

Modelling the effects of topographic heterogeneity on distribution of *Nitraria tangutorum* Bobr. species in deserts using LiDAR-data

Huoyan Zhou

Honghe University

Linyan Feng

Chinese Academy of Forestry

Liyong Fu

Chinese Academy of Forestry

Ram P. Sharma

Tribhuvan University

Xiao Zhou

International Network for Bamboo and Rattan

Xiaodi Zhao (✉ edison19890918@126.com)

Chinese Academy of Forestry

Article

Keywords: species distribution model, small scale, desert, lidar, random forest

Posted Date: March 8th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2566709/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations:

No competing interests reported.

Tables 1-2 is available in the Supplementary Files section.

Version of Record: A version of this preprint was published at Scientific Reports on August 22nd, 2023.

See the published version at <https://doi.org/10.1038/s41598-023-40678-5>.

Modelling the effects of topographic heterogeneity on distribution of *Nitraria tangutorum* Bobr. species in deserts using LiDAR-data

Huoyan Zhou^{1,2}, Linyan Feng³, Liyong Fu⁴, Ram P. Sharma⁵, Xiao Zhou⁶, and Xiaodi Zhao^{2,*}

1 College of Biological Sciences and Agronomy, Honghe university, Mengzi 661100, Yunnan province, P. R. of China; zhou.huoyan@163.com

2 Research Institute of Forestry Policy and Information, Chinese Academy of Forestry, Beijing 100091, China

3 Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, P. R. of China; linyan_feng@caf.ac.cn

4 Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, P. R. of China; fuly@ifrit.ac.cn

5 Institute of Forestry, Tribhuvan University, Kritipur, Kathmandu-44600, Nepal; ramsharm1@gmail.com

6 International Center for Bamboo and Rattan, Key Laboratory of National Forestry and Grassland Administration, Beijing 100102, China; szzx19960408@163.com

* Corresponding author Email: edison19890918@126.com; Tel: +86-10-62889196; Fax: +86-10-62888315

Abstract:

Microclimate ecology is attracting renewed attention because of its fundamental importance in understanding how organisms respond to climate change. A number of hot issues can be investigated in desert ecosystems, including the relationship between species distribution and environmental gradients (e.g., elevation, slope, topographic convergence index, and solar insolation). Species distribution models (SDMs) can be used to understand these relationships. We used the data acquired from the important desert plant *Nitraria tangutorum* Bobr. communities and desert topographic factors extracted from LiDAR (Light detection and ranging) data of one square kilometers in the Inner Mongolia region of China to construct the SDMs. We evaluated the performance of the SDMs constructed with both the variants of the parametric and nonparametric algorithms (bioclimatic Modelling (BIOCLIM), Domain, Mahalanobi, generalized linear model (GLM), generalized additive model (GAM), random forest (RF), and support vector machine (SVM)). The area under the receiver operating characteristic curve was used to evaluate the algorithms. The SDMs constructed with RF appeared to be the best based on the area under the curve (0.7733). We also generated *Nitraria tangutorum* Bobr. distribution maps with the constructed SDMs and the suitable habitat area of the Domain model. Based on the suitability map, we conclude that *Nitraria tangutorum* Bobr. is more suited to the southern part with a slope of 0-20 degree at an elevation of approximately 1010 m. This is the first attempt of modelling the effects of topographic heterogeneity on desert species distribution on a small scale. The presented SDMs will have important applications for predicting species distribution and will be useful for preparing conservation and management plans for desert ecosystems on a small scale.

Keywords: species distribution model; small scale; desert; lidar; random forest

Introduction

The Ulan Buh Desert is the 8th largest desert in China. It is located in the mid-latitudinal zone and is characterized by a typical continental climate with little precipitation. The desert is located in a transitional zone with a temperate climate ranging from semiarid to arid conditions ¹. It is one of the driest regions at similar latitudes in the world and has a very sensitive ecological environment, which creates a fragile desert ecotone. The desert ecotone is the key ecoregion and is suitable for conducting research on ecosystem degradation and recovery mechanisms. Desert ecotones may have high ecological significance, which quantitatively reveals the interactions between plants and environmental factors. The spatial distribution pattern of the desert plants is therefore the result of the interactions between the plant communities and the harsh environmental factors ².

Spatial distributions are fundamental to species ecology ³ and help identify threats to the conservation of plant species, manage the impacts of biological invasion ⁴, and plan conservation strategies. Climate, edaphic, and topographic factors are the main factors affecting species distributions ⁵. These factors are affected by a variety of biophysical processes including competitive interactions between the plant species and their dispersal history ⁶. SDMs, like habitat suitability or ecological models, are built with the observed species occurrence as a response variable and environmental factors as predictor variables ⁷. Many modelling techniques are available to construct the SDMs ⁸⁻¹⁰. Although previous SDMs studies have compared different modelling techniques for their precision ¹¹⁻¹⁶, most of them describe species distributions over a large scale and therefore have a low predictive performance in small localized conditions ^{6, 17-19}. Most SDMs are calibrated only with climate data at a low spatial resolution, such as the climate data acquired from WorldClim(<https://www.worldclim>), with a 1 km grid resolution. However, some studies have indicated that the predictive performance of SDMs can be substantially improved by incorporating other environmental factors in addition to climate factors ²⁰ or by considering finer-scale processes, such as topo-climate ²¹. Topographical heterogeneity is the most important factor affecting the spatial distribution of desert plants ²². In mountainous regions, topography may precisely describe local temperature, light, and humidity. Topography significantly affects light and water availability and soil development. Furthermore, the topography affects species composition, structure, and appearance and dynamics of the plant communities ²³⁻²⁵.

Building on the previous work done in this field, the current study intends to answer a few important questions including ‘what distribution patterns are followed by the desert plant species?’; and ‘which topography factors play a major role in affecting the species distribution?’ Intense solar radiation, high evapotranspiration, extreme temperatures, degraded soil and vegetation cover, low atmospheric moisture, and harsh topographic features are important factors in the desert environment. The desert is extremely arid with < 60-100 mm mean annual precipitation, and therefore, its moisture regime is the most critical factor for regulating biological processes and species distribution. The rainfall spatially varies at different scales, not only at the regional scale ranging from 0.1 km to 10 km but also at a local scale of a few meters. Rainfall is the main source of moisture for desert plants, which triggers species survival and growth ²⁶. Many studies have explored the appropriate methods for understanding the relationships between vegetation dynamics, soil moisture, and temperature in the desert ecosystem ²⁷⁻²⁹. However, under similar climate and desert environmental conditions, the influence of topography on species distribution has rarely been investigated on a small scale.

Additionally, several studies have been conducted on the topographic factors affecting animal communities in

the desert. In recent decades, some studies have been conducted to model the environmental variables influencing kit fox distribution. For example, Maxent fixed-effects and generalized linear mixed-effects models have been constructed to describe kit fox distribution²⁹. These modelling studies show the large effects of elevation on kit fox survival and growth, followed by that of the slope and canopy height of the vegetation at the landscape scale. The Maxent algorithm was used to predict the habitat in the Mojave Desert and parts of the Sonoran Desert considering the geographical and topographical factors at the spatial grid of one square kilometers³⁰. The field - validated site - and landscape - level SDMs were developed to identify the potentially rare and endemic plant habitats in the Great Basin of Western North America³¹, in which the potential habitat combined with the elevation, slope, aspect, rock type, and geologic processes is mapped with the resolution of 100 m². Similar to climatic factors, non-climate predictors (e.g., topography and habitat) are also important at a finer scale³². Thus, the analysis of these factors on a finer scale could be the most appropriate for constructing SDMs and increasing their predictive accuracy³³. As mentioned above, topography is one of the most important predictors of SDMs. It controls the habitat structure, biota, and all growing conditions locally³⁴, especially in the desert, where topographic features are more unfavorable than other land features. Thus, a high-resolution topographic map is necessary to construct the SDMs.

For dominant species, such as *Nitraria tangutorum* Bobr. in the desert, a small geographic range and narrow habitat may lead to acute habitat limitation, which is strongly correlated with other environmental factors. The microclimate, which influences ecosystem dynamics and process, is often disregarded in ecology and evolution³⁵. It is meaningful to investigate the geographical variations of *Nitraria tangutorum* Bobr. species in the desert on a small scale, such as within the scale of 100 m, where topographic factors significantly affect species distribution. Existing studies have shown that geographic and topographic predictors can improved the fitting performance of SDMs substantially³⁶.

High-resolution terrain data are rarely available. Remote sensing technology, which offers alternative solutions to various data sets, can supply accurate high-resolution terrain data using light detection and ranging (LiDAR) technology. The remote sensing data is important for determining species distribution^{37, 38}. In recent years, LiDAR datasets have become useful for modelling biodiversity and habitat analyses. For example, LiDAR data could improve the prediction of composition and changes in plant communities^{39, 40}, improve the descriptions of animal behaviors and their spatial distributions⁴¹, and enhance species richness prediction¹⁶.

To understand the relationships between desert plants and topographic factors affecting species distributions, this study developed SDMs that can predict the precise distribution of species covering an area of one square kilometers with a resolution of 10 m. Topographic variables derived from LiDAR data (resolution: 0.15 m) were used to construct SDMs. Different modelling algorithms were evaluated for 67 *Nitraria tanguotrum* species in the Ulan Buh Desert, an arid land in northern China. The objectives of this study were to: (1) compare various modelling algorithms, such as parametric and non-parametric variants, and choose the best to construct SDMs; (2) evaluate whether topographic factors included in algorithms would be useful for the precise prediction of the species distribution in the desert; (3) determine the effects of topographic heterogeneity on the distribution of *Nitraria tangutorum* Bobr. ; and (4) simulate the habitat suitability maps of *Nitraria tangutorum* Bobr. and suggest the management implications of the presented SDMs.

Results

●Model performances

Random sampling and each of the seven models were ran ten times, the AUC, true positive rate (TPR) and false positive rate (TNR) were calculated. With the presence of 67 and the absence of 100, The results indicated the AUC varied from low (0.59) to high (0.81), indicating different qualities of the prediction SDMs of *Nitraria tangutorum* *Bohr.* (Fig. 1). The RF produced the highest AUC value, and the Domain had the lowest AUC value (Table 1). All model performances were fair and satisfactory, except for BIOCLIM and Domain. RF produced the largest AUC, but not with maximum TPR and TNR; the TPR and TNR of RF were the third lowest of the seven models. Mahalanobis produced the highest TPR+TNR value.

Table 1

For the profile models, Mahalanobis produced the highest AUC, while AUC for Domain, which was the lowest, and AUC for BIOCLIM were more stable (Table 1). For the GLM and GAM, the AUC values were almost identical. For the machine learning models, the AUC for RF was higher than that for SVM, and the same was true for max TPR+TNR. RF proved to be better than SVM in modelling the distribution of *Nitraria tangutorum* *Bohr.* in the desert.

Figure 1

●Habitat suitability and prediction maps for *Nitraria tangutorum* *Bohr.* distribution

We produced maps showing habitat suitability for *Nitraria tangutorum* *Bohr.* . The potentially suitable habitat distribution maps of *Nitraria tangutorum* *Bohr.* in the study area were obtained from different algorithms. Furthermore, the corresponding species presence/absence with each algorithm was mapped. The *Nitraria tangutorum* *Bohr.* species distributed mainly in the eastern part of the study area. The spatial distributions of *Nitraria tangutorum* *Bohr.* simulated by the different models were significantly different. With regard to the most suitable area for *Nitraria tangutorum* *Bohr.* , the most suitable habitat area simulated by the Domain model was the largest, and the most suitable habitat area simulated by the BIOCLIM model was the smallest. The suitable habitat areas simulated by the GLM and GAM models were identical; the suitable habitat simulated by GLM and GAM was the largest, while that simulated by BIOCLIM was the smallest.

●The driven-factors of topography on the *Nitraria tangutorum* *Bohr.* distribution

The responses of the driven factor topography with different models were compared for *Nitraria tangutorum* *Bohr.* . Different models delineated the topography characteristics of the potential distribution areas. With BIOCLIM, the existence probability followed the trend of increase at first and then decreasing trend, and aspect of south part was on the top. For the DEM, the existence probability trend of the species was steeper, reaching a peak at an altitude of approximately 1010 m. For the profile factor, the species *Nitraria tangutorum* *Bohr.* appeared more suitable at the predicted value of zero, and was also suitable for the slope of 0°–20°. With the

Domain model, the existence probability trend of the species was similar to that of the BIOCLIM model, but seemed steeper. In the Malanhanobis model, at the aspect of north by east 10° , the species *Nitraria tangutorum* Bobr. exhibited minimum fitness, and at an altitude range of 1007 m to 1013 m, exhibited maximum fitness. Regarding the topography factor of the profile, there were two crests at the predicted value near zero. With the GLM and GAM models, the existence probability trend of the species was nearly the same as the linear trend line, but in the RF, the trend was nonlinear. Regarding the factor of aspect, the predicted value fluctuated between 0.4 and 0.8, but for the DEM factor, the predicted value was the highest at an altitude of approximately 1010 m. With the factor of profile and slope, at the degree of zero, the predicted value reached the top, at a slope of 20° to 60° , exhibiting a rising tendency.

Analysis of our maps showed that the distribution of the species *Nitraria tangutorum* Bobr. was more suitable in the south, with an altitude of approximately 1010 m, from a slope between 0° and 20° , and the general trend was the same with different models. The GLM or GAM model were found to be non-adequate for distribution modelling of *Nitraria tangutorum* Bobr. .

Discussion

●Importance of LiDAR data and its ecological implications with selected predictor variables

How does climate change affect organisms? In the past, Our limited ability to map and monitor the microclimatic variations using modelling through the incorporation of the microclimate. The microclimate sensors set in the networks provided point-based approaches and weather stations provided macroclimate data. However, means of interpolating and downscaling data were lacking ⁴². LiDAR is valuable for mapping microclimates, as it provides spatially continuous and sub-meter-scale information about ground topography ⁴³. Aerial photography provides an approach for assessing topography. However, these are less accurate than LiDAR at deriving terrain elevation ⁴⁴. An important advantage of UAVs is that they are flexible, enabling the collection of time series of aerial imagery over a period at a reasonable cost. In our study, LiDAR data were used to obtain DEM for SDMs, thereby supporting more effective conservation monitoring, management, and policy decision for a sustainable future ⁴⁵.

Our results showed that *Nitraria tangutorum* Bobr. species are specialists in the range of their habitat resources. In other words, *Nitraria tangutorum* Bobr. has a relatively narrow ecological niche.

It is limited to making the distribution by considering only topographic variables, the role of habitat heterogeneity in the desert ecosystem may be underestimated. If other environmental factors, such as light, soil, and water are considered, niche theory may be able to better explain the existence mechanism of species. Therefore, future studies need to consider more bio-variables to simulate the distribution of the species, and a permanent plot should be designed to contain the entire spatial gradient of the microclimate conditions in the study system. Long-term microclimate data series are required to complement the dynamics of the microclimate around the species. Gathering georeferenced microclimate data from different habitat types across the globe would significantly promote progress in microclimate ecology ⁴⁶. UAV equipped with multi-hyperspectral sensors and laser scanner systems are very promising, providing high-resolution data in the context of microhabitat, vegetation

structure, and topography ⁴⁷.

●Suggestion for improvement of SDMs

In our study, not only the AUC of the models but also the responses of the variables with the models were different. The parametric regression-based models (GLM and GAM) had lower mean performance, whereas nonparametric or machine learning-based models (RF and SVM) were more accurate. GLMs are widely used in SDMs studies ⁴⁸, but have poor nonlinear response performance. At the same time, the results confirmed the effectiveness and robustness of the machine learning techniques. Similarly, Marmion et al. found a single algorithm with lower accuracy than RF in predicting the plant distributions ⁴⁹. The parameters of the model affect its accuracy; for example, the performance of RF depends on two important parameters, *ntree* and *mtv*. As for the *ntree*, the optimal value using only the default, may not be the best; therefore, future research should test and adjust the model parameters to the optimum state. In this study, various parametric and nonparametric algorithms were evaluated for SDMs. In the next step, the semi-parametric models ⁵⁰ shall be used for better results.

A stable and reliable mathematical algorithm is required to construct more accurate SDMs. In previous research, for community-level predictions, the choice of how species and site-level occurrence probabilities are combined into species assemblage predictions is more important than the choice of the model type used ⁵¹. The following research shall use the ensemble models to predict the species distribution. As many researchers have expressed, no evidence has proven that one model is steadily accurate, but it is proven to be better than the other models.

The number of environmental factors that accurately represent the habitat characteristics of a species was sufficient SDM has some limitations, for example, over-prediction of species richness per site and sensitivity to methodological biases. In addition, sample size, sample prevalence, sample design, model techniques, imperfect detection of species, or the choice of environmental variables could affect the uncertainty of the predictions ⁵².

The relationships between species occurrence and sets of predictor variables explored by the models produce two kinds of useful outputs: estimates of the probability that species might occur at a given unrecorded location ⁴⁹, and estimates of an area's suitability for species ¹⁵. Climate determines the distribution of desert species. Previous research has shown that the roots of desert plant species develop and depend on the underwater ⁵³. As the study area was only one square kilometers, the variable of climate change could not be obvious, and species distribution could have been disturbed by the level of underground water. However, further studies are necessary in this regard. It is well known that biotic interactions play an important role in creating accurate SDMs for many species, particularly at small geographic scale, such as host requirements and competition ⁶. In this study, the model did not include the interaction between species and the influence of predators or human activities. More efforts should be dedicated to include fine-scale environmental measures, such as micro-temperature, underwater, moisture, and rainfall, for fine-scale species mapping and management.

●Management implication of SDMs

Species conservation and management are key to maintaining regional ecological balance, especially in ecologically fragile areas such as the Ulan Buh Desert. This study could distinguish the habitat requirements of *Nitraria tangutorum* *Bohr.*, especially for the topography, to provide a reference for protection. *Nitraria tangutorum* *Bohr.* was found to be distributed mainly in the eastern part of the study area. This finding is consistent

with the results of the actual survey. The suitable habitat area is larger than the real area, which indicates that the area of *Nitraria tangutorum* Bobr. is more widely distributed under current environmental conditions. Some of the suitable habitat characteristics shown in our study were consistent with those of other studies⁵⁴. For example, the population size of *Nitraria tangutorum* Bobr. increases with sand depth, making it suitable for sand dunes. To expand the area of *Nitraria tangutorum* Bobr., the land should be sand dune, and will be a good choice at an altitude of approximately 1010 m, better to face the aspect of south part, within the slope of 20°. Because of this characteristic, *Nitraria tangutorum* Bobr. can prevent quicksand from moving forward and land desertification. *Nitraria tangutorum* Bobr. is an excellent species for sand fixation.

Materials and Methods

•Study workflow

The study process shown in the flowchart (Fig. 2) consists of four parts: (1) preparing data, including species data and factors of topography; (2) fitting different models; (3) evaluating models by Area Under the Curve (AUC) and comparing the driving factors of different models; and (4) estimating areas suitable for habitat using the constructed SDMs.

Figure 2

•Study area and data analysis

Study area

The Ulan Buh Desert is located in the northern part of Inner Mongolia Province, 106°38'42"E—106°57'00"E, 40°17'24"N—40°28'36"N (Fig. 3), covering an area of 14905.13 km². The elevation ranges from 1009 to 1016 m, and the topographical profile on the surface varies from 1211.58 to 1339.63m⁵⁵. The climate is characterized by semi-arid to arid conditions, with a mean annual precipitation of 147.4 mm, mean annual evaporation of 2458.4 mm, mean annual relative humidity of 47.3%, and mean annual temperature of 6.8° C in the desert⁵⁶. The eastern edge of the desert is important for dividing the desert and grassland in Central Asia. Our study was based on control plots, such as desertification processes and control treatments⁵⁷.

Figure 3

Species data processing

Nitraria tangutorum Bobr. is a shrub species. It is 1–2 m tall, multi-branched, crooked, prone, and spreads with infertile needle-like apex branches and white tender branches. The leaves are clustered in 2–4 pieces on the young shoots, broadly lanceolate, 18–30 mm long, 6–8 mm wide, blunt apex, base narrowed into a wedge, and entire rare apex lobed. The flowers are densely arranged. The drupe is oval and dark red when ripe, juice rose colored, 8–12 mm long, and 6–9 mm in diameter. (<http://www.iplant.cn/>) (State Key Laboratory of Systems and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences).

(1) Species data collection

The ground surveys were conducted between August and September 2019. 100 sample plots of *Nitraria*

tangutorum Bobr. plant community were established in one square kilometers at the eastern edge of the study area. Each 100 × 100 m sample plot was divided into 100 small quadrats with an area of 10 × 10 m (Fig. 3). 335 occurrence records for *Nitraria tangutorum* Bobr. were collected in the area. The formal identification of the plant material used in my study have complied with Chinese Virtual Herbarium (CVH): (<https://www.cvh.ac.cn/spms/detail.php?id=f76aceff>), and a specimen of this material has been deposited in a publicly available herbarium. And we were employed by Chinese Academy of Forestry, which is also the State Forestry Administration Dengkou Desert ecosystem positioning observation station. To ensure that we have permission to collect *Nitraria tangutorum* Bobr. We confirm that all methods were performed in accordance with current guidelines and regulations in China.

(2) Processing species data

The data were cleaned for use in our modelling. The longitudes and latitudes of the sample plot center were recorded as coordinates, and if the record was out of the boundary of the study area, it was checked with the UAV-RGB (Unmanned Aerial Vehicle - Red Green Blue) data and overlaid on the map. If the canopy was outside the plot boundary, the root inside was preserved.

The coordinates were crosschecked by a visual or ‘overlay’ function using ARCGIS 10.5. This study applied the coordinate function from the ‘SP’ (spatial point) package to create a Spatial Point Data Frame, and then the over function from ‘SP’ to do a point-in-polygon query with the study boundary. Moreover, the occurrence of each grid would be the only once, without duplicate records; the function ‘duplicate’ could be used to remove duplicates in the software R3.4.3.

The species data included the presence and absence data, the resolution of the grid was 10 m, and each grid retained only presence or absence data, reducing the bias of the model predictions by space sampling⁵⁸.

Environment data processing

- LiDAR data

The UAV technique with the SPAN_IGM_STIM300 IMU (Fig. 4 (a)) was used to gather airborne LiDAR data during the study conducted on October 24, 2019. The LiDAR parameters were a pulse rate of 125 kHz, a scan angle (FOV(field of view)) of ±22.5°, and a laser beam divergence (IFOV) of 0.5 mrad. An average point density of 41.62 points per m² and footprint diameter of 25 cm were obtained.

Figure 4

- Topographic variables

The digital elevation model (DEM) was derived from the UAV LiDAR point cloud (Fig. 4(b)) with a resolution of 15 cm (Fig. 5(a)) and then resampled to 10 m (Fig. 5(b)) with image analysis tools. Other topographic variables, including slope, aspect, and profile (Fig. 6), were obtained using the spatial analysis model in ArcGIS 10.5. Predictor variables were organized as raster (grid) files for species distribution modelling. After generating the set of predictor variables (rasters) and the occurrence points, our next step was to extract the values of the predictors at the locations of the points using the raster tools of ‘extract’ in ArcGIS 10.5. To select the significant predictor variables, the variance inflation factor (VIF) was calculated. A VIF of 10 was acceptable as it showed

no significant collinearity among the predictor variables ⁵⁹. All variables were resampled using ArcGIS 10.5 to unify the space dimension of 10 m, which was the same value as the resolution of the species samples.

Figure 5

Figure 6

●Forms for SDMs

To predict species distribution, numbers of statistical and machine-learning approaches have been used. In conjunction with the geographic information systems and remote sensing, the methods used in species distribution modelling can be classified as profile, regression, and machine learning methods. The profile method only considers the presence (occurrence) data, whereas the regression and machine learning methods use both the presence and absence data. The profile method includes BIOCLIM ⁶⁰, Domain ⁶¹, and Mahalanobis ⁶². In addition to these methods, we also analyzed and compared the predictive performance of SDMs developed using regression methods, such as the GLM ⁶³ and GAM ^{64,65}, and some machine learning methods, such as RF ⁶⁶ and SVM.

Sampling bias may have been present in the occurrence data used ⁶⁷. Random sampling was performed to record the presence or absence data with the random points function in ‘dismo’ package. An ‘extent’ tool was used to further restrict the area from where random locations were drawn ^{7,68}. The ‘dismo’ package containing the k-fold function facilitates the data partitioning and creates a vector that assigns each row in the data matrix to a group (between 1 and k with k=5).

We used seven modelling algorithms, including both nonparametric and parametric algorithms, to construct the SDMs, seven modelling algorithms are briefly described below.

Bioclimatic modelling (BIOCLIM)

BIOCLIM is a classic climate envelope model ⁶⁰. Although it generally does not perform as well as some other modelling methods ⁶⁹, particularly in the context of climate change ⁶⁷, it is still being used, because the algorithm is easy to understand.

Domain

The Domain algorithm ⁶¹ has been extensively used for species distribution modelling. The Domain algorithm computes Gower’s metric ⁷⁰ between the environmental variables at any location and those at any of the known locations of occurrence, which were defined as training sites in this study.

A suitable means of quantifying the similarity between the two sites is provided by Gower’s metric. In Euclidean Q-dimensional space, the distance d between two points A and B is defined as

$$d_{AB} = \frac{1}{Q} \sum_{K=1}^Q \left(\frac{A_K - B_K}{range_K} \right) \quad (1)$$

To equalize the contribution from each climatic attribute, Gower’s metric uses range standardization. In this application, the standardization method is preferred over variance standardization because it is less susceptible to bias arising from dense clusters of sample points. The complementary similarity measure (R_{AB}) is defined as

$$R_{AB} = 1 - d_{AB} \quad (2)$$

R_{AB} is constrained between 0 and 1 for points within the ranges used in Equation 1 but may yield negative values for points outside this range. S_A is defined as the maximum similarity between candidate point A and the combination of known record site Tm:

$$S_A = \max_{j=1}^m R_{TjA} \quad (3)$$

S_A is evaluated for all grid points in a target area, where m denotes the number of known records, and a matrix of continuously varying similarity values is generated, which can be displayed as a grayscale, thematic, or contour map. As with all models discussed here, the values generated are not probability estimates but the degrees of classification confidence ⁶¹.

Mahalanobis

The Mahal function implements a SDMs based on Mahalanobis distance ⁶². The mahalanobis distance considers the correlations of the variables in the dataset, and it is not dependent on the scale of measurements.

$$PD_1^2 = \alpha_{\mu\nu} (da)^\mu (da)^\nu \quad (4)$$

where $\alpha_{\mu\nu}$ is the dispersion, $(\alpha)^\mu$ is the mean value, P is the independent variable, and D^2 is the sample value of the Δ^2 statistic. Both the dispersion and mean values are subject to sampling fluctuations.

Generalized linear models (GLM)

In GLM, $X_k (k = 1, \dots, r)$ is the vector of k predictor variables, which is related to the expected value $\mu = E(Y)$ of the response variable Y (denote the response variable) through a link function $g(\cdot)$ (Eq. 5). The use of GLM in species distribution modelling is provided in ⁶³.

$$g(E(Y)) = \alpha + X^T \beta \quad (5)$$

$$g(\mu_i) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \quad (6)$$

where α is a constant called the intercept, and β is the vector of k regression coefficients (one for each predictor). (Eq. (6)).

It is assumed that the topographic factor is normally distributed and is defined as X , and the species distribution probability as Y ; Therefore in the response distribution of GLM, which uses the "family=gaussian".

Generalized additive models (GAM)

GAMs ^{64, 65}) are extensions of GLMs. In GAMs, the linear predictor is the sum of the smoothing functions. This makes GAMs very flexible and they can fit complex functions. This also makes them very similar to machine-learning methods. The GAMs were implemented using the 'mgcv' package R 3.4.3.

$$\begin{aligned} g(\mu) &= s_0 + s_1(X_1) + s_2(X_2) + \dots + s_P(X_P) \\ n &= s_0 + \sum_{i=1}^P s_i(X_i) \end{aligned} \quad (7)$$

where $\mu = E(Y|X_1, X_2, \dots, X_p)$, n is for the linear prediction, and $s_i(\cdot)$ is the nonparametric smooth function. This model does not require any assumptions. It consists of a random component, additive component, and link function connecting the two parts. The distribution of the response variables (Y) belonging to the family of exponential distributions can be binomial.

Random forest (RF)

The RF method is an extension of the regression and classification trees (CART) ⁶⁶. This method can be implemented using the “random forest” function in the package with the same package name in R 3.4.3. The function RF can take a formula, or in two separate arguments, a data frame with the predictor variables and a vector with the response. If the response variable is a factor (categorical) type, RF performs classification; otherwise, it performs regression. In species distribution modelling, because RF showed remarkable modelling results, it is considered a viable option.

(1) Suppose there is a dataset $D = \{x_{i1}, x_{i2}, \dots, x_{in}, y_i\} (i \in [1, m])$ (x_{im} is the vector of m predictor variables, y_i are the response variables) with feature number n , sampling with replacement can generate a sampling space $(m * n)^{m*n}$.

(2) Build a basic learner (a decision tree): $d_i = \{x_{i1}, x_{i2}, \dots, x_{in}, y_i\} (i \in [1, m])$ generates a decision tree for each sample (where $K \ll M$) and records the results of each decision tree $h_j(x)$.

(3) Training times to T $H(x) = \max \sum_{t=1}^T \phi(h_t(x) = y)$, where $\phi(x)$ is an algorithm (absolute majority voting, majority voting, weighted voting method, etc.).

RF requires two parameters: 1) `mtry`, the number of predictor variables performing data partitioning at each node, and 2) `ntree`, the total number of trees to be grown in the model run. In this study, `ntree` (number of trees to grow) was set to 500 and `mtry` to 10 after some initial tuning experiments.

Support vector machines (SVM)

SVM (Vapnik, 1998) applies a simple linear method to data; however, in a high-dimensional feature space, nonlinearly exists in the input space. However, in practice, this does not involve any computations in a high-dimensional space. This simplicity, combined with state-of-the-art performance on many learning problems (classification, regression, and novelty detection), has contributed to the popularity of SVM ⁷¹. They were first used in species distribution modelling by Guo et al. ⁷². There are several implementations of SVM in R 3.4.3. The most useful implementations in our context are the function 'ksvm' in the package 'kernlab' and the 'svm' function in the package 'e1071'. The 'ksvm' includes many different SVM formulations and kernels, provides useful options, and features a method for plotting, but it lacks a proper model selection tool. The 'svm' function in package 'e1071' includes a model selection tool: the 'tune' function ⁷¹.

$$\hat{y} = \begin{cases} +1, \sum_{i=1}^m \lambda_i^* y_i K(x, x_i) \geq +1 \\ -1, \sum_{i=1}^m \lambda_i^* y_i K(x, x_i) \leq -1 \end{cases} \quad (8)$$

Given the training examples, $x_i (i = 1, 2, \dots, m)$ indicates the samples to be classified, $\hat{y} \in \{+1, -1\}$ indicates the marked label value for the samples, $+1$ denotes the point above the separating hyperplane, and -1 denotes the point under the separating hyperplane, where $K(x, x_i)$ is the Gaussian kernel function, $K(x, x_i)$ is nonlinearity, and λ_i^* is the optimal solution to the primal problem obtained by solving the dual problem. Note that if a

sample x_i is not a support vector, to maximize the Lagrangian function, there must be a $\lambda_i=0$ non-support vector corresponding to λ_i equal to zero, which theoretically shows that the decision boundary of the SVM is only related to the support vector.

●Evaluation and comparison of SDMs

●Accuracy assessment

Different measures may be used to evaluate the prediction ability of each model ⁷³⁻⁷⁵. Many measures for evaluating models based on presence-absence or presence-only data are threshold dependent. This implies that a threshold must first be set. However, many statistics are threshold independent, such as the correlation coefficients and the area under the receiver operator curve (AUROC, generally abbreviated as AUC). The AUC value is the area of ROC and the X-axis under the ROC curve; the larger the area, the higher the relationship, and the greater the creditability. When the AUC value is greater than 0.7, the result is fair, and when the AUC is higher than 0.9, the result is excellent ⁷⁶. We applied this measure to evaluate the SDMs.

In this framework, an algorithm is supposed to predict “positive” or “negative”. However, some concepts are confusing, as summarized in Table 2.

Table 2

Suitable habitat index (SHI)

The probability of species occurrence was extracted using ArcGIS. The lowest probability value was selected to distinguish the suitable zone from the non-suitable zone ⁷⁷. The habitats of the entire study area were divided into three categories using the frequency statistical method ⁷⁸: unsuitable habitat ($SHI \leq 0.3$), low suitable habitat ($0.3 < SHI \leq 0.5$), and suitable habitat ($SHI > 0.5$).

Conclusion

This is the first attempt of modelling the relationships between the spatial distribution of desert plant species and topographical factors extracted from the LiDAR data. The results indicated a general relationship between model performance and modelled species distribution on a small scale. Topographical factors should be contingent on the modelled species distribution. The LiDAR technique has certain applications in species distribution simulations. The RF produced the highest AUC when AUC was integrated into TNR + TPR (false positive rate plus true positive rate); Manhattanobis showed the best SDMs to predict the distribution of *Nitraria tangutorum* *Bohr.* in the desert on a small scale. The suitable habitat distribution area of *Nitraria tangutorum* *Bohr.* predicted in the study was larger than the actual area. A flat land with an aspect of south at an altitude near 1010 m was found to be the driving topography of the *Nitraria tangutorum* *Bohr.* distribution. Thus, this study provides a reliable reference for restoration and management of desert vegetation.

Author Contributions: H.Z., X.Z., and L.F. collected data, [undertook the formal identification of the plant material](#); H.Z. and X.Z. analyzed the data; H.Z., X.Z., L.F., and R.P.S. wrote the manuscript, contributed critically to improving the manuscript, and gave final approval for publication.

All the authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the 14th Five-Year Plan Pioneering Project of High Technology Plan of the National Department of Technology under Grant (Grant No. 2021YFD2200405), and National Natural Science Foundation of China (Grant Nos. 31470641, 31300534, and 31570628) .

Data Availability Statement: Data used in this study are available from the National Forestry and Grassland Science Data Center (<http://www.forestdata.cn/>). Data is available when needed.

Acknowledgments: We thank the 14th Five-Year Plan Pioneering Project of High Technology Plan of the National Department of Technology under Grant(Grant No. 2021YFD2200405), and the National Natural Science Foundation of China (Grant Nos. 31470641, 31300534, and 31570628) for financial support.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- 1.Zhang, X., Y. Liu, B. Xu, Z. Jiang, X. Wang, et al., *Discussion on the Method of High-resolution Remote Sensor Data Interpretation of Haloxylon Ammodendron in Ulan Buh Desert*. Journa. **25(6)**: 828-835 (2010)
- 2.Zhang, W., T. Lu; & K. Ma;, *Analysis of environmental and spatial factors of plant community distribution in arid valley of the upper Minjiang River*. Journa. **3**: 552-558 (2004)
- 3.Guisan, A., R. Tingley, J.B. Baumgartner, I. Naujokaitis-Lewis, P.R. Sutcliffe, et al., *Predicting species distributions for conservation decisions*. Journa. **16**: 1424-1435 (2013)
- 4.Walther, G.R., A. Roques, P.E. Hulme, M.T. Sykes, P. Pysek, et al., *Alien species in a warmer world: risks and opportunities*. Journa. **24(12)**: 86-93 (2009)
- 5.Willis, K.J. & S.A. Bhagwat, *Biodiversity and climate change*. Journa. **326(5954)**: 806-807 (2009)
- 6.Hao, T., G. Guillerá-Arroita, T.W. May, J.J. Lahoz-Monfort & J. Elith, *Using Species Distribution Models For Fungi*. (2020)
- 7.Elith, J. & J.R. Leathwick, *Species Distribution Models: Ecological Explanation and Prediction Across Space and Time*. 677-697 (2009)
- 8.Safaei, M., H. Rezayan, P. ZeaieanFirozabadi & J. Sadidi, *Optimization of species distribution models using a genetic algorithm for simulating climate change effects on Zagros forests in Iran*. Journa. **63**: 101-288 (2021)
- 9.Benkendorf, D.J. & C.P. Hawkins, *Effects of sample size and network depth on a deep learning approach to species distribution modeling*. Journa. **60**: 101-137 (2020)
- 10.Kaky, E., V. Nolan, A. Alatawi & F. Gilbert, *A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants*. Journa. **60**: 101-150 (2020)
- 11.Walker, P.O., *Modelling wildlife distribution using a geographic information system: kangaroos in relation to climate*. Journa. **17**: 279-289 (1990)
- 12.Bio, A.M.F., R. Alkemande & A. Barendregt, *Determining alternative models for vegetation response analysis – a non-parametric approach*. Journa. **9**: 5-16 (1998)
- 13.Huiller, W.T., *BIOMOD – Optimising predictions of species distributions and projecting potential future shifts under global change*. Journa. **9**: 1353-1362 (2003)

-
14. Pereira, J.M.C. & R.M. Itami, *GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham Red Squirrel*. *Journa*. **57**: 1474-1486 (1991)
 15. Pedro, S. & M.B. Araujo, *An evaluation of methods for modelling species distributions*. *Journa*. **31**: 1555–1568 (2004)
 16. Lopatin, J., K. Dolos, H.J. Hernández, M. Galleguillos & F.E.t.a. Fassnacht, *Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile*. *Journa*. **173**: 200-210 (2016)
 17. Randin, C.F., M.B. Ashcroft, J. Bolliger, J. Cavender-Bares, N. Coops, C. , et al., *Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models*. *Journa*. **239**: 111626 (2020)
 18. Milanesi, P., F.D. Rocca & R.A. Robinson, *Integrating dynamic environmental predictors and species occurrences: Toward true dynamic species distribution models*. *Journa*. **10**(2) (2020)
 19. Isaac, N.J.B., M.A. Jarzyna, P. Keil, L.I. Dambly, P.H. Boersch-Supan, et al., *Data Integration for Large-Scale Models of Species Distributions*. *Journa*. **35**(1): 56-67 (2020)
 20. Pradervand, J.-N., D. Anne, P. Loïc, G. Antoine & R. Christophe, *Very high resolution environmental predictors in species distribution models: Moving beyond topography?* *Journa*. **38**(1): 79-96 (2014)
 21. Karger, D.N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, et al., *Climatologies at high resolution for the earth's land surface areas*. *Journa*. **4**(1) (2017)
 22. Dixon, P.P.M., *Small - scale environmental heterogeneity and the analysis of species distributions along gradients*. *Journa*. **1**(1): 57-65 (2010)
 23. Mujuru, L. & A. Kundhlande, *Small - scale vegetation structure and composition of Chirinda Forest, southeast Zimbabwe*. *Journa*. **45** (2010)
 24. Punchi-Manage, R., S. Getzin, T. Wiegand, R. Kanagaraj, C.V.S. Gunatilleke, et al., *Effects of topography on structuring local species assemblages in a Sri Lankan mixed dipterocarp forest*. (2012)
 25. Swanson, F.J., T.K. Kratz, N. Caine & R.G. Woodmansee, *Landform effects on ecosystem patterns and processes*. *Journa*. **38**(2): 92-98 (1988)
 26. Noy-Meir, I., *Desert Ecosystems: Environment and Producers*. *Journa*. **4**(1): 25-51 (1973)
 27. Zhang, X., *A dissertation submitted to University of Chinese Academy of Sciences in partial fulfillment of the requirement C.A.o.S.M.o.E. Research Center for Eco-environments and Soils and Water Conversation, Editor: Chinese Academy of Science & Ministry of Education* (2018)
 28. Runquist, R.D.B., T. Lake, P. Tiffin & A.M. David, *Species distribution models throughout the invasion history of Palmer amaranth predict regions at risk of future invasion and reveal challenges with modeling rapidly shifting geographic ranges*. (2019)
 29. Dempsey, S.J., E.M. Gese, B.M. Kluever, R.C. Lonsinger & L.P. Waits, *Evaluation of Scat Deposition Transects versus Radio Telemetry for Developing a Species Distribution Model for a Rare Desert Carnivore, the Kit Fox*. *Journa*. **10**(10): 17 (2015)
 30. Nussear, K.E., T.C. Esque, R.D. Inman, L. Gass, K.A. Thomas, et al., *Modeling habitat of the desert tortoise (*Gopherus agassizii*) in the Mojave and parts of the Sonoran Deserts of California, Nevada, Utah, and Arizona*. *Journa*. **1102**: 18 (2009)
 31. Saatchi, S., W. Buermann, H.T. Steege, S. Mori & T.B. Smith, *Modeling distribution of Amazonian tree species and diversity using remote sensing measurements*. *Journa*. **112**(5): 2000-2017 (2008)
 32. Normand, S., U.A. Treier, C. Randin, P. Vittoz, A. Guisan, et al., *Importance of abiotic stress as a range-limit determinant for European plants: insights from species responses to climatic gradients*. *Journa*. **18**(4): 437-449 (2009)

-
33. Lembrechts, J.J., J. Lenoir, N. Roth, T. Hattab, A. Milbau, et al., *Comparing temperature data sources for use in species distribution models: From in-situ logging to remote sensing*. *Journa*. **28**(11): 1578-1596 (2019)
34. Austin, M.P. & Kimberly P. Van Niel, *Improving species distribution models for climate change studies: variable selection and scale*. *Journa*. **38**: 1-8 (2011)
35. Zellweger, F., P. De Frenne, J. Lenoir, D. Rocchini & D. Coomes, *Advances in Microclimate Ecology Arising from Remote Sensing*. *Journa*. **34**(4): 327-341 (2019)
36. Park, J., B. Choi & J. Lee, *Spatial Distribution Characteristics of Species Diversity Using Geographically Weighted Regression Model*. *Journa*. **31**(10): 3197-3213 (2019)
37. Leempoel, K., C. Parisod, G. Céline, D. Lucas, P. Vittoz, et al., *Data from: Very high resolution digital elevation models: are multi-scale derived variables ecologically relevant?* *Journa*. **6** (2016)
38. Mathys, L., N.E. Zimmermann & A. Guisan, *Spatial pattern of forest resources in a multifunctional landscape*. *Journa*. **36**: 340-342 (2004)
39. Hagar, J.C., A. Yost & P.K. Haggerty, *Incorporating LiDAR metrics into a structure-based habitat model for a canopy-dwelling species*. *Journa*. **236**: 11-49 (2020)
40. Wüest, R.O., A. Bergamini, K. Bollmann & A. Baltensweiler, *LiDAR data as a proxy for light availability improve distribution modelling of woody species*. *Journa*. **456**: 117-644 (2020)
41. Zhou, Z., R. Liu, S. Shi, Y. Su, W. Li, et al., *Ecological niche modeling with LiDAR data: A case study of modeling the distribution of fisher in the southern Sierra Nevada Mountains, California*. *Journa*. **26**(08): 878-891 (2018)
42. Zellweger, F., P. De Frenne, J. Lenoir, D. Rocchini & D. Coomes, *Advances in Microclimate Ecology Arising from Remote Sensing*. *Journa*. **34**(4): 327-341 (2019)
43. Lefsky, M.A., Cohen W.B. & P. G.G., *Lidar remote sensing for ecosystem studies*. *Journa*. **52**: 19-30 (2002)
44. Westoby, M.J., J. Brasington, N.F. Glasser, M.J. Hambrey & J.M. Reynolds, *Structure-from-motion photogrammetry: a low-cost, effective tool for geoscience applications*. *Journa*. **179**: 300-314 (2012)
45. Randin, C.F., M.B. Ashcroft, J. Bolliger, J. Cavender-Bares, N.C. Coops, et al., *Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models*. *Journa*. **239**: 11-62 (2020)
46. Lenoir, J., Hattab T. & P.G. . *Climatic microrefugia under anthropogenic climate change: implications for species re-distribution*. *Journa*. **40** (2017)
47. Manfreda, S., M.F. McCabe, P.E. Miller, R. Lucas, V.P. Madrigal, et al., *On the Use of Unmanned Aerial Systems for Environmental Monitoring*. *Journa*. **10**: 641 (2018)
48. Elith*, J., C.H. Graham*, R.P. Anderson, M. Dudík, S. Ferrier, et al., *Novel methods improve prediction of species' distributions from occurrence data*. *Journa*. **29**(2): 129-151 (2010)
49. Marmion, M., M. Parviainen, M. Luoto, R.K. Heikkinen & W. Thuiller, *Evaluation of consensus methods in predictive species distribution modelling*. *Journa*. **15**(1): 59-69 (2010)
50. Bonneau, M., M. Delecroix & E. Malin, *Semiparametric Versus Nonparametric Estimation in Single Index Models, a Computational Approach*. *Journa*. **8**(3): 207-222 (1993)
51. Zurell, D., N.E. Zimmermann, H. Gross, A. Baltensweiler, T. Sattler, et al., *Testing species assemblage predictions from stacked and joint species distribution models*. *Journa*. **47**(1): 101-113 (2019)
52. Fernandes, R.F., S. Daniel & A. Guisan, *How much should one sample to accurately predict the distribution of species assemblages? A virtual community approach*. *Journa*. **48**: 125-134 (2018)
53. Ma, S., X. Li, C. Luo & F. Sun, *The impact of climate change of Haloxylon ammodendron plants suitable for distribution*. *Journa*. **36**(05): 1243-1250 (2017)
54. Qiu-Yan, L.I., W.Z. Zhao & H.Y. Fang, *Adaptation of Nitraria sphaerocarpa to wind-blown sand environments at the edge of a desert oasis*. *Journa*. **19**(4): 482-487 (2007)

-
55. Tang, G.-a., X. Liu & G. Yan, *The principle and method of geological analysis and Digital elevation model*. 2005: Science press.
 56. Li, Z. & F. Li, *Interspecific associations of natural nitaria tangtorum population and its main companion species in ulanbuh desert*. *Journa*. **28**: 99 (2008)
 57. Luo, F., Gao, J., Xin, Z., Hao, Y., Wang, L., & Li, S., *Characteristics of Soil Temperature Variation and Influence Factors at Northeastern Margin Region of Ulan Buh Desert*. *Journa*. **39**(01): 179-186 (2019)
 58. Kramerschadt, S., J. Niedballa, J.D. Pilgrim, B. Schroder, J. Lindenborn, et al., *The importance of correcting for sampling bias in MaxEnt species distribution models*. *Journa*. **19**(11): 1366-1379 (2013)
 59. Zheng, J., *Statistical big dictionary*. Vol. 1. 1995, China statistical press.
 60. Booth, T.H., H.A. Nix, J.R. Busby & M.F. Hutchinson, *BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies*. *Journa*. **20**(1): 1-9 (2014)
 61. Carpenter, G., A.N. Gillison & J. Winter, *Domain: a flexible modelling procedure for mapping potential distributions of plants and animals*. *Journa*. **2**: 667-680 (1993)
 62. Mahalanobis, P.C., *On the generalised distance in statistics*. *Journa*. **2**: 49-55 (1936)
 63. Guisan, A., T.C. Edwards Jr & T. Hastie, *Generalized linear and generalized additive models in studies of species distributions: setting the scene*. *Journa*. **157**(2-3): 89-100 (2002)
 64. Hastie, T. & T. Robert, *Generalized additive models*, in *Statistical models in S* Routledge. 249-307 (2017)
 65. Wood, S.N., *Generalized additive models: an introduction with R*. 2006: CRC press.
 66. Breiman, L., J. Friedman, R.A. Olshen & J.S. Charles, *Classification and regression trees*. 1984: CRC press.
 67. Hijmans, R., J. & C. Graham, H., *Testing the ability of climate envelope models to predict the effect of climate change on species distributions*. *Journa*. **12**: 2272-2281 (2006)
 68. Elith, J., S.J. Phillips, T. Hastie, M. Dudík, Y.E. Chee, et al., *A statistical explanation of MaxEnt for ecologists*. *Journa*. **17**(1): 43-57 (2011)
 69. Elith, J., C.H. Graham, R.P. Anderson, S. Ferrier, A. Guisan, et al. *Novel methods improve prediction of species' distributions from occurrence data*. 2010. *Ecography*
 70. Gower, J.C.t.a., *A general coefficient of similarity and some of its properties*. *Journa*. **27**: 57-71 (1971)
 71. Karatzoglou, A., D. Meyer & K. Hornik, *Support vector machines in R*. *Journa*. **15**(9): 1-28 (2006)
 72. Guo, Q., M. Kelly, Maggi, Graham & C. H., *Support vector machines for predicting distribution of Sudden Oak Death in California*. *Journa*. **182**(1): 75-90 (2005)
 73. Friedman, J.H., *Greedy function approximation: a gradient boosting machine*. 1189-1232 (2001)
 74. Liu, C., M. White & G. Newell, *Measuring and comparing the accuracy of species distribution models with presence-absence data*. *Journa*. **34**: 232-243 (2011)
 75. Potts, J.M. & J. Elith, *Comparing species abundance models*. *Journa*. **199**: 153-163 (2006)
 76. Guo, C., X. Li, Z. Zhao & Z. Nawaz, *Predicting the impacts of climate change, soils and vegetation types on the geographic distribution of Polyporus umbellatus in China*. *Journa*. **648**: 1-11 (2019)
 77. Pearson, R.G., C.J. Raxworthy, M. Nakamura & A.T. Peterson, *ORIGINAL ARTICLE: Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar*. *Journa*. **34**(1) (2007)
 78. Chen, H., L. Chen & T.P. Albright, *Predicting the potential distribution of invasive exotic species using GIS and information-theoretic approaches: A case of ragweed (Ambrosia artemisiifolia L.) distribution in China*. *Journa*. **52**(9): 1223-1230 (2007)

Figures

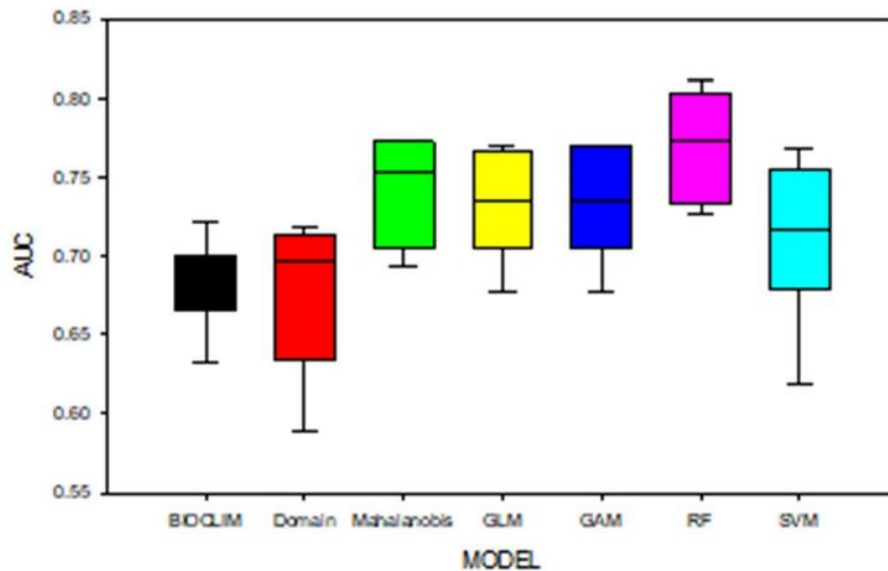


Figure 1. The boxplots of different models (BIOCLIM: classic climate envelope model; Domain computes the Gower distance between environmental variables and locations of occurrence; Mahalanobis model based on the mahalanobis distance; GLM: generalized linear model; GAM: generalized additive models; RF: random forest; SVM: support vector machines)

Figure 1

See image above for figure legend.

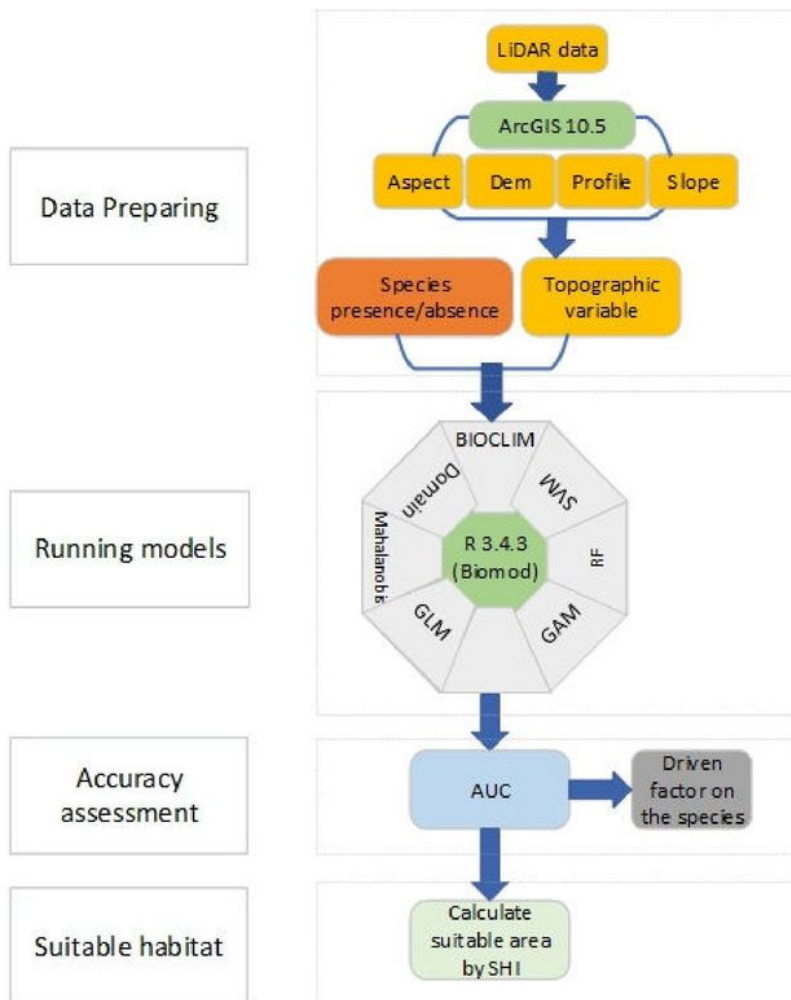


Figure 2. Workflow of the study (ARCGIS 10.5 is the software of geographic information system, R3.4.3 is the data processing platform, AUC is the area under the receiver operator curve, which evaluates the fitting of the models. Species distribution models include Bioclimatic Modelling (BIOCLIM), Domain, Mahalanobis, Generalized linear models (GLM), Generalized additive models (GAM), Random forest (RF), Support vector machine (SVM). SHI denotes suitable area index)

Figure 2

See image above for figure legend.

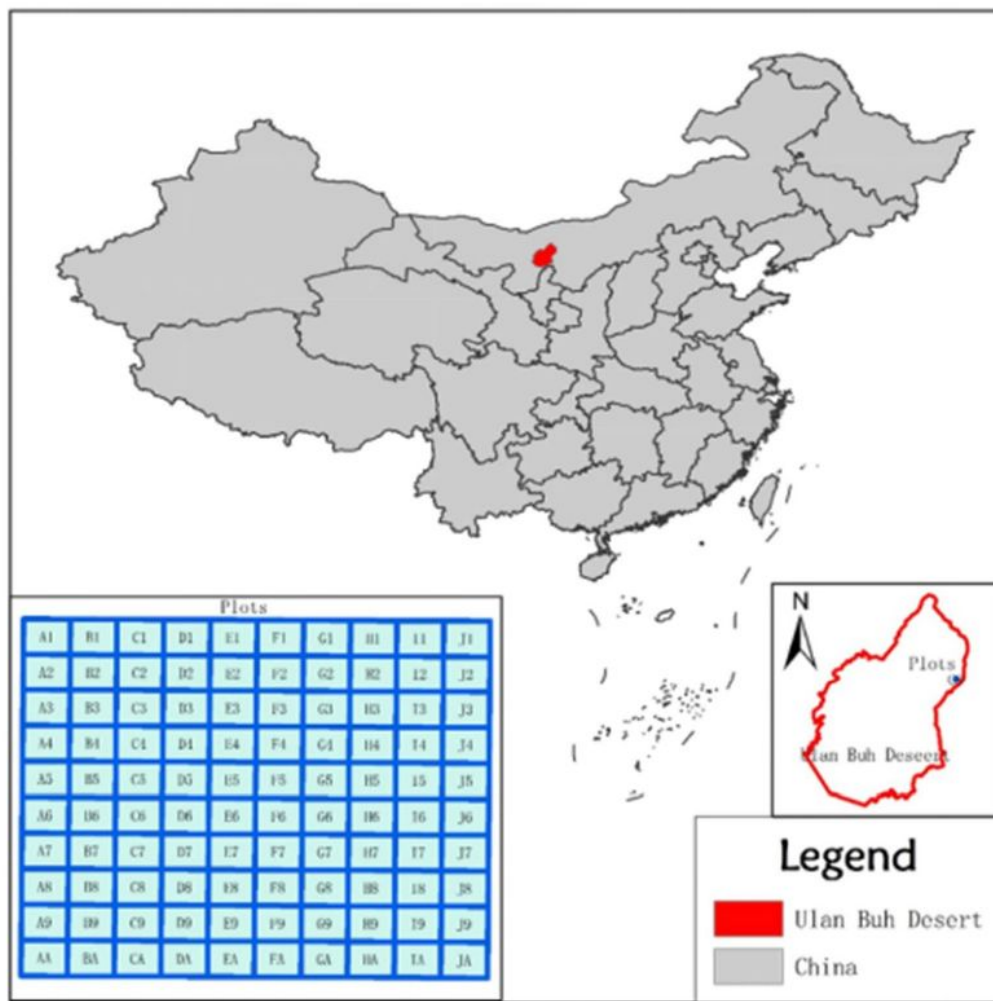


Figure 3. Location of study site in the Ulan Buh desert (the top of the figure is the northwest arid area in China; the bottom left is the study area located in Ulan Buh desert; and the bottom right is the plot of Ulan Buh desert. The number of rows are 0,1,2,3,...and the columns are A,B,C,...in the proper sequence)

Figure 3

See image above for figure legend.

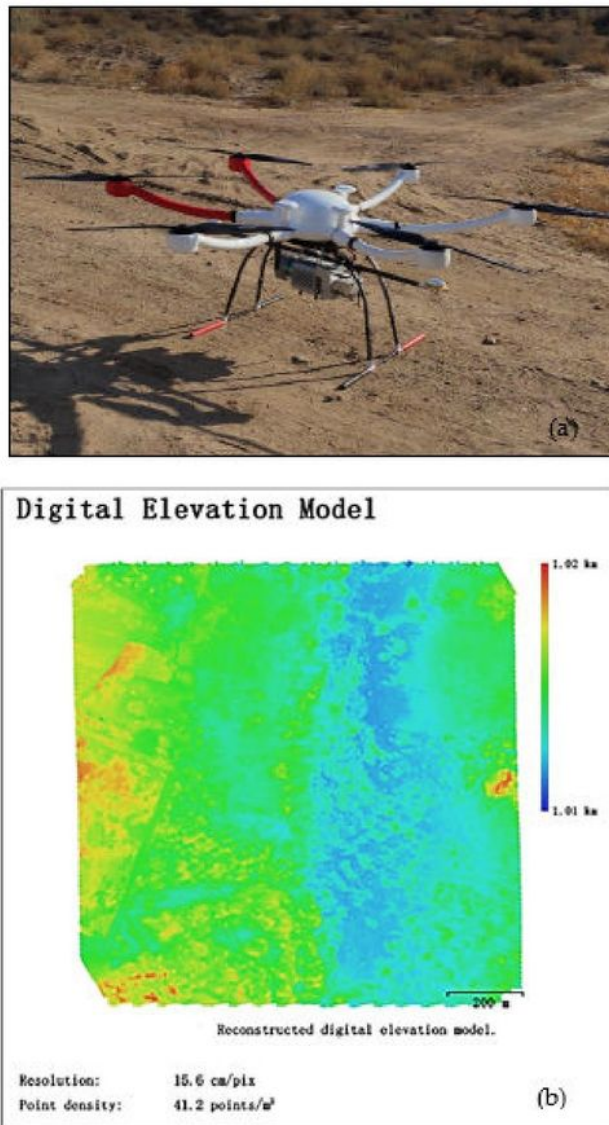


Figure 4. The Unmanned Aerial Vehicle (UAV) Light Detection And Ranging (LiDAR); (a) is the UAV LiDAR; (b) is the digital elevation model (DEM), and the legend describes the elevation of the point

Figure 4

See image above for figure legend.

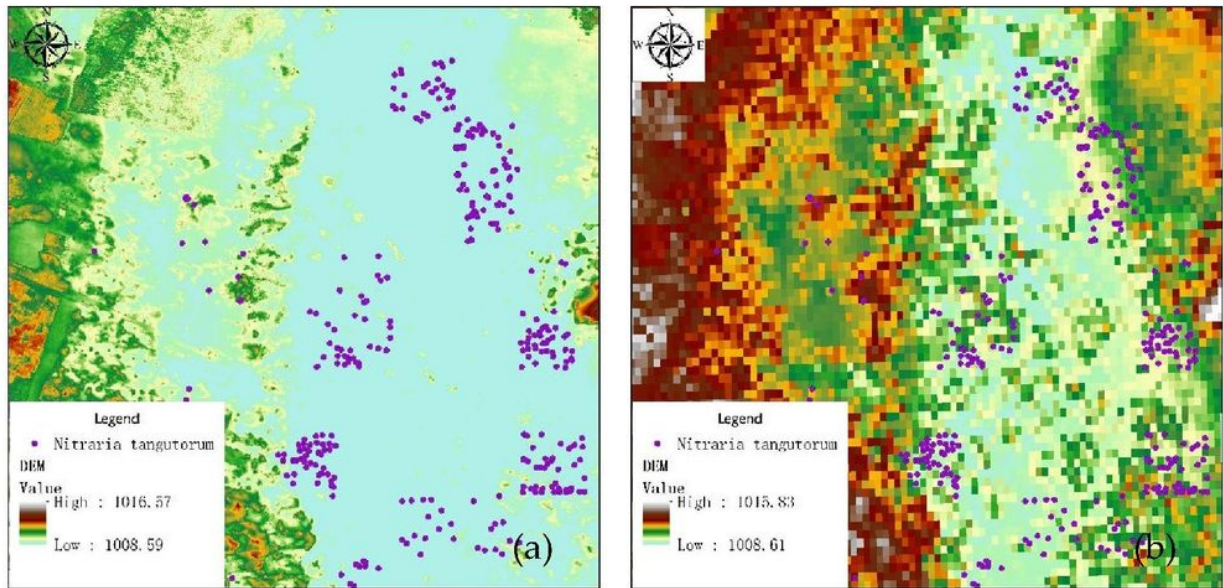


Figure 5. The original and rescaled LiDAR data with different resolution, (a) is LiDAR DEM with a resolution of 0.15 m, (b) is LiDAR DEM with a resolution of 10 m, and the purple points indicate the occurrence of *Nitraria tangutorum* Bobr.

Figure 5

See image above for figure legend.

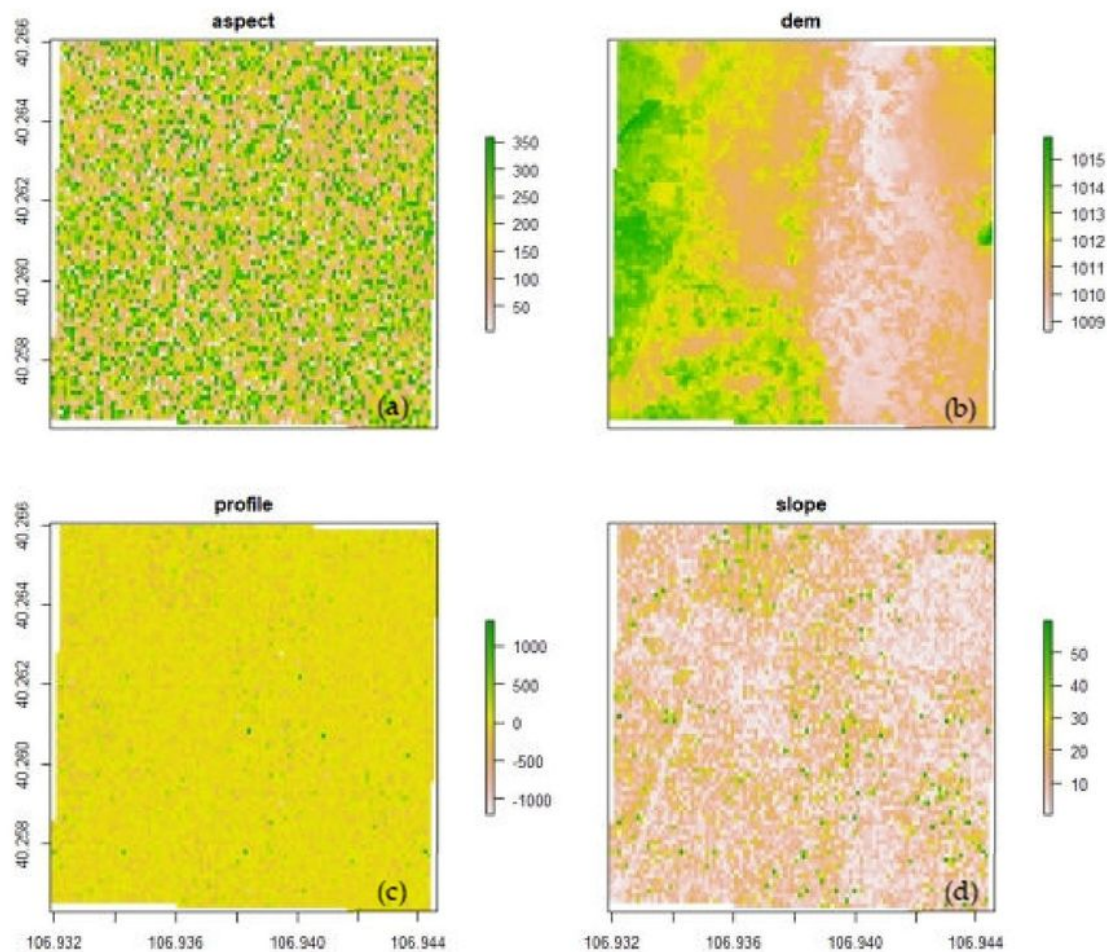


Figure 6. Maps of the predictor variables. The horizontal and vertical coordinates represent latitude and longitude, respectively. The figure 5 (a) indicates different aspect of the study area, figure 5 (b) indicates different elevation of the study area, figure 5 (c) indicates different profile of the study area, and figure 5 (d) indicates different slope of the study area

Figure 6

See image above for figure legend.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.pdf](#)