Postprint: Vegetation Community Diversity and Spatial Distribution Patterns in the Ganjiahu Haloxylon Forest National Nature Reserve, Xinjiang

Authors: Wang Anqi, Zhang Yu, Liu Jiawei, Zhao Wei, Liu Xinyu, Li Kai, Li Kai

Date: 2024-12-04T00:00:00+00:00

Abstract

Xinjiang Ganjiahu Haloxylon ammodendron Forest National Nature Reserve is situated in a typical oasis-desert ecotone. Identifying the main vegetation types and their distribution within the reserve is essential for biodiversity conservation and restoration. Based on remote sensing imagery from the same period in 2022 for sample plot establishment, a field survey was conducted in August 2023. Two-Way Indicator Species Analysis (TWINSPAN) was employed to classify vegetation quadrats within the reserve, alpha diversity indices were calculated for each vegetation type, and a Random Forest model was subsequently constructed to evaluate and visualize the clustering results. The results indicate that: (1) In the Ganjiahu Reserve, Salsola spp. is the most widely distributed among herbaceous plants, while shrubs such as Haloxylon ammodendron, Suaeda spp., Reaumuria songarica, and Ceratocarpus arenarius constitute the primary vegetation composition in this region. (2) TWINSPAN classified 62 vegetation quadrats into 14 types, among which the Salsola spp. + Suaeda spp. association (Ass. Salsola spp. + Suaeda spp., G5) occupies the largest area and exhibits the highest vegetation diversity; the Salsola spp. + Ceratocarpus arenarius association (Ass. Salsola spp. + Ceratocarpus arenarius, G7) demonstrates relatively high species diversity and a substantial area proportion; the single vegetation type association (G14) exhibits the lowest Shannon-Wiener diversity index and Simpson dominance index, with a Pielou evenness index of 1, primarily representing cropland and highly salinized bare land within the reserve. (3) The Random Forest model achieved an overall accuracy of 87.10% and a Kappa coefficient of 0.8553, with relatively small commission and omission errors. (4) Haloxylon ammodendron and Haloxylon persicum are concentrated in the northwestern core zone and the south-central buffer zone of the reserve. These findings can

provide a foundation for subsequent precision management of the reserve.

Full Text

Vegetation Community Diversity and Spatial Distribution Pattern in Ganjiahu Saxoul National Nature Reserve, Xinjiang

WANG Anqi[1], **ZHANG Yu**[1], **LIU Jiawei**[1], **ZHAO Wei**[2], **LIU Xinyu**[2], **LI Kai**[1]
[1]School of Ecology and Nature Conservation, Beijing Forestry University, Beijing 100083, China
[2]Jinghe Branch of Ganjiahu Saxoul National Nature Reserve Administration, Jinghe 833300, Xinjiang, China

Abstract

The Ganjiahu Saxoul National Nature Reserve in Xinjiang, China, is situated in a typical oasis-desert transition zone. Clearly identifying the main vegetation types and their distribution within the reserve is essential for biodiversity protection and restoration. Based on remote sensing images from 2022, a field survey was conducted in August 2023. Vegetation samples from the Ganjiahu Saxoul National Nature Reserve were classified using the two-way indicator species analysis (TWINSPAN) method. Alpha diversity indices of each vegetation type were calculated, followed by the development of a random forest model to evaluate and visualize the clustering results of vegetation quadrats. The results are as follows: (1) In the Ganjiahu Saxoul National Nature Reserve, the dominant herbaceous plants are primarily *Salsola* spp., while the principal shrubs include *Haloxylon ammodendron*, *Suaeda* spp., *Reaumuria songarica*, and *Ceratocarpus arenarius*. (2) Sixty-two vegetation sample plots were categorized into 14 types in TWINSPAN. Among these, the Ass. *Salsola* spp.+*Suaeda* spp. (G5) covered the largest area and exhibited the highest vegetation diversity. The Ass. *Salsola* spp. + *Ceratocarpus arenarius* (G7) showed both high species diversity and extensive coverage. The G14 type represented a single vegetation community with the lowest Shannon-Wiener and Simpson's diversity indices and had a Pielou's evenness index of 1, predominantly comprising farmland and highly saline bare land in the Ganjiahu Saxoul National Nature Reserve. (3) The random forest model achieved an overall accuracy of 87.10%, with a Kappa coefficient of 0.8553, and demonstrated relatively low commission and omission errors. (4) *H. ammodendron* and *H. persicum* are mainly distributed in the core area in the northwest and the buffer zone extending from the central to southern parts of the Ganjiahu Saxoul National Nature Reserve. These findings provide a foundation for future precise management of the Ganjiahu Saxoul National Nature Reserve.

Keywords: vegetation types; spatial distribution; two-way indicator species analysis method; species diversity; random forest

1 Introduction

Vegetation comprises all plant species within a region formed and developed under specific natural conditions. As primary producers, vegetation constitutes the most fundamental biological component of the biosphere. Comprehensive understanding of vegetation types and their distribution characteristics facilitates rational utilization of vegetation and related resources. Arid desert ecosystems are particularly fragile and among the regions most susceptible to climate change and human activities, where vegetation diversity can exert positive effects on ecosystem function changes in arid regions.

The Ganjiahu Saxoul National Nature Reserve (hereinafter referred to as Ganjiahu Reserve) was upgraded to a national-level nature reserve in 2007 and represents an important intersection distribution area for *Haloxylon ammodendron* and *H. persicum*. *Haloxylon ammodendron* and *H. persicum* are the main constructive species in arid desert regions, exhibiting excellent characteristics such as drought and salt-alkali tolerance, and play crucial ecological roles in soil and water conservation, windbreak and sand fixation, and carbon sequestration and oxygen release. Ganjiahu Reserve possesses a relatively complete desert ecosystem, and due to its unique geographic location and important ecological functions, it is essential to periodically and comprehensively understand the current status of biodiversity within the reserve.

The two-way indicator species analysis (TWINSPAN) method, proposed in 1979, has become one of the most frequently applied methods in vegetation community classification from the 1980s to present, and has been widely used in vegetation studies of forests, grasslands, deserts, arid river valleys, and wetlands. In recent years, with rapid scientific and technological development, big data algorithms have become increasingly mature, and more efficient and convenient vegetation classification methods have emerged, such as combining hyperspectral remote sensing with machine learning algorithms for plant classification. Random forest is an ensemble algorithm that utilizes multiple tree classifiers for classification. Through resampling methods, it extracts multiple samples from the original dataset, constructs decision trees for each sample, and finally integrates the results from multiple decision trees, offering advantages such as high accuracy and overfitting prevention. In ecological research, random forest has demonstrated superior fitting accuracy compared to other machine learning algorithms, becoming an important research tool in the field.

Traditional vegetation surveys generally employ the belt transect method for comprehensive investigation and uniform sampling of reserve vegetation. Although this approach offers high accuracy, the final results are difficult to express intuitively through images. Therefore, this study selected survey sample points based on vegetation patches preliminarily divided from remote sensing images in August 2022, combined machine learning algorithms with field survey data to analyze vegetation community distribution characteristics and diversity in Ganjiahu Reserve, and visually presented the spatial distribution patterns of vegetation communities in image form, providing important foundations for subsequent scientific construction and development of the reserve.

[Figure 1: see original paper] Location of the study area and distribution of survey sample sites

Note: Unsupervised classification of summer 2022 remote sensing data for Ganjiahu Reserve yielded 6 patch types, which served as the basis for random sampling selection of survey sites.

1.1 Study Area Overview

Ganjiahu Reserve (83°18 –83°52 E, 44°46 –44°58 N) is located in the hinterland of the Eurasian Continental Bridge, in the southwestern part of the Junggar Basin in Xinjiang. It spans Jinghe County of Bortala Mongol Autonomous Prefecture and Wusu City of Tacheng Region. The reserve's climate is temperate continental, characterized by dryness and low rainfall throughout the year, with an average annual temperature of 5.9–7.8°C, annual precipitation of approximately 140 mm, and enormous annual evaporation far exceeding precipitation. Zonal soils are gray-brown desert soil and gray desert soil, with extensive desert areas. Saline-alkali soil, meadow soil, and other intrazonal soils are also widely distributed. The reserve is traversed by seasonal rivers fed by snowmelt and rainfall, forming various landscape types including wetlands, deserts, and saline-alkali lands, which nurture diverse vegetation such as tall trees dominated by poplar, shrubs including *Haloxylon* and *Tamarix*, and herbaceous plants such as reed (*Phragmites australis*).

1.2 Survey Methods

Using Sentinel-2 satellite data as the data source, vegetation patches within the reserve were divided through unsupervised classification based on natural geographic elements. Sixty-two sample sites were selected for field survey, recording vegetation species, quantity, height, coverage, crown width, ground diameter, and other parameters within quadrats. The reserve belongs to desert grassland habitats, mostly composed of shrubs and herbaceous vegetation. According to the compilation specifications of *Vegetation of China*, large quadrat size was uniformly set at 10 m × 10 m to record all shrub information. Combined with the five-point sampling method, 2 m × 2 m small quadrats (q1–q62) were established within large quadrats to record herbaceous vegetation information. Since poplar is the only tall tree species in this region, its quantity was recorded within a 20 m × 20 m range in areas where poplar was distributed. This study included a total of 62 large quadrats, 62 shrub quadrats, and 62 herbaceous quadrats.

1.3 Data Processing

1.3.1 TWINSPAN Classification

By calculating the relative density, relative coverage, and relative frequency of each vegetation type within quadrats, the importance value (IV) of each species in each quadrat was comprehensively calculated to obtain the original species importance value matrix. The importance value is a comprehensive numerical index reflecting the role and status of a certain plant in its community. Species with IV ≥ 20 are considered constructive species, $10 \leq$ IV < 20 as companion species, and IV < 10 as rare species. Using the original importance value matrix, TWINSPAN was applied to classify vegetation quadrats within the study area. All calculations were performed in R Studio. The formulas for relative density, relative coverage, and relative frequency are as follows:

$$RD_i = \frac{D_i}{\sum D_i} \times 100\%$$

$$RC_i = \frac{C_i}{\sum C_i} \times 100\%$$

$$RF_i = \frac{F_i}{\sum F_i} \times 100\%$$

where RD_i, RC_i, and RF_i are the relative density, relative coverage, and relative frequency of species i in each quadrat, respectively; D_i, C_i, and F_i are the average density (individuals/100 m^2), average coverage (%), and average frequency (times) of species i in each quadrat, respectively.

The importance value calculation formula is:

$$IV = \frac{RD_i + RC_i + RF_i}{3}$$

1.3.2 Species α Diversity Analysis

The Shannon-Wiener index is considered the most common indicator for measuring community species diversity, used to explain species richness of sample communities. The Simpson dominance index reflects changes in species quantity within communities. The Pielou evenness index describes the uniformity of species distribution within communities. Based on TWINSPAN vegetation community classification results, the Shannon-Wiener diversity index (H'), Simpson dominance index (D), and Pielou evenness index (J) were calculated for each community using the following formulas:

Shannon-Wiener diversity index:

$$H' = -\sum_{i=1}^{S} P_i \times \ln P_i$$

Simpson dominance index:

$$D = 1 - \sum_{i=1}^{S} P_i^2$$

Pielou evenness index:

$$J = \frac{H'}{\ln S}$$

where P_i is the proportion of individuals of species i to the total number of individuals of all species, and S is the number of species.

1.3.3 Random Forest Model Construction and Accuracy Evaluation

To classify vegetation types across Ganjiahu Reserve and visualize the classification results, this study stratified and sampled 62 quadrats based on remote sensing classification and TWINSPAN results. Seventy percent of the quadrats were selected as the training set for random forest model construction, with the remaining 30% used as the test set for model accuracy verification. The verification results were overlaid with classification patches to represent the overall spatial layout of *Haloxylon* vegetation across the region.

A confusion matrix was generated by comparing the results with classification patches, and the Kappa coefficient was calculated to evaluate the accuracy of this classification method in vegetation classification. A higher Kappa coefficient indicates better model fitting.

1.3.4 Delineation of *Haloxylon* Plant Distribution Patterns

Among the classified vegetation types, high, medium, and low-level distribution areas of *Haloxylon* plants were delineated based on the importance values of *Haloxylon ammodendron* and *H. persicum*. These were used as qualitative data displayed through color intensity. The Kriging interpolation method in ArcGIS was applied to calculate the density of *Haloxylon* plants per unit area from field survey data as quantitative data. The density of *Haloxylon* plants was calculated from 42.4% of the field survey data.

2 Results

2.1 Overview of Reserve Vegetation and TWINSPAN Classification Results

The main vegetation information in Ganjiahu Reserve from the survey is shown in Table 1. In vegetation identification, except for *Salsola* spp., *Suaeda* spp., *Artemisia* spp., and *Calligonum* spp., which were identified to genus level, all other plants were identified to species level. Within Ganjiahu Reserve, herbaceous plants are most widely distributed as *Salsola* spp., appearing in approximately 62.5% of quadrats and serving as the dominant species in 8 associations with the highest average importance value. Shrubs such as *Haloxylon ammodendron*, *Suaeda* spp., *Reaumuria songarica*, and *Tamarix ramosissima* have relatively high average importance values and coverage, representing the main vegetation components in this region.

According to the classification principles of *Vegetation of China* and based on importance values, 62 large quadrats in Ganjiahu Reserve were classified using TWINSPAN and named according to community dominant species, resulting in 14 associations (Figure 2). The composition of each vegetation association is as follows:

G1: *Artemisia* spp. Association
Contains 2 large quadrats with total community coverage of 23.7%. *Artemisia* spp. importance value is 51.2%. Main companion plant groups are *Salsola* spp. Rare species in this association include *Calligonum* spp., *Alhagi camelorum*, *Haloxylon persicum*, *Agriophyllum pungens*, *Apocynum pictum*, and *Echinops sphaerocephalus*.

G2: *Suaeda* spp. + *Ceratocarpus arenarius* Association
Contains 6 large quadrats with total community coverage of 48.25%. *Suaeda* spp. importance value is 6.65%; *Ceratocarpus arenarius* importance value is 4.95%. Main companion plant groups include *Salsola* spp. and *Iljinia regelii*. Rarely distributed species include *Alhagi camelorum*, *Populus euphratica*, *Krascheninnikovia ceratoides*, *Haloxylon ammodendron*, and *Capsella bursa-pastoris*.

G3: *Haloxylon persicum* + *Salsola* spp. Association
Contains 4 large quadrats with total community coverage of 17.5%. *Haloxylon persicum* importance value is 14.84% with coverage of 16.1%; *Salsola* spp. importance value is 10.72% with coverage of 3.95%. Main companion plant is *Populus euphratica*. Rare species include *Alhagi camelorum*, *Reaumuria songarica*, *Tamarix ramosissima*, and *Calligonum* spp.

G4: *Salsola* spp. + *Haloxylon persicum* Association
Contains 2 large quadrats with total community coverage of 51.2%. *Salsola* spp. importance value is 40.23%; *Haloxylon persicum* importance value is 8.75%. Main companion plant groups are *Suaeda* spp. and *Ceratocarpus arenarius*. Rare species include *Iljinia regelii*, *Caragana halodendron*, *Limonium gmelinii*,

Haloxylon ammodendron, and *Anabasis salsa*.

G5: *Salsola* spp. + *Suaeda* spp. Association
Contains 8 large quadrats with total community coverage of 62.5%. *Salsola* spp. importance value is 13.48%; *Suaeda* spp. importance value is 47.5% with coverage of 9.15%. Main companion plants are *Tamarix ramosissima* and *Iljinia regelii*. Rare species include *Haloxylon ammodendron* and *Eremopyrum orientale*.

G6: *Salsola* spp. + *Reaumuria songarica* Association
Contains 4 large quadrats with total community coverage of 43.75%. *Salsola* spp. importance value is 17.52%; *Reaumuria songarica* importance value is 7.65%. Main companion plant groups are *Suaeda* spp. and *Caragana halodendron*. Rare species include *Haloxylon ammodendron*.

G7: *Salsola* spp. + *Ceratocarpus arenarius* Association
Contains 4 large quadrats with total community coverage of 81.26%. *Salsola* spp. importance value is 83.71%; *Ceratocarpus arenarius* importance value is 13.23%. Main companion plant groups are *Suaeda* spp. and *Tamarix ramosissima*. Rare species include *Haloxylon ammodendron*, *Eremopyrum orientale*, *Calligonum* spp., *Anabasis salsa*, and *Limonium gmelinii*.

G8: *Haloxylon ammodendron* + *Salsola* spp. Association
Contains 6 large quadrats with total community coverage of 83.71%. *Haloxylon ammodendron* importance value is 13.23%; *Salsola* spp. importance value is 81.26%. Main companion plant group is *Suaeda* spp. Rare species include *Halostachys caspica*, *Kalidium foliatum*, *Grubovia dasyphylla*, *Glycyrrhiza uralensis*, *Nitraria sibirica*, *Limonium gmelinii*, and *Apocynum pictum*.

G9: *Phragmites australis* + *Salsola* spp. Association
Contains 2 large quadrats with total community coverage of 13.23%. *Phragmites australis* importance value is 81.26%; *Salsola* spp. importance value is 13.23%. Main companion plant is *Reaumuria songarica*. Rare species include *Phragmites australis*, *Nitraria sibirica*, *Calligonum* spp., and *Cistanche deserticola*.

G10: *Salsola* spp. + *Halocnemum strobilaceum* Association
Contains 2 large quadrats with total community coverage of 81.26%. *Salsola* spp. importance value is 13.23%; *Halocnemum strobilaceum* importance value is 81.26%. Main companion plant group is *Suaeda* spp.

G11: *Salsola* spp. + *Haloxylon ammodendron* Association
Contains 4 large quadrats with total community coverage of 13.23%. *Salsola* spp. importance value is 81.26%; *Haloxylon ammodendron* importance value is 13.23%. Main companion plant group is *Suaeda* spp. Rare species include *Karelinia caspia* and *Haloxylon persicum*.

G12: *Karelinia caspia* Association
Contains 2 large quadrats with total community coverage of 81.26%. *Karelinia caspia* importance value is 13.23%; *Salsola* spp. importance value is 81.26%.

Main companion plant is *Reaumuria songarica*. Rare species include *Phragmites australis*, *Lycium ruthenicum*, *Grubovia dasyphylla*, *Krascheninnikovia ceratoides*, *Cistanche deserticola*, etc.

G13: *Lycium ruthenicum* + *Phragmites australis* Association
Contains 2 large quadrats with total community coverage of 13.23%. *Lycium ruthenicum* importance value is 81.26% with coverage of 13.23%; *Phragmites australis* importance value is 13.23%. Companion plants include *Tamarix ramosissima* and *Haloxylon ammodendron*.

G14: Single Vegetation Type Association
Contains 2 large quadrats with total community coverage of 1.67%. This association has very low wild plant coverage. Field surveys show it mainly consists of two situations: (1) Farmland in the experimental zone on the periphery of the reserve; (2) Highly saline bare land with extremely fragile habitats in the reserve, where the surface is covered by thick saline crusts with only minimal vegetation growth.

Overview of the main vegetation in Ganjiahu Nature Reserve

2.2 Vegetation Community Diversity Analysis

Species diversity effectively characterizes the complexity of biological community and ecosystem structure. Different vegetation community types represented by different dominant species are formed under the combined effects of topography, substrate, moisture, salinity, and other factors. Alpha diversity indices were analyzed for classified quadrats, expressed through the Shannon-Wiener diversity index, Simpson dominance index, and Pielou evenness index.

In Ganjiahu Reserve, the alpha diversity indices of vegetation communities show that the *Salsola* spp.+*Suaeda* spp. association (G5) has the highest Shannon-Wiener diversity index, indicating this vegetation association contains the richest species with up to 31 species. The *Salsola* spp.+*Ceratocarpus arenarius* association (G7) also has relatively high diversity. The single vegetation type association (G14) has the lowest Shannon-Wiener and Simpson's diversity indices, with a Pielou evenness index of 1, and is predominantly represented by farmland and highly saline bare land in the reserve.

Although the *Salsola* spp.+*Haloxylon persicum* association (G3) and *Salsola* spp.+*Haloxylon persicum* association (G4) share the same dominant species (*Salsola* spp. and *Haloxylon persicum*), their companion and rare species differ significantly. The G3 association mainly accompanies *Populus euphratica*, with rare species including *Alhagi camelorum*, *Reaumuria songarica*, *Tamarix ramosissima*, and *Calligonum* spp. The G4 association mainly accompanies *Ceratocarpus arenarius* and *Suaeda* spp., with rare species including *Iljinia regelii*, *Caragana halodendron*, *Limonium gmelinii*, *Haloxylon ammodendron*, and *Anabasis salsa*. This difference may be attributed to variations in soil salinity and groundwater conditions among these vegetation communities, ultimately resulting in the G4 association having higher Shannon-Wiener diversity index, Simpson dominance index, and Pielou evenness index than the G3 association.

α diversity indices of different vegetation clusters

2.3 Vegetation Type Classification Based on Random Forest

Using vegetation classification data, a random forest model was constructed to obtain the vegetation type distribution map of Ganjiahu Reserve (Figure 3). A confusion matrix was used to evaluate classification accuracy (Table 3). The overall classification accuracy reached 87.10% with a Kappa coefficient of 0.8553, indicating good overall classification accuracy based on random forest. Producer's accuracy and user's accuracy for each vegetation type were higher than 70%, with most concentrated between 70%–100%, and both commission errors and omission errors were relatively small.

By calculating the area of each vegetation community classification, the *Salsola* spp.+*Suaeda* spp. association (G5) covered the largest area of 124.87 km^2, accounting for 18.91% of the total Ganjiahu area. The *Salsola* spp.+*Ceratocarpus arenarius* association (G7) covered 116.15 km^2, accounting for 13.63%. Other vegetation communities had smaller distribution areas (Table 3).

[Figure 3: see original paper] Distribution of vegetation types in Ganjiahu Nature Reserve based on random forest model

Verification of the area and accuracy of each vegetation type based on random forest model

2.4 Delineation of *Haloxylon* Vegetation Distribution Areas

For Ganjiahu Reserve, *Haloxylon ammodendron* and *H. persicum* are local constructive species and nationally protected plants with important ecological significance. Therefore, this study focused on these two *Haloxylon* species, combining field surveys with model construction using qualitative and quantitative methods to determine their spatial distribution (Figure 4). The study found that *Haloxylon* species are mainly concentrated in the northwestern core area and the central-to-southern buffer zone of the reserve, with fewer distributions in the southeastern and northeastern experimental zones, forming a strip-shaped geographic distribution pattern.

[Figure 4: see original paper] Spatial distribution pattern of *Haloxylon* plants in Ganjiahu Nature Reserve

3 Discussion

As an oasis-desert transition zone on the edge of an oasis, Ganjiahu Reserve plays a crucial role in blocking wind-sand intrusion. Therefore, research on vegetation diversity analysis and vegetation community spatial distribution patterns in the reserve is imperative. In Ganjiahu Reserve, 62 sample plots were classified into 14 vegetation community types based on TWINSPAN. Among them, the *Salsola* spp.+*Suaeda* spp. association (G5) covered the largest area and exhibited the highest vegetation diversity. The *Salsola* spp.+*Ceratocarpus arenarius* association (G7) showed high species diversity while also occupying a relatively large area. The single vegetation type association (G14) had the lowest Shannon-Wiener diversity index and Simpson dominance index, with a Pielou evenness index of 1, and was predominantly represented by farmland and highly saline bare land in the reserve. Areas with high vegetation diversity were mainly distributed in the northwestern, southern, and eastern edges of the reserve, showing a penetrating distribution pattern closely related to topography and water source conditions.

The *Phragmites australis* association has relatively high diversity, likely because most are located in riparian edges and other areas with good groundwater recharge. Although the *Salsola* spp.+*Haloxylon persicum* associations (G3 and G4) share the same dominant species, their companion and rare species differ dramatically: G3 mainly accompanies *Populus euphratica* with rare species including *Alhagi camelorum*, *Reaumuria songarica*, *Tamarix ramosissima*, and *Calligonum* spp.; G4 mainly accompanies *Ceratocarpus arenarius* and *Suaeda* spp. with rare species including *Iljinia regelii*, *Caragana halodendron*, *Limonium gmelinii*, *Haloxylon ammodendron*, and *Anabasis salsa*. This situation may arise from differences in soil salinity and groundwater among these vegetation communities, ultimately resulting in G4 having higher diversity indices than G3. This study lacks correlation analysis between soil types and vegetation community classification, a limitation that will be addressed in future research.

Random forest models have gained popularity in recent years due to their excellent classification results and efficient processing speed. Using randomly selected training samples and variable subsets to generate multiple decision trees, random forest can learn from satellite remote sensing data, UAV monitoring data, and field survey data, making it applicable not only for vegetation classification but also for vegetation identification, dynamic monitoring, and land use change research in ecology. In typical arid landscapes such as sand dunes and marshes, random forest classifiers have demonstrated high accuracy for shrub identification. Barrett et al. evaluated the use of medium spatial resolution optical and radar satellite data combined with soil and terrain data to identify and map upland vegetation using random forest algorithms, with classification results being universally applicable across different study areas. This study used remote sensing to divide Ganjiahu Reserve into 6 vegetation types, randomly sampled field survey data based on this for vegetation classification, obtained 14 vegetation community types, and established a random forest model to correct remote sensing classification results. The model achieved an overall accuracy of 87.10% and a Kappa coefficient of 0.8553, confirming the feasibility of combining remote sensing imagery with field survey data for vegetation classification.

Haloxylon species, as constructive and pioneer species in the Junggar Basin, play an irreplaceable role in maintaining local community stability and ecosystem balance, such as consolidating sand particles, promoting soil formation, reducing soil water evaporation, providing shelter for desert animals, altering direct solar radiation absorption and loss processes to regulate local microclimate, and achieving dynamic balance of deep soil moisture in desert areas. Recent studies have shown that reserve vegetation has degraded seriously due to climate change and human activities, with rapid increases in sand and saline-alkali land areas and substantial decreases in grassland area since 2000, presenting a concerning situation. Clarifying the distribution of *Haloxylon* in the reserve is fundamental for many research efforts and enables managers to conduct targeted monitoring and protection. Cao Jiarui's study on the spatial pattern of *Haloxylon persicum* in Ganjiahu suggested that *H. persicum* distribution is relatively dense in the northwestern part of the study area with a trend of gradual diffusion to edge areas, while distribution density is smaller in the southeastern *H. persicum* communities. This study shows that *Haloxylon ammodendron* and *H. persicum* are mainly distributed in the northwestern core area and the central-to-southern buffer zone of the reserve with high density, while the southeastern and northeastern experimental zones have fewer distributions, forming a strip-shaped distribution pattern. These findings are consistent with previous research while adding delineation of *Haloxylon* distribution range and further refining the distribution pattern of *Haloxylon* species in the study area.

4 Conclusion

1) In Ganjiahu Reserve, herbaceous plants are most widely distributed as *Salsola* spp., appearing in approximately 62.5% of quadrats and serving as the dominant species in 8 associations. Shrubs such as *Haloxylon ammodendron*, *Suaeda* spp., *Reaumuria songarica*, and *Tamarix ramosissima* have relatively high average importance values, representing the main vegetation components in this region.

2) The 62 sample plots were classified into 14 vegetation types using TWINSPAN. Among them, the *Salsola* spp.+*Suaeda* spp. association (G5) covered the largest area and exhibited the highest vegetation diversity. The *Salsola* spp.+*Ceratocarpus arenarius* association (G7) showed high species diversity while also occupying a relatively large area. The single vegetation type association (G14) had the lowest Shannon-Wiener diversity index and Simpson dominance index, with a Pielou evenness index of 1, and was predominantly represented by farmland and highly saline bare land in the reserve.

3) Based on remote sensing classification, random forest models were constructed using Ganjiahu vegetation classification data. The overall classification accuracy reached 87.10% with a Kappa coefficient of 0.8553. Producer's accuracy and user's accuracy for vegetation types were mostly concentrated between 70%–100%, confirming the feasibility of combining remote sensing imagery with field survey data for vegetation classification.

4) The spatial pattern of the two *Haloxylon* species shows that *Haloxylon ammodendron* and *H. persicum* are mainly distributed in the northwestern core area and the central-to-southern buffer zone of the reserve, with fewer distributions in the southeastern and northeastern experimental zones, forming a strip-shaped geographic distribution pattern.

Acknowledgments

We thank the Jinghe and Wusu branches of the Ganjiahu Saxoul National Nature Reserve Administration for their strong support of this study, and we acknowledge the contributions and assistance of staff at the Wudaoquan Management Station of the reserve.

References

[1] Fang Jingyun, Guo Ke, Wang Guohong, et al. Vegetation classification system and classification of vegetation types used for the compilation of *Vegetation of China* [J]. *Chinese Journal of Plant Ecology*, 2020, 44(2): 96-110.

[2] Li C J, Fu B J, Wang S, et al. Drivers and impacts of changes in China's drylands[J]. *Nature Reviews Earth & Environment*, 2021, 2(12): 858-873.

[3] Liu Bingqiang. Causes of environmental degradation and control strategies in the Ganjiahu desert forest area[J]. *Protection Forest Science and Technology*, 2014(10): 85, 126.

[4] Chen Yuting, Ma Songmei, Zhang Dan, et al. Diversity pattern and formation mechanism of sympatric *Haloxylon ammodendron* and *Haloxylon persicum* in Xinjiang, China[J]. *Chinese Journal of Plant Ecology*, 2024, 48(1): 56-67.

[5] Yu Tian, Zhang Junping, Ma Hongyu, et al. Advances of study on the adversity resistance of *Haloxylon*[J]. *Xinjiang Agricultural Sciences*, 2009, 46(4): 907-912.

[6] Cai Xinbin, Wu Junxia. Characteristics and dynamics analysis of *Haloxylon persicum* populations in the Nature Reserve of Gan Jia Lake[J]. *Journal of Arid Land Resources and Environment*, 2016, 30(7): 90-94.

[7] Li Yinmin. Last three decades forest degradation process of Ganjiahu Nature Reserve in Xinjiang[D]. Urumqi: Xinjiang Normal University, 2013.

[8] Hill M O. TWINSPAN: A fortran program for arranging multivariate data in an ordered two way table by classification of the individuals and attributes[D]. New York: Cornell University, 1979.

[9] Wang Guohong, Fang Jingyun, Guo Ke, et al. Contents and protocols for the classification and description of Vegetation Formations, Alliances and Associations of *Vegetation of China* [J]. *Chinese Journal of Plant Ecology*, 2020, 44(2): 128-178.

[10] Swierkosz K, Reczyńska K, Boublík K. Variability of *Abies alba*-dominated forests in central Europe[J]. *Open Life Sciences*, 2014, 9(5): 495-518.

[11] Zhu Y J, Shan D, Wang B Z, et al. Floristic features and vegetation classification of the Hulun Buir steppe in north China: Geography and climate driven steppe diversification[J]. *Global Ecology and Conservation*, 2019, 20: e00741.

[12] Nkosi S E, Barrett A S, Brown L R. Vegetation ecology of the Nooitgedacht section of Loskop Dam Nature Reserve, Mpumalanga[J]. *South African Journal of Botany*, 2016, 105: 79-88.

[13] Sallam H, Alzain M N, Abuzaid A O, et al. Wild plant diversity and soil characteristics of desert roadside vegetation in the eastern desert[J]. *Diversity*, 2023, 15(7): 874.

[14] Zeng Y, Zhao C Y, Shi F Z, et al. Impact of groundwater depth and soil salinity on riparian plant diversity and distribution in an arid area of China[J]. *Scientific Reports*, 2020, 10(1): 7272.

[15] Manhas R K, Gautam M K, Kumari D. Two way indicator species analysis (TWINSPAN) of the herbaceous vegetation in an inland wetland ecosystem of Doon Valley Himalaya, India[J]. *Journal of Wetlands Ecology*, 2009, 3: 23-31.

[16] Rollet R, Benie G B, Li W, et al. Image classification algorithm based on the RBF neural network and means[J]. *International Journal of Remote Sensing*, 1998, 19(15): 3003-3009.

[17] Wen Tong, Liu Xiaoni, Ji Tong, et al. Studying on plant classification and recognition method for Three River Source alpine grassland plant based on vegetation index[J]. *Acta Agrestia Sinica*, 2022, 30(7): 1811-1818.

[18] Belgiu M, Drăguţ L. Random forest in remote sensing: A review of applications and future directions[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 114: 24-31.

[19] Zhou J H, Lai L M, Guan T Y, et al. Comparison modeling for alpine vegetation distribution in an arid area[J]. *Environmental Monitoring and Assessment*, 2016, 188(7): 408.

[20] Han F S, Fu G, Yu C Q, et al. Modeling nutrition quality and storage of forage using climate data and normalized difference vegetation index in alpine grasslands[J]. *Remote Sensing*, 2022, 14(14): 3410.

[21] Caddy Retalic S, Wardle G M, Leitch E J, et al. Vegetation change along a Mediterranean to arid zone bioclimatic gradient reveals scale dependent ecotone patterning[J]. *Australian Journal of Botany*, 2020, 68(8): 574-586.

[22] Sarker S K, Nur Nabi M, Mohasinul H M, et al. Tree assemblages and diversity patterns in tropical Juri forest, Bangladesh[J]. *Journal of Forestry Research*, 2015, 26(1): 159-169.

[23] Liu Liyan, Cai Xinbin, Jiang Xiaoheng, et al. Types and characteristics of vegetation in Ganjiahu *Haloxylon* Forest National Nature Reserve[J]. *Protection Forest Science and Technology*, 2016(6): 54-57.

[24] Ma Qian, Chu Xinzheng, Sun Huilan, et al. Study on land use/cover dynamic change in Ganjia Lake Ecotone[J]. *Journal of Xinjiang Normal University (Natural Sciences Edition)*, 2014, 33(1): 8-13, 37.

[25] Wang Yusong, Shangguan Tieliang. Discussion on calculating method of important values[J]. *Journal of Shanxi University (Natural Sciences Edition)*, 2010, 33(2): 312-316.

[26] Zang S, Cao Q, Keremu A, et al. Food patch particularity and foraging strategy of reintroduced Przewalski's horse in north Xinjiang, China[J]. *Turkish Journal of Zoology*, 2017, 41: 924-930.

[27] Li Junxiao, Li Chaokui, Yin Zhihui. ArcGIS based Kriging interpolation method and its application[J]. *Bulletin of Surveying and Mapping*, 2013(9): 87-90.

[28] Liu Liyan, Cai Xinbin, Jiang Xiaoheng, et al. Species composition and floral components of vascular plants in Ganjiahu Nature Reserve of Xinjiang[J]. *Plant Science Journal*, 2016, 34(5): 695-704.

[29] Xu Xianying, Yan Ping, Guo Shujiang, et al. The interception loss of rainfall by three sand fixing shrubs at the fringe of Minqin Oasis[J]. *Journal of Desert Research*, 2013, 33(1): 141-145.

[30] Li X R, Xiao H L, Zhang J G, et al. Long term ecosystem effects of binding vegetation in the Tengger Desert, northern China[J]. *Restoration Ecology*, 2004, 3(12): 376-390.

[31] Li Yinmin, Qu Ning. A brief analysis of the changes in *Haloxylon* forest in Ganjiahu[J]. *Journal of Hetian Normal College*, 2014(5): 77-82.

[32] Zhu Jialong, Zhou Zhibin, Wang Lisheng, et al. Coupling relationship between growth and soil moisture change of *Haloxylon ammodendron* plantation without irrigation[J]. *Arid Land Geography*, 2022, 45(5): 1579-1590.

[33] Cao Jiarui. Analysis of the spatial pattern of *Haloxylon persicum* in Ganjiahu[J]. *South China Agriculture*, 2020, 14(27): 186-188, 218.

[34] Abdollahi A, Pradhan B, Alamri A. Regional scale analysis of vegetation dynamics using satellite data and machine learning algorithms: A multi factorial approach[J]. *International Journal on Smart Sensing and Intelligent Systems*, 2023, 16(1): 264378840.

[35] Tikuye B G, Rusnak M, Manjunatha B R, et al. Land use and land cover change detection using the random forest approach: The case of the upper Blue Nile River Basin, Ethiopia[J]. *Global Challenges*, 2023, 7(10): 2300155.

[36] Francis R J, Kingsford R T, Moseby K, et al. Tracking landscape scale vegetation change in the arid zone by integrating ground, drone and satellite data[J]. *Remote Sensing in Ecology and Conservation*, 2023, 10(3): 374-387.

[37] Barrett B, Raab C, Cawkwell F, et al. Upland vegetation mapping using random forests with optical and radar satellite data[J]. *Remote Sensing in Ecology and Conservation*, 2016, 2(4): 212-231.

[38] Gao Xiaoyu, Hao Haichao, Zhang Xueqi, et al. Responses of vegetation water use efficiency to meteorological factors in arid areas of northwest China: A case of Xinjiang[J]. *Arid Land Geography*, 2023, 46(7): 1111-1120.

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