ISSN 2616-5910 Vol. 6, Issue 1: 17-24, DOI: 10.25236/AJALS.2025.060103

Soil Salinity Inversion in the Anjihai Irrigation Zone of Northwest China Based on Landsat Data

Zhi Wu[1,2], Jinzhi Gu[1,2,*], Junwei Xuan[1,2], Xinshuo Liu[1,2], Yu Wei[1,2]

[1]College of Resources and Environment, Xinjiang Agricultural University, Xinjiang Uygur Autonomous Region, Urumqi, 830052, China
[2]Xinjiang Key Laboratory of Soil and Plant Ecological Processes, Urumqi, 830052, China
**Corresponding author: gujinzhi001@163.com*

Abstract: *Soil salinization in Northwest China, caused by its arid climate, high evaporation, and inappropriate irrigation methods, is a severe environmental issue affecting resources and the ecosystem. The **st**udy area is the Anjihai Irrigation District in Northwest China, where regression analysis was applied to soil salinity data and Landsat TM/ETM+ imagery, with the multi-year average vegetation index as the independent variable, to model the spatial distribution of soil salinity. The results indicate that the best correlation between soil salinity and vegetation index occurs in the 30–60 cm soil layer. Using the multi-year average vegetation index as a covariate significantly improves the accuracy of soil salinity remote sensing inversion in the study area. The soil in the region is primarily non-salinized and lightly to moderately salinized, with salinized soils mainly distributed in the northwestern, central, and eastern parts of the study area. This study provides valuable insights into the spatial distribution of soil salinity in the Anjihai Irrigation District. It demonstrates the potential of using vegetation indices to improve remote sensing inversion accuracy for soil salinity assessment.*

Keywords: *Soil salinity, Remote sensing, Vegetation index, Inversion model*

1. Introduction

Soil salinization is a serious global issue that affects resources and the ecological environment, primarily occurring in arid and semi-arid regions with low precipitation, high evaporation, high groundwater levels, and high concentrations of soluble salts[1]. Northwest China, located in the northwest inland of China, has a temperate continental climate characterized by low annual precipitation and high evaporation[2]. Harsh natural conditions and inappropriate agricultural irrigation methods have led to a severe soil salinization problem in Northwest China[3].

Traditional methods for monitoring and evaluating soil salinization mainly involve field sampling combined with laboratory analysis of physical and chemical properties to assess the degree of salinization in a given area[4]. However, this method is time-consuming and labor-intensive, and it cannot meet the demands for large-scale, rapid, and dynamic monitoring and evaluation of salinization. Remote sensing, with its large scale, wide coverage, and high timeliness, provides a more efficient, cost-effective, and faster means for monitoring soil salinization[5]. Currently, remote sensing monitoring of soil salinization typically uses annual remote sensing data and crop covariates to invert soil salinity. However, recent studies have shown that using only single-year remote sensing data for soil salinity inversion increases the risk of obtaining erroneous conclusions. Using multi-year average data as the dependent variable in model construction can yield more accurate and objective results.

This study, building on previous research, takes the Anjihai irrigation district in the Manas River Basin as an example. It uses seven years of Landsat TM/ETM+ remote sensing data, global soil data, and measured soil salinity data, employing regression analysis to construct an inversion model for quantitatively remote sensing soil salinity in the study area. This aims to provide scientific evidence for the sustainable development of oasis agriculture.

ISSN 2616-5910 Vol. 6, Issue 1: 17-24, DOI: 10.25236/AJALS.2025.060103

2. Materials and Methods

2.1 Overview of the Study Area

The Manas River Basin is located in the central section of the northern Tianshan Mountains in the Northwest region of China , in the heart of the Eurasian continent. To the north, it borders the Gurbantunggut Desert, while the southern part is mountainous[6]. The oasis area within the basin is composed of alluvial fans, spring overflow zones, alluvial plains, deltas, and lakefront plains. The study area, the Anjihai Irrigation District, is located in the western part of the middle section of the Manas River Basin. It lies in the saline-alkali accumulation zone on the edge of the alluvial fan in the Manas River Basin, with geographic coordinates ranging from 85°10′E to 85°37′E and from 44°20′N to 44°40′N. The average altitude is between 300 and 500 meters. The climate is a typical temperate continental climate, with an average annual temperature of 5-7°C, a frost-free period of 147-191 days, annual precipitation of 110-200 mm, and annual evaporation of 1500-2000 mm. Major crops cultivated in the area include cotton, corn, and vegetables such as peppers and onions.

2.2 Soil Sample Collection and Chemical Analysis

Soil samples were collected in July 2017. Sampling points were arranged based on local geological structure, topography, land use, and other conditions, with a sampling interval of 2-3 km. Samples were taken at three depths: 0-30 cm, 30-60 cm, and 60-100 cm, resulting in a total of 49 soil samples. The soil samples were brought back to the laboratory, naturally air-dried, ground, and passed through a 20-mesh sieve. A 1:5 soil-to-water ratio extraction solution was prepared, and the total salt content was determined using the drying residue method.

2.3 Remote Sensing Data Acquisition and Processing

The remote sensing imagery used in this study includes seven years of Landsat TM/ETM+ data from 2014 to 2020. The imaging period was from June to September each year, during which the study area in the Manas River Basin experiences summer and autumn. During this period, vegetation growth is lush, making it more favorable for vegetation index extraction.

2.4 Soil Property Data Acquisition and Processing

Soil property data were obtained from the Harmonized World Soil Database (version 1.1) provided by the Food and Agriculture Organization (FAO) of the United Nations. For data within China, the source was the second national land survey, with soil data provided by the Nanjing Soil Institute at a 1:1,000,000 scale.

2.5 Selection and Construction of Vegetation Indices

Using vegetation indices to indirectly infer soil salinity is an effective method[7]. Based on previous research, this study selected and constructed four vegetation indices. These indices were then correlated with measured soil salinity at three different soil depths. The most sensitive vegetation index to soil salinity was selected for model construction. The formulas for calculating the vegetation indices are shown in Table 1.

Table 1 Calculation formula of vegetation index required for the study

Vegetation Index	Formula
Normalized Difference Vegetation Index (NDVI)	$NDVI=\frac{(NIR-R)}{(NIR+R)}$
Canopy-Scaled Vegetation Index (CRSI)	$CRSI=\sqrt{\frac{NIR\times R-G\times B}{NIR\times R+G\times B}}$
Ratio Vegetation Index (RVI)	$RVI=\frac{NIR}{R}$
Difference Vegetation Index (DVI)	$DVI=NIR-R$

Note: Landsat TM/ETM+ sensor bands: band 1 = blue (B); band 2 = green (G); band 3 = red (R); band 4 = near-infrared (NIR)

ISSN 2616-5910 Vol. 6, Issue 1: 17-24, DOI: 10.25236/AJALS.2025.060103

2.6 Selection and Construction of Vegetation Indices

In this study, cross-validation is used for model construction and optimization. Random numbers were generated and sorted for the 49 sampling points in Excel, dividing the points into 7 groups, each containing 7 samples. One group of samples is selected as the validation set (Validation Data), while the remaining 6 groups are used as the training set (Train Data). The model is then constructed sequentially using each of the 7 training sets, and the average R^2 of the 7 models is used as the evaluation criterion to optimize the model selection. Afterward, the validation set is used to test the models, and the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are calculated for each model. This process results in 7 sets of error values, and the average of these values is taken as the model's prediction error. This value is then used to assess the performance of the model.

$$\mathrm{MAE} = \frac{1}{n}\sum |y_i - y_j| \tag{1}$$

$$RMSE = \sqrt{\frac{\sum (y_i - y_j)^2}{n}} \tag{2}$$

In the formula, y_i represents the measured soil salinity, and y_j represents the predicted soil salinity. A smaller RMSE indicates less fluctuation in the salinity error, suggesting better inversion results. Similarly, a smaller MAE means the absolute error between the actual and predicted salinity values is smaller, indicating better inversion performance.

After optimization, the best model is selected, and then all 49 sampling points are used as the training set to train the model. This results in the final function for soil salinity inversion, which is applied to the study area.

2.7 Soil salinity classification

Soil salinity inversion is carried out based on the optimal inversion model. The salinity levels are then classified according to the "Northwest China Soil" standards (see Table 2).

Table 2: Soil salinity classification standard

Level	Non-Salinized	Slightly Salinized	Moderately Salinized	Heavily Salinized	Salt Soil
Salt Content (g/kg)	<3	3~6	6~10	10~20	>20

3. Results and Discussion

3.1 Correlation analysis between soil salinity and vegetation index

The five vegetation indices obtained for each year were analyzed for their correlation with soil salinity at three depths: 0-30 cm, 30-60 cm, and 60-100 cm (Figure 1). The results show that the correlation between soil salinity in the 0-30 cm layer and the vegetation indices is generally low, which does not meet the modeling requirements. However, the correlation between soil salinity at depths of 30-60 cm and 60-100 cm with the vegetation indices is relatively better. Overall, the 30-60 cm soil layer exhibits the strongest correlation with the vegetation indices, suggesting that the middle soil layer in the study area is more sensitive to changes in vegetation indices.

NDVI, CRSI, RVI, and DVI all show negative correlations with soil salinity. Among these, NDVI and CRSI exhibit a stronger sensitivity to soil salinity compared to the other indices.

The correlation between soil salinity at varying depths and vegetation indices exhibits significant interannual variability, primarily due to the influence of numerous factors such as soil salinity, precipitation, temperature, and biotic stressors (e.g., pests and diseases). These factors impact vegetation growth differently each year, resulting in fluctuating responses of vegetation indices to soil salinity.

Research suggests that using single-year remote sensing data for soil salinity inversion increases the likelihood of erroneous conclusions[8, 9]. In contrast, employing multi-year averages as the dependent

ISSN 2616-5910 Vol. 6, Issue 1: 17-24, DOI: 10.25236/AJALS.2025.060103

variable in model construction leads to more reliable and objective results. This is because salinity conditions tend to stabilize over several years, and averaging vegetation indices over time mitigates the influence of transient factors, thereby more accurately reflecting the relationship between vegetation and soil salinity[10].

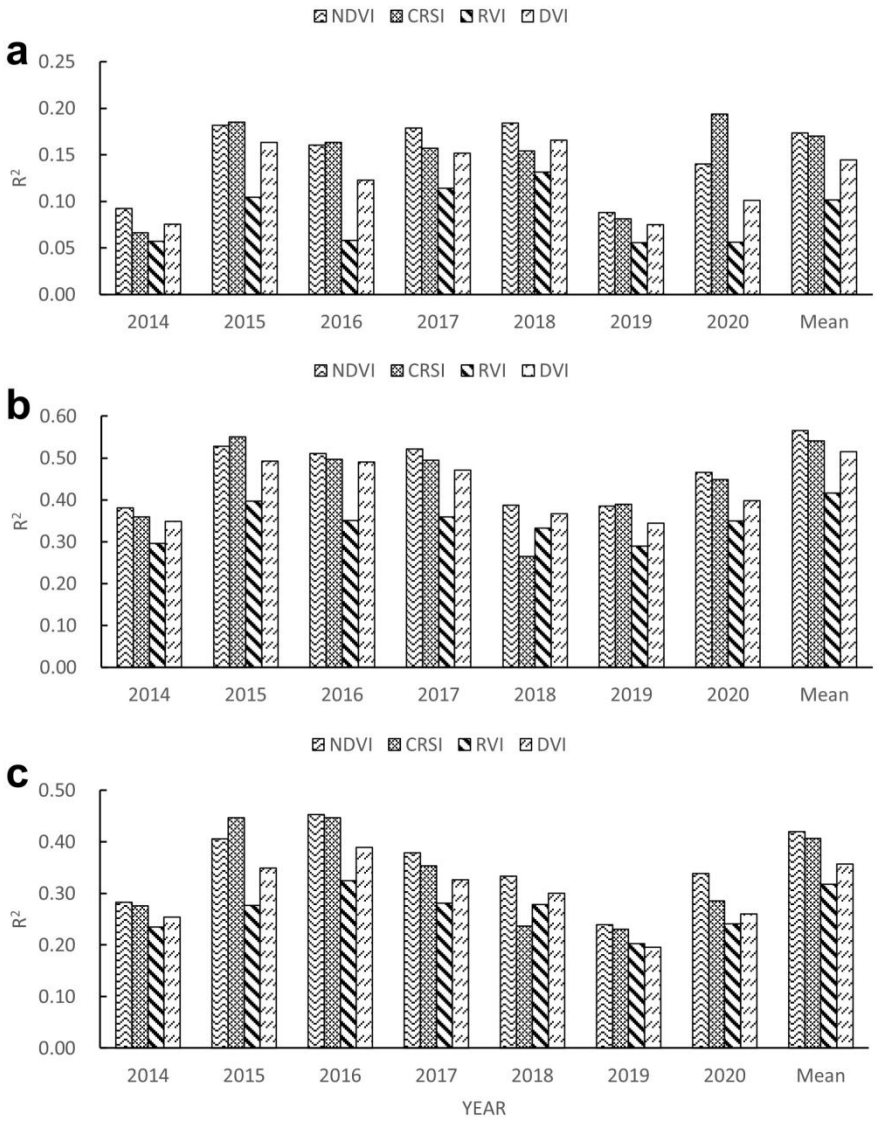

Figure 1: Correlation between soil salinity at different depths and vegetation index. (a) surface soil; (b) middle soil; (c) deep soil.

In this study, the multi-year average vegetation index was computed, and a correlation analysis was conducted between this average and soil salinity. The results demonstrate that the multi-year average vegetation index shows a stronger and more consistent correlation with soil salinity at various depths compared to single-year data. As a result, NDVI was selected as the key vegetation index for constructing the soil salinity inversion model for the two soil layers in the study area.

3.2 Inversion model construction

Through correlation analysis, it was found that the correlation between surface soil salinity and vegetation index was poor, with the maximum determination coefficient being only 0.1937. Therefore, this paper only inverted the soil salinity in the middle and lower layers.

3.2.1 Inversion model construction

Table 3 presents four models constructed from seven different modeling datasets. Except for Modeling Set 3, the R^2 values for the other six sets are all greater than 0.48, indicating good model fit and statistical significance. However, it is also evident that the models constructed from different datasets

ISSN 2616-5910 Vol. 6, Issue 1: 17-24, DOI: 10.25236/AJALS.2025.060103

exhibit substantial differences. Taking the binary linear model as an example, the model constructed from Modeling Set 5 has an R^2 of 0.704, while the model from Modeling Set 3 only has an R^2 of 0.409, showing a considerable discrepancy. This suggests that there is some randomness in the sample selection process. Therefore, to objectively evaluate the predictive capability of the models, it is necessary to use multiple sets of modeling and testing datasets for a more comprehensive assessment. In this study, the mean R^2 value across the seven sets of models is chosen for comparison, allowing for an initial selection of the most suitable model.

The mean R^2 values of the four models are calculated and presented in Table 4. Among the four models, the binary linear model has the best fit, with an R^2 of 0.595. However, as shown in Table 4, the F-value of the logarithmic model is higher than that of the binary linear model, indicating that the logarithmic model is more statistically significant. To obtain an accurate and reasonable model, it is necessary to conduct precision validation by performing regression analysis between the predicted and observed values. The evaluation criteria, including R^2, RMSE, and MAE, will be used to identify the optimal model.

Table 3: Middle layer soil salinity inversion model

Modeling set	Model	R^2	F-value	Significance
M1	Y=-12.705NDVI+2.025ln(sand)+2.388	0.607	30.17	P<0.01
	Y=-4.73ln(NDVI)-0.244	0.571	53.32	P<0.01
	Y=-13.221NDVI+9.933	0.553	49.543	P<0.01
	$Y=12.28e^{-3.052NDVI}$	0.530	45.127	P<0.01
M2	Y=-13.202NDVI+1.744ln(sand)+3.709	0.593	28.444	P<0.01
	Y=-4.788ln(NDVI)-0.142	0.571	53.242	P<0.01
	Y=-13.755NDVI+10.29	0.554	49.772	P<0.01
	$Y=12.722e^{-3.03NDVI}$	0.512	42.029	P<0.01
M3	Y=-11.014NDVI+1.205ln(sand)+4.281	0.409	13.52	P<0.01
	Y=-4.066ln(NDVI)+0.083	0.404	27.109	P<0.01
	Y=-10.77NDVI+8.491	0.386	25.186	P<0.01
	$Y=9.607e^{-2.655NDVI}$	0.354	21.959	P<0.01
M4	Y=-12.955NDVI+1.909ln(sand)+2.864	0.649	36.069	P<0.01
	Y=-4.8ln(NDVI)-0.374	0.633	69.071	P<0.01
	Y=-13.453NDVI+10.004	0.601	60.303	P<0.01
	$Y=12.331e^{-3.115NDVI}$	0.551	49.146	P<0.01
M5	Y=-15.562NDVI+0.738ln(sand)+8.404	0.704	46.314	P<0.01
	Y=-5.423ln(NDVI)-0.75	0.695	91.094	P<0.01
	Y=-15.929NDVI+11.269	0.700	93.38	P<0.01
	$Y=16.527e^{-3.692NDVI}$	0.669	80.748	P<0.01
M6	Y=-13.413NDVI+2.094ln(sand)+2.305	0.588	27.776	P<0.01
	Y=-5.257ln(NDVI)-0.719	0.580	55.257	P<0.01
	Y=-14.037NDVI+10.205	0.530	45.098	P<0.01
	$Y=12.842e^{-3.23NDVI}$	0.488	38.192	P<0.01
M7	Y=-13.696NDVI+2.048ln(sand)+2.759	0.613	30.943	P<0.01
	Y=-4.915ln(NDVI)-0.291	0.583	55.837	P<0.01
	Y=-14.275NDVI+10.542	0.578	54.760	P<0.01
	$Y=13.266e^{-3.209NDVI}$	0.513	42.058	P<0.01

Table 4: Middle layer soil salinity inversion model

Model	Average R^2
Binary Linear Model	0.595
Logarithmic Model	0.577
Simple Linear Model	0.557
Exponential Model	0.517

3.2.2 Construction of deep soil salinity inversion model

Table 5 shows three deep inversion models constructed by seven modeling sets. Except for modeling sets 3 and 6, the determination coefficients R2 of the other five models are all higher than 0.4, and the model fit is good.

ISSN 2616-5910 Vol. 6, Issue 1: 17-24, DOI: 10.25236/AJALS.2025.060103

Table 5: Deep soil salinity inversion model

Modeling set	Model	R^2	F-value	Significance
M1	Y=-10.384NDVI+8.566	0.414	28.262	P<0.01
	Y=-3.653ln(NDVI)+0.625	0.413	28.190	P<0.01
	$Y=9.667e^{-2.489NDVI}$	0.401	26.811	P<0.01
M2	Y=-10.529NDVI+8.651	0.437	31.069	P<0.01
	Y=-3.583ln(NDVI)+0.736	0.430	30.224	P<0.01
	$Y=9.941e^{-2.472NDVI}$	0.441	31.581	P<0.01
M3	Y=-9.439NDVI+7.993	0.354	21.919	P<0.01
	Y=-3.667ln(NDVI)+0.546	0.392	25.768	P<0.01
	$Y=8.67e^{-2.284NDVI}$	0.298	16.948	P<0.01
M4	Y=-9.939NDVI+8.140	0.475	36.250	P<0.01
	Y=-3.549NDVI+0.471	0.501	40.223	P<0.01
	$Y=8.91e^{-2.407NDVI}$	0.449	32.591	P<0.01
M5	Y=-11.297NDVI+9.12	0.470	35.403	P<0.01
	Y=-3.785ln(NDVI)+0.647	0.451	32.909	P<0.01
	$Y=11.022e^{-2.683NDVI}$	0.452	32.985	P<0.01
M6	Y=-9.598NDVI+8.044	0.341	20.703	P<0.01
	Y=-3.387ln(NDVI)+0.747	0.331	19.821	P<0.01
	$Y=9.001e^{-2.394NDVI}$	0.345	21.051	P<0.01
M7	Y=-10.363NDVI+8.599	0.423	29.332	P<0.01
	Y=-3.537ln(NDVI)+0.762	0.419	28.844	P<0.01
	$Y=9.753e^{-2.46NDVI}$	0.415	28.333	P<0.01

The average R^2 of the three models was calculated, as shown in Table 6. The univariate linear model and the logarithmic model have a better fit, but in order to obtain an accurate and reasonable model, the model accuracy needs to be verified and evaluated based on R^2, RMSE, and MAE to select the optimal model.

Table 6: Deep layer soil salinity inversion model

Model	Average R^2
Simple Linear Model	0.416
Logarithmic Model	0.420
Exponential Model	0.400

3.3 Model accuracy verification and optimization

The validation set was used to verify the binary linear model and univariate logarithmic model of soil salinity in the middle and lower layers. It was found that the univariate logarithmic model had higher inversion accuracy, so this paper selected the univariate logarithmic model as the inversion model of soil salinity in the middle and lower layers. The full set was used as the training set for training, and the middle layer soil salinity model Y=-4.902ln(NDVI)-0.379, R^2 was 0.577, and the lower layer soil salinity model Y=-3.601ln(NDVI)+0.641, R^2 was 0.482.

3.4 Soil salinity remote sensing inversion

Based on the optimal inversion model, remote sensing inversion of soil salinity in the middle and deep layers of the study area was carried out, and the inversion map of soil salinity in the middle and deep layers of the Anjihai irrigation area was obtained (Figure 2).

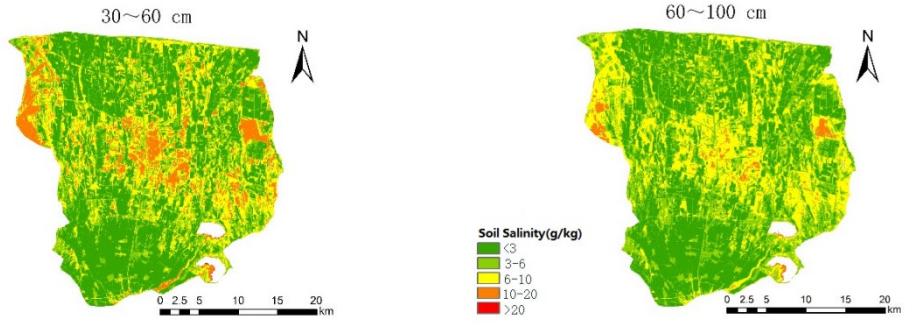

Figure 2: Soil salinity distribution map of the Anjihai irrigation area

ISSN 2616-5910 Vol. 6, Issue 1: 17-24, DOI: 10.25236/AJALS.2025.060103

As shown in Figure 2 and Table 7, the two layers of soil in the study area are mainly non-salinized and light-to-moderately salinized, with less severely salinized soil.

Table 7: Statistics of soil salinity in Anjihai Irrigation Area

Salinity Level	Middle soil		Deep soil	
	Pixel Count	Percentage	Pixel Count	Percentage
Non-salinized (<3 g/kg)	413843	42.05%	394770	40.11%
Lightly Salinized (3~6 g/kg)	237411	24.12%	315568	32.07%
Moderately Salinized (6~10 g/kg)	235284	23.91%	253819	25.79%
Heavily Salinized (10~20 g/kg)	97163	9.87%	19845	2.02%
Saline Soil (>20 g/kg)	439	0.04%	138	0.01%

In terms of spatial distribution, salinized soil is mainly distributed in the northwest, central, and eastern parts of the irrigation area, and the soil salt content in most areas of the south and north is low. In summary, the threat of soil salinization is prevalent in the study area, and soil salinization prevention and control should be strengthened, and irrigation, planting, and reclamation should be carried out reasonably to prevent the aggravation of soil salinization.

4. Conclusion

This study conducted a correlation analysis between three layers of measured soil salinity data and four vegetation indices—NDVI, CRSI, RVI, and DVI—constructed from multi-year Landsat data. The results showed that, except for the surface soil layer, there is a high correlation between soil salinity in the middle and deep layers and the vegetation indices. Further analysis revealed that using the multi-year average of vegetation indices as a covariate significantly improves the accuracy of soil salinity remote sensing inversion in the study area.

The soil in the study area is primarily non-salinized and lightly to moderately salinized, with heavily salinized soils being relatively rare. In terms of spatial distribution, salinized soils are mainly found in the northwestern, central, and eastern parts of the irrigation area, while most regions in the south and north have low soil salinity levels.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China, grant number 52079136.

References

[1] ONDRASEK G, RENGEL Z. Environmental salinization processes: Detection, implications & solutions[J]. Science of the Total Environment, 2021, 754: 142432.
[2] HE B, SHENG Y, CAO W, et al. Characteristics of climate change in northern Xinjiang in 1961–2017, China[J]. Chinese Geographical Science, 2020, 30: 249-265.
[3] HOU J, RUSULI Y. Assessment of soil salinization risk by remote sensing-based ecological index (RSEI) in the Bosten lake watershed, Xinjiang in northwest China[J]. Sustainability, 2022, 14(12): 7118.
[4] ZAMAN M, SHAHID S A, HENG L, et al. Introduction to soil salinity, sodicity, and diagnostics techniques [J]. Guideline for salinity assessment, mitigation and adaptation using nuclear and related techniques, 2018: 1-42.
[5] SAHBENI G, NGABIRE M, MUSYIMI P K, et al. Challenges and opportunities in remote sensing for soil salinization mapping and monitoring: A review [J]. Remote Sensing, 2023, 15(10): 2540.
[6] Zhang Tianyou, Wang Ling, Wang Hui, et al. Study on remote sensing monitoring of the ecological environment in salinization irrigation area of Manas River Basin[J]. Acta Ecologica Sinica, 2017,37(9): 3009-3018.
[7] WANG F, SHI Z, BISWAS A, et al. Multi-algorithm comparison for predicting soil salinity[J]. Geoderma, 2020, 365: 114211.
[8] LIU H, GUO B, YANG X, et al. High spatiotemporal resolution vegetation index time series can facilitate enhanced remote sensing monitoring of soil salinization[J]. Plant and Soil, 2024: 1-23.

ISSN 2616-5910 Vol. 6, Issue 1: 17-24, DOI: 10.25236/AJALS.2025.060103

[9] LEVIN J, ARANGO H G, LAUGHLIN B, et al. The impact of remote sensing observations on cross-shelf transport estimates from 4D-Var analyses of the Mid-Atlantic Bight[J]. Advances in space research, 2021, 68(2): 553-570.
[10] HE B, DING J, HUANG W, et al. Spatiotemporal Variation and Future Predictions of Soil Salinization in the Werigan–Kuqa River Delta Oasis of China[J]. Sustainability, 2023, 15(18): 13996.