

# Predicting Suitable Habitats of *Melia Azedarach* L. Using Data Mining

**Lei Feng**

Nanjing Forestry University

**Xiangni Tian**

Yunnan University

**Yousry A. El-Kassaby**

University of British Columbia

**Jian Qiu**

Nanjing Forestry University

**Ze Feng**

Kangda College of Nanjing Medical University

**Jiejie Sun**

University of British Columbia

**Guibin Wang**

Nanjing Forestry University

**Tongli Wang** (✉ [tongli.wang@ubc.ca](mailto:tongli.wang@ubc.ca))

University of British Columbia

---

## Research Article

**Keywords:** *Melia azedarach* L., Climate change, Habitat suitability, Data mining models

**Posted Date:** October 27th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1004808/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** *Melia azedarach* L. is a globally distributed tree species of economic importance; however, it is unclear how the species distribution will respond to future climate changes.

**Methods:** We aimed to select the most accurate one among seven data mining models to predict the species suitable contemporary and future habitats. These models include: maximum entropy (MaxEnt), support vector machine (SVM), generalized linear model (GLM), random forest (RF), naive bayesian model (NBM), extreme gradient boosting (XGBoost), and gradient boosting machine (GBM). A total of 906 *M. azedarach* locations were identified, and sixteen climate predictors were used for model building. The models' validity was assessed using three measures (Area Under the Curves (AUC), kappa, and accuracy).

**Results:** We found that the RF provided the most outstanding performance in prediction power and generalization capacity. The top climate factors affecting the species distribution were mean coldest month temperature (MCMT), followed by the number of frost-free days (NFFD), degree-days above 18°C (DD>18), temperature difference between MWMT and MCMT, or continentality (TD), mean annual precipitation (MAP), and degree-days below 18°C (DD<18). We projected that future suitable habitat of this species would increase under both the RCP4.5 and RCP8.5 scenarios for the 2020s, 2050s, and 2080s.

**Conclusion:** Our findings are expected to assist in better understanding the impact of climate change on the species and provide scientific basis for its planting and conservation.

## 1 Introduction

*Melia azedarach* L., Meliaceae, is a fast-growing species with good timber attributes of multiple-use, such as construction and furniture, farm tools, boats, vehicles, and musical instruments manufacturing<sup>1</sup>. The species roots, bark, flowers, and fruits are of high medicinal values<sup>2,3</sup>. Additionally, its fruit and leaf extracts can control numerous agricultural pests and are commonly used as biological pesticides raw materials<sup>4</sup>. The species is an excellent urban greening tree that is resistant to smoke and dust and can absorb many toxic and harmful gases. At present, it is planted in more than 50 countries<sup>5,6</sup>. The species productivity is climate-dependent, and it is expected that climate change will reshape its future suitable habitat<sup>7</sup>.

The intensification of global warming, accompanied by the frequent occurrence of extreme natural disturbances, such as wind storms, droughts, fires, and floods, will undoubtedly impact the global forest ecosystem<sup>8</sup>. Different tree species respond differently to climate change, with positive and negative effects in different areas. For example, climate change is expected to increase the suitable habitats of Mediterranean oaks in the western temperate areas<sup>9</sup> as well as the total suitable habitat for *Cypripedium japonicum*<sup>10</sup>. Conversely, eucalyptus species are expected to face future challenges due to their poor

spread capability<sup>11</sup>, and Persian oak (*Quercus macranthera*) will experience a reduction in its contemporary range and is expected to move to higher altitudes<sup>12</sup>. Consequently, assessing the impact of climate change on the potential distribution of species and formulating sustainable forest management strategies are critical to maintaining forest ecosystems integrity.

With climate change challenges, species distribution models (SDMs) have become essential tools for projecting plants adaptation to a changing climate<sup>11</sup>. At present, a variety of data mining techniques have been applied to model species distribution data. These include: 1) Generalize linear model (GLM), a common regression model first introduced by Austin et al. (1983), to simulate tree species distribution and subsequently was used to predict the spread of Emerald Ash Borer (*Agilus planipennis*) in southern Ontario, Canada<sup>13</sup>; 2) Gradient boosting machine (GBM), a machine-learning technology used to generate predictive models in the form of a collection of weak predictive models<sup>14</sup> and currently is being used to predict invasive plant species distribution, high-resolution, high-precision multi-type vegetation mapping, and species distribution models<sup>15,16</sup>; 3) Random Forest (RF), a machine-learning technology through building a large number of decision trees during the program's training phase<sup>17</sup> and presently is being used to predict invasive species range, classification of tree species based on hyperspectral information, and prediction of stands basal areas and the distribution of plantation forests<sup>18,19</sup>; 4) Support Vector Machine (SVM), a supervised learning model used for data classification and regression analysis and is widely used to classify invasive species and detect the presence of farmland weeds<sup>20,21</sup>; 5) Maximum Entropy (MaxEnt), a machine-learning technology by finding the maximum entropy of the probability distribution of the species through the species distribution and environmental data to estimate and predict future species distribution<sup>22</sup>, and mainly is being used in crop niches, plant diseases and insect pests, and species invasion prediction<sup>23,24</sup>; 6) Extreme Gradient Boosting (XGBoost), an open-source software library algorithm effective in predicting species abundance and identifying critical environmental factors<sup>25</sup>, and is playing an essential role in designing new drugs to treat related diseases; and 7) Naive Bayesian Model (NBM), a series of simple probabilistic classifiers based on Bayes' theorem and independent assumptions between features<sup>26</sup>, and mainly is being applied in forestry for predicting the potential distribution areas of *Taxus chinensis* and identifying plant long non-coding RNA and predicting its functions<sup>27</sup>.

Understanding the potential distribution of *M. azedarach* is of great significance to its cultivation and conservation. Studies conducted on *M. azedarach* were mainly focused on tree and stand productivity, extraction of active ingredients, and pest resistance potential<sup>3,28</sup>. Research on *M. azedarach* potential distribution as affected by climate change is lacking and thus, the present study is aimed at exploring the above-mentioned seven data mining techniques to establish climate-based distribution prediction models and select the best model in predictions of the species future suitable habitat. Our specific objectives were to: 1) compare the prediction accuracy of the seven modeling algorithms and select the one with the best performance; 2) determine the key climatic factors related to the species distribution; 3) develop

current and future species suitable habitat maps highlighting the areas of change; and 4) assess the potential impact of future climate change on the species suitable habitat.

## 2 Material And Methods

### 2.1 Species location data

Here, we used the Chinese presence and absence *M. azedarach* data to establish the prediction models. First, we found 1,432 presence data (data source: Global Biodiversity Information Facility (GBIF), <https://www.gbif.org>, and the Chinese Virtual Herbarium (CVH), <http://www.cvh.ac.cn/>). To avoid redundant sampling, we deleted those sample points with similar longitude and latitude<sup>29</sup>. Then a 0.01° mesh thinning was performed, and the actual distance corresponding to 0.01° was about 1km and only one distribution point was reserved in each grid so that the distance between sample points was more than 1 km<sup>30</sup>. Finally, a total of 906 samples were included for model building (Figure. 1).

### 2.2 Environment variables

We established a GLM model with *M. azedarach* presence-absence data as dependent variables and 16 climatic factors derived from ClimateAP\_v221 software (<http://ClimateAP.net>) as predictors (Table S1)<sup>31</sup>. Correlation analysis showed that there was strong data collinearity among the 16 climate variables. Then we used stepwise regression analysis to eliminate those variables causing the observed multicollinearity<sup>32</sup> and ultimately reduced the climate variable to ten<sup>33</sup>.

### 2.3 Model development and prediction

We used seven models (Generalize Linear Model (GLM), Gradient Boosting Machine (GBM), Random Forest (RF), Support Vector Machine (SVM), Maximum Entropy (MaxEnt), Extreme Gradient Boosting (XGBoost), and Naive Bayesian Model (NBM)) to associate the distribution of *M. azedarach* with climate variables. Firstly, we used the “dismo” package in R to randomly generate 2,000 “pseudo-nonexistent” records in the study area. Models were established with species presence-absence data as the dependent variable and climate variable as the independent variables. In order to evaluate the models’ prediction accuracy, we randomly selected 70% data for training and the remaining 30% data for testing (validation). We used the “caret” package to train and adjust the parameters for all the seven models except Maxent, since it facilitates the process of building, evaluating, as well as selecting features. Then, ten cross-verifications were carried out, and each model was repeated three times. At the same time, the Maxent model was executed using the Maxent version 3.4.4 software in R-package.

### 2.4 Model validation

To assess the performance of the seven predictive models, we compared their area under receiver operating character curve (AUC), Kappa, and accuracy. The AUC is the probability value, with evaluation criteria were: 0.5-0.6 = fails, 0.6-0.7 = poor, 0.7-0.8 = fair, 0.8-0.9 = good, 0.9-1.0 = excellent<sup>34</sup>. Kappa coefficient is an index to measure classification accuracy. The calculation result of kappa is -1 to 1, but

usually, kappa falls between 0 and 1, which can be divided into five groups: 0.0-0.2 means very low consistency, 0.21-0.40 means general consistency, 0.41-0.60 means moderate consistency, 0.61-0.80 means high consistency, 0.81-1 means almost perfect<sup>35</sup>. Accuracy refers to the proximity of measured values to specific values, and it is reported as the average cross-validated accuracy<sup>36</sup>.

## 2.5 Habitat Classification

Appropriate habitat evaluation index values were determined as follows: predicted values of 0-0.2, 0.2-0.4, 0.4-0.6, and >0.6 were deemed unsuitable, low-, medium-, and highly-suitable habitat, respectively<sup>37</sup>.

## 3 Results

### 3.1 Models performance evaluation

Through the cross-validation evaluation of the tested models, Accuracy, Kappa, and AUC values were obtained for the training and testing portions of each model (Figure. 2). All models performed well (AUC>0.8, Kappa>0.5, and Accuracy>0.7). For the training data, the seven models AUC values varied from 0.8825 (NBM) to 1 (RF), Kappa values varied from 0.5754 (GLM) to 0.9988 (RF), and Accuracy values ranged from 0.7887 (NBM) to 0.9995 (RF). While the testing data produced AUC values varied from 0.8471 (NBM) to 0.9039 (RF), Kappa values varied from 0.5335 (SVM) to 0.5896 (MaxEnt), and Accuracy values ranged from 0.7678 (NBM) to 0.8138 (XGBoost). Overall, the three evaluation metrics all indicated that the Random Forest (RF) model provided the best predictive performance and while the Naive Bayesian Model (NBM) was the worst, thus, we selected the RF model to establish *M. azedarach* distribution patterns.

### 3.2 Important climate variables and their response curves in Random Forest (RF)

The top three climate variables contributing to the RF model include MCMT (189.24), NFFD (180.69), and DD>18 (104.77), followed by TD (72.82), MAP (69.43), DD<18 (64.12), DD>5 (56.27), and AHM (54.88); and finally DD<0 (44.01) and PAS (28.54) also played some roles in the determining the potential distribution of *M. azedarach* (Table 1).

Figure. 3 displayed the relationships between the top six climate variables and *M. azedarach* suitability according to the predictions of RF algorithms. The habitat suitable range was between -10 and -28°C for MCMT (Figure. 3a), between 0 and 175 days for NFFD (Figure. 3b), between 0 and 250 for DD>18 (Figure. 3c), between 5 and 21°C for TD (Figure. 3d), between 0 and 480 mm for MAP (Figure. 3e), and between 0 and 1750 for DD<18 (Figure. 3f).

Table 1  
Contributions of the most influencing climate variables to the *M. Azedarach* Random Forest (RF) model

Variable <sup>1</sup>	Units	Overall contribution
MCMT	°C	189.24
NFFD	day	180.69
DD>18	°C-days	104.77
TD	°C	72.82
MAP	mm	69.43
DD<18	°C-days	64.12
DD>5	°C-days	56.27
AHM		54.88
DD<0	°C-days	44.01
PAS	mm	28.54
<sup>1</sup> see Table S1 for variables abbreviations.		

### 3.3 RF model prediction of *M. azedarach* contemporary habitats distribution

The spatial distributions of *M. azedarach* and areas of suitable habitats under current climatic conditions as predicted by the RF algorithm are shown Figure. 4. The overall suitable habitat was mainly distributed between 18 and 40°N (Figure. 4a). These habitats were classified as: 1) highly-suitable habitats (mainly scattered in Shandon (SD), Jiangsu (JS), Shanghai (SH), Zhejiang (ZJ), Guangdong (GD), Hunan (HN), Hainan (HI), South Jiangxi (JX), the junction of the three provinces of Hubei (HB), Anhui (AH), Jiangxi (JX), and the junction of Chongqing (CQ) and Sichuan (SC), covering  $9.3 \times 10^5$  km<sup>2</sup> (9.6%; Figure. 4b); 2) medium-suitable habitats (scattered around the high-suitable habitats, covering  $6.8 \times 10^5$  km<sup>2</sup> (7%; Figure. 4b) and specifically concentrated in eastern Sichuan (SC), northern and western Shandong (SD), and the junction of Hubei (HB) and Hunan (HN)); and 3) low-suitable habitats (slightly larger than the medium-suitable habitats, covering  $7.1 \times 10^5$  km<sup>2</sup> (7.4%; Figure. 4b)), and it is distributed in Yunnan (YN), central Guangxi (GX), eastern and northern Guizhou (GZ), southern Shaanxi (SN), western and northern Henan (HA), and southern Hebei (HE)).

### 3.4 RF model prediction of *M. azedarach* projected suitable habitats future changes

Future projections using the RF model with two different climate scenarios (RCP 8.5 and RCP 4.5) indicated a clear graphical expansion of *M. azedarach* in the future periods with an increasing magnitude over time (Figure. 5). The projected range increase was greatest under RCP 8.5 as compared to RCP 4.5 (Figure. 5). More specifically, the expanded area would increase by  $562.6 \times 10^3 \text{ km}^2$  and  $584.5 \times 10^3 \text{ km}^2$  by 2020s,  $807.4 \times 10^3 \text{ km}^2$  and  $930.3 \times 10^3 \text{ km}^2$  by 2050s, and  $906.1 \times 10^3 \text{ km}^2$  and  $1486.3 \times 10^3 \text{ km}^2$  by 2080s under the RCP4.5 and RCP8.5 scenarios, respectively (Figure. 5g). The main expanded area will be located in Yunnan (YN), Anhui (AH), Henan (HA), Shanxi (SX), Shaanxi (SN), central Guangxi (GX), central Jiangxi (JX), and northern Guizhou (GZ). Interestingly, based on the RCP8.5 climate scenario, Xinjiang (XJ) would see a larger magnitude of area expansion in 2080s (Figure. 5f). Additionally, the species stable range area showed the same change pattern as that of the expanded area (Figure. 5g). The main stable area included Guangdong (GD), Guangxi (GX), Guizhou (GZ), Hunan (HN), Chongqing (CQ), Fujian (FJ), Zhejiang (ZJ), Jiangsu (JS), southwestern Jiangxi (JX), and eastern Sichuan (SC) (Figure. 5a-f). Furthermore, the species area loss exhibited an opposite trend to that of expansion and stable range areas (Figure. 5f) and most of the loss area was mainly distributed in eastern coastal provinces near 30-38°N (e.g., Shandong (SD)) (Figure. 5).

## 4 Discussion

### 4.1 Model performance

Here, we used the Area Under the Curves (AUC), Kappa statistic, and Accuracy to evaluate the performance of seven species range prediction models (Generalized Linear Models (GLM), Gradient Boosting Machine (GBM), Maximum Entropy (MaxEnt), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Naive Bayesian Model (NBM), and Random Forest (RF)) to predict *M. azedarach* contemporary and future ranges under two climate scenarios (RCP 8.5 and RCP 4.5). The results showed that Random Forest (RF) and Extreme Gradient Boosting (XGBoost) were the top-performing models with RF being the best, while Naive Bayesian Model (NBM) and Generalized Linear Models (GLM) were the low-performing with the NBM being the worst. Multiple lines of evidence support the superiority of the RF algorithm<sup>38</sup>. The RF is an ensemble machine-learning model that could handle data with multi-dimensional, non-linear relationships, high-order correlations, and missing values<sup>39</sup>. Additionally, the RF model is capable of avoiding the accuracy reduction problem caused by missing and noisy data in the training sample when predicting the relationship between a large number of predictor variables and the response variable<sup>40</sup>, attributes supporting the present study results. In contrast, while the NBM like RF is also a machine learning algorithm, it was proven to be not very sensitive to missing data, and the algorithm is relatively simple<sup>41</sup>. Studies have demonstrated that more complex species distributions models provided better predictive performance demonstrating the suitability of the RF model in processing complex high-dimensional data such as the data used in the present study<sup>42</sup>. Moreover, the NBM is a linear classifier and similar to the traditional linear statistical methods, all are insufficient in revealing the complex relationship among environmental variables<sup>43</sup>. In our case, the two linear models, GBM and GLM, demonstrated this with their poor predictive power. Additionally, we observed that the

prediction accuracy of the Extreme Gradient Boosting (XGBoost) was very close to that of RF as the XGBoost has good generalization performance<sup>44</sup>. Although, previous studies have shown that Maximum Entropy (MaxEnt), Support Vector Machine (SVM), and Gradient Boosting Machine (GBM) models performed well in simulating species suitability distribution<sup>45,46</sup>, our results have shown that the prediction accuracy of these models was intermediate relative to the performance of the seven tested models. These phenomena may indicate that species characteristics and sample size also have influence on the accuracy of species distribution models<sup>47</sup>.

## 4.2 The importance of climate variables

Our study along with several others<sup>48–50</sup> were based on the assumption that species distribution is mainly determined by climate<sup>51,52</sup>. It is well documented that climatic factors are key elements for most species' population regeneration<sup>53</sup>. Here, our results indicated that temperature-associated climate factors have greater influence on *M. azedarach* suitable habitats than precipitation factors. Specifically, the top three temperature-related climate variables included mean coldest month temperature (MCMT), the number of frost-free days (NFFD), and degree-days below 18°C (DD>18)), with MCMT contributing the most. This shows that low temperature was the main climatic factor that restricted *M. azedarach* distribution, which is consistent with previous studies, as low-temperature stress imparted a negative impact on plant physiological and biochemical responses (e.g., plant membrane system disorder, photosynthetic rate decline, harmful active oxygen increased, and osmotic adjustment substances increase)<sup>54</sup>. *M. azedarach* is known to prefer warm and humid climates, suitable temperature and abundant precipitation were conducive to the species growth and biomass accumulation. Research has demonstrated that *M. azedarach* ground diameter shown increasing trend with precipitation increase<sup>55</sup>. The extension of the number of frost-free days (NFFD) was beneficial to increasing *M. azedarach* seed size and quality<sup>56</sup>.

## 4.3 Range shift in response to climate change

Our study showed that *M. azedarach* would benefit from the anticipated climate change. More specifically, we found the RCP 8.5 scenario to be more favorable for the species habitat suitability expansion as compared to the RCP 4.5 scenario (Figure. 5g). The RCP 8.5 scenario predicted a greater increase in future temperature warming and precipitation, providing climatic conditions favorable to the species growth<sup>56</sup>. From the species geographic range change point of view, it is expected that the future suitable habitat distribution to expand north- and west-ward. Compared with the RCP4.5 scenario, the predicted trend of suitable habitats changes of the RCP8.5 scenario was more significant in the plateau area near 40 °N (Figure. 5), including the Xinjiang Tarim Basin (RCP8.5) (Figure. 5f). Under the RCP4.5 and RCP8.5 scenarios, the future temperature is envisaged to rise by 1.4 - 1.8 and 2.0 - 3.7°C, respectively, making high latitude areas warmer, resulting in a contemplated rise of mountains tree line, which would ultimately provide the species with a potential of geographic range expansion<sup>57</sup>. At the same time, we noted that the suitable habitat in the Shandong region would experience substantial range loss (Figure. 5), caused by a drastic change in climatic conditions from mainly dry continental airflow with little

precipitation to a future warmer climate associated with intensified precipitation reduction<sup>58</sup>. Additionally, the impact of subtropical high pressure could not be overlooked as the Shandong is often affected by sinking air currents with long periods of high temperature and low precipitation. This subtropical high pressure is expected to gradually moved northward, followed by anticipated clear trend of northward movement associated with precipitation pattern change in the Shandong<sup>59</sup>. To a certain extent, the contemplated climate changes are expected to exacerbate the dryland climate in the Shandong, creating predominantly drought conditions that is unsuitable for the drought-intolerant *M. azedarach*<sup>60</sup>.

## 4.4 Management strategies

Rapid climate change causes most tree populations to exist in unsuitable environmental conditions, threatening their growth and survival and even leading to population extinction<sup>61</sup>. Some tree species adapted to the new climatic conditions by migrating to the same environmental gradient or evolving<sup>62</sup>; however, other tree species would benefit from climate change<sup>63</sup>. *M. azedarach* belongs to those species who would benefit from future climate change leading to anticipated range expansion. The wide distribution of *M. azedarach* harbours abundant phenotypic variation and most of the species phenotypic diversity is mainly distributed in the southwest and south regions and to a lesser extent in other regions<sup>64</sup>. It is worth noting that if a widely distributed species could not track the changing climate due to long-term local adaptation, they would become more vulnerable<sup>65</sup>. Therefore, to prevent this uncertainty, we suggest taking proactive *in-situ* conservation measures for Yunnan, Guizhou, Sichuan, Guangdong, and Guangxi regions, as they are rich in phenotypic diversity which will help in coping with future environmental uncertainty<sup>66</sup>. Assisted migration initiatives should applied to presently unsuitable habitats that are expected to be suitable in the future. For example, the northern regions of Jiangxi, Hubei, Anhui, Henan, and areas near 40°N are reasonable targets for assisted migration conservation measures<sup>67</sup>. We recommend for areas that would be negatively affected by future climate as Shandong, taking *ex-situ* measures through establishing botanical gardens and seed banks in suitable habitats to protect their resources. Therefore, analyzing the *ex-situ* target areas' climate ecology could provide reference for breeding programs and seed transfer guidelines/policies. At the same time, we suggest that other biological factors along with climate should also be considered in the species future research, such as species interaction (allelopathy, soil nutrient competition), land-use change (bio-energy farmland expansion), and the influence of human activities<sup>68,69</sup>, these factors collectively affect the contemporary and future distribution of *M. azedarach*.

## 5 Conclusion

Here, we used three common model accuracy evaluation indicators to compare the suitability of seven data mining techniques for predicting *M. azedarach* distribution. The RF model, with its strong robustness and stability, provided the highest accuracy in establishing a climate niche model. Based on this model, maps of contemporary and future suitable habitats were developed. The RF prediction results indicated that *M. azedarach* would benefit from future climate change through range expansion and this

has tendency towards north- and west-ward expansion. In order to maximize the species protection and development, we recommend taking a proactive *in-situ* conservation measures to conserve genetic variation for adaptation to uncertainties and *ex situ* conservation to protect genetic resources under risk, and assisted migration to better use the areas with good potential in future climates.

## Declarations

Author Contributions: L.F., Y.E., W.G., and T.W. participated in the data analysis, and wrote the paper. X.T., J.Q., Z.F., and J.S. participated in the data preparation and processing. W.G. and T.W. designed the study, and T.W. also provided the paper editing. All the authors read and approved the final manuscript.

Funding: This work was jointly funded by the National Key Research and Development Program of China (No. 2017YFD0600700), and the Special Admission for Postgraduate Study Abroad Program by China Scholarship Council (No. 202008320472).

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

## References

1. Chen, L. *et al.* Geographic variation in traits of fruit stones and seeds of *Melia azedarach*. *Journal of Beijing Forestry University*, **36**, 15–20 (2014).
2. Angamuthu, D., Purushothaman, I., Kothandan, S. & Swaminathan, R. Antiviral study on *Punica granatum* L., *Momordica charantia* L., *Andrographis paniculata* Nees, and *Melia azedarach* L., to human herpes virus-3. *Eur. J. Integr. Med*, **28**, 98–108 <https://doi.org/10.1016/j.eujim.2019.04.008> (2019).
3. Wang, N. *et al.* Selective ERK1/2 agonists isolated from *Melia azedarach* with potent anti-leukemic activity. *BMC Cancer*, **19**, 1–9 <https://doi.org/10.1186/s12885-019-5914-8> (2019).
4. Khoshraftar, Z., Safekordi, A., Shamel, A. & Zaefizadeh, M. Evaluation of insecticidal activity of nanoformulation of *Melia azedarach* (leaf) extract as a safe environmental insecticide. *International Journal of Environmental Science and Technology*, **17**, 1159–1170 <https://doi.org/10.1007/s13762-019-02448-7> (2020).
5. Al-Rubae, A. Y. The potential uses of *Melia azedarach* L. as pesticidal and medicinal plant, review. *American-Eurasian Journal of Sustainable Agriculture*, **3**, 185–194 (2009).
6. Nakatani, M. *et al.* Three new antifeeding meliacarpinins from Chinese *Melia azedarach* Linn., **51**, 11731–11736 [https://doi.org/10.1016/0040-4020\(95\)00724-M](https://doi.org/10.1016/0040-4020(95)00724-M) (1995).
7. Wu, L., Kaewmano, A., Fu, P., Wang, W. & Fan, Z. Intra-annual Radial Growth of *Melia azedarach* in a Tropical Moist Seasonal Forest and its Response to Environmental Factors in Xishuangbanna, Southwest China *Acta Ecol. Sin*, **40**, 6831–6840 <https://doi.org/10.5846/stxb202003120508> (2020).

8. Hoegh-Guldberg, O. *et al.* The human imperative of stabilizing global climate change at 1.5 C. *Science* **365**; <https://doi.org/10.1126/science.aaw6974> (2019).
9. López-Tirado, J., Vessella, F., Schirone, B. & Hidalgo, P. J. Trends in evergreen oak suitability from assembled species distribution models: assessing climate change in south-western Europe. *New For*, **49**, 471–487 <https://doi.org/10.1007/s11056-018-9629-5> (2018).
10. Xu, Y. *et al.* Modelling the Effects of Climate Change on the Distribution of Endangered *Cypripedium japonicum* in China. *Forests* **12**, 429; <https://doi.org/10.3390/f12040429>(2021).
11. Booth, T. H. Species distribution modelling tools and databases to assist managing forests under climate change. *For. Ecol. Manag*, **430**, 196–203 <https://doi.org/10.1016/j.foreco.2018.08.019> (2018).
12. Dyderski, M. K., Paź, S., Frelich, L. E. & Jagodziński, A. M. How much does climate change threaten European forest tree species distributions? *Global change biology*, **24**, 1150–1163 <https://doi.org/10.1111/gcb.13925> (2018).
13. Zhong, Y. *et al.* A Generalized Linear Mixed Model Approach to Assess Emerald Ash Borer Diffusion. *ISPRS International Journal of Geo-Information*, **9**, 414 <https://doi.org/10.3390/ijgi9070414> (2020).
14. Cho, G., Yim, J., Choi, Y., Ko, J. & Lee, S. H. Review of machine learning algorithms for diagnosing mental illness. *Psychiatry investigation*, **16**, 262 <https://doi.org/10.30773/pi.2018.12.21.2> (2019).
15. Brun, P. *et al.* Model complexity affects species distribution projections under climate change. *J. Biogeogr*, **47**, 130–142 <https://doi.org/10.1111/jbi.13734> (2020).
16. Shiferaw, H., Bewket, W. & Eckert, S. Performances of machine learning algorithms for mapping fractional cover of an invasive plant species in a dryland ecosystem. *Ecology and evolution*, **9**, 2562–2574 <https://doi.org/10.1002/ece3.4919> (2019).
17. Ho, T. K. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, **20**, 832–844 <https://doi.org/10.1109/34.709601> (1998).
18. Sothe, C. *et al.* Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. *GIScience & Remote Sensing*, **57**, 369–394 <https://doi.org/10.1080/15481603.2020.1712102> (2020).
19. Sung, S., Kwon, Y. S., Lee, D. K. & Cho, Y. Predicting the potential distribution of an invasive species, *Solenopsis invicta* Buren (Hymenoptera: Formicidae), under climate change using species distribution models. *Entomol. Res*, **48**, 505–513 <https://doi.org/10.1111/1748-5967.12325> (2018).
20. Chen, Y., Wu, Z., Zhao, B., Fan, C. & Shi, S. Weed and Corn Seedling Detection in Field Based on Multi Feature Fusion and Support Vector Machine. *Sensors*, **21**, 212 <https://doi.org/10.3390/s21010212> (2021).
21. Sabat-Tomala, A., Raczko, E. & Zagajewski, B. Comparison of support vector machine and random forest algorithms for invasive and expansive species classification using airborne hyperspectral data. *Remote Sensing*, **12**, 516 <https://doi.org/10.3390/rs12030516> (2020).

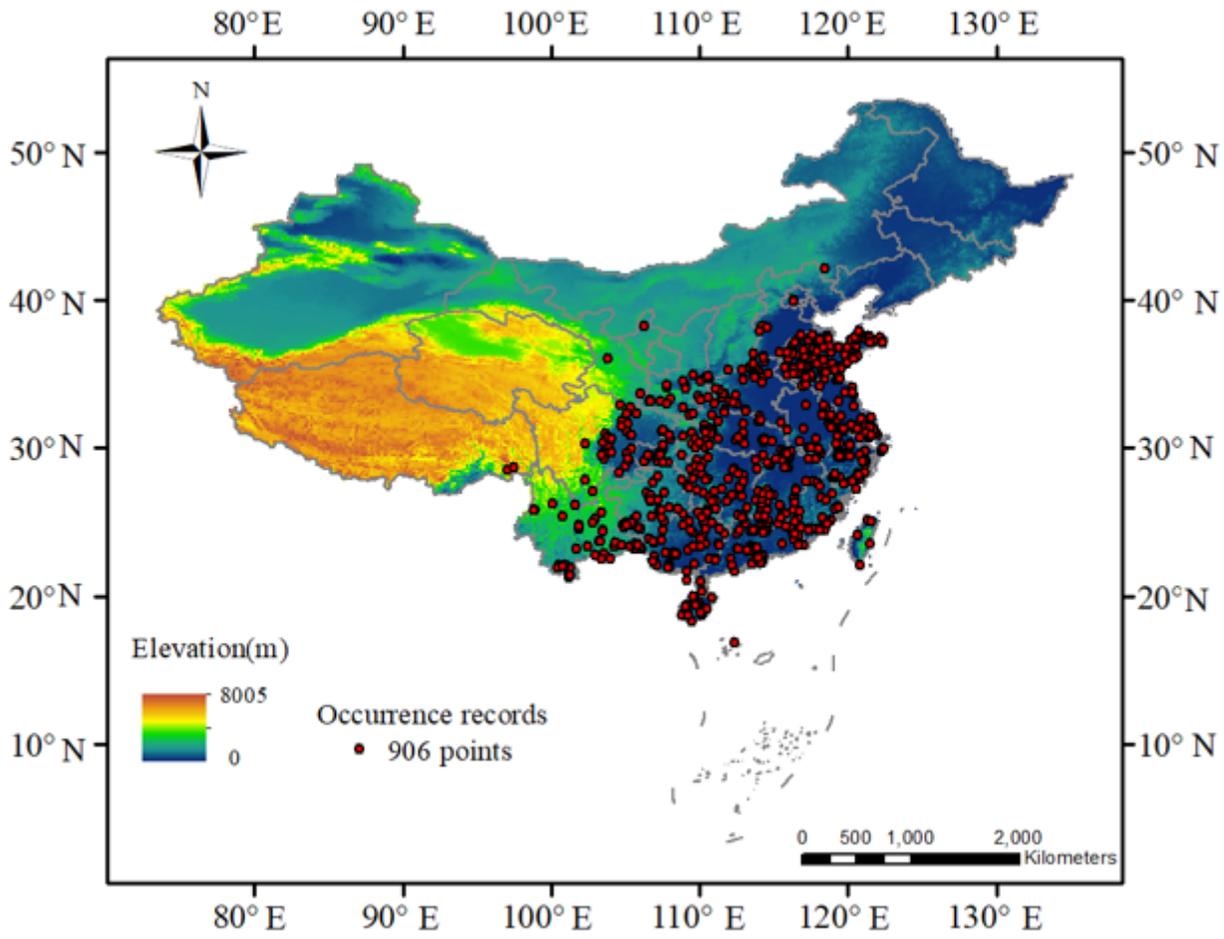
22. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model*, **190**, 231–259 <https://doi.org/10.1016/j.ecolmodel.2005.03.026> (2006).
23. Tang, X., Yuan, Y., Li, X. & Zhang, J. Maximum entropy modeling to predict the impact of climate change on pine wilt disease in China. *Frontiers in Plant Science*, **12**, 764 <https://doi.org/10.3389/fpls.2021.652500> (2021).
24. Chhogyel, N., Kumar, L., Bajgai, Y. & Jayasinghe, L. S. Prediction of Bhutan's ecological distribution of rice (*Oryza sativa* L.) under the impact of climate change through maximum entropy modelling. *The Journal of Agricultural Science*, **158**, 25–37 <https://doi.org/10.1017/S0021859620000350> (2020).
25. Smith, E. H. Using extreme gradient boosting (XGBoost) to evaluate the importance of a suite of environmental variables and to predict recruitment of young-of-the-year spotted seatrout in Florida. *bioRxiv*, 543181; <https://doi.org/10.1101/543181> (2019).
26. Do, B. H., Langlotz, C. & Beaulieu, C. F. Bone tumor diagnosis using a naïve Bayesian model of demographic and radiographic features. *J. Digit. Imaging*, **30**, 640–647 <https://doi.org/10.1007/s10278-017-0001-7> (2017).
27. Chang, Z., Meng, J., Shi, Y. & Mo, F. Lnc RNA recognition by fusing multiple features and its function prediction. *CAAI Transactions on Intelligent Systems*, **13**, 928–934 <https://doi.org/10.11992/tis.201806008> (2018).
28. Ahmad, Z. *et al.* Melia Azedarach impregnated Co and Ni zero-valent metal nanoparticles for organic pollutants degradation: validation of experiments through statistical analysis. *J. Mater. Sci.: Mater. Electron*, **31**, 16938–16950 <https://doi.org/10.1007/s10854-020-04250-5> (2020).
29. Hijmans, R. J., Huaccho, L. & Zhang, D. in *I International Conference on Sweetpotato. Food and Health for the Future 583*. 41-49.
30. Luo, M., Wang, H. & Lyu, Z. Evaluating the performance of species distribution models Biomod2 and MaxEnt using the giant panda distribution data. *The journal of applied ecology*, **28**, 4001–4006 <https://doi.org/10.13287/j.1001-9332.201712.011> (2017).
31. Wang, T., Wang, G., Innes, J. L., Seely, B. & Chen, B. ClimateAP: An application for dynamic local downscaling of historical and future climate data in Asia Pacific. *Frontiers of Agricultural Science and Engineering*, **4**, 448–458 (2017).
32. Yang, X. Q., Kushwaha, S., Saran, S., Xu, J. & Roy, P. Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in Lesser Himalayan foothills. *Ecological engineering*, **51**, 83–87 <https://doi.org/10.1016/j.ecoleng.2012.12.004> (2013).
33. Ravisankar, P., Navya, C. N., Pravallika, D. & Sri, D. N. A review on step-by-step analytical method validation. *IOSR J Pharm*, **5**, 7–19 (2015). <https://www.researchgate.net/publication/338166852>
34. Pepe, M. S., Cai, T. & Longton, G. Combining predictors for classification using the area under the receiver operating characteristic curve., **62**, 221–229 <https://doi.org/10.1111/j.1541-0420.2005.00420.x> (2006).
35. McHugh, M. L. Interrater reliability: the kappa statistic. *Biochemia medica*, **22**, 276–282 (2012).

36. Powers, D. M. What the F-measure doesn't measure: Features, Flaws, Fallacies and Fixes. arXiv preprint arXiv:1503.06410(2015).
37. Lu, C. Y., Gu, W., Dai, A. H. & Wei, H. Y. Assessing habitat suitability based on geographic information system (GIS) and fuzzy: a case study of Schisandra sphenanthera Rehd. et Wils. in Qinling Mountains, China. *Ecol. Model*, **242**, 105–115 <https://doi.org/10.1016/j.ecolmodel.2012.06.002> (2012).
38. Zhang, L. *et al.* The basic principle of random forest and its applications in ecology: A case study of Pinus yunnanensis. *Acta Ecol. Sin*, **34**, 650–659 (2014).
39. Lin, H. Y. *et al.* Climate-based approach for modeling the distribution of montane forest vegetation in Taiwan. *Appl. Veg. Sci*, **23**, 239–253 <https://doi.org/10.1111/avsc.12485> (2020).
40. Zhang, L. *et al.* Consensus forecasting of species distributions: The effects of niche model performance and niche properties. *PLoS ONE*, **10**, e0120056 <https://doi.org/10.1371/journal.pone.0120056> (2015).
41. Zhang, H. The optimality of naive Bayes. *American Association for Artificial Intelligence*, **1**, 3 (2004).
42. Wang, Q., Nguyen, T. T., Huang, J. Z. & Nguyen, T. T. An efficient random forests algorithm for high dimensional data classification. *Advances in Data Analysis and Classification*, **12**, 953–972 <https://doi.org/10.1007/s11634-018-0318-1> (2018).
43. Zheng-tao, Y., Bin, D., Bo, H. & Lu, H. Jian-yi, G. Word Sense Disambiguation Based on Bayes Model and Information Gain. the Proceedings of the International Journal of Advanced Science and Technology 3(2009).
44. Yu, B. *et al.* SubMito-XGBoost: predicting protein submitochondrial localization by fusing multiple feature information and eXtreme gradient boosting., **36**, 1074–1081 <https://doi.org/10.1093/bioinformatics/btz734> (2020).
45. Hailu, B. T., Siljander, M., Maeda, E. E. & Pellikka, P. Assessing spatial distribution of Coffea arabica L. in Ethiopia's highlands using species distribution models and geospatial analysis methods. *Ecological Informatics*, **42**, 79–89 <https://doi.org/10.1016/j.ecoinf.2017.10.001> (2017).
46. Ramirez-Reyes, C. *et al.* Embracing Ensemble Species Distribution Models to Inform At-Risk Species Status Assessments. *Journal of Fish and Wildlife Management*, **12**, 98–111 <https://doi.org/10.3996/JFWM-20-072> (2021).
47. Wisz, M. S. *et al.* Effects of sample size on the performance of species distribution models. *Divers. Distrib*, **14**, 763–773 <https://doi.org/10.1111/j.1472-4642.2008.00482.x> (2008).
48. Feng, L., Sun, J., Shi, Y., Wang, G. & Wang, T. Predicting Suitable Habitats of Camptotheca acuminata Considering Both Climatic and Soil Variables. *Forests*, **11**, 891 <https://doi.org/10.3390/f11080891> (2020).
49. Wang, T., Campbell, E. M., O'Neill, G. A. & Aitken, S. N. Projecting future distributions of ecosystem climate niches: uncertainties and management applications. *For. Ecol. Manag*, **279**, 128–140 <https://doi.org/10.1016/j.foreco.2012.05.034> (2012).

50. Wang, T., Hamann, A., Spittlehouse, D. L. & Murdock, T. Q. ClimateWNA—high-resolution spatial climate data for western North America. *Journal of Applied Meteorology and Climatology*, **51**, 16–29 <https://doi.org/10.1175/JAMC-D-11-043.1> (2012).
51. Feng, L. *et al.* Predicting suitable habitats of Ginkgo biloba L. fruit forests in China. *Climate Risk Management*, 100364 <https://doi.org/10.1016/j.crm.2021.100364> (2021).
52. Wang, T., Hamann, A., Spittlehouse, D. & Carroll, C. Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS ONE*, **11**, e0156720 <https://doi.org/10.1371/journal.pone.0156720> (2016).
53. Guo, Y. *et al.* Spatial prediction and delineation of Ginkgo biloba production areas under current and future climatic conditions. *Industrial Crops and Products*, **166**, 113444 <https://doi.org/10.1016/j.indcrop.2021.113444> (2021).
54. Jiao, C., Lan, G., Sun, Y., Wang, G. & Sun, Y. Dopamine Alleviates Chilling Stress in Watermelon Seedlings via Modulation of Proline Content, Antioxidant Enzyme Activity, and Polyamine Metabolism. *J. Plant Growth Regul.* **40** (2021).
55. Xu, L. *et al.* Effect of salt stress on growth and physiology in Melia azedarach seedlings of six provenances. *International Journal of Agriculture and Biology*, **20**, 471–480 <https://doi.org/10.17957/IJAB/15.0618> (2018).
56. Thakur, S., Thakur, I. & Sankanur, M. Assessment of Genetic Diversity in Drek (Melia azedarach) Using Molecular Markers. *Journal of Tree Sciences*, **36**, 78–85 <https://doi.org/10.5958/2455-7129.2017.00011.5> (2017).
57. Lenoir, J., Gégout, J. C., Marquet, P., De Ruffray, P. & Brisse, H. A significant upward shift in plant species optimum elevation during the 20th century. *Science*, <https://doi.org/10.1126/science.1156831> (2008).
58. Ou-Yang, C. F. *et al.* Impact of equatorial and continental airflow on primary greenhouse gases in the northern South China Sea. *Environmental Research Letters*, **10**, 065005 <https://doi.org/10.1088/1748-9326/10/6/065005> (2015).
59. Liu, B., Zhu, C., Su, J., Ma, S. & Xu, K. Record-breaking northward shift of the western North Pacific subtropical high in July 2018. *Journal of the Meteorological Society of Japan. Ser. II*, <https://doi.org/10.2151/jmsj.2019-047> (2019).
60. Huang, J. *et al.* Dryland climate change: Recent progress and challenges. *Reviews of Geophysics*, **55**, 719–778 <https://doi.org/10.1002/2016RG000550> (2017).
61. Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T. & Prentice, I. C. Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences* **102**, 8245–8250; <https://doi.org/10.1073/pnas.0409902102> (2005).
62. Waldvogel, A. M. *et al.* Evolutionary genomics can improve prediction of species' responses to climate change. *Evolution Letters*, **4**, 4–18 <https://doi.org/10.1002/evl3.154> (2020).
63. Vilà-Cabrera, A., Coll, L., Martínez-Vilalta, J. & Retana, J. Forest management for adaptation to climate change in the Mediterranean basin: A synthesis of evidence. *For. Ecol. Manag.*, **407**, 16–22

- <https://doi.org/10.1016/j.foreco.2017.10.021> (2018).
64. He, X., Li, J., Wang, F., Zhang, J. & Chen, X. Variation and selection of *Melia azedarach* provenances and families. *Journal of Northeast Forestry University*, **47**, 1–7 <https://doi.org/10.13332/j.1000-1522.20170321> (2019).
65. Smith, A. B., Alsdurf, J., Knapp, M., Baer, S. G. & Johnson, L. C. Phenotypic distribution models corroborate species distribution models: A shift in the role and prevalence of a dominant prairie grass in response to climate change. *Glob. Change Biol*, **23**, 4365–4375 <https://doi.org/10.1111/gcb.13666> (2017).
66. Bellon, M. R., Dulloo, E., Sardos, J., Thormann, I. & Burdon, J. J. In situ conservation—harnessing natural and human-derived evolutionary forces to ensure future crop adaptation. *Evolutionary applications*, **10**, 965–977 <https://doi.org/10.1111/eva.12521> (2017).
67. Bidak, L. M., Heneidy, S. Z., Halmy, M. W. A. & El-Kenany, E. T. Sustainability potential for *Ginkgo biloba* L. plantations under climate change uncertainty: An ex-situ conservation perspective. *Acta Ecol. Sin*, <https://doi.org/10.1016/j.chnaes.2021.09.012> (2021).
68. Qin, F., Liu, S. & Yu, S. Effects of allelopathy and competition for water and nutrients on survival and growth of tree species in *Eucalyptus urophylla* plantations. *For. Ecol. Manag*, **424**, 387–395 <https://doi.org/10.1016/j.foreco.2018.05.017> (2018).
69. Zabel, F. *et al.* Global impacts of future cropland expansion and intensification on agricultural markets and biodiversity. *Nat. Commun*, **10**, 1–10 <https://doi.org/10.1038/s41467-019-10775-z> (2019).

## Figures



**Figure 1**

Distributions of the 906 *M. azedarach* occurrence records.

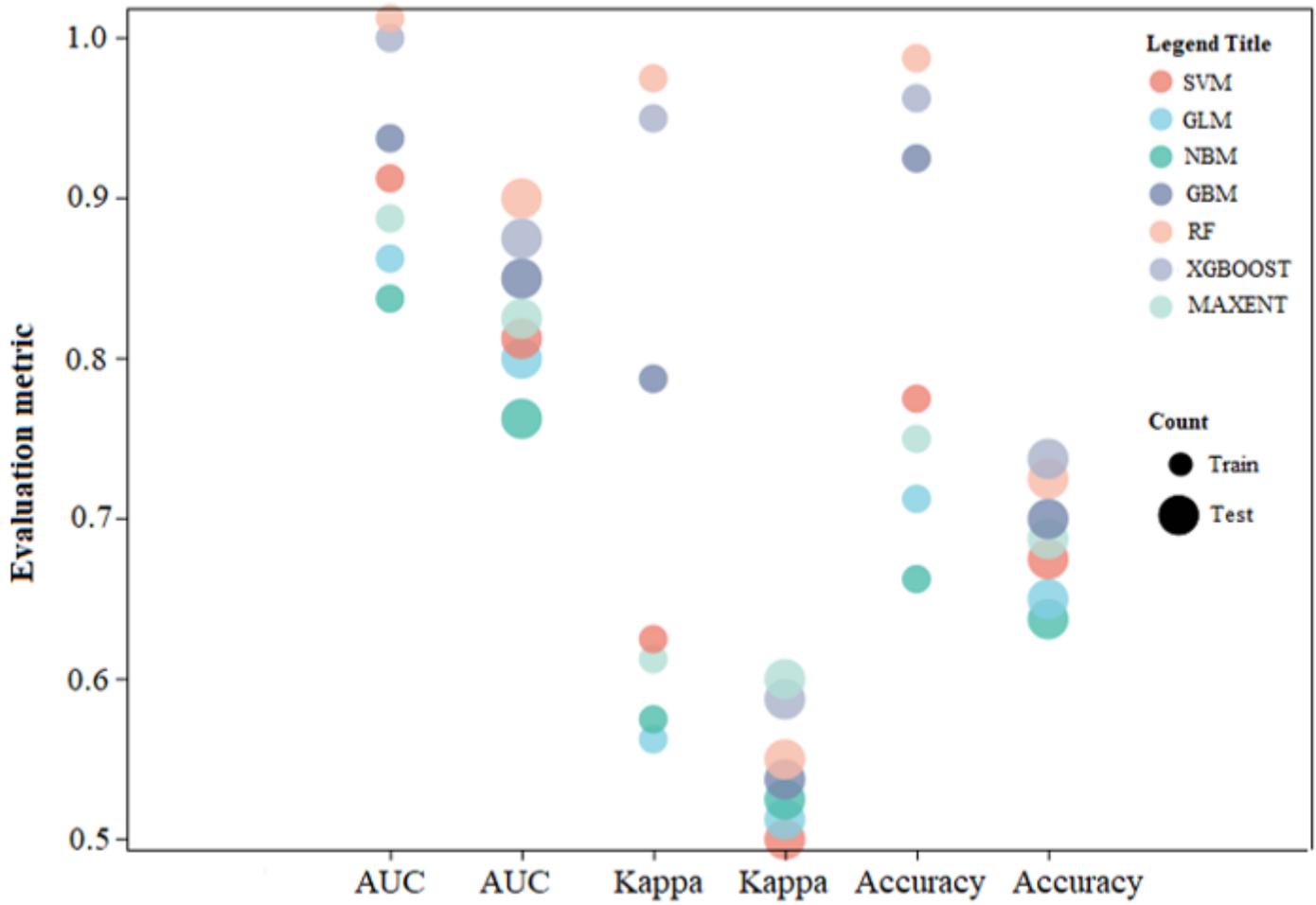
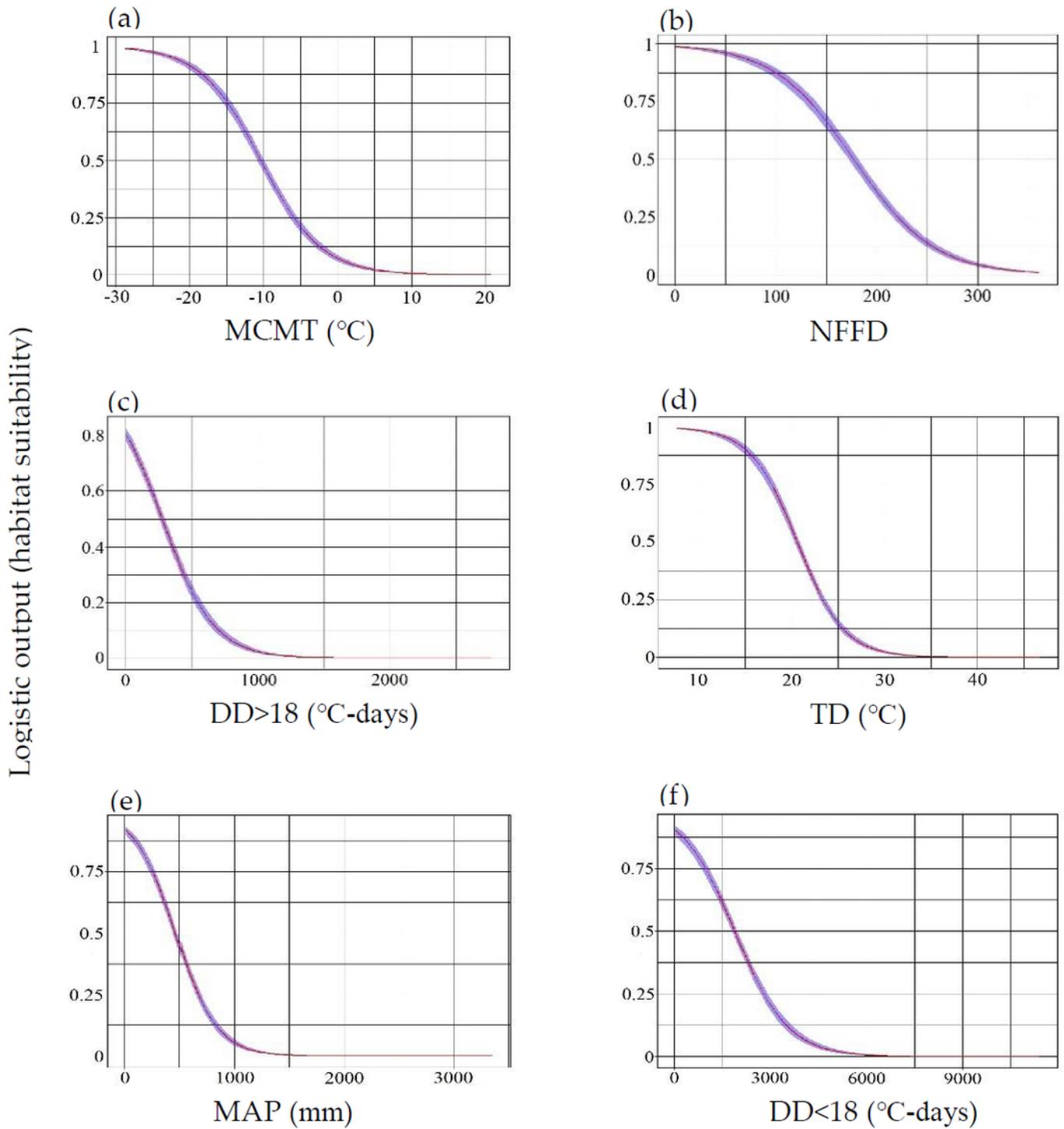


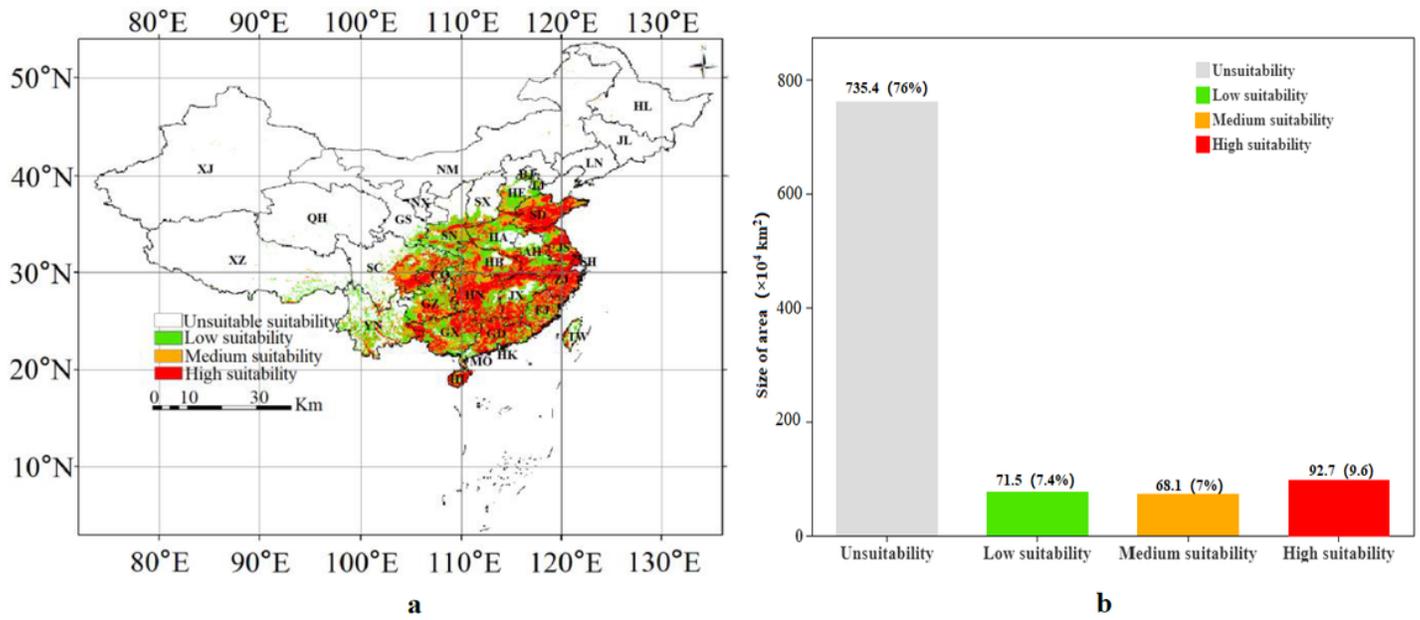
Figure 2

Bubble diagram of evaluation metrics for both the training and testing data, where small bubble represents the training and big bubble represents the testing. Different color bubbles represent different models.



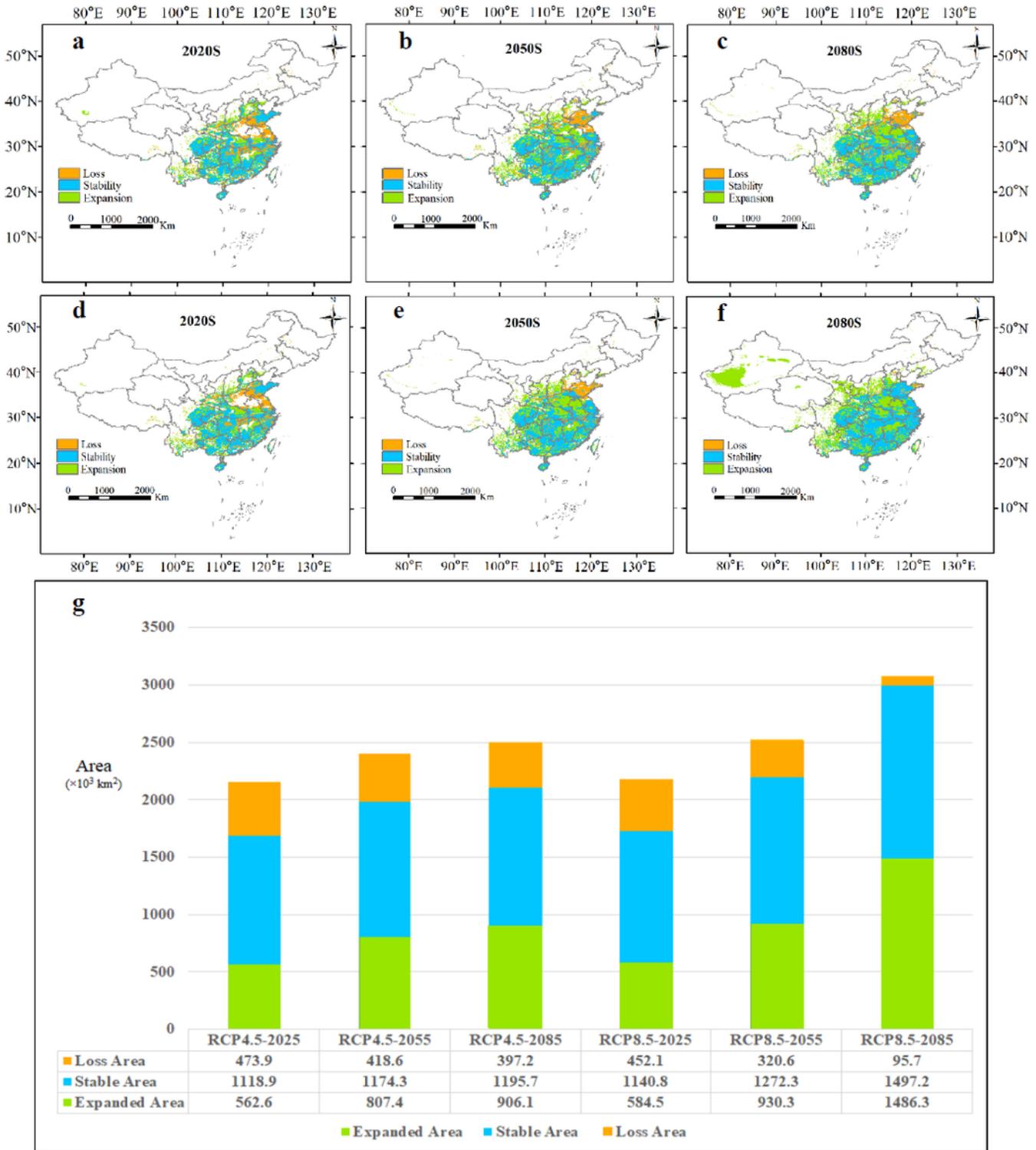
**Figure 3**

Response curves of the top six important climate variables (a-f) in the RF model. When the logical output > 0.5, the probability of species presence under this condition is higher than that under a typical condition, indicating that the condition is suitable for tree species.



**Figure 4**

(a) *M. azedarach* contemporary suitable habitats distributions (1960 – 1990) and (b) their percentage representations.



**Figure 5**

RF projected range changes for *M. azedarach* under RCP 8.5 and RCP 4.5 climate change scenarios (a-f) (g shows areas of habitat change).

## Supplementary Files

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

- [AttachmentTableS1.docx](#)