



ARTICLE

Environmental Drivers and Spatial Prediction of the Critically Endangered Species *Thuja sutchuenensis* in Sichuan-Chongqing, China

Liang Xie^{1,2,5}, Peihao Peng^{1,*}, Haijun Wang^{1,3} and Shengbin Chen⁴

¹College of Earth Science, Chengdu University of Technology, Chengdu, 610059, China

²School of Geoscience and Technology, Southwest Petroleum University, Chengdu, 610500, China

³Engineering and Technical College, Chengdu University of Technology, Leshan, 614000, China

⁴College of Ecology and Environment, Chengdu University of Technology, Chengdu, 610041, China

⁵Spatial Information Technology and Big Data Mining Research Center, Southwest Petroleum University, Chengdu, 610500, China

*Corresponding Author: Peihao Peng. Email: pozi2340590743@163.com

Received: 18 August 2021 Accepted: 04 January 2022

ABSTRACT

Identifying the ecological environment suitable for the growth of *Thuja sutchuenensis* and predicting other potential distribution areas are essential to protect this endangered species. After selecting 24 environmental factors that could affect the distribution of *T. sutchuenensis*, including climate, topography, soil and Normalized Difference Vegetation Index (NDVI), we adopted the Random Forest-MaxEnt integrated model to analyze our data. Based on the Random Forest study, the contribution of the mean temperature of the warmest quarter, mean temperature of the coldest quarter, annual mean temperature and mean temperature of the driest quarter was large. Based on MaxEnt model prediction outputs, the potential distribution map not only identified areas that originally recorded *T. sutchuenensis*, such as Xuanhan County, Kai County and Chengkou County, but also identified highly suitable distribution areas where *T. sutchuenensis* may exist, including Wanyuan County, Sichuan Province, and the junction of Chongqing and Hubei Province. This provides a more explicit geographic range for *ex situ* conservation and reintroduction of *T. sutchuenensis*. Our results also indicate that, in addition to climate factors, topography and soil factors are also important environmental factors that affect distribution. This provides a theoretical basis for subsequent laboratory construction to simulate the indoor growth of *T. sutchuenensis*.

KEYWORDS

Thuja sutchuenensis; environmental drivers; spatial prediction; conservation; Random Forest-MaxEnt model

1 Introduction

Thuja sutchuenensis (*T. sutchuenensis*) is an evergreen tree of the family Cupressaceae [1]. In 1892, *T. sutchuenensis* was first discovered by the French missionary and botanist Paul Guillaume Farges during a botanical survey in Chengkou County, Chongqing, China [2,3]. Over the next 100 years, despite repeated field surveys in the Daba Mountain area, there were no further records of this species, resulting in its recording as being “wild extinct” in the English version of “Flora of China” (gymnosperm) volume 4 (<http://www.iplant.cn/frps>). It was thought to be extinct, and the type of habitat that Farges had found it or found growing in the wild was unknown [4]. In October 1999, the Chongqing Forestry Bureau



organized an expedition team that conducted a comprehensive survey on rare and endangered plants in Chengkou County. The status of *T. sutchuenensis* is described as “Extinct in the Wild” (EW) in the global Red List of Conifers [5].

During this expedition, a wild *T. sutchuenensis* tree was found in the Daba Mountain area, which yielded cone samples, where Farges had first found *T. sutchuenensis* [3]. After the year 2000, research teams that explored the remote areas of Daba Mountain Nature Reserve noted the existence of *T. sutchuenensis*. Since the rediscovery of *T. sutchuenensis*, studies on its morphological characteristics, community structure, biological characteristics of seeds, geographical distribution, habitat characteristics and genetic diversity have been conducted [2,3,5–22]. Momentous research results indicated that *T. sutchuenensis* has important significance for studies in plant system development, paleontology, and paleoclimate, which are owing to its special status. It plays an irreplaceable role in global biodiversity protection [11,23]. In 2012, China launched the second national survey on wild plant resources, resulting in the identification of 12 *T. sutchuenensis* trees (from which samples were collected) in Xuanhan County, Sichuan, in November 2014 (Fig. 1). Based on the results of this field study, the pattern of distribution of *T. sutchuenensis* in the local region becomes our concern, which is of substantial significance for the analysis of habitat characteristics and formulation of field protection measures for this tree. The distribution of *T. sutchuenensis* effectively correlates with climate, since it is hardy to the zone 6 cold hardiness limit between -23.2°C and -17.8°C [24].



Figure 1: Field investigation of *Thuja sutchuenensis*

Few studies have explored the mechanism of impact of macro-environmental factors from a geographical spatial scale, and the large-scale prediction of distribution areas has not been undertaken. To the best of our knowledge, Ma et al. [2] ventured deep into the southern and southeastern tip of Daba Mountain and conducted field observations that provided a detailed description for the geographic distribution, climate characteristics, terrain and soil characteristics, and vegetation characteristics, and inferred the cause of pattern of the geographic distribution of *T. sutchuenensis* in Daba Mountain [2]. However, this result neither predicted the potential distribution of this tree nor demonstrated the weight ranking of the many factors that influence the distribution of *T. sutchuenensis*. Qin et al. [14] adopted the

Jackknife method to screen the main impact factors of *T. sutchuenensis* and utilized the MaxEnt model to predict the potential area of distribution of *T. sutchuenensis* in macroscale. In this study [14], important environmental factors that affect the growth of *T. sutchuenensis* were identified as the mean of diurnal temperature range, minimum temperature of coldest month, annual precipitation, precipitation of the driest month, precipitation seasonality and precipitation of the warmest quarter. These environmental factors were used to analyze the temporal and spatial distribution variation of *T. sutchuenensis*. The important conclusions from this study highlight climate as the main factor that determines the distribution in a regional context, a finding that provides an important reference for subsequent investigations. However, predicting the potential distribution of *T. sutchuenensis* by only relying on climatic factors does not comprehensively reflect the distribution pattern of *T. sutchuenensis*. For example, to accurately predict the potential distribution area of *Justicia adhatoda* L., Yang et al. [25] adopted Worldclim [1.4] bioclimatic variables, in addition to the slope, aspect, elevation and land use/land cover, as significant impact factors in the model [25]. The distribution of rubber trees was predicted by Ray et al. [26] by adopting the MaxEnt model, using climatic, soil, topographical and socioeconomic factors. Although they concluded that climate plays a major role in the distribution of rubber trees (*Ficus elastica*), their results indicated that topography, soil and socioeconomic factors were also significant. Based on previous studies, we believe that when predicting the potential distribution of rare plants, the prediction ability of the model should be revised and improved in combination with natural environment factors, such as topography and soil factors. The distribution pattern of *T. sutchuenensis* plays a significant role at the local regional scale, which had been ignored by Qin et al. [14] in her study. Our study attempted to improve this, which is of substantial significance to the implementation of measures to protect *T. sutchuenensis*.

Ecologists commonly adopt Species Distribution Models (SDM) to determine the pattern of distribution and develop protection measures for selected species. Common models that have been previously used include GARP, MaxEnt and ENFA [27,28]. Wang and Huang predicted the potential distribution of chestnut phylloxera (*Mortzella castanaivora*) using a GAPR model, and their results indicated that this species is most likely distributed in northeastern, eastern, northern and central areas in China. Farashi et al. [29] utilized the ENFA model to analyze the populations of racoons and found that the most important factor that affects their distribution is vegetation and predicted that most areas in Gilan Province in Iran are potential distribution areas and areas with the highest risk of invasion of racoons [29]. Yi et al. [30] utilized the MaxEnt model to discover the pattern of distribution pattern of *Homonoia riparia* in Yunan, China. They found that having *H. riparia* mostly distributed along the river and the area closer to the riverbank are more favorable for the survival of this species.

In this study, we used the Random Forest model to reveal the significant impact factors that influence the pattern of distribution of *T. sutchuenensis*. The study area is focused on seven counties around Daba Mountain. We used the Maxent model to predict the potential distribution of *T. sutchuenensis* in this area. We also analyzed the habitat characteristics of *T. sutchuenensis* based on the response curve and discussed the cause of the habitat at the local regional scale to provide a theoretical basis for subsequent ex-situ conservation and reintroduction of *T. sutchuenensis*.

2 Methods

2.1 Study Area and Data

The samples used in this study were located between 108.2° to 108.9° E and 31.6° to 31.8° N, and most of the study was located in the National Nature Reserve of Daba Mountain, and a small portion was located in Xuanhan County, Sichuan. The altitude in the study area ranged between 320 and 2,813 m (Fig. 2). The climate in the study region belongs to the transitional zone of mid-subtropical and north-subtropical zones

with subtropical evergreen broad-leaved forest zonal vegetation. We obtained habitat parameters, such as geographic location, altitude, slope, aspect, soil pH, and canopy density, during a field study in Xuanhan County. In addition, we also acquired information on the age structure and population composition of *T. sutchuenensis*. Based on the climatic and site conditions of known geographical locations of *T. sutchuenensis*, we concluded that the tree is heliophilous, has slow growth, and is primarily located in calcareous limestone mountainous areas. Owing to the large distribution range of limestone in Sichuan and Chongqing, the study area was expanded to 107.3°–109.8° E and 30.4°–32.4° N during the field study. We attempted to examine suitable distribution range and habitat characteristics of *T. sutchuenensis* in this area and study the main reasons for the lack of survival of *T. sutchuenensis* in non-suitable distribution areas.

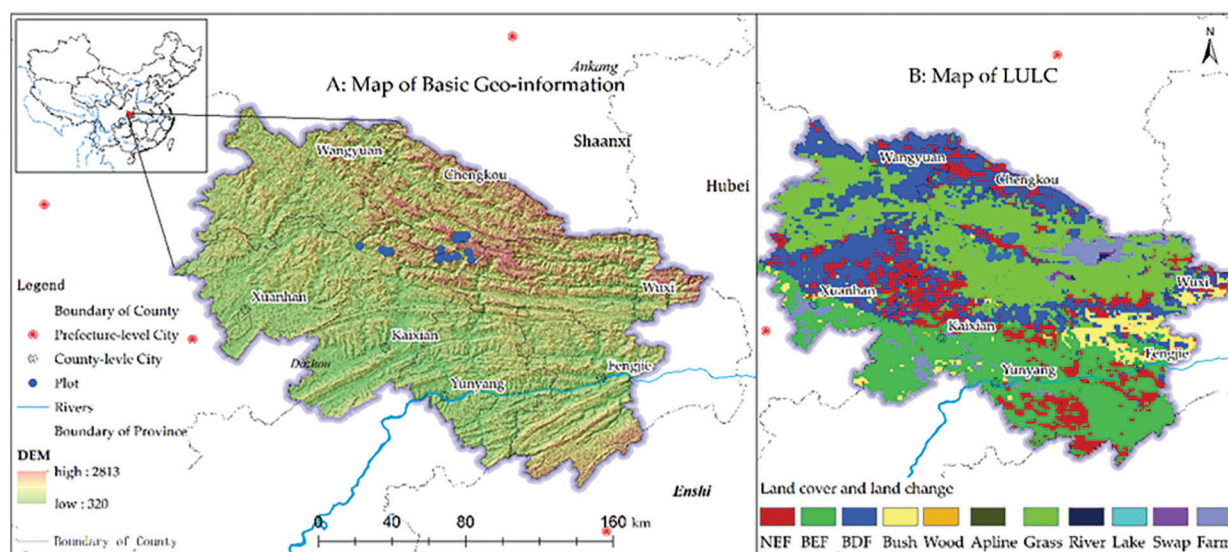


Figure 2: Basic geo-information and vegetation map (GLC2000) in the study area

The data for site distribution for *T. sutchuenensis* were derived from China's second plant resources survey, the China Digital Plant Specimen Museum (<http://www.cvh.ac.cn>) and published literature [7]. The data from Fig. 2 are displayed using the WGS84 coordinates system. The climatic data (<http://www.worldclim.org>) consisted of 19 independent data that corresponded to 19 climatic factors, which were all raster data structures with a resolution of one square kilometer. The value of each grid corresponded to the average value of each climate factor during the period from 1970 to 2000. The topographic factor data contains elevation, slope and aspect. The elevation data was obtained from ASTER GDEM V2 (<http://www.gscloud.cn>) and had a spatial resolution of 30 m. ArcGIS10.2 software (ESRI, Redlands, CA, USA) was previously used to produce slope and slope direction data based on elevation data. The soil data used in this study were collected from the National Earth System Science Data Center (<http://www.geodata.cn/>). To consider the relationship between the growth of *T. sutchuenensis* and surrounding plants, we utilized the Normalized Difference Vegetation Index (NDVI), a data product of MODIS13Q1 in the USA. The MODIS13Q1 data were sourced from NASA with a spatial resolution of 250 m and a temporal resolution of 16 days. This data was pre-processed, so that the final NDVI data could represent the optimal growth status of vegetation. All the climatic, topographic factor, elevation, and soil data and NDVI that were used are shown in Table 1. RF was used to screen these factors, and ArcGIS10.2 was

used to uniformly process the screened factor data into the same coordinate system, resolution and boundary range before using the MaxEnt model.

Table 1: Environmental variables used in the study

Factor	Environmental variables
BIO1	Annual mean temperature
BIO2	Mean diurnal range (Mean of monthly (max temp–min temp))
BIO3	Isothermality (BIO2/BIO7) ($\times 100$)
BIO4	Temperature seasonality (standard deviation $\times 100$)
BIO5	Max temperature of warmest month
BIO6	Min temperature of coldest month
BIO7	Temperature annual range (BIO5-BIO6)
BIO8	Mean temperature of wettest quarter
BIO9	Mean temperature of driest quarter
BIO10	Mean temperature of warmest quarter
BIO11	Mean temperature of coldest quarter
BIO12	Annual precipitation
BIO13	Precipitation of wettest month
BIO14	Precipitation of driest month
BIO15	Precipitation seasonality (Coefficient of variation)
BIO16	Precipitation of wettest quarter
BIO17	Precipitation of driest quarter
BIO18	Precipitation of warmest quarter
BIO19	Precipitation of coldest quarter
H	Elevation
Slope	
Aspect	
Soil	Spatial distribution data of soil types
NDVI	Normalized difference vegetation index

2.2 Integrated Statistical Modeling

RF and MaxEnt are machine learning models that are used to filter impact factors and predict species distribution, respectively [31–33]. By combining the two models, we initially used the RF model to screen impact factors with a maximum weight for the sample and environmental data. The patterns of distribution of *T. sutchuenensis* were then predicted by inputting the data from RF into the MaxEnt model.

RF is an integrated learning method based on decision trees. This method was adapted to further screen core explanatory variables by calculating two indicators, representing the ranking of a factor's relative importance and contribution to model fitness, respectively. The first screening indicator, the mean decrease impurity (MDI, IncNodePurity), quantifies the change of Gini impurity (information gain)

generally used to measure information uncertainty or the degree of the confusion system. The second screening indicator was the mean decrease accuracy (MDA, %IncMSE). The main idea is to disrupt the order of the eigenvalues of each feature and then measure the effect of change in order on the accuracy of model. For a candidate variable, higher MDI and MDA values are more important to improve the performance of model. Before RF is performed, the data are frame data in CSV format, which are composed of presence/absent samples of *Thuja* as the dependent variable and 24 environmental factors that influence selection (i.e., Bioclimatic variables, those are BIO1-BIO19, slope, aspect, NDVI, and soil) as independent variables. Environmental variables with higher rankings of MDI and MDA were retained and used in a next step regression model. All the RF modeling was undertaken using the “Random Forest” package in R [34–37]. The value of the sample data set and corresponding impact factors were input into the RF model in the form of a data frame based on the grammatical rules of R language, thus, obtaining MDI and MDA.

Maximum entropy theory (MaxEnt; [38]) is based on partial knowledge and matched unknown probability distribution with the most uncertain or random inference. An integrated MaxEnt algorithm based on the entropy theory and statistical method was later proposed by Phillips et al. [39]. This algorithm, using only existing data from the n-dimensional environment variable space to simulate the geographical distribution of species, enables the target distribution to be estimated by finding the distribution of maximum entropy [31,39–41]. In this study maximum entropy was expressed as follows:

$$H(\varphi) = \sum_{i=1}^n p_i \log \frac{1}{p_i} = - \sum_{i=1}^n p_i \log p_i \quad (1)$$

where, φ denotes a hypothesized random variable, which has n different potential results X_1, X_2, \dots, X_n , and its corresponding occurrence probability is designated p_1, p_2, \dots, p_n , and $H(\varphi)$ denotes the calculation result of (1), namely entropy.

The MaxEnt model has many features: (1) it only requires data and environmental information from the study area [40,42,43]; (2) it can take advantage of both continuous and categorical data; (3) it proposes the deterministic algorithm that guarantees convergence to optimal (maximum entropy) probability distribution; (4) the output result is a probability value, and the patterns of distribution of species can be interpreted based on the probability value [31,39], and (5) under small sample size conditions, the MaxEnt model is the best at predicting [44]. Wisz et al. [44] not only considered the MaxEnt model to be much less sensitive to sample size, but they also showed that this model has the best prediction ability across all the sample sizes, thus, rendering this model ideal for modeling data with a sample size that spans 10–30 samples.

Based on our understanding of these MaxEnt features, we introduced 30 samples into the MaxEnt model to predict the potential distribution area of *T. sutchuenensis* and quantitatively reveal its habitat characteristics. Using this method, we also believe that suitable predictions should be obtained for local regions.

3 Results

3.1 Environmental Factors

Based on RF modeling, the results for %IncMSE and IncNodePurity evaluation indicators were obtained (Fig. 3). The contribution of indicators was evaluated based on results from the perspective of effect of changes in order on the accuracy of model (%IncMSE) and the degree of the confusion system (IncNodePurity). By comparing the ranking of indicators of the two results [45], we found that evaluation results for the top nine factors were consistent. BIO10, BIO11, BIO1 and BIO9 recorded higher rankings, thus, indicating that these environmental impact factors are indispensable (red). BIO3, BIO5, BIO6, BIO8, and H were also found to have an important effect on the distribution of species (orange). These

factors were all derived from continuous data, and the only categorical factor was soil. A literature review showed that many researchers claimed to consider limiting factors of soil when assessing the growth range of species [26,46,47]. Therefore, the soil was retained as the impact factor of the prediction model. Based on these results, 10 environmental factors were selected (BIO1, BIO3, BIO5, BIO6, BIO8, BIO9, BIO10, BIO11, H and Soil; Fig. 4).

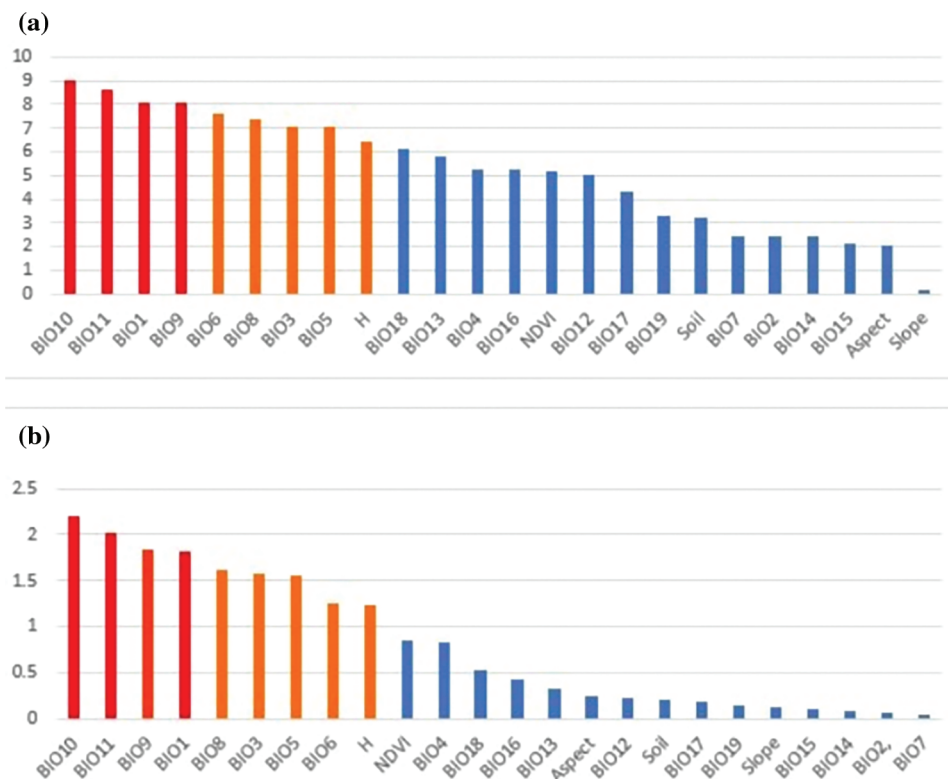


Figure 3: Rankings of variable importance using the evaluation indicators of (a) %IncMSE and (b) IncNodePurity calculated using random forest

3.2 Model Evaluation

The fitness performance of MaxEnt was evaluated using the Receiver Operating Characteristic Curve (ROC). The ROC curve was drawn according to the sensitivity of the y axis and 1–specificity of the x axis. The main advantage of ROC analysis is that the area under the ROC curve (AUC) provides a single measurement of model performance, which is not related to any specific threshold selection [31]. Therefore, ROC is an excellent parameter index to measure the performance of a model [48,49]. Typically, the AUC value is in the range of 0.5–1.0 and the performance of the corresponding model can be divided into five levels: failing (0.5–0.6), poor (0.6–0.7), fair (0.7–0.8), good (0.8–0.9) and excellent (0.9–1) [30,50]. The ROC results indicated that the AUC values for the training data set (80% from samples) and the test data sets (20% from samples) were 0.960 and 0.963, respectively. These results indicate that the model can be classified as satisfactory (Fig. 5).

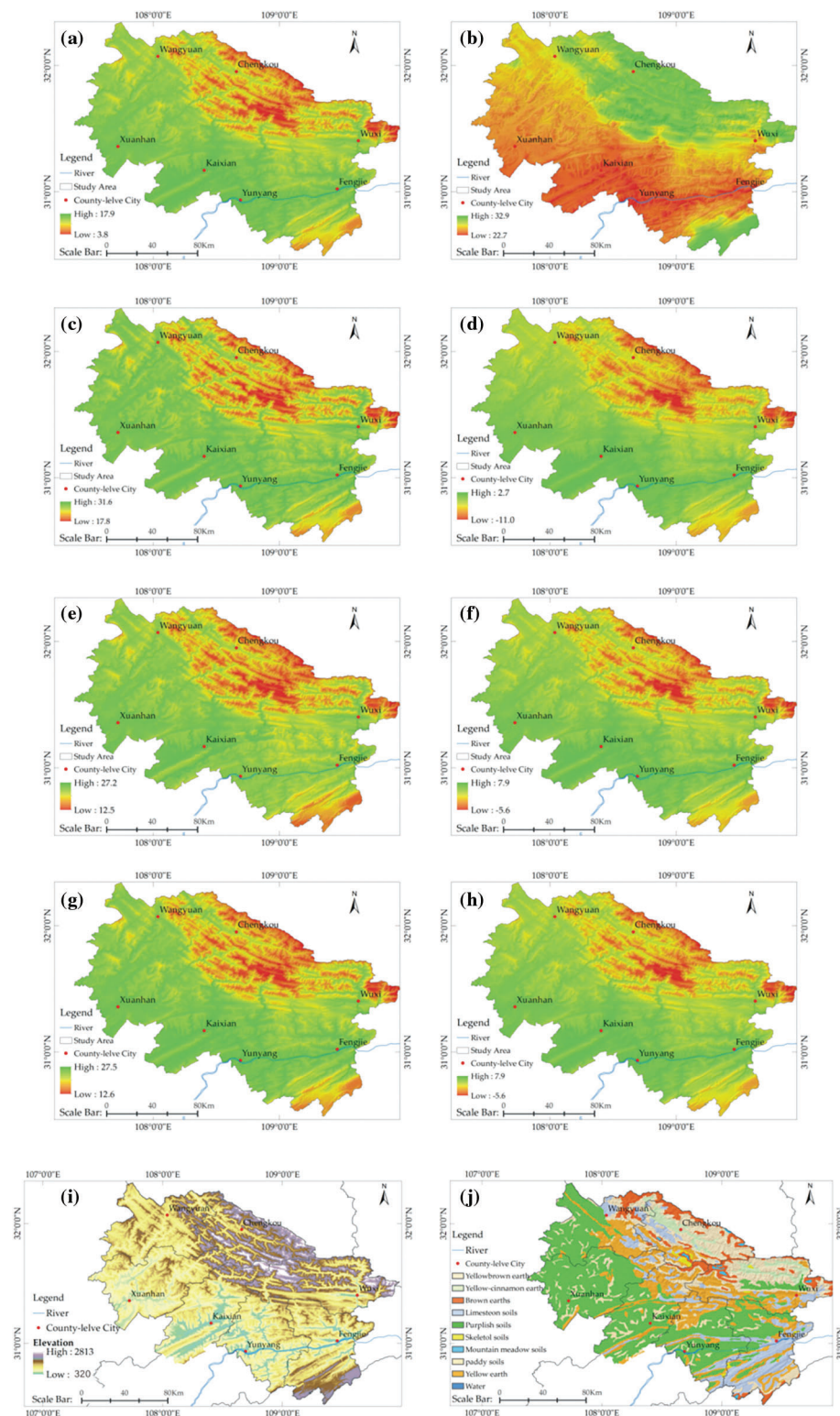


Figure 4: Spatial distribution of the selected environmental factors for predicting/ modeling *Thuja sutchuenensis*: (a) BIO1, (b) BIO3, (c) BIO5, (d) BIO6, (e) BIO8, (f) BIO9, (g) BIO10, (h) BIO11, (i) H, and (j) Soil

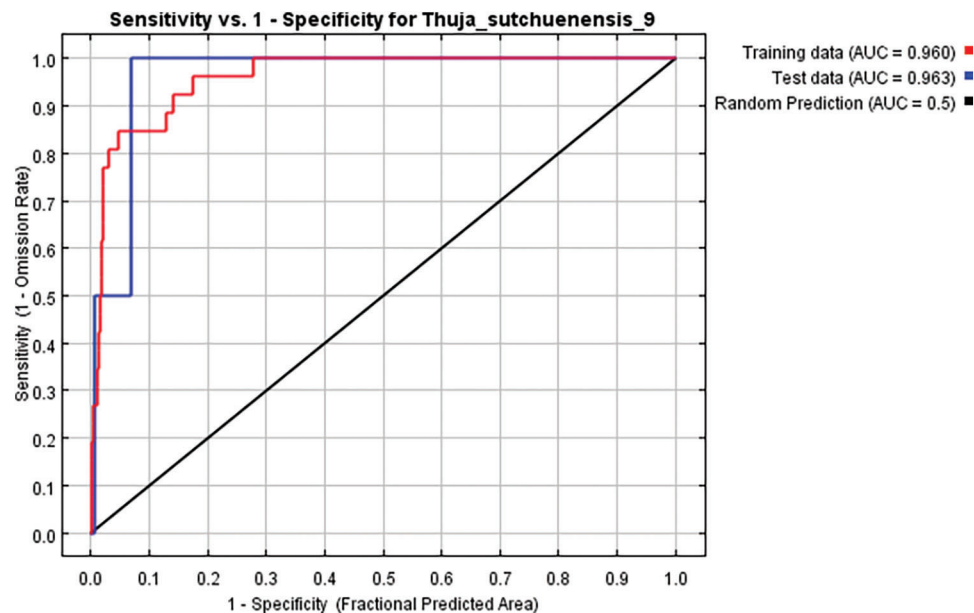


Figure 5: AUC curves for developing *Thuja sutchuenensis* Franch habitat suitability models. AUC, area under the curve

3.3 Nonlinear Relationships between Environmental Factors and *T. sutchuenensis*

Fig. 6 shows the quantitative relationship of probability between the environmental variables and the logic existence, which provides a deeper understanding of the suitability of the environment. The habitat characteristics of *T. sutchuenensis* were analyzed using the average results from 10 MaxEnt runs (Fig. 6).

The temperature results indicated that the annual mean temperature (BIO1) was 7°C–9.8°C; the mean temperature of the warmest quarter (BIO10) was between 15.8°C and 18.2°C, and the mean temperature of the coldest quarter (BIO11) was –2°C to 0.5°C. Since the possibility of *T. sutchuenensis* growing in areas with these temperature conditions was very high, these areas were designated as distribution areas with a high level of suitability. Under extreme temperature conditions, the areas with the warmest month (BIO5; 21°C–23.5°C) and the coldest month (BIO6; –8°C to –5°C) were the most suitable for the growth of *T. sutchuenensis*. When the temperature was >26°C or <–10°C, there was little possibility of the presence of *T. sutchuenensis*.

Isothermality (BIO3) is the ratio of mean diurnal range to annual temperature range, namely BIO2/BIO7. Our results indicated that a highly suitable distribution area of *T. sutchuenensis* was present when $32 < \text{BIO3} < 32.7$. A decrease in BIO3 suggests that the gap between the mean diurnal range and temperature annual range was gradually widening. Areas with $\text{BIO3} < 29$ were noted as having a very low probability of the presence of *T. sutchuenensis*.

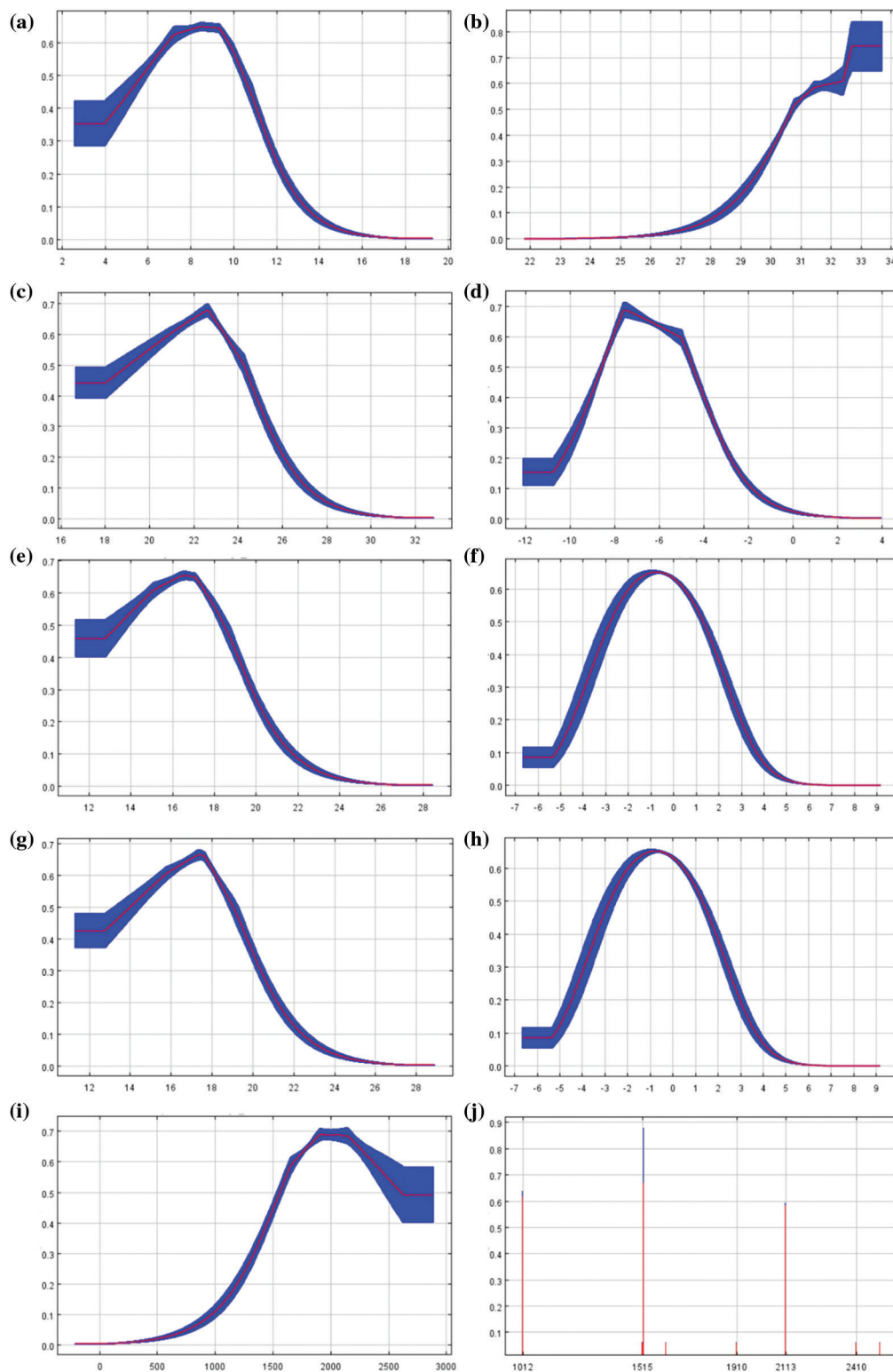


Figure 6: Response curves of 10 environmental variables of the *T. sutchuenensis* habitat distribution model. (a) BIO1, (b) BIO3, (c) BIO5, (d) BIO6, (e) BIO8, (f) BIO9, (g) BIO10, (h) BIO11, (i) H, and (j) Soil

The mean temperature of the wettest (BIO8) and driest (BIO9) months reflect precipitation and heat synchronization status. When BIO8 was between 15°C and 17.5°C, it corresponded to the highest volume of precipitation in the highly suitable distribution area of *T. sutchuenensis*. When BIO9 ranged from −2°C to 0.5°C, precipitation in the highly suitable distribution area of *T. sutchuenensis* recorded minimum levels. By comparing the response curves of BIO8 with BIO10 and BIO9 and BIO11, the curve morphology was similar, and the corresponding peaks were basically the same.

As a topographic factor, altitude response curves indicated that areas with the highest frequency of *T. sutchuenensis* had an altitude range between 1,600 and 2,300 m. However, when the altitude was <1,200 m, the probability of the presence of *T. sutchuenensis* was very small.

The soil curve results indicated that our samples predominantly occurred in areas with yellow-brown soil (code 1012), limestone soil (1515) and yellow earth (2113).

3.4 Potential Distribution Maps for *T. sutchuenensis*

With reference to the classification criteria proposed by Yang et al. [25], we divided the probability (0–1) of the potential species distribution into four levels: high potential (>0.6), good potential (0.4–0.6), moderate potential (0.2–0.4) and lowest potential (<0.2). Areas in the study site with good and high potential had an area of about 1,297.5138 and 873.6066 square kilometers, respectively. According to the spatial location difference, we divided the suitable distribution area into four parts (A, B, C and D; Fig. 7). The results of the MaxEnt calculation indicate that the area that has a high potential for *T. sutchuenensis* presence was primarily located in the middle section of Daba Mountain (A). The distribution area crosses the Northeastern Corner of Xuanhan County, Southern Chengkou County, Northern Kai County, and extends to Western Wuxi County. The northern part of Chengkou County belongs to the northern part of Daba Mountain, an area that has a possibility of the occurrence of *T. sutchuenensis* (B). The results of MaxEnt calculation also revealed that a suitable distribution area was present in the eastern area of Wuxi County and the adjacent areas of Hubei Province as shown in C. An important finding appears in Fengjie County and the mountainous area bordering Hubei Province. Here, modeling results indicated that area D belonged to a habitat with a high potential, providing an area of land that merits study in the future.

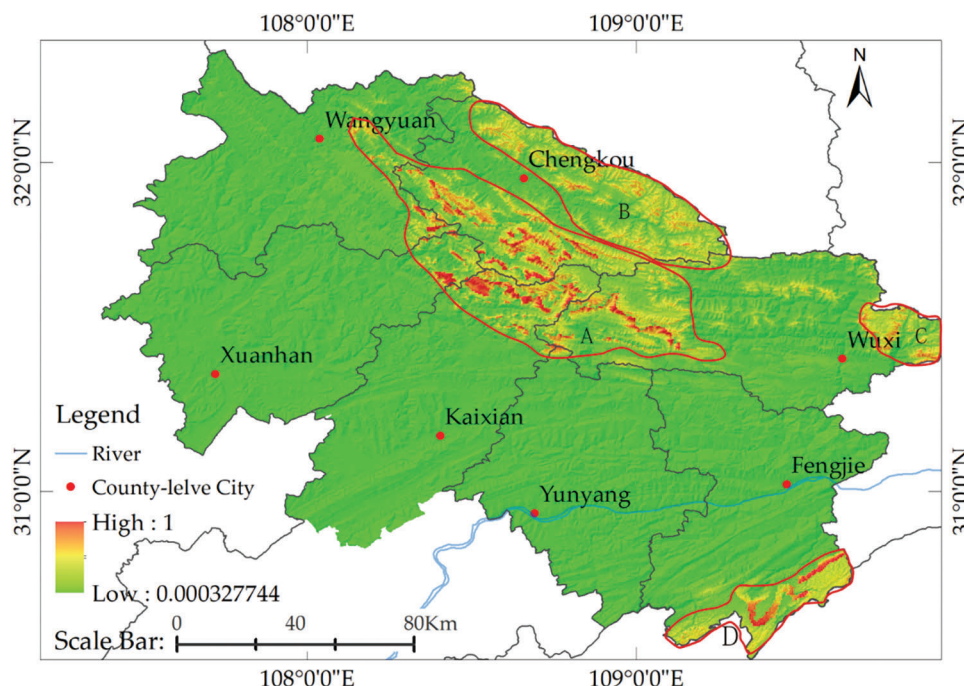


Figure 7: Prediction of potential distribution areas of *T. sutchuenensis*

4 Discussion

The results from our study indicated that 10 environmental factors affect the distribution of *T. sutchuenensis*, and eight were climatic factors. These findings are consistent with those of Qin et al. [14], with the climate serving as the main impact factor that determines the distribution of *T. sutchuenensis*. Our results also indicated that elevation and soil environmental factors play an important role in its distribution.

Regarding meteorological factors, differences between the research on Qin et al. [14] and findings from this study are due to the screened climate factors in our study that were mainly concentrated on the temperature indicators (BIO1, BIO3, BIO5, BIO6, BIO8, BIO9, BIO10 and BIO11). These results indicate that temperature is the main factor that affects the growth of *T. sutchuenensis*. Zhu et al. [9] examined the variation in characteristics of leaves in the seasonal natural temperature reduction process in autumn and winter using *T. sutchuenensis* seedlings. Their results showed that when the daily mean temperature decreased to -7.77°C , and the daily minimum temperature reached -13.96°C , a curling phenomenon occurred in the *T. sutchuenensis* leaves, recording obvious symptoms of injury due to freezing [9]. An analysis of *T. sutchuenensis* cones and seeds during the cold period in spring indicated that cone and seed abortion was particularly serious [10]. These results confirm that *T. sutchuenensis* is sensitive to changes in temperature, highlighting that, in combination with the response curve, this tree prefers a cool environment.

In terms of geographical environment factors, it is important to consider the restrictive effect of soil factors in predicting the suitability distribution range of plants [30,47]. Field studies have shown that *T. sutchuenensis* is mainly distributed in mountainous limestone areas enriched with calcium, and the corresponding soil type of samples were yellow-brown earths, limestone soils and yellow earths, which can all develop from limestone weathering. The results of chemical and physical analyses for the soil samples taken near *T. sutchuenensis* trees indicate the soil to be slightly alkaline, and physical indices, such as soil density and soil porosity, were generally in good condition [2].

Elevation is also an important factor that determines the distribution pattern of plants on a regional scale [51]. The response curve showed that an elevation zone between 1,200 and 2,600 m was suitable (probability > 0.2) for the growth of *T. sutchuenensis*, indicating that temperature, wind speed, precipitation, humidity, solar radiation and other conditions in the altitude range comprehensively form a special suitable environment for survival.

The factor analysis described above showed that the geological environment in Daba Mountain region is peculiar. Qin et al. [52] and Wang et al. [53] both claimed in published studies that *T. sutchuenensis* is a unique endangered species that is only found in Southwest China.

An important prerequisite for factor analysis and potential distribution prediction is whether the screening of impact factors is reasonable [54]. This is the first time that the RF model was adopted to quantitatively assess the habitat variables of *T. sutchuenensis*. Bootstrap samples were used by the RF approach to establish multiple classification of trees. The purpose was to determine the optimal model while analyzing the relationship between the independent and dependent variables to determine the weight of each factor. This modeling method has the following advantages: (a) it has no over fitness; (b) since the value is an average of multiple trees, it has a lower standard deviation and a smaller variance; (c) since it is based on a sampling method to construct the forest with return, and the diversity of trees increases, it has a lower correlation between individual trees, and (d) a higher prediction performance was obtained via an error estimate based on out-of-bag error (OOB) data [55]. Finally, the screened factors in our study could reasonably explain that the habitat characteristics of *T. sutchuenensis* are consistent with the biological features of *T. sutchuenensis* having “good” prediction results.

The results from our study also indicate potential new areas for the presence and natural distribution of *T. sutchuenensis*. By focusing on seven counties in Sichuan and Chongqing, and undertaking regional division to predict the results (Fig. 7), we identified new suitable distribution areas for *T. sutchuenensis*. Area A, which has the most complete population and abundant plants of *T. sutchuenensis*, is located in the central area of Daba Mountain. The mountain range runs from NW to SE, and the prevailing winds are from the southwest in the summer and northwest in the winter. This area is dominated by winds from the southwest in spring and autumn with some deflections in the wind direction.

Qinling Mountain ranges to the North of Daba Mountain and blocks the cold northwestern winds in the winter to some degree. This area has a unique climatic condition in the four seasons. Additionally, limestone soil developed from carbonate rocks, the specific altitude, complex terrain conditions and the meteorological conditions highlighted above resulted in this area becoming a refuge area to some species during the Quaternary ice age. This area retains many rare and endangered species, including *T. sutchuenensis*. Area B is located to the north of Area A, and is part of the middle section of Daba Mountain. Although the soil conditions and altitude in Area B are similar to those in Area A, it is located to the south of Qinling Mountain and in the northern part of the middle section of Daba Mountain. Thus, there is a slight difference in climatic conditions. Although the conditions are suitable for the survival of *T. sutchuenensis*, this species has yet to be found in this area. Area C is located in western Daba Mountain in the junction between Daba Mountain and Wushan Mountain. Although modeling results indicate that this area has environmental conditions suitable for the growth of *T. sutchuenensis*, the occurrence of various geological disasters in the area, frequent human activities engaged in infrastructure construction and other factors that interfere with the natural habitat. Thus, it is believed that there are possible difficulties for endangered plants to survive. Area D is mainly located in Fengjie County, Chongqing, on the eastern edge of Sichuan Basin, the front edge of Daba Mountain and the border area of the western Hubei mountainous area, predominantly located in areas with terrain that fluctuate significantly. Compared with area A, the climate of Area D is also affected by the southwest and northwest monsoons. The mean annual temperature in this area above 1,400 m is lower than 10.8°C, and the soils are mainly yellow and yellow-brown earth. The results of the model show that there is a very high possibility of the presence of wild *T. sutchuenensis* in this area. Although there are no reports of new discoveries, we believe that Area D could be used as an important area of *ex situ* conservation and the reintroduction of *T. sutchuenensis*, which merits further study. In the rest of the study area, there is a low possibility of finding *T. sutchuenensis* owing to the low altitude and high river density and a high level of urbanization and human interference that result from continuous infrastructure construction over the last four decades.

The results of prediction are consistent with the protection measures that have been implemented. In Chongqing, the local government has constructed two national natural reserves, Daba Mountain Natural Reserve and Xuebao Mountain Natural Reserve, in Chengkou County and Kai County, respectively. The department in charge has conducted missions to protect *T. sutchuenensis* from four perspectives: 1) the forest is closed to the public and associated tree species are screened to restore the ecosystem of *T. sutchuenensis*; 2) an artificial propagation experimental base had been constructed to study domestication and breeding techniques; 3) an *ex situ* conservation is being utilized to rebuild the *T. sutchuenensis* population, and 4) germplasm resources are being collected, organized, and preserved. The Xuanhan County Forestry Bureau in Sichuan has convened ecology experts to discuss and develop plans to protect *T. sutchuenensis*. These plans have been implemented, and the construction of a protection station and the artificial propagation experimental base have been completed. In regard to the potential distribution area newly found by the model, it is suggested to conduct field studies as soon as possible to launch the experiment of introducing *T. sutchuenensis* and provide a practical basis for the *ex situ* conservation of this rare and precious tree.

Finally, there are still some limitations in our study. Owing to limitations associated with the method of acquiring data, the 24 candidate factors did not fully represent the growth environment of *T. sutchuenensis*. In subsequent studies, several factors, such as light, wind, animal and human interferences, and the invasion of alien species need to be considered. We forecast that the border area of Chongqing and Hubei is a suitable distribution area of *T. sutchuenensis*. However, the field study could not be substituted by results that predict whether *T. sutchuenensis* is present in this area, and whether it could be used for the *ex situ* conservation of *T. sutchuenensis*. These few scientific difficulties require proper study in the future. In addition, our study on impact factor analysis aimed at environmental factors without taking into account the local temporal and spatial nonstationarity [56].

5 Conclusions

The distribution of *T. sutchuenensis* in the Sichuan-Chongqing region in China was examined by constructing a flexible and advanced integrated model Random Forest-MaxEnt. Our results indicate that the most important meteorological impact factors on the distribution of the studied species were annual mean temperature, mean temperature of the driest quarter, mean temperature of the warmest quarter, and mean temperature of the coldest quarter, and the role of altitude and soil could not be neglected. Our findings show that *T. sutchuenensis* mainly grows in limestone mountainous areas (1,600–2,300 m) with cool environments (mean annual temperature of 7°C to 9.8°C). Suitable environments are synchronized between precipitation and temperature.

Although the results from our model inferred an area of high distribution, field studies recorded a narrow distribution area for *T. sutchuenensis* owing to poor habitat conditions and a significant impact from anthropogenic activities. The population's age structure has a concave phenomenon, and *T. sutchuenensis* is still in an extremely endangered state. To change the current dilemma faced by *T. sutchuenensis*, laboratory studies, including habitat simulation, cutting propagation and tissue culture, have been undertaken based on the current understanding of its habitats. The geographical distribution range and habitat changes of *T. sutchuenensis* will continue to be monitored particularly in the highly suitable distribution areas identified in our study. To establish *in situ* protection and *ex situ* conservation of *T. sutchuenensis*, it is also important that relevant government departments are strongly urged to establish corresponding nature reserves, develop protection measures, and avoid interference from humans and other factors. These factors are important for maintaining the balance between *T. sutchuenensis* populations and preserving species diversity.

Limitations

Owing to the shortage of funding, the team could not carry out field verification in the border of Chongqing and Hubei. This is also an important question to solve on of subsequent studies for this team, i.e., to explore new samples of *T. sutchuenensis*, thus improving the understanding about habitat and population characteristics of *T. sutchuenensis*. This will also contribute to provide a more comprehensive theoretical basis for the formulation of protection measures of *T. sutchuenensis*.

Author Contributions: Study concept and design: Peihao Peng. Acquisition of data: Peihao Peng. Methodology: Haijun Wang. Analysis and interpretation of data: Liang Xie. Drafting of the manuscript: Liang Xie. Visualization: Haijun Wang. Critical revision of the manuscript for important intellectual content: Shengbin Chen.

Acknowledgement: We would like to thank Professor Peihao Peng, Senior Engineer Tinglun Long, Dr. Xian'an Liu, and Senior Engineer Xiaodong Wang. Their detailed and accurate work in the field has provided real and reliable data for this article.

Data Availability: The climate data, DEM data, NDVI data, and Soil data were all collected from an online sharing platform (<http://www.worldclim.org>, <http://www.gscloud.cn/>, <http://www.geodata.cn/>,

nasa.gov/). The adopted sample data of *T. sutchuenensis* were from the corresponding author, the China Digital Plant Specimen Museum (<http://www.cvh.ac.cn>), and published Chinese articles.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Song, J., Zhang, Z., Lv, R. (2021). Study on SFE-CO₂ and chemical composition of *Thuja* essential oil. *Agronomy Journal*, 1–12. DOI 10.1002/agj2.20923.
2. Ma, F., Qin, A., Guo, Q., Jian, Z., Pei, S. et al. (2017). Geographical distribution and habitat characteristics of critically endangered species, *Thuja sutchuenensis*. *Chinese Journal of Ecology*, 36(7), 1777–1784. DOI 10.13292/j.1000-4890.201707.034.
3. Xiang, Q., Fajon, A., Li, Z., Fu, L., Liu, Z. (2002). *Thuja sutchuenensis*: A rediscovered species of the cupressaceae. *Botanical Journal of the Linnean Society*, 139, 305–310. DOI 10.1046/j.1095-8339.2002.00055.x.
4. Fu, L. K., Jin, J. M. (1992). *China plant red data book—rare and endangered plants*, vol. 1. Beijing, China: Science Press.
5. Farjon, A., Page, C. N. (1999). *Conifers. Status survey and conservation action plan*. IUCN-SSC Conifer Specialist Group. Gland, Switzerland and Cambridge, UK: IUCN.
6. Jin, J. Q., Ren, F. M., Xia, Y., Liu, Z. Y., Chen, Y. H. et al. (2020). Research on reproductive phenology, pollination, and embryonic development of *Thuja sutchuenensis* Franch, a plant species with extremely small populations. *Plant Science Journal*, 38(5), 696–706. DOI 10.11913/PSJ.2095-0837.2020.50696.
7. Guo, Q., Wang, X., Ba, G., Wan, Q. (2007). Interspecific relationships of dominant tree species in *Thuja sutchuenensis* community. *Chinese Journal of Ecology*, 26 (12), 1911–1917.
8. Liu, J., Shi, S., Chang, E., Yang, W., Jiang, Z. (2013). Genetic diversity of the critically endangered *Thuja sutchuenensis* revealed by ISSR markers and the implications for conservation. *International Journal of Molecular Sciences*, 14, 14860–14871. DOI 10.3390/ijms140714860.
9. Zhu, L., Guo, Q., Jin, J., Liu, J., Li, Z. et al. (2013). Physiological and biochemical responses of *Thuja sutchuenensis* and *Platycladus orientalis* seedlings to natural cooling. *Forest Research*, 26(2), 220–226. DOI 10.3969/j.issn.1001-1498.2013.02.014.
10. Zhu, L., Guo, Q., Zhu, N., Qin, A., Xu, G. et al. (2014). Study on the cones and seeds biological characteristic of a critically endangered species, *Thuja sutchuenensis*, in the world. *Seed*, 33, 56–63. DOI 10.3969/j.issn.1001-4705.2014.07.014.
11. Guo, Q., Qin, A., Ma, F., Jian, Z., Pei, S. (2015). Research progress on *Thuja sutchuenensis*: A critically endangered species in the world. *World Forestry Research*, 28(6), 18–22. DOI 10.13348/j.cnki.sjlyyj.2015.0026.y.
12. Tang, C. Q., Yang, Y., Ohsawa, M., Momohara, A., Yi, S. et al. (2015). Community structure and survival of tertiary relict *Thuja sutchuenensis* (Cupressaceae) in the subtropical daba mountains, Southwestern China. *PLoS One*, 10(4), 1–19. DOI 10.1371/journal.pone.0125307.
13. Wang, X., Zhang, H., Li, Z., Zhang, S., Wang, G. et al. (2016). Community structure and population regeneration of an endangered plant, *Thuja sutchuenensis*. *Journal of Beijing Forestry University*, 38, 28–37. DOI 10.13332/j.1000—1522.20160028.
14. Qin, A., Liu, B., Guo, Q., Bussmann, R. W., Ma, F. et al. (2017). Maxent modeling for predicting impacts of climate change on the potential distribution of *Thuja sutchuenensis* Franch., an extremely endangered conifer from Southwestern China. *Global Ecology and Conservation*, 10, 139–146. DOI 10.1016/j.gecco.2017.02.004.
15. Wang, X., Deng, H., Huang, Q., Zhang, S., Rao, D. (2017). Flora analysis and vertical structure optimization of the community of *Thuja sutchuenensis* Franch. *Acta Botanica Boreali-Occidentalia Sinica*, 37(1), 0181–0190. DOI 0.7606/j.issn.1000-4025.2017.01.0181.

16. Jian, Z., Ma, F., Guo, Q., Pei, S., Qin, A. et al. (2017). Responses of survival and growth of *Thuja sutchuenensis* reintroduction seedlings to altitude gradient. *Scientia Silvae Sinicae*, 53(11), 1–11. DOI 10.11707/j.1001-7488.20171101.
17. Jin, J., Guo, Q., Han, S., Zhu, L., Liu, Y. et al. (2018). *In vitro* propagation of ‘Lazarus’ species *Thuja sutchuenensis* Franch. *Propagation of Ornamental Plants*, 18(3), 77–86.
18. Guo, X., Wang, M., Wu, J., Wu, G., Zhang, X. et al. (2019). Chemical constituents of the trunks and roots of *Thuja sutchuenensis*. *Fitoterapia*, 134, 264–269. DOI 10.1016/j.fitote.2019.02.025.
19. Qin, A., Guo, Q., Ma, F., Jian, Z., Zheng, X. (2020). Effects of temperature, light and water conditions on seed germination of *Thuja sutchuenensis* Franch. *Seed*, 39(2), 15–20. DOI 10.16590/j.cnki.1001-4705.2020.02.015.
20. Wang, M., Zhao, L., Chen, K., Shang, Y., Wu, J. et al. (2020). Antibacterial sesquiterpenes from the stems and roots of *Thuja sutchuenensis*. *Bioorganic Chemistry*, 96. DOI 10.1016/j.bioorg.2020.103645.
21. Zhou, X., Deng, H., Tang, S. (2020). *In vitro* rapid propagation of endemic, rare and endangered *Thuja sutchuenensis* in China. *Plant Physiology Journal*, 56(5), 990–996.
22. Yu, T., Huang, B. H., Zhang, Y., Liao, P. C., Li, J. Q. (2020). Chloroplast genome of an extremely endangered conifer *Thuja sutchuenensis* Franch.: Gene organization, comparative and phylogenetic analysis. *Physiology and Molecular Biology of Plants*, 26(3), 409–418. DOI 10.1007/s12298-019-00736-7.
23. Zhao, Z., Zhao, C., Deng, S., Shen, G., Xie, Z. et al. (2020). Community structure and dynamics of a remnant forest dominated by *Thuja sutchuenensis* after deforestation. *Biodiversity Science*, 28(3), 333–339. DOI 10.17520/biods.2019235.
24. Bannister, P., Neuner, G. (2001). Frost resistance and the distribution of conifers. In: Bigras, F. J., Colombo, S. J. (Eds.) *Conifer cold hardiness*, pp. 3–22. Dordrecht: Kluwer Academic Publishers.
25. Yang, X. Q., Kushwaha, S. P. S., Saran, S., Xu, J., Roy, P. S. (2013). Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in lesser himalayan foothills. *Ecological Engineering*, 51, 83–87. DOI 10.1016/j.ecoleng.2012.12.004.
26. Ray, D., Behera, M. D., Jacob, J. (2015). Predicting the distribution of rubber trees (*Hevea brasiliensis*) through ecological niche modelling with climate, soil, topography and socioeconomic factors. *Ecological Research*, 31(1), 75–91. DOI 10.1007/s11284-015-1318-7.
27. Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., Kadmon, R. (2007). A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, 13, 397–405. DOI 10.1111/j.1472-4642.2007.00346.x.
28. Hengl, T., Sierdsema, H., Radović, A., Dilo, A. (2009). Spatial prediction of species’ distributions from occurrence-only records: Combining point pattern analysis, ENFA and regression-kriging. *Ecological Modelling*, 220(24), 3499–3511. DOI 10.1016/j.ecolmodel.2009.06.038.
29. Farashi, A., Kaboli, M., Karami, M. (2013). Predicting range expansion of invasive raccoons in Northern Iran using ENFA model at two different scales. *Ecological Informatics*, 15, 96–102. DOI 10.1016/j.ecoinf.2013.01.001.
30. Yi, Y., Cheng, X., Yang, Z., Zhang, S. (2016). Maxent modeling for predicting the potential distribution of endangered medicinal plant (*H. riparia* Lour) in Yunnan, China. *Ecological Engineering*, 92, 260–269. DOI 10.1016/j.ecoleng.2016.04.010.
31. Phillips, S. J., Anderson, R. P., Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259. DOI 10.1016/j.ecolmodel.2005.03.026.
32. Yuan, H., Wei, Y., Wang, X. (2015). Maxent modeling for predicting the potential distribution of *Sanghuang*, an important group of medicinal fungi in China. *Fungal Ecology*, 17, 140–145. DOI 10.1016/j.funeco.2015.06.001.
33. Guo, X., Bian, Z., Wang, S., Wang, Q., Zhang, Y. et al. (2020). Prediction of the spatial distribution of soil arthropods using a random forest model: A case study in Changtu County, Northeast China. *Agriculture, Ecosystems and Environment*, 292, 1–13. DOI 10.1016/j.agee.2020.106818.
34. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. DOI 10.1023/A:1010933404324.
35. Liaw, A., Wiener, M. (2002). Classification and regression by randomForest. *R News*, 3(2), 18.

36. Genuer, R., Michel, V., Eger, E., Thirion, B. (2010). Random forests based feature selection for decoding fMRI data. *Proceedings of COMPSTAT'2010-19th International Conference on Computational Statistics*, pp. 1071–1078. Paris.
37. Genuer, R., Poggi, J. M., Tuleau-Malot, C. (2010). Variable selection using random forests. *Elsevier*, 31(14), 2225–2236. DOI 10.1016/j.patrec.2010.03.014.
38. Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106(4), 620–630. DOI 10.1103/PhysRev.106.620.
39. Phillips, S. J., Dudík, M., Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. *Proceedings of the Twenty-First International Conference on Machine Learning*, pp. 655–662. Princeton.
40. Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E. et al. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17, 43–57. DOI 10.1111/j.1472-4642.2010.00725.x.
41. Liu, Y., Zhou, K., Xia, Q. (2017). A MaxEnt model for mineral prospectivity mapping. *Natural Resources Research*, 27(3), 299–313. DOI 10.1007/s11053-017-9355-2.
42. Davis, A. P., Gole, T. W., Baena, S., Moat, J. (2012). The impact of climate change on indigenous arabica coffee (*Coffea arabica*): Predicting future trends and identifying priorities. *PLoS One*, 7(11), 1–13. DOI 10.1371/journal.pone.0047981.
43. Wang, Y. Q., Ma, J. F., Li, X. Q., Wang, Y. F., Cao, S. et al. (2017). The distribution of *Athetis lepigone* and prediction of its potential distribution based on GARP and MaxEnt. *Journal of Applied Entomology*, 141(6), 431–440. DOI 10.1111/jen.12347.
44. Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H. et al. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14(5), 763–773. DOI 10.1111/j.1472-4642.2008.00482.x.
45. Harley, J. R., Lanphier, K., Kennedy, E., Whitehead, C., Bidlack, A. (2020). Random forest classification to determine environmental drivers and forecast paralytic shellfish toxins in Southeast Alaska with high temporal resolution. *Harmful Algae*, 99, 1–9. DOI 10.1016/j.hal.2020.101918.
46. Coudun, C., Gégout, J. C., Piedallu, C., Rameau, J. C. (2006). Soil nutritional factors improve models of plant species distribution: An illustration with *Acer campestre* (L.) in France. *Journal of Biogeography*, 33(10), 1750–1763. DOI 10.1111/j.1365-2699.2005.01443.x.
47. Ehrenfeld, J. G., Ravit, B., Elgersma, K. (2005). Feedback in the plant-soil system. *Annual Review of Environment and Resources*, 30, 75–115. DOI 10.1146/annurev.energy.30.050504.144212.
48. Yang, Y., Yang, J., Xu, C., Xu, C., Song, C. (2019). Local-scale landslide susceptibility mapping using the B-GeoSVC model. *Landslides*, 16(7), 1301–1312. DOI 10.1007/s10346-019-01174-y.
49. Vanagas, G. (2004). Receiver operating characteristic curves and comparison of cardiac surgery risk stratification systems. *Interactive CardioVascular and Thoracic Surgery*, 3(2), 319–322. DOI 10.1016/j.icvts.2004.01.008.
50. Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293. DOI 10.1126/science.3287615.
51. Adhikari, D., Barik, S. K., Upadhaya, K. (2012). Habitat distribution modelling for reintroduction of *Ilex khasiana* Purk., a critically endangered tree species of northeastern India. *Ecological Engineering*, 40, 37–43. DOI 10.1016/j.ecoleng.2011.12.004.
52. Qin, A., Ding, Y., Jian, Z., Ma, F., Worth, J. R. P. (2021). Low genetic diversity and population differentiation in *Thuja sutchuenensis* Franch., an extremely endangered rediscovered conifer species in Southwestern China. *Global Ecology and Conservation*, 25, 1–9. DOI 10.1016/j.gecco.2020.e01430.
53. Wang, M., Huo, L., Liu, H., Zhao, L., Xu, Z. et al. (2020). *Thujasutchins* n and o, two new compounds from the stems and roots of *Thuja sutchuenensis*. *Natural Product Research*, 10, 1–7. DOI 10.1080/14786419.2020.1836627.
54. Song, C., He, Y., Bo, Y., Wang, J., Ren, Z. et al. (2019). Disease relative risk downscaling model to localize spatial epidemiologic indicators for mapping hand, foot, and mouth disease over China. *Stochastic Environmental Research and Risk Assessment*, 33(10), 1815–1833. DOI 10.1007/s00477-019-01728-5.

55. Rahmati, O., Pourghasemi, H. R., Melesse, A. M. (2016). Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. *Catena*, 137, 360–372. DOI 10.1016/j.catena.2015.10.010.
56. Song, C., Shi, X., Bo, Y., Wang, J., Wang, Y. et al. (2019). Exploring spatiotemporal nonstationary effects of climate factors on hand, foot, and mouth disease using Bayesian spatiotemporally varying coefficients (STVC) model in Sichuan, China. *Science of The Total Environment*, 648, 550–560. DOI 10.1016/j.scitotenv.2018.08.114.