation indices include SAVI, GDVI, CSI, and NDSI. The calculation formulas for each spectral index are shown in . Based on the study area imagery, spectral indices were constructed by calling the getIndex function on the GEE platform, selecting red, blue, green, NIR, and SWIR bands according to the formulas for calculation and extraction. All spectral index extraction operations were performed using JavaScript in the GEE platform' s online code editor.

2.2 Sensitive Spectral Parameter Screening

Due to collinearity among independent variables, directly inputting numerous variables into modeling can cause regression coefficient errors due to minor differences in sample data, reducing model stability and affecting accuracy. To eliminate adverse effects from excessive input variables and multicollinearity, Pearson correlation analysis and significance testing of the 10 spectral parameters under different data sources were conducted using the Hmisc package (version 4.5-0) in R software. Parameters passing the significance test were selected as sensitive spectral parameters for model construction.

2.3 Model Construction and Evaluation

2.3.1 Random Forest Model Construction The core concept of random forest is a machine learning algorithm that performs bootstrap sampling on training sets to generate multiple training sets, creates a decision tree from each training set, and combines all decision trees to form a random forest for sample training and prediction. The method combines multiple decision trees to create a random forest. The steps are as follows: First, N training sets S_1, S_2, ···, S are randomly generated from the training sample set D. Second, a corresponding decision tree f_1, f_2, ···, f is generated for each training set. Before selecting attributes at each non-leaf node, m attributes (m « M) are randomly extracted from all M attributes as the current node' s splitting attribute set, and the best splitting attribute is selected for node splitting. Third, the generated multiple decision trees form a random forest. For test set sample X, each decision tree is used for testing, and the prediction result f(x) is obtained according to the formula.

Based on reference [45], the prediction formula is: $f(x) = (1/N)\sum_n f_n(x)$, where $f(x)$ represents the prediction result of the random forest model, and $f_n(x)$ represents the prediction result of a single decision tree. In this study, the ee.Classifier.smileRandomForest function was called on the GEE platform to implement regression prediction. Model performance was adjusted through two parameters: the number of decision trees (n) and the number of features used at nodes (m). Excessive decision trees affect model efficiency, while too few affect accuracy. After weighing the number of decision trees and considering both precision and efficiency, the number of decision trees was set to 100. The number of features used at nodes was set to the default value, i.e., the square root of the total number of input features.

2.3.2 Model Evaluation Metrics To quantify the prediction performance of the soil salinity inversion model, two commonly used metrics were calculated and output on the GEE platform: coefficient of determination (R^2) and root mean square error (RMSE). R^2 closer to 1 indicates better model fit and prediction performance, while smaller RMSE indicates better model performance.

3. Results

3.1 Soil Salinity Statistical Characteristics

Following Brady' s classification method [58], the 105 soil samples were divided into four salinization grades based on salt content. Descriptive statistical analysis of the samples is shown in . The results indicate relatively uniform distribution across different salinization grades, with an overall coefficient of variation exceeding 85%, showing strong variability and high dispersion, which demonstrates the samples' universal applicability. Soil samples were sorted by salt content from low to high, and 84 samples were selected as the training set for inversion model establishment using a 4:1 ratio, with the remaining 21 samples as the validation set for model verification.

3.2 Correlation Analysis Between Soil Salinity and Spectral Indices

The correlation matrix of 10 spectral parameters under different data sources was plotted using the GGally package in R software based on parameters passing significance tests. [Figure 2: see original paper] shows that for Landsat-8 OLI data, all spectral indices except NDSI have strong correlations with soil salinity. For Sentinel-2A data, all spectral indices show strong correlations with soil salinity, with SI_1 having the largest correlation coefficient, indicating strong correlation with soil salinity and suitability for soil salinity model construction.

3.3 Model Construction

Based on the random forest model with soil salinity content as the output layer, SI_1, GDVI, CSI, and NDSI were selected as input layer data for the model established with Landsat-8 OLI as the data source. For the model established with Sentinel-2A as the data source, SI_1, SI_2, SRI, GDVI, CSI, and NDSI were selected as input layer data. The soil salinity content was simulated, with results shown in . The R^2 value of the model established with Sentinel-2A is larger than that with Landsat-8 OLI, and the RMSE is smaller, indicating that the soil salinity prediction model based on Sentinel-2A is superior.

The scatter plot between measured and predicted values [Figure 3: see original paper] shows that models based on both data sources have good correlation between predicted and measured values. The random forest model based on Sentinel-2A shows greater R^2 and smaller RMSE, indicating better prediction accuracy than the model based on Landsat-8 OLI.

3.4 Soil Salinization Prediction

Using the random forest prediction model based on Sentinel-2A, soil salinization in the Yinchuan Plain was predicted and classified according to salinization degree. [Figure 4: see original paper] shows that soil salinity in the Yinchuan Plain generally presents a pattern of lighter salinity in the south and heavier in the north. This may be due to the plain' s topography being higher in the south and lower in the north, with the Yellow River flowing from south to north. The northern region has shallow groundwater depth, allowing salts to easily accumulate in the soil surface. Saline soils are mainly concentrated in the western part of Pingluo County, Dawukou District, and Huinong District in the northern plain, accounting for 89.85% of the total saline soil area. The southern region, as the Yellow River' s inlet carrying salt-based ions outward, has less soil salt accumulation. Non-salinized and lightly salinized soils are concentrated in Jinfeng District, Xingqing District, Yongning County, Lingwu City, Qingtongxia City, and Litong District in the southern plain, accounting for 71.59% of the total Yinchuan Plain area. The salinization prediction results are consistent with actual sampling conditions, indicating that the random forest model can be applied to soil salinity prediction in this region.

4. Discussion

Using spectral indices calculated from satellite image bands has become the most common method for monitoring soil salinization spatial distribution. Most scholars use soil spectral indices from satellite data' s visible bands and select those with high correlation with ground measurement data as input variables for model establishment, achieving good prediction results. This demonstrates that establishing spectral indices has certain advantages in soil salinity monitoring applications. Currently, two main modeling approaches exist: directly estimating soil salinity using salinity indices, or indirectly estimating soil salinity using vegetation indices. However, for more accurate monitoring of soil salinization, the spectral reflectance of halophytes must be considered, especially in severely salinized arid and semi-arid regions. Therefore, this study comprehensively considered both salinity and vegetation indices, establishing a random forest model based on Sentinel-2A data to predict soil salinity across the Yinchuan Plain. The prediction results show good consistency with field sampling conditions, demonstrating that the random forest model established with Sentinel-2A can be used for soil salinization prediction and generate results on the GEE platform.

Compared with commonly used linear regression methods for analyzing soil salinity distribution, machine learning algorithms can provide more reliable predictions of soil salinization spatial distribution, especially when field measurement data are limited. Moreover, compared with other machine learning algorithms, random forest can more accurately predict soil salinity distribution in salinized lands. However, current research still faces uncertainties from two main sources. First, the uncertain relationship between soil salinity characteristics and remote sensing data: due to the complex mechanism of soil salinization, soil salinity information obtained from satellite images may be affected by soil type, surface roughness, and other soil properties. The fundamental reason is the inherent limitation of satellite sensors in expressing qualitative soil indicators. Although remote sensing images have undergone radiometric and atmospheric correction, captured spectral information may still be affected by terrain factors. Second, uncertainty in random forest model establishment: for machine learning models, dataset quality and quantity significantly impact prediction results. This study' s relatively small sample size ($n = 105$) may introduce uncertainty in model establishment. Therefore, whether the established estimation model can be applied to other regions requires further research and validation.

5. Conclusions

This study compared the spectral characteristics of Landsat-8 OLI and Sentinel-2A images and evaluated their performance in predicting soil salt content using random forest models. The main findings are as follows:

1. The GEE platform provides large amounts of analytical data and algorithms from different sensors, facilitating accurate estimation of soil salinity distribution in the study area and promoting large-scale regional soil salinity monitoring.
2. The random forest model established with Sentinel-2A data demonstrated good prediction accuracy and model performance (R^2 = 0.789, RMSE = 1.487), with soil salinity prediction results superior to those based on Landsat-8 OLI data.
3. The random forest model based on Sentinel-2A can be used for soil salinization prediction, and its prediction results on the GEE platform are consistent with actual sampling conditions, indicating its applicability for soil salinity monitoring in this region.

References

[1] Li Xing, Li Kaifeng. Research progress of soil salinization development and utilization[J]. Agricultural Science and Technology Bulletin, 2018, 23(6): 244-246.

[2] Nachshon U. Cropland soil salinization and associated hydrology: Trends, processes and examples[J]. Water, 2018, 10(8): 1030.

[3] Zaman M, Shahid S A, Heng L. Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques[M]. Springer Nature, 2018.

[4] Yang Zhen, Wang Baoshan. Present situation of saline soil resources in China and its improvement and utilization strategy[J]. Shandong Agricultural Sciences, 2015, 47(4): 125-130.

[5] Cao Lei, Ding Jianli, Coranti Hallick, et al. Extraction and modeling of regional soil salinization based on domestic Gaofen-1 satellite data[J]. Acta Pedologica Sinica, 2016, 53(6): 1399-1409.

[6] Meng L, Zhou S, Zhang H, et al. Estimating soil salinity in different landscapes of the Yellow River Delta through Landsat OLI/TIRS and ETM+ Data[J]. Journal of Coastal Conservation, 2016, 20(4): 271-279.

[7] Farifteh J, Farshad A, George R J. Assessing salt affected soils using remote sensing, solute modelling, and geophysics[J]. Geoderma, 2006, 130(3-4): 191-207.

[8] Hu J, Peng J, Zhou Y, et al. Quantitative estimation of soil salinity using UAV borne hyperspectral and satellite multispectral images[J]. Remote Sensing, 2019, 11(7): 736.

[9] Metternicht G I, Zinck J A. Remote sensing of soil salinity: Potentials and constraints[J]. Remote sensing of Environment, 2003, 85(1): 1-20.

[10] Fan X, Weng Y, Tao J. Towards decadal soil salinity mapping using Landsat time series data[J]. International Journal of Applied Earth Observation and Geoinformation, 2016, 52: 32-41.

[11] Ivushkin K, Bartholomeus H, Bregt A K, et al. Global mapping of soil salinity change[J]. Remote Sensing of Environment, 2019, 231: 111-160.

[12] Huang Xiaoyu, Wang Xuemei, Kawuciati Baishan. Inversion of soil salt content in arid oasis based on Landsat8 OLI image[J]. Remote Sensing of Natural Resources, 2023, 35(1): 189-197.

[13] Yang Ning, Cui Wenxuan, Zhang Zhitao, et al. Inversion of soil salinity at different depths by UAV multi-spectral remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(22): 13-21.

[14] Chen Junying, Wang Xintao, Zhang Zhitao, et al. Soil salinization monitoring method based on UAV Satellite remote sensing scaling[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 161-169.

[15] Li Yanling, Zhao Gengxing, Chang Chunyan, et al. Soil salinity inversion model based on OLI and HSI image fusion[J]. Transactions of the CSAE, 2017, 33(21): 533-544.

[16] Zhang Zhitao, Wei Guangfei, Yao Zhihua, et al. Inversion model of soil salt content based on UAV multi-spectral remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 151-160.

[17] Fathizad H, Ardakhani M A H, Sodaiezadeh H, et al. Investigation of the spatial and temporal variation of soil salinity using random forests in the central desert of Iran[J]. Geoderma, 2020, 365: 114-226.

[18] Liu X, Hu G, Chen Y, et al. High resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine platform[J]. Remote Sensing of Environment, 2018, 209: 227-239.

[19] Simonetti D, Simonetti E, Szantoi Z, et al. First results from the phenology based synthesis classifier using Landsat 8 imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(7): 1496-1500.

[20] Xu Hanzeyu, Liu Chong, Wang Junbang, et al. Remote sensing extraction of citrus orchards in Southern Jiangxi Province supported by Google Earth Engine[J]. Journal of Geoinformation Science, 2018, 20(3): 396-404.

[21] Dong J, Xiao X, Menarguez M A, et al. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine[J]. Remote Sensing of Environment, 2016, 185: 142-154.

[22] Lobell D B, Thau D, Seifert C, et al. A scalable satellite-based crop yield mapper[J]. Remote Sensing of Environment, 2015, 164: 324-333.

[23] Tamiminia H, Salehi B, Mahdianpari M, et al. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review[J]. ISPRS Journal

of Photogrammetry and Remote Sensing, 2020, 164: 152-170.

[24] Hansen M C, Potapov P V, Moore R, et al. High-resolution global maps of 21st-century forest cover change[J]. Science, 2013, 342(6160): 850-853.

[25] Pekel J F, Cottam A, Gorelick N, et al. High-resolution mapping of global surface water and its long-term changes[J]. Nature, 2016, 540(7633): 418-422.

[26] Trianni G, Lisini G, Angiuli E, et al. Scaling up to national/regional urban extent mapping using Landsat data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(7): 3710-3719.

[27] Zhang Yu' anpei, Hu Keling, Li Baoguo, et al. Spatial distribution pattern of soil salinity and saline soil in Yinchuan Plain[J]. Transactions of the Chinese Society of Agricultural Engineering, 2009, 25(7): 19-24.

[28] He Wenshou, Liu Yangchun, He Jinyu. Relationship between water-soluble salt content and electrical conductivity of different salinized soils in Ningxia[J]. Agricultural Research in the Arid Areas, 2010, 28(1): 111-116.

[29] Xiao Wenjuan. Remote Sensing Monitoring of Cultivated Land Change in Yunnan Province in Recent 10 Years Based on GEE[D]. Kunming: Yunnan Normal University, 2020.

[30] Ding Jianli, Wu Manchun, Liu Haixia, et al. Monitoring of regional soil salinization based on comprehensive hyperspectral index[J]. Spectroscopy and Spectral Analysis, 2012, 32(7): 1918-1922.

[31] Jia Pingping, Shang Tianhao, Zhang Junhua, et al. Inversion of soil salt content in wet and dry seasons in Yinbei Region of Ningxia using multi-source spectral information[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(17): 125-134.

[32] Yang Xiaohu, Luo Yanqin, Yang Haichang, et al. Inversion and spatial distribution of soil salinity in oasis farmland in Manas River Basin[J]. Journal of Arid Land Resources and Environment, 2021, 35(2): 156-161.

[33] Liu Xuhui, Bai Yungang, Chai Zhongping, et al. Inversion and verification of soil salinity in spring in typical oasis cotton fields based on multispectral remote sensing[J]. Arid Land Geography, 2022, 45(4): 1165-1175.

[34] Wang Fei, Yang Shengtian, Ding Jianli, et al. Optimization of environmental sensitive variables and prediction of oasis soil salinity by machine learning algorithm[J]. Transactions of the CSAE, 2018, 34(22): 102-110.

[35] Allbed A, Kumar L, Aldakheel Y Y. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-resolution imageries: Applications in a date palm dominated region[J]. Geoderma, 2014, 230: 1-8.

[36] Fan Yanguo, Zhang Weikang, Liu Jingyi. Remote sensing monitoring of salinization in Yellow River Delta based on vegetation index and salinity index feature space[J]. Shandong Agricultural Sciences, 2016, 48(5): 137-141.

[37] Yao Yuan, Ding Jianli, Zhang Fang, et al. Regional soil salinization monitoring model based on hyperspectral index and electromagnetic induction technology[J]. Spectroscopy and Spectral Analysis, 2013, 33(6): 1658-1664.

[38] Khan S, Abbas A. Using remote sensing techniques for appraisal of irrigated soil salinity[J]. Int. Congr. Model. Simul.(MODSIM), Model. Simul. Soc. Aust. New Zealand, Bright, 2007 (January): 2632-2638.

[39] Nicolas H, Walter C. Detecting salinity hazards within a semiarid context by means of combining soil and remote sensing data[J]. Geoderma, 2006, 134(1-2): 217-230.

[40] Wu Dan, Jia Keli, Zhang Xiaodong, et al. Soil salinization disaster prediction model based on heterogeneous SVM neural network[J]. Hydrogeology and Engineering Geology, 2018, 45(5): 143-149, 157.

[41] Huete A R. A soil-adjusted vegetation index (SAVI)[J]. Remote Sensing of Environment, 1988, 25(3): 295-309.

[42] Wu W. The generalized difference vegetation index (GDVI) for dry-land characterization[J]. Remote Sensing, 2014, 6(2): 1211-1233.

[43] Scudiero E, Skaggs T H, Corwin D L. Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA[J]. Geoderma Regional, 2014, 2: 82-90.

[44] Parolo G, Rossi G, Ferrarini A. Toward improved species niche modelling: Arnica montana in the Alps as a case study[J]. Journal of Applied Ecology, 2008, 45(5): 1410-1418.

[45] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

[46] Cao Zhengfeng. Research on Optimization of Random Forest Algorithm[D]. Beijing: Capital University of Economics and Business, 2014.

[47] Gieseke F, Igel C. Training big random forests with little resources[C]//Proceedings of the 24th Acmsigkdd International Conference on Knowledge Discovery & Data Mining, 2018: 1445-1454.

[48] Wang Mingkuan, Mo Hongwei, Chen Hongyan. Modeling method of soil salinity inversion based on multi-spectral image[J]. Chinese Journal of Soil Science, 2016, 47(5): 1036-1041.

[49] Yang Lianbing, Chen Chunbo, Zheng Hongwei, et al. Soil salinization inversion based on optimized stochastic forest regression model[J]. Journal of Geo-Information Science, 2021, 23(9): 1662-1674.

[50] Morgan R S, El Hady M A, Rahim I S. Soil salinity mapping utilizing sentinel-2 and neural networks[J]. Indian Journal of Agricultural Research, 2018, 52(5): 524-529.

[51] Allbed A, Kumar L. Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: A review[J]. Advances in Remote

Sensing, 2013, 2: 373-385.

[52] Gorji T, Yildirim A, Hamzehpour N, et al. Soil salinity analysis of Urmia Lake Basin using Landsat-8 OLI and Sentinel-2A based spectral indices and electrical conductivity measurements[J]. Ecological Indicators, 2020, 112: 106-173.

[53] Han L, Liu D, Cheng G, et al. Spatial distribution and genesis of salt on the saline playa at Qehan Lake, Inner Mongolia, China[J]. Catena, 2019, 177: 22-30.

[54] Wang J, Ding J, Yu D, 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[J]. Geoderma, 2019, 353: 74-85.

[55] Zhang T T, Qi J G, Gao Y, et al. Detecting soil salinity with MODIS time series VI data[J]. Ecological Indicators, 2015, 52: 480-489.

[56] Scudiero E, Skaggs T H, Corwin D L. Regional scale soil salinity assessment using Landsat ETM+ canopy reflectance[J]. Remote Sensing of Environment, 2015, 169: 335-343.

[57] 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[J]. Geoderma, 2014, 235: 316-322.

[58] Brady N C. The Nature and Properties of Soils[M]. New York: Macmillan, 1984.

Note: Figure translations are in progress. See original paper for figures.