Soil Salinization Susceptibility Assessment Based on Logistic Regression Analysis: A Case Study of the Tarim Irrigation District in Southern Xinjiang (Postprint)

Authors: Jiang Lei

Date: 2023-12-06T00:00:00+00:00

Abstract

Southern Xinjiang (hereinafter referred to as Southern Xinjiang) possesses tremendous potential for land resource utilization and serves as a crucial reserve cultivated land resource supply area in China. Conducting soil salinization susceptibility evaluation holds significant reference value for water and land increase site selection and zoned and classified land management in Southern Xinjiang. This study takes the Tarim irrigation district in Southern Xinjiang as the research area, and based on field survey data including total dissolved solids (TDS) of groundwater, groundwater depth, soil texture, geomorphological type, land use type, and soil salt content, employs a data-driven weights-of-evidence method for single-factor spatial correlation analysis of soil salinization, and conducts multi-factor comprehensive quantitative evaluation of soil salinization susceptibility through a Logistic regression model. The results indicate: (1) High groundwater TDS, shallow groundwater depth, fine-grained soil texture, alluvial (lacustrine) plains among geomorphological types, grassland among land use types, and areas with high soil salt content demonstrate relatively high positive spatial correlation. (2) Paleochannels within the alluvial plains of the irrigation district and trough-shaped, closed depressions such as ancient lake marshes within the lacustrine plains constitute high soil salinization susceptibility zones, covering approximately 29 km^2 and accounting for 1.3% of the total irrigation district area; abandoned cultivated land surrounding farmland in the irrigation district represents moderate soil salinization susceptibility zones, covering approximately 453 km^2 and accounting for 20.5% of the total irrigation district area; the extensive desert and water bodies surrounding the irrigation district constitute low soil salinization susceptibility zones, covering approximately 1726 km^2 and accounting for 78.2% of the total irrigation district area. (3) The comprehensive model evaluation confidence results and field survey status demonstrate that the Logistic regression model's comprehensive quantitative evaluation of soil salinization susceptibility is accurate and reliable, providing a new approach for soil salinization susceptibility evaluation and offering a scientific basis for soil salinization prevention and reserve cultivated land resource site selection in Southern Xinjiang.

Full Text

Evaluation of Soil Salinization Susceptibility Based on Logistic Regression Analysis: A Case Study of the Tarim Irrigation Area in Southern Xinjiang

JIANG Lei[1,2], LIU Xiaolong[1], GUO Shuai[1], HE Liang[1], XING Jianlei[1], GUO Junjie[1]

[1]Urumqi Center of Natural Resources Comprehensive Survey, China Geological Survey, Urumqi 830057, Xinjiang, China
[2]School of Environmental Studies, China University of Geosciences, Wuhan 430078, Hubei, China

Abstract

Southern Xinjiang possesses tremendous potential for land resource utilization and serves as a crucial reserve area for cultivated land resources in China. Evaluating soil salinization susceptibility provides an essential reference for water and land resource allocation and zonal land management in this region. This study focuses on the Tarim irrigation area in southern Xinjiang, utilizing field survey data including total dissolved solids (TDS) in groundwater, groundwater depth, soil texture, geomorphic type, land use type, and soil salt content. A data-driven weights-of-evidence approach was employed to analyze the spatial correlation of individual factors contributing to soil salinization, followed by a multifactor comprehensive quantitative evaluation using a Logistic regression model. The results demonstrate that: (1) High groundwater TDS, shallow groundwater depth, fine-grained soil texture, alluvial (lacustrine) plain geomorphology, and grassland land use exhibit strong positive spatial correlations with high soil salt content areas. (2) High susceptibility zones are primarily located in trough-shaped and closed depressions such as ancient river channels within alluvial plains and ancient lake marshes within lacustrine plains, covering approximately 29 km^2 (1.3% of the total area). Moderate susceptibility zones occur mainly in abandoned farmland surrounding irrigated fields, covering about 453 km^2 (20.5% of the total area). Low susceptibility zones are distributed across the vast desert and water areas around the irrigation district, covering approximately 1726 km^2 (78.2% of the total area). (3) Validation against field investigation data confirms that the Logistic regression model yields accurate and reliable results, providing a novel approach for soil salinization susceptibility evaluation and offering scientific guidance for salinization prevention and reserve cultivated land selection in southern Xinjiang.

Keywords: soil salinization; Logistic regression analysis; susceptibility evaluation; Tarim irrigation area; southern Xinjiang

1 Study Area Overview

The study area is located at the eastern end of the Tarim Basin (86°34 ~88°00 E, 40°30 ~41°05 N), encompassing the Tarim irrigation district of the Xinjiang Production and Construction Corps' Second Division along the lower reaches of the Tarim River [Figure 1: see original paper]. This region features a typical continental desert arid climate with scarce precipitation (average annual rainfall of 34.7 mm) and high evaporation (average annual evaporation of 2408.6 mm). The northern part of the study area comprises marine-terrestrial strata at the southern foothills of the Tianshan Mountains, including Proterozoic, Paleozoic Jurassic, Cenozoic Neogene, and Quaternary systems, with bedrock dominated by clastic and carbonate rock formations.

The terrain is flat with low hydraulic gradients, and the fine-grained aquifer exhibits weak permeability and slow runoff, resulting in groundwater stagnation and high salt content in both surface water and groundwater. This area represents a convergence and discharge zone for water systems in the Tarim Basin and its surrounding mountains, as well as a salt accumulation area [Figure 2: see original paper]. The fine-soil plain zone is dominated by salt accumulation processes, while the low-lying central areas are characterized by evaporation and concentration effects, leading to high soil salt content. These geological conditions indicate that the study area's native substrate is salt-bearing parent material, creating a primary environment prone to saline-alkali land development. In recent years, increased agricultural irrigation water use in the lower Tarim River region, coupled with infiltration from canal systems and field drip irrigation, has raised groundwater levels, causing soil salts to migrate vertically to the surface and readily inducing soil salinization problems within the irrigation district.

2 Data Sources and Methodology

2.1 Data Collection and Processing

Data were obtained from 1:50,000-scale ecological geological surveys conducted across the study area, covering approximately 2208 km^2. The survey employed a transect method with point and line spacing adjusted according to geomorphic units—sparse in desert areas and dense in fine-soil plain zones—while meticulously documenting geomorphic variations. At each survey point, soil samples were collected from 0–50 cm depth using a soil auger, and groundwater samples were obtained from wells drilled to the aquifer whenever possible. Field measurements included coordinates, geomorphic types, soil properties, thickness, and groundwater depth.

A total of soil samples were collected and analyzed for soil texture, total salt content, and ion composition at the Xinjiang Nonferrous Metals Geological Exploration Bureau Testing Center. Groundwater samples were analyzed for major ions at the Laboratory of the Third Geological Brigade of the Xinjiang Bureau of Geology and Mineral Resources Development. Soil texture was determined using the international soil texture classification standard via laser particle size analysis, while total salt content was measured by gravimetric methods. Land use type data were derived from Landsat TM/ETM imagery (30 m spatial resolution) from July 2020, preprocessed and classified through supervised classification according to the Technical Regulations for Current Land Use Survey.

Data processing involved: (1) vector-to-raster conversion to transform polygon features such as geomorphic type and soil texture into raster data; (2) interpolation analysis to generate continuous raster surfaces from point data including groundwater depth and TDS; and (3) reclassification to convert continuous raster data into integer-valued raster data with a cell size of 25 m $\times$ 25 m, meeting the requirements for soil salinization susceptibility evaluation.

2.2 Analytical Methods

2.2.1 Single-Factor Spatial Correlation Analysis This evaluation employed a data-driven weights-of-evidence method to analyze spatial correlations between known high soil salt content points and various influencing factors. The method quantifies spatial relationships through two metrics: W^+ and W^-. W^+ represents the influence degree when a specific category of a factor is present, while W^- represents the influence degree when a category is absent. The comprehensive impact can be assessed through their combination.

$W^+ > 0$ or $W^- < 0$ indicates positive correlation with soil salinization, with larger absolute values indicating stronger spatial positive correlation. Conversely, $W^+ < 0$ or $W^- > 0$ indicates negative correlation, with larger absolute values reflecting stronger negative correlation. The metrics are calculated as follows:

$W^+ = \ln(D/D\ )$
$W^- = \ln((B\text{-}D)/(B\ \text{-}D\ ))$

where D is the number of grid cells containing known soil salinization points, D is the number without soil salinization, B is the total number of grid cells where the factor exists, and B is the total where the factor is absent.

To facilitate comparison, the contrast value $C = W^+ - W^-$ provides an effective measure of correlation strength between soil salinization distribution and influencing factors, while also enabling data reduction from multiclass to fewer categories.

2.2.2 Multifactor Comprehensive Quantitative Evaluation The evaluation utilized a Logistic regression model, a generalized linear regression approach that establishes multivariate relationships between a dependent variable and multiple independent variables to predict event occurrence probability. Assuming P influencing factors participate in the evaluation, with a response factor (spatial distribution of existing salinization points) containing values of 1 (salinization occurred) or 0 (no salinization), the probability of salinization occurrence at any pixel i can be expressed as a function of the P influencing factors:

$$\text{Logit}(\ i) = \ln[\ i/(1\text{-}\ i)] = b_0 + b_1X_{\ 1} + b_2X_{\ 2} + \dots + b\ X$$

where b_0 is a constant vector and b_1 through b are regression parameters estimated via maximum likelihood estimation. The final probability map of soil salinization occurrence was reclassified into susceptibility zones: low (0.00–0.04), moderate (0.04–0.16), and high (0.16–0.44).

3 Results

3.1 Single-Factor Spatial Correlation Analysis

3.1.1 Groundwater TDS Shallow groundwater TDS in the study area ranges from 3–10 $g \cdot L^{-1}$, with a freshwater zone along both sides of the Tarim River channel receiving river water recharge. Within 500 m north of the Tarim River, TDS ranges from 1–3 $g \cdot L^{-1}$, while local lake marsh and desert-covered areas exhibit high TDS (>10 $g \cdot L^{-1}$). Overall, TDS increases from river floodplains to desert-covered areas. Spatial correlation analysis reveals low correlation between groundwater TDS and high soil salt content points, with positive correlation only observed near reservoirs and shallow groundwater areas [Figure 3a: see original paper].

3.1.2 Groundwater Depth Groundwater depth ranges from 2–15 m, being relatively shallow in oasis areas near the Tarim River banks and reservoir peripheries, and deeper in desert-covered areas. Depth generally increases downstream along the Tarim River. Analysis shows significant correlation between shallow groundwater depth and high soil salt content, with the highest positive correlation index reaching 2.42. High positive correlation zones are mainly distributed in upstream areas with shallow groundwater depth [Figure 3b: see original paper].

3.1.3 Soil Texture Soil texture varies across the study area: sandy loam dominates near the Tarim River channel; floodplains feature thin silt clay over sandy loam; alluvial plains have silt clay surfaces interlayered with sandy loam; lacustrine deposits contain clay or silt clay over sandy loam; and desert areas consist of thick sand layers. Spatial correlation analysis demonstrates significant correlation between fine-grained soils (silt clay, clay) and high soil salt content, with the highest positive correlation index reaching 3.12. High positive correlation zones are distributed in fine-grained soil areas within alluvial and lacustrine plains [Figure 3c: see original paper].

3.1.4 Geomorphic Type The study area comprises three geomorphic types: alluvial plains along both sides of the Tarim River containing braided ancient channels; lacustrine plains in low-lying, flat areas north of the Tarim River consisting of dried oxbow lakes and ancient marshes; and aeolian deserts outside the Tarim River oasis. Analysis reveals significant correlation between alluvial/lacustrine plains and high soil salt content, with the highest positive correlation index reaching 2.88. Numerous high positive correlation zones are distributed throughout the irrigation district, while desert margins show negative correlation [Figure 3d: see original paper].

3.1.5 Land Use Type Land use types are relatively simple: cultivated land, orchard, forest, and minimal construction land within farm clusters; grassland around reservoirs; and sandy land and saline-alkali land in other areas. Spatial correlation analysis shows significant correlation between natural grassland and high soil salt content, with the highest positive correlation index reaching 2.56. High positive correlation zones are mainly distributed in wetlands and grasslands along both sides of the Tarim River [Figure 3e: see original paper].

3.2 Multifactor Comprehensive Quantitative Evaluation

Based on the single-factor analysis results, the Logistic regression model integrated five factors (groundwater TDS, groundwater depth, soil texture, geomorphic type, and land use type) to generate a soil salinization probability map. The evaluation results [Figure 4: see original paper] indicate that high susceptibility zones (29 km^2, 1.3% of total area) are concentrated in trough-shaped and closed depressions such as ancient river channels in alluvial plains and ancient lake marshes in lacustrine plains, where poor drainage and convergent groundwater flow exacerbate secondary salinization, occasionally forming small salt lakes. Moderate susceptibility zones (453 km^2, 20.5%) occur mainly in abandoned farmland around cultivated fields, where blocked and abandoned drainage channels impair water discharge, allowing saline-alkali water to accumulate and infiltrate. Low susceptibility zones (1726 km^2, 78.2%) are distributed across desert areas and water bodies with sparse vegetation, undulating terrain, and relatively deep groundwater, where thick surface sand accumulation prevents salinization.

Significance testing [Figure 5: see original paper] demonstrates that when significance index values exceed 2.5, confidence levels surpass 95%, validating the model's accuracy and reliability against field investigation data.

4 Discussion

Soil salinization reduces crop yields, degrades surface and groundwater quality, decreases biodiversity, and increases flood and desertification risks. The primary cause is poor "soil-water" relationships, a finding consistent with other arid region studies. Seydehmet et al. identified high soil salinization risk zones associated with low-lying topography, high-salt parent material, high groundwater salinity, and shallow groundwater depth. Bouksila et al. found that poor drainage, fine texture, and low soil hydraulic conductivity increase salinization trends, while Kosmas emphasized soil water storage capacity as a key indicator, with finer-textured soils being more susceptible.

Previous studies using artificial neural networks, principal component analysis, and analytic hierarchy process have limitations: neural networks cannot consider factors deeply enough; principal component analysis focuses only on soil ion relationships; and analytic hierarchy process involves subjective weighting. This study innovatively combines data-driven weights-of-evidence with Logistic regression, considering both influencing factors (salinization drivers) and response factors (existing salinization distribution). The approach avoids artificial interference and information redundancy while providing accurate, reliable results.

The Tarim irrigation district's native geological background of salt-bearing parent material creates an inherently saline environment. Prevention is more cost-effective than remediation, making it crucial to identify susceptibility zones and implement targeted management strategies based on local conditions to promote cultivated land quality protection and improvement.

5 Conclusions

This study evaluated soil salinization susceptibility in a typical arid irrigation district of southern Xinjiang by analyzing geological background and innovatively combining weights-of-evidence with Logistic regression modeling. Using five single-factor indicators (groundwater TDS, groundwater depth, soil texture, geomorphic type, and land use type), the comprehensive quantitative analysis yielded the following conclusions:

1. The study area's native geological background consists of salt-bearing parent material in an inherently saline environment. Geological conditions determine salinization development degree. Meaningful "soil-water" indicators selected through investigation reveal that high groundwater TDS, shallow groundwater depth, fine-grained soil texture, alluvial/lacustrine plain geomorphology, and grassland land use show strong positive spatial correlations with high soil salt content areas.
2. Multifactor comprehensive evaluation identifies: (a) high susceptibility zones in ancient river channels of alluvial plains and ancient lake marshes of lacustrine plains (1.3% of total area); (b) moderate susceptibility zones in abandoned farmland around cultivated fields (20.5% of total area); and (c) low susceptibility zones in desert areas and water bodies (78.2% of total area).
3. Model validation through significance testing and field investigation confirms the accuracy and reliability of the Logistic regression approach, pro-

viding a novel method for soil salinization susceptibility evaluation. This approach offers farmers, rural planners, and land resource managers a practical tool for identifying salinization-prone areas and implementing appropriate agricultural strategies and management solutions to prevent further soil degradation.

Acknowledgments

We thank Prof. Tian Changyan and Prof. Zhao Zhenyong from the Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, for their guidance during field investigations, and Prof. Chen Chuan from Xinjiang University for assistance with data analysis and mapping.

References

[1] Li Lijun, Zhang Xiaoqing, Chen Changqing, et al. Ecological effects of water conveyance on the lower reaches of Tarim River in recent twenty years[J]. Arid Land Geography, 2018, 41(2): 238-247.

[2] Yang Hongmei, Xu Hailiang, Fan Zili, et al. Spatial variability and pattern of surface soil salinity in the lower reaches of the Tarim River[J]. Journal of Desert Research, 2010, 30(3): 564-570.

[3] Tian Changyan, Mai Wenxuan, Zhao Zhenyong. Study on key technologies of ecological management of saline alkali land in arid area of Xinjiang[J]. Acta Ecologica Sinica, 2016, 36(22): 7064-7068.

[4] Zhao Zhenyong, Tian Changyan, Zhang Ke, et al. Biological improvement of saline alkali land and comprehensive utilization of halophyte resources[J]. High Technology & Commercialization, 2020(9): 64-66.

[5] Tomaz A, Palma P, Fialho S, et al. Risk assessment of irrigation related soil salinization and sodification in Mediterranean areas[J]. Water, 2020, 12(12): 3569, doi: 10.3390/w12123569.

[6] Castro F C, Araújo J F, dos Santos A M. Susceptibility to soil salinization in the quilombola community of Cupira Santa Maria da Boa Vista Pernambuco Brazil[J]. Catena, 2019, 179: 175-183.

[7] Ma L G, Ma F L, Li J D, et al. Characterizing and modeling region-scale variations in soil salinity in the arid oasis of Tarim Basin, China[J]. Geoderma, 2017, 305: 1-11.

[8] Alexandre C, Borralho T, Durão A. Evaluation of salinization and sodification in irrigated areas with limited soil data: Case study in southern Portugal[J]. Spanish Journal of Soil Science, 2018, 8: 130-140.

[9] Xie Conghui, Wu Shixin, Lin Juan, et al. Analysis of cultivated land salinization in Kashgar Oasis based on PSO-PNN model[J]. Arid Land Geography, 2022, 45(5): 1547-1558.

[10] Yao Yuan, Ding Jianli, Zhang Fang, et al. Monitoring of soil salinization in northern Tarim Basin, Xinjiang of China in dry and wet seasons based on remote sensing[J]. Chinese Journal of Applied Ecology, 2013, 24(11): 3213-3220.

[11] Grundy M J, Silburn D M, Chamberlain T. A risk framework for preventing salinity[J]. Environmental Hazards, 2007, 7(2): 97-105.

[12] Zhou D, Lin Z L, Liu L M, et al. Assessing secondary soil salinization risk based on the PSR sustainability framework[J]. Journal of Environmental Management, 2013, 128: 642-654.

[13] Niu Zengyi, Ding Jianli, Li Yanhua, et al. Soil salinization information extraction method based on GF-1 image[J]. Arid Land Geography, 2016, 39(1): 171-181.

[14] Jiang Lei, Zhao Yi, Zhang Pengwei, et al. Study on influence degree of phreatic evaporation based on hydrogen and oxygen isotope characteristics[J]. Arid Zone Research, 2022, 39(6): 1793-1800.

[15] Zhou Long, Yang Pengnian, Wang Yongpeng, et al. Characteristics of water consumption and the evolution of water delivery methods in the lower reach of the Tarim River[J]. Arid Zone Research, 2022, 39(1): 144-154.

[16] Huang Xiang, Chen Yaning, Ma Jianxin. Analysis of the ecosystem services value of the typical river basin in desert areas of northwest China[J]. Journal of Natural Resources, 2011, 26(8): 1364-1376.

[17] Ma Songzeng, Shi Mingchang, Yang Guisen, et al. Analysis on spatiotemporal change of land use based on GIS technology: Taking Xinjiang Tarim Basin as an example[J]. Research of Soil and Water Conservation, 2013, 20(1): 177-181.

[18] Chen Yaning, Cui Weicheng, Li Weihong, et al. Utilization of water resources and ecological protection in the Tarim River[J]. Acta Geographica Sinica, 2003, 58(2): 215-222.

[19] Xie Haixia, Gong Jiang, He Shuai. Influence of irrigation amount of mulched drip irrigation on type of soil salt[J]. Journal of Henan Agricultural Sciences, 2013, 42(8): 45-47.

[20] Chen Yaning, Chen Zhongsheng. Analysis of oasis evolution and suitable development scale for arid regions: A case study of the Tarim River Basin[J]. Chinese Journal of Eco-Agriculture, 2013, 21(1): 134-140.

[21] Seydehmet J, Lü G H, Abliz A, et al. Irrigation salinity risk assessment and mapping in arid oasis, northwest China[J]. Water, 2018, 10(7): 966, doi: 10.3390/w10070966.

[22] Bouksila F, Bahri A, Berndtsson R, et al. Assessment of soil salinization risks under irrigation with brackish water in semiarid Tunisia[J]. Environmental and Experimental Botany, 2013, 92: 176-185.

[23] Kosmas C, Kairis O, Karavitis C, et al. Evaluation and selection of indicators for land degradation and desertification monitoring: Methodological approach[J]. Environmental Management, 2014, 54: 951-970.

[24] Shi Xiaojie, Zhang Yu, Sun Tao, et al. Soil salinization sensitivity analysis model based on genetic artificial neural network[J]. Hydrogeology & Engineering Geology, 2008(4): 116-119.

[25] Wang Shuixian, Dong Xinguang, Liu Feng, et al. Application of analytic hierarchy process on the soil salinization in irrigation areas of Xinjiang plain[J]. Journal of Arid Land Resources and Environment, 2007, 21(4): 111-116.

[26] Guo Xinbao, Tumaerbai Hudan, Zhai Yongxian, et al. Application of salinization evaluation to soil upper layer of cotton field under film irrigation by main component analytical method in different years[J]. Hydrogeology & Engineering Geology, 2009, 46(5): 935-940.

[27] Gao Hui, Tan Limei, Liu Peng, et al. Estimation of arable land resources potential in hilly area of Taihang Mountain based on binary Logistic regression model[J]. Chinese Journal of Eco-Agriculture, 2017, 25(4): 490-497.

[28] Fan Zhiyin, Gou Xiaofeng, Qian Jing, et al. Information and Logistic regression models based coupling analysis for susceptibility of geological hazards[J]. Journal of Engineering Geology, 2018, 26(2): 340-347.

[29] Wang Xuedong, Zhang Chaobiao, Wang Cui, et al. Geological disaster susceptibility in Helong City based on Logistic regression and random forest[J]. Journal of Jilin University (Earth Science Edition), 2022, 52(6): 1957-1970.

[30] Aruhan, Du Lai, Sheng Yan, et al. Driving forces on land desertification in Duolun County Inner Mongolia based on Logistic regression model[J]. Arid Land Geography, 2019, 42(1): 137-143.

[31] Chen Caoliang, Zhang Wenjun, Qian Jing, et al. Application of improved Logistic regression model in geological hazard evaluation[J]. Environmental Science & Technology, 2019, 42(4): 188-193, 213.

Table 1 Results of spatial correlation analysis between different single factors and high soil salt content in the study area

Table 2 Confidence degree of significance index of soil salinization susceptibility evaluation results

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