

# Land Suitability Assessment Dataset of Six Economically Important Desert Plants in Xinjiang Uygur Autonomous Region

Zhang, P.<sup>1,2</sup> Wang, X. P.<sup>3</sup> Fan, J. L.<sup>1,2</sup> Zhang, G. F.<sup>4</sup> Zhang, H.<sup>1,2\*</sup> Yu, X. X.<sup>1,2</sup>

1. National Engineering Technology Research Center for Desert-Oasis Ecological Construction, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China;  
2. Taklimakan Desert Ecosystem Field Observation and Research Station of Xinjiang, Bayingolin Mongolian Autonomous Prefecture 841900, China; 3. Zhongtan Energy (Shandong) Co., Ltd., Weifang 262700, China;  
4. Central Taklimakan Desert Meteorological Station, Bayingolin Mongolian Autonomous Prefecture 841000, China

**Abstract:** 6 plant species (*Isatis tinctoria* L., *Glycyrrhiza uralensis* Fisch., *Lycium ruthenicum* Murray, *Carthamus tinctorius* L., *Hippophae rhamnoides* L., and *Lycium dasystemum* Pojark.) were selected to explore the areas that are potentially suitable for growing economically important desert plants in Xinjiang. Species distribution models (SDMs) were established for the 6 plants using the MaxEnt model using the points these 6 species were sample in Xinjiang, along with 31 environmental variables categorized as bioclimate, soil, and topography factors. The outputs of the constructed models, were used to map and visualize the areas that are potentially suitable for growing the 6 desert plants under the current climate conditions in Xinjiang. The areas under the curve of the suitability assessment models for each plant species were larger than 0.8, indicating high prediction accuracy. Each dataset includes 2 data files: (1) the land suitability evaluation

**Received:** 25-03-2025; **Accepted:** 20-05-2025; **Published:** 25-06-2025

\***Corresponding Author:** Zhang, H., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, zhangheng@ms.xjb.ac.cn

**Data Citation:** [1] Zhang, P., Wang, X. P., Fan, J. L., *et al.* Land suitability assessment dataset of six economically important desert plants in Xinjiang Uygur Autonomous Region [J]. *Journal of Global Change Data & Discovery*, 2025, 9(2): 209–221. <https://doi.org/10.3974/geodp.2025.02.09>.

[2] Zhang, P., Yu, X. X., Chang, C., *et al.* Land suitability assessment dataset for *Isatis tinctoria* L. in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.02.V1>.

[3] Zhang, P., Yu, X. X., Chang, C., *et al.* Land suitability assessment dataset for *Glycyrrhiza uralensis* Fisch. in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.03.V1>.

[4] Zhang, P., Yu, X. X., Chang, C., *et al.* Land suitability assessment dataset for *Lycium dasystemum* Pojark. in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.04.V1>.

[5] Zhang, P., Yu, X. X., Chang, C., *et al.* Land suitability assessment dataset for *Lycium ruthenicum* Murray in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.05.V1>.

[6] Zhang, P., Yu, X. X., Chang, C., *et al.* Land suitability assessment dataset for *Carthamus tinctorius* L. in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.06.V1>.

[7] Zhang, P., Yu, X. X., Chang, C., *et al.* Land suitability assessment dataset for *Hippophae rhamnoides* L. in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.07.V1>.

results for each plant (.txt and .tif formats) and (2) spatial distribution point data (reference data) for each plant (.xlsx and .shp formats).

**Keywords:** MaxEnt model; economic desert plants; spatial distribution; suitable habitat

**DOI:** <https://doi.org/10.3974/geodp.2025.02.09>

**Dataset Availability Statement:**

The dataset supporting this paper was published and is accessible through the *Digital Journal of Global Change Data Repository* at: <https://doi.org/10.3974/geodb.2025.02.02.V1>. <https://doi.org/10.3974/geodb.2025.02.03.V1>. <https://doi.org/10.3974/geodb.2025.02.04.V1>. <https://doi.org/10.3974/geodb.2025.02.05.V1>. <https://doi.org/10.3974/geodb.2025.02.06.V1>. <https://doi.org/10.3974/geodb.2025.02.07.V1>.

## 1 Introduction

Predicting suitable habitats for various species is a key focus in the field of ecology, playing an important role in controlling pests<sup>[1]</sup>, conserving rare and endangered species<sup>[2,3]</sup>, as well as introducing and cultivating crops<sup>[4,5]</sup>. Species distribution models (SDMs) are tools for predicting the potential geographic distribution of habitats that are suitable for a species<sup>[6-8]</sup>. The maximum entropy model (MaxEnt) is one of SDMs providing a general machine learning method that uses known species distribution records and their associated environmental variables to infer ecological requirements, predicting potential suitable habitats within a selected area<sup>[9,10]</sup>. The MaxEnt model has been widely used because of only requiring a small number of samples, ease of operation, and high prediction accuracy<sup>[11]</sup>. For example, Zhang<sup>[12]</sup> used the MaxEnt model to predict areas suitable for *Lycium ruthenicum* Murray in China, whereas Chen<sup>[13]</sup> predicted the geographic distribution of *Hippophae rhamnoides* L. in Xinjiang using the MaxEnt model. Additionally, the MaxEnt model has been applied for predicting crop distribution areas<sup>[14]</sup>, assessing the impact of climate change on species growth<sup>[15]</sup>, and optimizing the layout of species planting structures<sup>[16]</sup>, etc.

Studies identifying the habitats that are suitable for *Isatis tinctoria* L., *Carthamus tinctorius* L. and *Lycium dasystemum* Pojark. are lacking, and these suitable areas have not been visually analyzed in the Xinjiang region. Predictions of areas suitable for *Glycyrrhiza uralensis* Fisch., *Lycium ruthenicum* Murray, and *Hippophae rhamnoides* L. have primarily been conducted at the national scale, whereas the predictions of suitable areas within Xinjiang require refinement<sup>[12,17,18]</sup>. Overall, additional research is needed to identify the habitats that are suitable for these 6 economically important desert plants in Xinjiang.

This study focused on *Isatis tinctoria* L., *Glycyrrhiza uralensis* Fisch., *Lycium barbarum* Pojark., *Lycium ruthenicum* Murray, *Carthamus tinctorius* L., and *Hippophae rhamnoides* L. Data on plant sample points and environmental variables were obtained from the Global Biodiversity Information Facility (GBIF), WorldClim 2.1, and World Soil Databases using the MaxEnt model and ArcGIS. The spatial distribution of the habitats suitable for these 6 species of desert plants in Xinjiang was determined to provide a reference for rationally planning the use of these plants in the region.

## 2 Metadata of the Dataset

Table 1 presents the meta data of land suitability assessment dataset for *Isatis tinctoria* L.<sup>[19]</sup>, *Glycyrrhiza uralensis* Fisch.<sup>[20]</sup>, *Lycium dasystemum* Pojark.<sup>[21]</sup>, *Lycium ruthenicum* Murray<sup>[22]</sup>, *Carthamus tinctorius* L.<sup>[23]</sup>, and *Hippophae rhamnoides* L.<sup>[24]</sup> in Xinjiang Uygur Autonomous Region of China. The table includes details such as the full names, short names, authors, year of the datasets, temporal resolution, spatial resolution, data formats, etc.

**Table 1** Metadata summary of the datasets

Items	Description
Dataset full name	Land suitability assessment dataset for <i>Isatis tinctoria</i> L. in Xinjiang Uygur Autonomous Region of China
Dataset short name	<i>Isatis tinctoria</i> L._XJSuitable
Authors	Zhang, P., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, z1571824849@163.com Yu, X. X., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, yuxiangxiang@ms.xjb.ac.cn Chang, C., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, changcun@ms.xjb.ac.cn Zhang, H., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, zhangheng@ms.xjb.ac.cn Fan, J. L., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, fanjl@ms.xjb.ac.cn
Geographical region	Xinjiang
Year	Bioclimatic variables 1970–2000, topographic data 2000, and soil characteristics data 2009
Temporal resolution	Year
Spatial resolution	30"
Data format	.shp, .tif, .xlsx, .txt
Data size	292 KB
Data files	Land suitability assessment data for <i>Isatis tinctoria</i> L., coordinate points of <i>Isatis tinctoria</i> L. (cited data)
Dataset full name	Land suitability assessment dataset for <i>Glycyrrhiza uralensis</i> Fisch. in Xinjiang Uygur Autonomous Region of China
Dataset short name	<i>Glycyrrhiza uralensis</i> Fisch._XJSuitable
Authors	Zhang, P., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, z1571824849@163.com Yu, X. X., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, yuxiangxiang@ms.xjb.ac.cn Chang, C., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, changcun@ms.xjb.ac.cn Ma, Q., Tarim Oilfield branch of China National Petroleum Corporation, 462920056@qq.com Wang, L., Tarim Oilfield branch of China National Petroleum Corporation, 41063307@qq.com Fan, J. L., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, fanjl@ms.xjb.ac.cn
Geographical region	Xinjiang
Year	Bioclimatic variables 1970–2000, topographic data 2000, and soil characteristics data 2009
Temporal resolution	Year
Spatial resolution	30"
Data format	.shp, .tif, .xlsx, .txt
Data size	313 KB
Data files	Land suitability assessment data for <i>Glycyrrhiza uralensis</i> Fisch., coordinate points of <i>Glycyrrhiza uralensis</i> Fisch. (cited data).
Dataset full name	Land suitability assessment dataset for <i>Lycium dasystemum</i> Pojark. in Xinjiang Uygur Autonomous Region of China
Dataset short name	<i>Lycium dasystemum</i> Pojark._XJSuitable
Authors	Zhang, P., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, z1571824849@163.com Yu, X. X., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, yuxiangxiang@ms.xjb.ac.cn Chang, C., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, changcun@ms.xjb.ac.cn Zhang, H., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, zhangheng@ms.xjb.ac.cn Fan, J. L., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, fanjl@ms.xjb.ac.cn

(To be continued on the next page)

(Continued)

Items	Description
Geographical region	Xinjiang
Year	Bioclimatic variables 1970–2000, topographic data 2000, and soil characteristics data 2009
Temporal resolution	Year
Spatial resolution	30"
Data format	.shp, .tif, .xlsx, .txt
Data size	285 KB
Data files	Land suitability assessment data for <i>Lycium dasystemum</i> Pojark., coordinate points of <i>Lycium dasystemum</i> Pojark. in Xinjiang (cited data).
Dataset full name	Land suitability assessment dataset for <i>Lycium ruthenicum</i> Murray in Xinjiang Uygur Autonomous Region of China
Dataset short name	<i>Lycium ruthenicum</i> Murray_XJSuitable
Authors	Zhang, P., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, z1571824849@163.com Yu, X. X., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, yuxiangxiang@ms.xjb.ac.cn Chang, C., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, changcun@ms.xjb.ac.cn Zhang, H., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, zhangheng@ms.xjb.ac.cn Fan, J. L., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, fanjl@ms.xjb.ac.cn
Geographical region	Xinjiang
Year	Bioclimatic variables 1970–2000, topographic data 2000, and soil characteristics data 2009
Temporal resolution	Year
Spatial resolution	30"
Data format	.shp, .tif, .xlsx, .txt
Data size	332 KB
Data files	Land suitability assessment data for <i>Lycium ruthenicum</i> Murray, coordinate points of <i>Lycium ruthenicum</i> Murray (cited data)
Dataset full name	Land suitability assessment dataset for <i>Carthamus tinctorius</i> L. in Xinjiang Uygur Autonomous Region of China
Dataset short name	<i>Carthamus tinctorius</i> L._XJSuitable
Authors	Zhang, P., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, z1571824849@163.com Yu, X. X., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, yuxiangxiang@ms.xjb.ac.cn Chang, C., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, changcun@ms.xjb.ac.cn Zhang, H., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, zhangheng@ms.xjb.ac.cn Fan, J. L., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, fanjl@ms.xjb.ac.cn
Geographical region	Xinjiang
Year	Bioclimatic variables 1970–2000, topographic data 2000, and soil characteristics data 2009
Temporal resolution	Year
Spatial resolution	30"
Data format	.shp, .tif, .xlsx, .txt
Data size	231 KB
Data files	Land suitability assessment data for <i>Carthamus tinctorius</i> L. coordinate points of <i>Carthamus tinctorius</i> L. (cited data)
Dataset full name	Land suitability assessment dataset for <i>Hippophae rhamnoides</i> L. in Xinjiang Uygur Autonomous Region of China
Dataset short name	<i>Hippophae rhamnoides</i> L._XJSuitable
Authors	Zhang, P., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, z1571824849@163.com

(To be continued on the next page)

(Continued)

Items	Description
Authors	Yu, X. X., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, yuxiangxiang@ms.xjb.ac.cn Chang, C., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, changcun@ms.xjb.ac.cn Zhang, H., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, zhangheng@ms.xjb.ac.cn Fan, J. L., Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, fanjl@ms.xjb.ac.cn
Geographical region	Xinjiang
Year	Bioclimatic variables 1970–2000, topographic data 2000, and soil characteristics data 2009
Temporal resolution	Year
Spatial resolution	30"
Data format	.shp, .tif, .xlsx, .txt
Data size	378 KB
Data files	Land suitability assessment data for <i>Hippophae rhamnoides</i> L., coordinate points of <i>Hippophae rhamnoides</i> L. in Xinjiang (cited data).
Foundation	Xinjiang Uygur Autonomous Region, China (2022B03030)
Data computing environment	MaxEnt, ArcGIS
Data publisher	Global Change Research Data Publishing & Repository, <a href="http://www.geodoi.ac.cn">http://www.geodoi.ac.cn</a>
Address	No. 11A, Datun Road, Chaoyang District, Beijing 100101, China
Data sharing policy	(1) <i>Data</i> are openly available and can be free downloaded via the Internet; (2) End users are encouraged to use <i>Data</i> subject to citation; (3) Users, who are by definition also value-added service providers, are welcome to redistribute <i>Data</i> subject to written permission from the GCDatPR Editorial Office and the issuance of a <i>Data</i> redistribution license; and (4) If <i>Data</i> are used to compile new datasets, the “ten percent principal” should be followed, such that <i>Data</i> records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset <sup>[25]</sup> .
Communication and searchable systems	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS, GEOSS, PubScholar, CKRSC

### 3 Methods

#### 3.1 Data Sources

##### 3.1.1 Plant Distribution Data

The plant distribution data were obtained from the GBIF<sup>1</sup>, with sampling spanning from 1949 to 2020. We acquired precise latitudinal and longitudinal information on the distribution of the sampling points of *Isatis tinctoria* L., *Glycyrrhiza uralensis* Fisch., *Lycium ruthenicum* Murray, *Carthamus tinctorius* L., *Hippophae rhamnoides* L., and *Lycium dasystemum* Pojark. within the Xinjiang region using this platform<sup>[26]</sup>. These data served as the foundation for determining the spatial distribution for each plant.

##### 3.1.2 Environmental Factor Data

A total of 31 environmental datasets were selected for modeling, which included 19 bioclimatic (Bio1–Bio19) and 9 soil variables, as well as elevation, slope, and aspect. The data for the 19 bioclimatic variables and elevation were obtained from the WorldClim2.1 database<sup>[27]</sup>. The slope and aspect data were derived from elevation data using ArcGIS, and surface soil data were sourced from the National Cryosphere Desert Data Center<sup>[28]</sup>. The environmental variables are listed in Table 2.

<sup>1</sup> GBIF. <http://www.gbif.org>.

**Table 2** List of environmental variables used in MaxEnt model

Type	Variable	Description	Unit
Bioclimate variables	Bio1	Annual mean temperature	°C
	Bio2	Mean diurnal range	°C
	Bio3	Isothermality ((Bio2/Bio7)×100)	\
	Bio4	Temperature seasonality (standard deviation ×100)	\
	Bio5	Max. temperature of the warmest month	°C
	Bio6	Min. temperature of the coldest month	°C
	Bio7	Temperature annual range (Bio5–Bio6)	°C
	Bio8	Mean temperature of the wettest quarter	°C
	Bio9	Mean temperature of the driest quarter	°C
	Bio10	Mean temperature of the warmest quarter	°C
	Bio11	Mean temperature of the coldest quarter	°C
	Bio12	Annual precipitation	mm
	Bio13	Precipitation of the wettest month	mm
	Bio14	Precipitation of the driest month	mm
	Bio15	Precipitation seasonality	\
	Bio16	Precipitation of the wettest quarter	mm
	Bio17	Precipitation of the driest quarter	mm
	Bio18	Precipitation of the warmest quarter	mm
	Bio19	Precipitation of the coldest quarter	mm
Terrain	Elev	Elevation	m
	Slope	Slope	°
	Aspect	Aspect	\
Soil variables	T_CaCO <sub>3</sub>	Topsoil calcium carbonate	%
	T_cec_soil	Topsoil CEC (soil)	cmol/kg
	T_clay	Topsoil clay fraction	%
	T_esp	Topsoil sodicity (ESP)	\
	T_gravel	Topsoil gravel content	%
	TOC	Topsoil organic carbon	%
	T_pH_H <sub>2</sub> O	Topsoil pH (H <sub>2</sub> O)	–log(H <sup>+</sup> )
	T_sand	Topsoil sand fraction	%
	T_texture	Topsoil texture	

## 3.2 Methodology

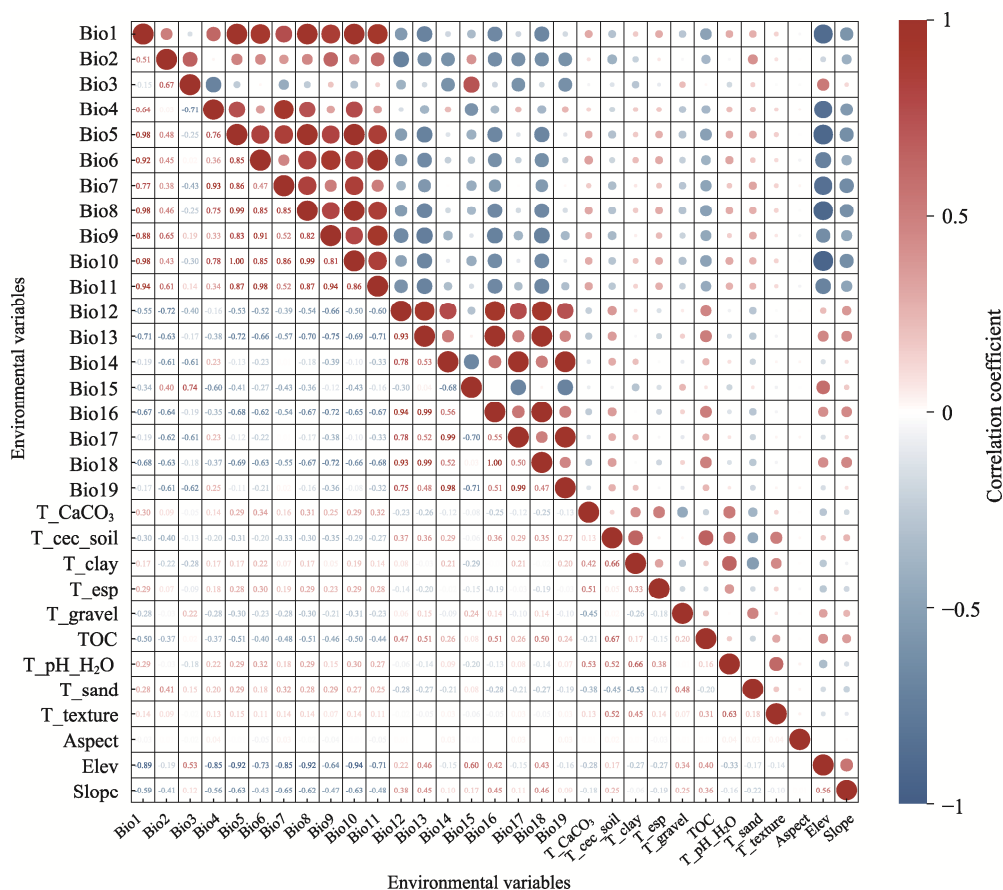
### 3.2.1 Data Processing

We focused on Xinjiang for the 6 plants species. The spatial resolution of the environmental factor layer was 30 arcseconds, and the climatic factor data within the same grid cell were identical. Redundant or duplicate distribution data within the same grid cell caused overfitting. Therefore, ENMTools software was used to filter the collected distribution data. ENMTools is open-source software for modeling and predicting species distribution, providing a suite of tools that help researchers extract information from large species distribution datasets. This tool was used to automatically match the resolution of the environmental factor layer used in the analysis and remove the redundant data within the same grid cell, enhancing the accuracy of the model-fitting results.

### 3.2.2 Environmental Variables

Correlations exist among environmental factors, as such, incorporating all the environmental factors into the model would have resulted in overfitting. Precisely quantifying the

relationships between known locations of species and their corresponding environmental variables forms the foundation for modeling species distribution. Environmental variables that are ecologically relevant should be selected to minimize the collinearity among variables and prevent model overfitting during distribution simulations<sup>[29]</sup>. The correlation among the variables was analyzed using ENMTools software to generate pairwise correlation coefficients among the environmental factors to address collinearity-induced overfitting (Figure 1). The correlation coefficient quantitatively measures the linear relationships between variables, ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation). Absolute values approaching 1 indicate stronger correlations. All 31 environmental variables were incorporated into the initial modeling to assess the contribution rates of the individual variables. Subsequently, the factors demonstrating both high collinearity (absolute correlation coefficient  $\geq 0.8$ ) and low ecological relevance (contribution rate  $< 0.5$ ) were excluded based on established criteria<sup>[30]</sup>. A refined set of environmental variables was then used to construct the final model.



**Figure 1** Results of correlation analysis of environmental factors

### 3.2.3 Maximum Entropy Modeling

The species distribution and environmental data were processed and then imported into MaxEnt for training and validating the model. A cross-validation approach was implemented for each species by partitioning all occurrence records into 10 subsets, with 1 and 9 subsets used for testing and training, respectively. The modeling parameters were configured for a maximum of 10,000 iterations and a logistic output format. Variable importance was evaluated using the jackknife method with response curve generation. The modeling process

was repeated through 10-fold cross-validation to minimize uncertainty, with the final predictions derived by averaging all the replicate results.

The accuracy of the model was evaluated using the area under the receiver operating characteristic curve (ROC) of the MaxEnt outputs. The ROC curve demonstrates the relationship between the correctly predicted distribution points and false predictions at different thresholds. The area under the curve (AUC) ranges from 0 to 1, with higher values corresponding to higher prediction reliability. AUC values below 0.8, between 0.8 and 0.9, and from 0.9 to 1.0 represent low, moderate, and high accuracy, respectively<sup>[31]</sup>.

### 3.2.4 Delineation and Spatiotemporal Dynamics of Suitable Areas

The operational results of the MaxEnt model were imported into ArcGIS 10.8 software, where the reclassification function in the spatial analysis module was used to process the generated raster data files. Area distribution maps were created for the 6 plants species across Xinjiang. The suitability values predicted using the MaxEnt model were continuous raster data ranging from 0 to 1. Habitat suitability was classified into 4 categories: inappropriate, low suitable, medium suitable, and highly suitable based on the natural break method<sup>[32]</sup>. Table 3 lists the specific natural break thresholds for the suitability of each plant species.

**Table 3** Classification criteria determined using natural break method

Plant name	Inappropriate area	Low suitable area	Medium suitable area	Highly suitable area
<i>Isatis tinctoria</i> L.	0–0.06	0.06–0.23	0.23–0.49	0.49–1
<i>Glycyrrhiza uralensis</i> Fisch.	0–0.09	0.09–0.26	0.26–0.51	0.51–1
<i>Lycium ruthenicum</i> Murray	0–0.08	0.08–0.25	0.25–0.52	0.52–1
<i>Carthamus tinctorius</i> L.	0–0.07	0.07–0.22	0.22–0.49	0.49–1
<i>Hippophae rhamnoides</i> L.	0–0.09	0.09–0.26	0.26–0.54	0.54–1
<i>Lycium dasystemum</i> Pojark.	0–0.07	0.07–0.23	0.23–0.50	0.50–1

The technical roadmap of this study is shown in Figure 2.

## 4 Data Results and Validation

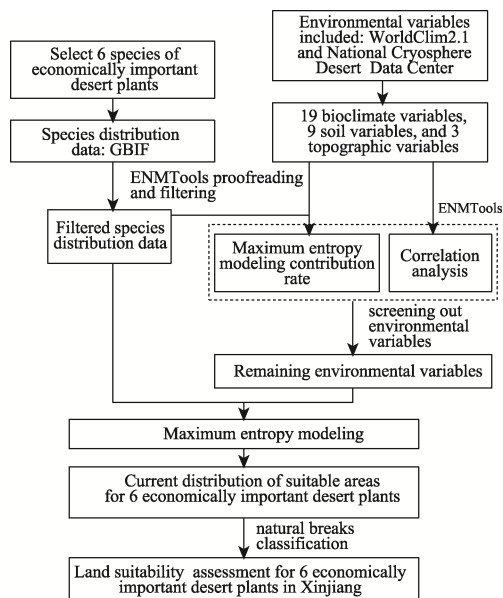
### 4.1 Dataset Composition

6 datasets were constructed, each containing 2 data files: (1) land suitability assessment data for economically important desert plants in Xinjiang (*Isatis tinctoria* L., *Glycyrrhiza uralensis* Fisch., *Lycium ruthenicum* Murray, *Carthamus tinctorius* L., *Hippophae rhamnoides* L., and *Lycium dasystemum* Pojark.), archived in .txt and .tif formats; (2) coordinate points of the 6 plants in Xinjiang (cited data), which were archived in .xlsx and .shp formats.

### 4.2 Data Products

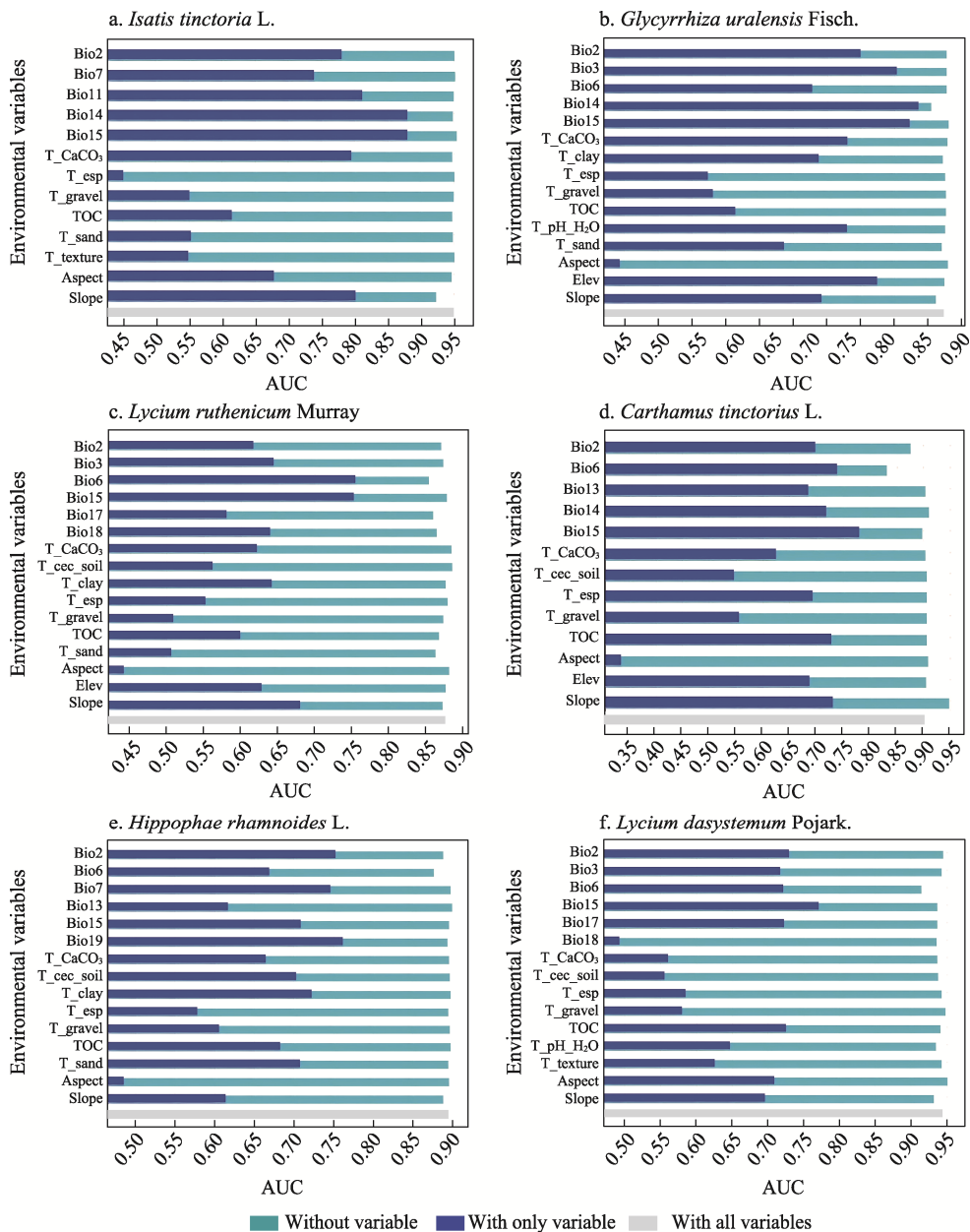
#### 4.2.1 Filtering Data and Evaluating Accuracy of MaxEnt Model

The distribution data were filtered using ENMTools, resulting in the final selection of 22, 36, 43, 19, 66, and 16 sample points for *Isatis tinctoria* L., *Glycyrrhiza uralensis* Fisch., *Lycium*



**Figure 2** Flowchart of the dataset development

*ruthenicum* Murray, *Carthamus tinctorius* L., *Hippophae rhamnoides* L., and *Lycium dasystemum* Pojark, respectively. The environmental variables were screened based on the results of correlation analysis and the contribution rates, retaining 13, 15, 16, 13, 15, and 15 variables for each species, as illustrated in Figure 3. The AUCs of the suitability assessment model for these species (Figure 3) were 0.950, 0.874, 0.877, 0.905, 0.896, and 0.943, respectively. All AUC values exceeded 0.8, demonstrating the reliability of the MaxEnt model. The AUC values for *Isatis tinctoria* L., *Carthamus tinctorius* L., and *Lycium dasystemum* Pojark. were higher than 0.9, indicating high prediction accuracy.



**Figure 3** AUC values of predictive models for 6 economically important desert plants

#### 4.2.2 Distribution Ranges of 6 Plants Under Current Climatic Conditions

The areas potentially suitable for *Isatis tinctoria* L., *Glycyrrhiza uralensis* Fisch., *Lycium ruthenicum* Murray, *Carthamus tinctorius* L., *Hippophae rhamnoides* L., and *Lycium dasystemum* Pojark. for Xinjiang under current climatic conditions are shown in Figure 4. These 6 plants predominantly inhabit the oasis regions in Xinjiang, and their potential distributions widely vary. The specific distribution patterns are described in detail below.

(1) Northwestern Xinjiang is the area highly suitable for *Isatis tinctoria* L. cultivation, mainly including the Ili Kazakh Autonomous Prefecture, Tacheng, and Altay. The areas that are medium suitable include the Tacheng, Ili Kazakh Autonomous Prefecture, and Changji Hui Autonomous Prefecture. Low suitable area mainly covers the Hami Diqu (Region), certain oasis zones in southern Xinjiang, and the Gurbantunggut Shamo (Desert). Southern Xinjiang is mostly inappropriate for growing *Isatis tinctoria* L. (Figure 4a).

(2) The areas that are highly suitable for cultivating *Glycyrrhiza uralensis* Fisch. include northern Xinjiang, where the distribution pattern is fragmented. Medium suitable areas surround these high-suitability zones, whereas the low suitable areas mainly include the Hami Diqu (Region), Bayingolin Mongol Autonomous Prefecture, Aksu, and Kashgar. Southern Xinjiang, particularly in the southern Bayingolin Mongol Autonomous Prefecture and Hotan Diqu (Region), is inappropriate (Figure 4b).

(3) The areas highly suitable for cultivating *Lycium ruthenicum* Murray are more prevalent in southern than in northern Xinjiang, with belt-shaped formations. This species is predominantly found in the oasis zones in southern Xinjiang spanning the Hotan-Kashgar-Kizilsu Kirghiz-Aksu, whereas the populations in northern Xinjiang are mainly distributed along the Bortala-Tacheng-Changji-Urumqi-Turpan oasis belt. The medium suitable areas mainly include the Tacheng; and the Bayingolin Mongol Autonomous Prefecture and the eastern Altay are low suitable areas (Figure 4c).

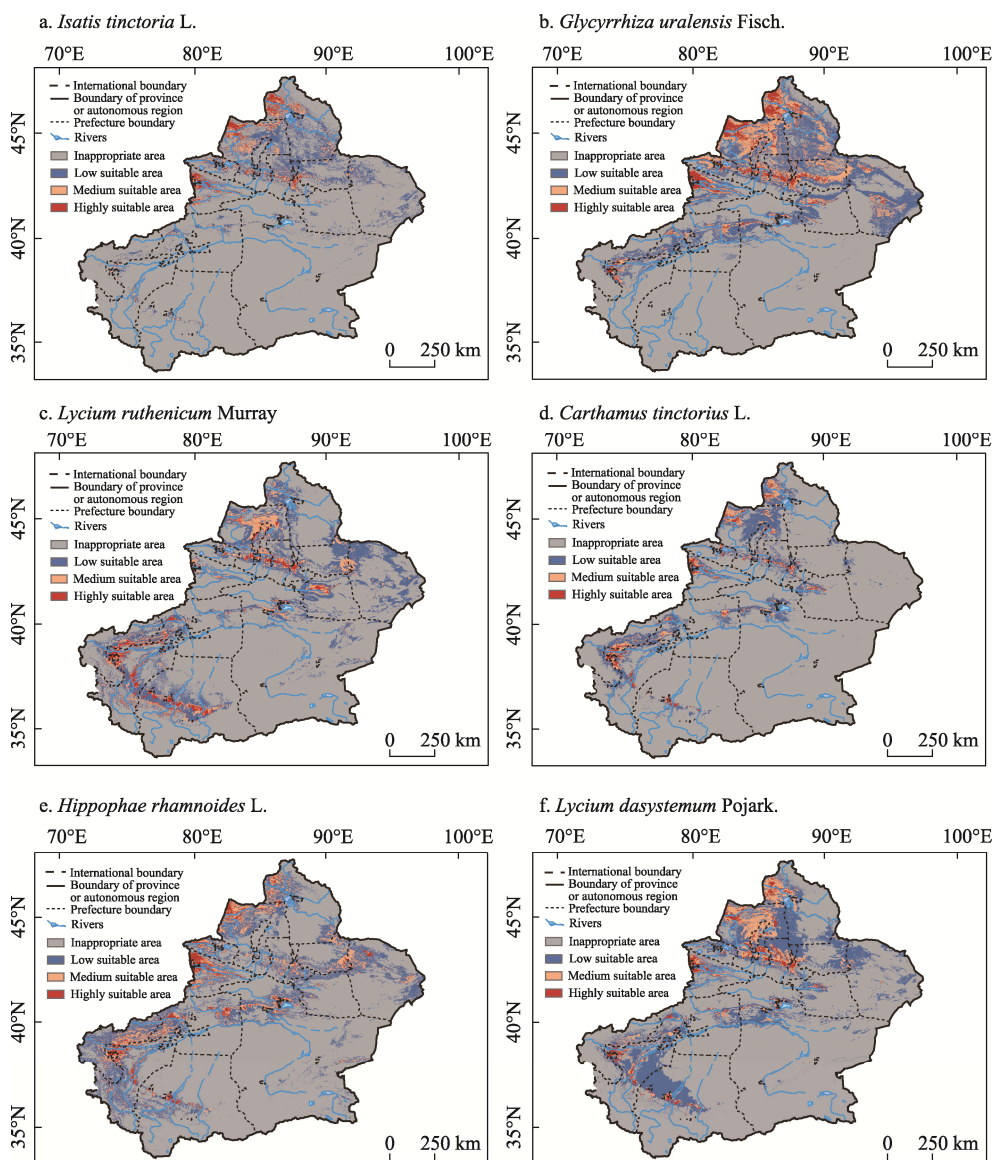
(4) The main areas that are highly suitable for cultivating *Carthamus tinctorius* L. include the oasis regions throughout various prefectures, with the distribution being sparse and fragmented. The medium suitable areas cluster in the northwestern Tacheng and western Altay. The less-suitable areas primarily occur within the Tacheng. The inappropriate areas include eastern Xinjiang, with notable concentrations in Altay, Hami, Bayingolin Mongol Autonomous Prefecture, and Hotan (Figure 4d).

(5) Areas highly suitable for growing *Hippophae rhamnoides* L. are the oasis regions throughout various prefectures, with the distribution being patchy and primarily concentrated in the Ili Kazakh Autonomous Prefecture and Kashgar. The medium suitable area is the in Tacheng, whereas the low suitable areas mainly include the Hami, Kizilsu Kirgiz Autonomous Prefecture, and the southern Kashgar. Areas inappropriate for *Hippophae rhamnoides* L. cultivation are concentrated in southern Xinjiang, particularly in the southern Bayingolin Mongol Autonomous Prefecture and northern Hotan (Figure 4e).

(6) Northern Xinjiang is highly suitable for *Lycium dasystemum* Pojark., with a patchy distribution pattern, with scattered occurrences in the Hotan and Kashgar in southern Xinjiang. The medium suitable area mainly includes the Tacheng and Altay, whereas the low suitable areas are primarily located in western Hotan, eastern Kashgar, western Altay, and Hami. The main area that is inappropriate is the southern Bayingolin Mongol Autonomous Prefecture (Figure 4f).

Overall, the highly and medium suitable areas for cultivating the 6 plants are limited in Xinjiang, with the highly suitable areas covering less than 5% of the region. The areas highly suitable for growing *Glycyrrhiza uralensis* Fisch., *Hippophae rhamnoides* L., and *Lycium dasystemum* Pojark. cover 3.61%, 2.58%, and 2.46% of the region, respectively. Most areas were classified as inappropriate, with 86.27%, 84.40%, and 75.01% of the region being unsuitable for *Carthamus tinctorius* L., *Isatis tinctoria* L., and *Lycium ruthenicum*

Murray, respectively. The proportion of inappropriate land was smallest for *Glycyrrhiza uralensis* Fisch., at 69.91%. The proportions of low suitable area was largest and smallest for *Lycium dasystemum* Pojark. (17.98%) and *Carthamus tinctorius* L. (9.39%), respectively.



**Figure 4** Land suitability maps for the 6 economically important desert plants in Xinjiang

## 5 Discussion and Conclusion

A model was developed for predicting the habitats that are suitable for economically important desert plants in Xinjiang using MaxEnt and ArcGIS. Various environmental variables, such as bioclimate, topography, and soil characteristics, were incorporated into the model. The habitats were categorized into 4 suitability grades using the natural break method, with the AUC of all plant suitability assessment models exceeding 0.8, confirming the ecological validity of the predictions. The areas that are potentially suitable for the 6

plants in Xinjiang are spatially heterogeneous, with the distribution patterns strongly correlating with locations of oasis ecosystems. Highly suitable areas accounted for less than 5% of the total region, with 3.61%, 2.58%, and 2.46% of the region being highly suitable for *Glycyrrhiza uralensis* Fisch., *Hippophae rhamnoides* L., and *Lycium dasystemum* Pojark., respectively.

A dataset was constructed of the habitats that are suitable for these plants in Xinjiang, which provides a reference for the spatial planning of desert-based agriculture and provides scientific support for sustainable plant resource use. Future studies should investigate the spatiotemporal dynamics of the habitats of economically important desert plants under projected climate change scenarios, quantify future centroid migration rates across different suitability zones, and evaluate the long-term sustainability of the plant resources in deserts.

### Author Contributions

Zhang, P. collected, processed the data, and drafted the manuscript. Wang, X. P., Zhang, G. F. and Zhang, H. contributed to the data processing. Fan, J. L. designed the overall framework for developing the dataset. Zhang, H. and Yu, X. X. supervised the writing of this paper.

### Conflicts of Interest

The authors declare no conflicts of interest.

### References

- [1] Qin, X. S., Li, M. X. Predicting the potential distribution of *Oxalis debilis* Kunth, an invasive species in China with a maximum entropy model [J]. *Plants*, 2023, 12(23): 3999.
- [2] Zhao, D., Wang, J., Dai, W., *et al.* Effects of climate warming and human activities on the distribution patterns of *Fritillaria unibracteata* in eastern Qinghai-Tibetan Plateau [J]. *Scientific Reports*, 2023, 13: 15770.
- [3] Lin, H. D., Zheng, Q. Z., Shen, L. Q., *et al.* Assessment of habitat suitability in autumn for wild alpine musk deer in Xinglongshan National Nature Reserve with MaxEnt model [J]. *Chinese Journal of Ecology*, 2024, 43(1): 299–304.
- [4] Deng, C. R., Zhong, Q. W., Shao, D. K., *et al.* Potential suitable habitats of chili pepper in China under climate change [J]. *Plants*, 2024, 13(7): 1027.
- [5] Li, J. L., Deng, C. R., Duan, G. Z., *et al.* Potentially suitable habitats of Daodi goji berry in China under climate change [J]. *Frontiers in Plant Science*, 2024, 14: 1279019.
- [6] Evans, J. M., Fletcher, R. J., Alavalapati, J. Using species distribution models to identify suitable areas for biofuel feedstock production [J]. *Global Change Biology Bioenergy*, 2010, 2(2), 63–78.
- [7] Zhao, Y., Deng, X. W., Xiang, W. H., *et al.* Predicting potential suitable habitats of Chinese fir under current and future climatic scenarios based on MaxEnt model [J]. *Ecological Informatics*, 2021, 64:101393.
- [8] Zhao, Z. Y., Xiao, N. W., Shen, M., *et al.* Comparison between optimized MaxEnt and random forest modeling in predicting potential distribution: a case study with *Quasipaa boulengeri* in China [J]. *Science of the Total Environment*, 2022, 842: 156867.
- [9] Kang, J. F., Liu, M. S., Qu, M. K., *et al.* Identifying the potential soil pollution areas derived from the metal mining industry in China using MaxEnt with mine reserve scales (MaxEnt\_MRS) [J]. *Environmental Pollution*, 2023, 329: 121687.
- [10] Phillips, S. J., Anderson, R. P., Schapire, R. E. Maximum entropy modeling of species geographic distributions [J]. *Ecological Modelling*, 2006, 190(3–4): 231–259.
- [11] Phillips, S. J., Dudik, M. Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation [J]. *Ecography*, 2008, 31(2): 161–175.
- [12] Zhang, L. Wei, Y. Q., Wang, J. N., *et al.* The potential geographical distribution of *Lycium ruthenicum* Murr. under different climate change scenarios [J]. *Chinese Journal of Applied and Environmental Biology*, 2020, 26(4): 969–978.

- [13] Chen, S. Q., Dong, H. F., Yue, Y. F., et al. Geographical distribution and dynamic change prediction of *Hippophae rhamnoides* subsp. *sinensis* under different climate scenarios [J]. *Arid Zone Research*, 2024, 41(9): 1560–1571.
- [14] Zhou, D. M., Zhang, R. Z., Sun, W. C., et al. Study on climatic suitability for winter rapeseed planting in arid and cold regions in north China [J]. *Scientia Agricultura Sinica*, 2014, 47(13): 2541–2551.
- [15] Gan, X. L., Chang, Y. P., Jiang, Y., et al. Impact of climate change on potential distribution of *Amygdalus mongolicain* the Qilian Mountains [J]. *Acta Ecologica Sinica*, 2023, 43(2): 768–776.
- [16] Yang, R. Z., Zhou, G. S. A comprehensive risk assessment of the main maize agrometeorological disasters in the three provinces of Northeast China [J]. *Acta Meteorologica Sinica*, 2015, 73(6): 1141–1153.
- [17] Xue, H. B., Liu, M. L., Li, L. S., et al. Prediction of suitable regions suitable for *Glycyrrhiza uralensis* based on MaxEnt model [J]. *Shaanxi Forest Science and Technology*, 2024, 52(3): 19–25.
- [18] Zhang, X. W., Jiang, Y. M., Bi, Y., et al. Identification of potential distribution area for *Hippophae rhamnoidessubsp.sinensis* by the MaxEnt model [J]. *Acta Ecologica Sinica*, 2022, 42(4): 1420-1428.
- [19] Zhang, P., Yu, X. X., Chang, C., et al. Land suitability assessment dataset for *Isatis tinctoria* L. in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.02.V1>.
- [20] Zhang, P., Yu, X. X., Chang, C., et al. Land suitability assessment dataset for *Glycyrrhiza uralensis* Fisch. in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.03.V1>.
- [21] Zhang, P., Yu, X. X., Chang, C., et al. Land suitability assessment dataset for *Lycium dasystemum* Pojark. in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.04.V1>.
- [22] Zhang, P., Yu, X. X., Chang, C., et al. Land suitability assessment dataset for *Lycium ruthenicum* Murray in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.05.V1>.
- [23] Zhang, P., Yu, X. X., Chang, C., et al. Land suitability assessment dataset for *Carthamus tinctorius* L. in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.06.V1>.
- [24] Zhang, P., Yu, X. X., Chang, C., et al. Land suitability assessment dataset for *Hippophae rhamnoides* L. in Xinjiang Uygur Autonomous Region of China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2025. <https://doi.org/10.3974/geodb.2025.02.07.V1>.
- [25] GCdataPR Editorial Office. GCdataPR data sharing policy [OL]. <https://doi.org/10.3974/dp.policy.2014.05> (Updated 2017).
- [26] GBIF.org User. Occurrence download [OL]. The Global Biodiversity Information Facility. 2024. <https://doi.org/10.15468/dl.s9sy5y>.
- [27] Fick, S. E., Hijmans, R. J. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas [J]. *International Journal of Climatology*, 2017, 37 (12): 4302–4315.
- [28] Lu, L., Liu, C. Chinese soil dataset based on the World Soil Database (HWSD) (v1.1) [DB/OL]. National Cryosphere Desert Data Center (<http://www.ncdc.ac.cn>), 2020. <https://cstr.cn/CSTR:11738.11.ncdc.Westdc.2020.585>.
- [29] Fang, J. Q., Shi, J. F., Zhang, P., et al. Potential distribution projections for *Senegalia senegal* (L.) Britton under climate change scenarios [J]. *Forests*, 2024, 15(2): 379.
- [30] Tarnian, F., Kumar, S., Azarnivand, H., et al. Assessing the effects of climate change on the distribution of *Daphne mucronata* in Iran [J]. *Environmental Monitoring and Assessment*, 2021, 193(9): 562.
- [31] Cecilia, A. C., Francisco, N. L., Ricardo, C. T., et al. Effects of environmental changes on the occurrence of *Oreomunnea mexicana* (Juglandaceae) in a biodiversity hotspot cloud forest [J]. *Forests*, 2017, 8(8): 261.
- [32] Shao, M. H., Wang, L., Li, B. W., et al. Maxent modeling for identifying the nature reserve of *Cistanche deserticola* Ma under effects of the host (*Haloxylon* Bunge) forest and climate changes in Xinjiang, China [J]. *Forests*, 2022, 13(2): 189.