Postprint of Analysis of Vertical Characteristics of Oasis Salinization Integrating Machine Learning and Geographical Detectors

Authors: Haibin Xiong, Hao Xingming, Yanfeng Di, Ci Mengtao, Liang Qixiang, Zhang Jingjing, Fan Xue

Date: 2026-03-24T18:53:18+00:00

Abstract

Soil salinization is a key obstacle restricting the sustainable development of oasis agriculture in arid regions. This study aims to reveal the vertical variation characteristics of soil salinity and its main control mechanisms in the Aksu Oasis. The research integrates machine learning algorithms with the Optimal Parameters-based Geographical Detector (OPGD), utilizing Sentinel-2 imagery to construct soil total salt content models for four soil depths (0–10 cm, 10–20 cm, 20–30 cm, and 30–50 cm). The results indicate that the XGBoost model achieved the highest prediction accuracy ($R^2 \geq 0.6$, $RMSE \leq 5.97$ g·kg^{-1}); the model is overall robust, although uncertainty is slightly higher in severely salinized areas. Regarding spatial distribution, salinity shows a decreasing trend with increasing depth, and significant low values are observed near rivers due to the leaching effect during the high-water period. Attribution analysis demonstrates that the driving mechanisms exhibit vertical stratification: surface (0–10 cm) salinity is dominated by the bivariate synergistic enhancement of local factors such as human activities, soil, and climate, while deep-layer (30–50 cm) salinity is determined by the coupling mechanism of "groundwater-climatic evaporation." This study elucidates the vertical differentiation mechanisms of the salinization process, providing a scientific basis for the three-dimensional monitoring and prevention of oasis salinization.

Full Text

Analysis of Vertical Characteristics of Oasis Salinization Integrating Machine Learning and Geographical Detectors

Xiong Haibin[1,2,3,4], Hao Xingming[1,2,3,4], Di Yanfeng[1], Ci Mengtao[1,2,3,4], Liang Qixiang[1,2,3,4], Zhang Jingjing[1,2,3,4], Fan Xue[1,2,3,4]

([1]State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China; [2]University of Chinese Academy of Sciences, Beijing 100049, China; [3]Aksu Oasis Farmland Ecosystem National Field Scientific Observation and Research Station, Aksu 843017, China; [4]Xinjiang Key Laboratory of Water Cycle and Water Utilization in Arid Zones, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China)

Abstract

Soil salinization is a critical barrier restricting the sustainable development of oasis agriculture in arid regions. This study aims to reveal the vertical distribution characteristics of soil salinity in the Aksu Oasis and to evaluate the potential of multi-source remote sensing data for quantifying soil salt content (SSC) at different depths. By integrating multispectral data from Sentinel-2, backscattering coefficients from Sentinel-1, and topographic information, we developed a series of machine learning models to estimate SSC across various soil profiles.

Our results indicate that soil salinity in the study area exhibits significant vertical heterogeneity, with surface accumulation being a prominent feature. The integration of SAR (Synthetic Aperture Radar) data and topographic indices significantly improved the prediction accuracy compared to models relying solely on optical data, particularly for deeper soil layers where optical sensors have limited penetration. Among the tested algorithms, the XGBoost model achieved the highest predictive accuracy, characterized by an $R^2 \geq 0.6$ and an $RMSE \leq 5.97\ \mathrm{g \cdot kg^{-1}}$. Attribution analysis revealed that the driving mechanisms of salinization exhibit vertical stratification. For the surface layer (0–10 cm), salt content is primarily driven by the synergistic enhancement of human activities, soil properties, and climate. In contrast, the deeper layers (30–50 cm) are determined by the coupling mechanism between groundwater and climatic evaporation. This research provides a scientific basis for the precise management of saline-alkali land and the optimization of water resource allocation in the Aksu Oasis.

Keywords: Soil Salinization; Remote Sensing; Machine Learning; Optimal Parameter Geographic Detector; Soil Soluble Salt Content

---

1. Introduction

Soil salinization is one of the most prevalent forms of land degradation in arid and semi-arid regions globally. In the oasis ecosystems of Northwest China, the combination of high evaporation rates and low precipitation leads to the accumulation of salts in the upper soil layers, threatening food security and regional ecological safety. Traditional methods for monitoring soil salinity often rely on extensive field sampling and laboratory analysis, which are time-consuming, costly, and difficult to implement over large spatial scales. Consequently, remote sensing technology has become a vital tool for assessing soil salinity.

However, surface reflectance captured by satellites primarily reflects the characteristics of the topsoil (0–20 cm), while the vertical distribution of salts in deeper layers remains a challenge for traditional inversion methods. Accurately estimating soil salinity in the complex surface environments of arid regions requires accounting for the non-linear relationships between spectral signatures and actual salt content. While significant progress has been made in surface monitoring, there is a lack of systematic evaluation regarding the vertical heterogeneity of salinity within deeper soil profiles [**?**].

This study utilizes the Aksu Oasis as a representative case. Located on the northwestern edge of the Tarim Basin, the region faces severe secondary salinization due to intense evaporation and irrational irrigation practices. By integrating machine learning with the Optimal Parameters-based Geographical Detector (OPGD), we aim to elucidate the mechanisms underlying vertical variations in the salinization process, providing a scientific basis for three-dimensional monitoring and control.

2. Materials and Methods

2.1 Study Area and Data Collection The Aksu Oasis (40°9′–41°41′N, 78°46′–82°3′E) is characterized by a temperate continental arid climate. Annual precipitation is approximately 80.6 mm, while potential evaporation reaches 1200–1500 mm [**?**]. Field sampling was conducted in June 2024. A total of 71 sampling points were established, with composite soil samples collected from four depths: 0–10 cm, 10–20 cm, 20–30 cm, and 30–50 cm [FIGURE:1]. Soil soluble salt content (SSC) was determined using the 1:5 soil-to-water extraction method [**?**].

presents the descriptive statistics of SSC. The results indicate significant spatial variability and high coefficients of variation (C_v) across all layers, confirming the non-uniform distribution of salt influenced by micro-topography and irrigation.

2.2 Remote Sensing and Environmental Covariates Sentinel-2 Multi-Spectral Instrument (MSI) imagery was processed on the Google Earth Engine (GEE) platform. We selected 29 representative indices, including 10 vegetation indices (e.g., NDVI, EVI), 15 salinity indices (e.g., SI, NDSI), and 4 soil indices (e.g., BI, CLEX). Additionally, environmental covariates were integrated based on the SCORPAN framework, including topographic factors (slope, aspect), climatic factors (precipitation, potential evapotranspiration), and groundwater depth.

2.3 Modeling and Statistical Analysis We compared four machine learning models: Random Forest (RF), Support Vector Machine (SVM), LightGBM, and XGBoost. To address the "curse of dimensionality," Pearson correlation

Figure 5

Figure 1: Figure 5

Figure 6

Figure 2: Figure 6

analysis and Best Subset Regression (BSR) were employed to select optimal predictor combinations for each depth.

The Optimal Parameters Geographical Detector (OPGD) was used to quantify the driving factors. The factor detector measures the explanatory power (q-statistic) of individual variables:

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2}$$

Interaction detectors were used to evaluate whether the combined influence of factors was enhanced or weakened .

3. Results and Analysis

3.1 Model Performance and Uncertainty A comparative analysis based on R^2 and RMSE demonstrates that the XGBoost model provided the most accurate predictions across all four soil depths [FIGURE:3]. XGBoost achieved an R^2 of up to 0.70 in the 20–30 cm layer. Uncertainty was quantified using a 95% confidence interval (CI):

$$CI = P_i \pm 1.96 \times \sqrt{MSE + \sigma^2}$$

Spatial mapping of uncertainty [FIGURE:4] shows that while the model is generally stable, uncertainty increases in areas of extreme salinization and at greater soil depths.

3.2 Spatial and Vertical Distribution of Salinity The spatial distribution of SSC

reveals a ring-shaped pattern, with higher values along the oasis-desert ecotone and lower values in the interior near the Aksu River. Vertically, soil salinity exhibits distinct surface accumulation. The 0–10 cm layer is the most severely salinized

, driven by intense evaporation. As depth increases, the salt content generally decreases, though subsurface layers (30–50 cm) remain influenced by groundwater dynamics.

3.3 Driving Mechanisms and Interactions Factor detection results [FIGURE:8] indicate that the control mechanisms are vertically stratified. Surface salinity (0–10 cm) is primarily determined by the interaction between potential evapotranspiration and human activities (human footprint). In the deeper layers (30–50 cm), the "groundwater-evaporation" coupling mechanism becomes the decisive force, with groundwater table depth accounting for a significant portion of the total importance. Interaction analysis shows that all dual-factor interactions result in "bi-factor enhancement" or "nonlinear enhancement," meaning the combined effect of environmental variables is greater than their individual impacts.

4. Discussion and Conclusion

This study demonstrates that integrating machine learning with geographical detectors provides a robust framework for three-dimensional soil salinity monitoring. The XGBoost model proved superior in capturing the non-linear relationships between multi-source remote sensing data and vertical soil profiles.

The findings highlight that salinization in the Aksu Oasis is not a uniform process. The surface layer is highly sensitive to climatic and anthropogenic factors, while deeper layers are governed by hydrogeological conditions. This vertical stratification suggests that land management strategies must be depth-specific. For instance, surface-level prevention should focus on reducing evaporation and optimizing irrigation, while deeper-layer control requires managing groundwater levels and improving drainage infrastructure. This research provides essential technical support for precision agriculture and sustainable soil management in arid regions.