



OPEN

## Predicting suitable habitats of *Melia azedarach* L. in China using data mining

Lei Feng<sup>1,2</sup>, Xiangni Tian<sup>3</sup>, Yousry A. El-Kassaby<sup>2</sup>, Jian Qiu<sup>1</sup>, Ze Feng<sup>4</sup>, Jiejie Sun<sup>2,5</sup>, Guibin Wang<sup>1</sup>✉ & Tongli Wang<sup>2</sup>✉

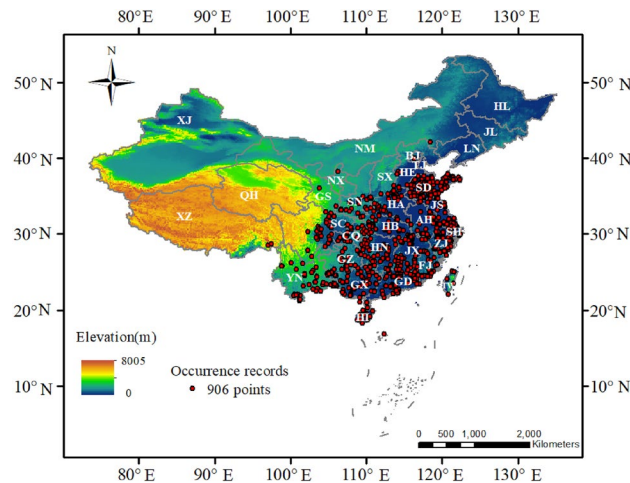
*Melia azedarach* L. is an important economic tree widely distributed in tropical and subtropical regions of China and some other countries. However, it is unclear how the species' suitable habitat will respond to future climate changes. We aimed to select the most accurate one among seven data mining models to predict the current and future suitable habitats for *M. azedarach* in China. 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 overall accuracy (OA)). We found that the RF provided the most outstanding performance in prediction power and generalization capacity. The top climate factors affecting the species' suitable habitats 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 2011–2040 (2020s), 2041–2070 (2050s), and 2071–2100 (2080s). 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.

*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. In China, it is widely distributed in mixed forests, fields, and roadsides between 18 and 40° N of China, occupying about one-third of the country's land area<sup>5</sup>. It is mainly used in plant pesticides, timber, medicinal and ecological restoration<sup>6</sup>. However, uncertain climate factors may reshape their future suitable habitats in China<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 suitable habitat 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

<sup>1</sup>Co-Innovation Centre for Sustainable Forestry in Southern China, College of Forestry, Nanjing Forestry University, Nanjing 210037, China. <sup>2</sup>Department of Forest and Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. <sup>3</sup>School of Mathematics and Statistics, Yunnan University, Kunming 650504, China. <sup>4</sup>College of Pharmacy, Kangda College of Nanjing Medical University, Lianyungang 222000, China. <sup>5</sup>College of Biology and the Environment, Nanjing Forestry University, Nanjing 210037, China. ✉email: gbwang@njfu.com.cn; tongli.wang@ubc.ca



**Figure 1.** Distributions of the 906 *M. azedarach* occurrence records. The map was created in ArcMap 10.2 of the Environmental System Resource Institute, Inc. (<https://www.esri.com/zh-cn/arcgis/products/arcgis-desktop/resources>).

model species distribution data. For example, GLM was used to predict the spread of Emerald Ash Borer (*Agrilus planipennis*) in southern Ontario, Canada<sup>13</sup>. The NBM predicts the potential distribution areas of *Taxus chinensis* and identifying plant long non-coding RNA and predicting its functions<sup>14</sup>. Hailu Shiferaw et al. selected the best performing algorithm for mapping the coverage of *Prosopis juliflora* (Swartz DC.) in Afar, Ethiopia by comparing GBM, RF and SVM algorithms<sup>15</sup>. Maxent models have been widely used in the fields of crop niches, plant diseases and insect pests, and species invasion prediction<sup>16,17</sup>. Therefore, we compared these modeling approaches through data mining techniques to identify the most effective modeling approach to predict the suitable habitat of *M. azedarach* species distribution based on the relationship between its occurrence and climate variables.

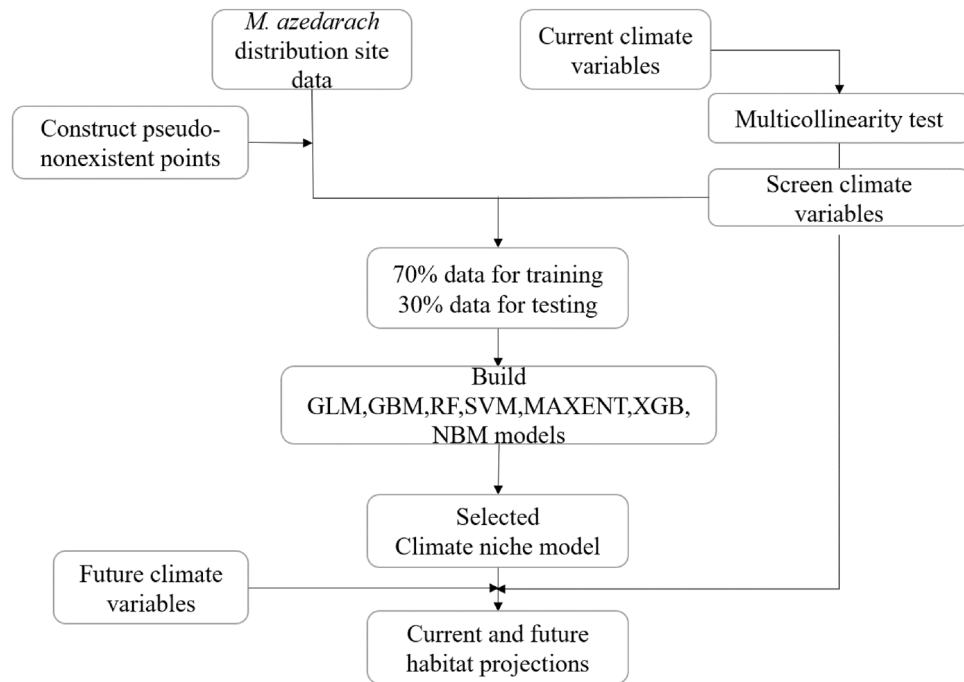
Understanding the potential impact of climate change on the suitable habitat of *M. azedarach* is of great significance to its cultivation and conservation in China. Studies conducted on *M. azedarach* were mainly focused on tree and stand productivity, extraction of active ingredients, and pest resistance potential<sup>3,18</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 suitable habitat; (3) develop current and future suitable habitat maps for *M. azedarach* in China and highlighting the areas of change; and (4) assess the potential impact of future climate change on the species suitable habitat.

## Material and methods

**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://doi.org/10.15468/dl.3t8r62>, accessed on 17 May 2022, and the Chinese Virtual Herbarium (CVH), <http://www.cvh.ac.cn/>, accessed on 17 May 2022). All the *M. azedarach* distribution data have been licensed. To avoid redundant sampling, we deleted those sample points with similar longitude and latitude<sup>19</sup>. Then a 0.01° mesh thinning was performed, and the actual distance corresponding to 0.01° was about 1 km and only one distribution point was reserved in each grid so that the distance between sample points was more than 1 km<sup>20</sup>. A total of 906 samples were included for model building. Finally, we used ArcMap 10.2 to overlay the asc result file generated by the model with the map of China to generate the final result map (Fig. 1). In addition, all maps in our study were created in ArcMap 10.2.

**Environment variables.** We used *M. azedarach* presence-absence data as the dependent variable and 16 climatic factors derived from ClimateAP\_v221 software (<http://ClimateAP.net>) as predictors to build the model (Table S1)<sup>21</sup>. We used the following tests to avoid the effect of multicollinearity among the climate variables. First, the variance inflation factor (VIF) was calculated for each of the 16 variables (Table S2). Second, the correlation analysis was conducted for each pair of the 16 variable (Figure S1). Finally, we used stepwise regression analysis to eliminate the variables that led to an observed multicollinearity<sup>22</sup>.

**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 (Max-Ent), Extreme Gradient Boosting (XGBoost), and Naive Bayesian Model (NBM)) to associate the distribution of *M. azedarach* with climate variables. We used a data-driven approach to select the number of pseudo-existent points, and started with 1000, 2000, 10,000, 30,000, and 100,000 pseudo-nonexistent points. It was found that



**Figure 2.** Flowchart for modeling of *M. azedarach*.

the most models had the highest prediction accuracy with 2000 pseudo-nonexistent points. Therefore, we used the “dismo” package in R to randomly generate 2000 “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 10 times. At the same time, the Maxent model was executed using the Maxent version 3.4.4 software in R-package (Fig. 2).

**Model validation.** To assess the performance of the seven predictive models, we compared their area under receiver operating character curve (AUC), Kappa, and overall accuracy (OA). 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>23</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>24</sup>. Both Kappa and AUC consider the true positive rate and true negative rate to avoid an overestimation or underestimation error (Sahin 2020).

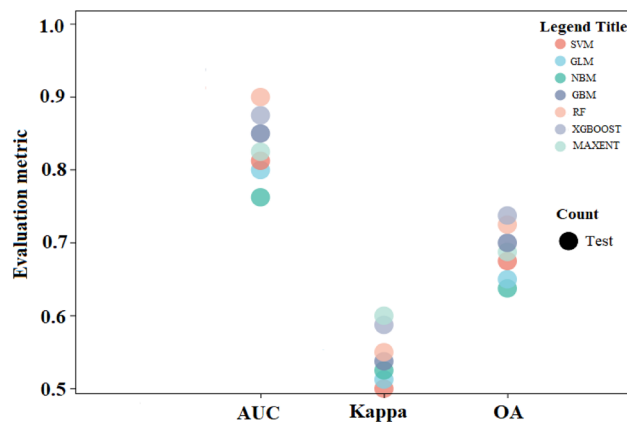
**Habitat classification.** Appropriate habitat evaluation index values were determined as follows: predicted values of 0–0.2, 0.2–0.4, 0.4–0.6, and >0.6 were deemed unsuitable, low-, medium-, and high-suitable habitat, respectively<sup>25</sup> (All methods were performed in accordance with the relevant guidelines and regulations).

## Results

**Models performance evaluation.** Through the cross-validation evaluation of the tested models, Kappa, AUC, and OA values were obtained for the testing portions of the dataset (Fig. 3). All models performed well (AUC > 0.8, Kappa > 0.5, and OA > 0.7). The AUC values of the seven models varied from 0.85 (NBM) to 0.90 (RF), Kappa values varied from 0.53 (SVM) to 0.59 (MaxEnt), and Overall accuracy values ranged from 0.77 (NBM) to 0.81 (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.

**Important climate variables and their response curves in random forest (RF).** Through multicollinearity analysis of the variables, we finally identified ten key climate variables. 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 4 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 3.** Bubble diagram of evaluation metrics for testing data. Different color bubbles represent different models.

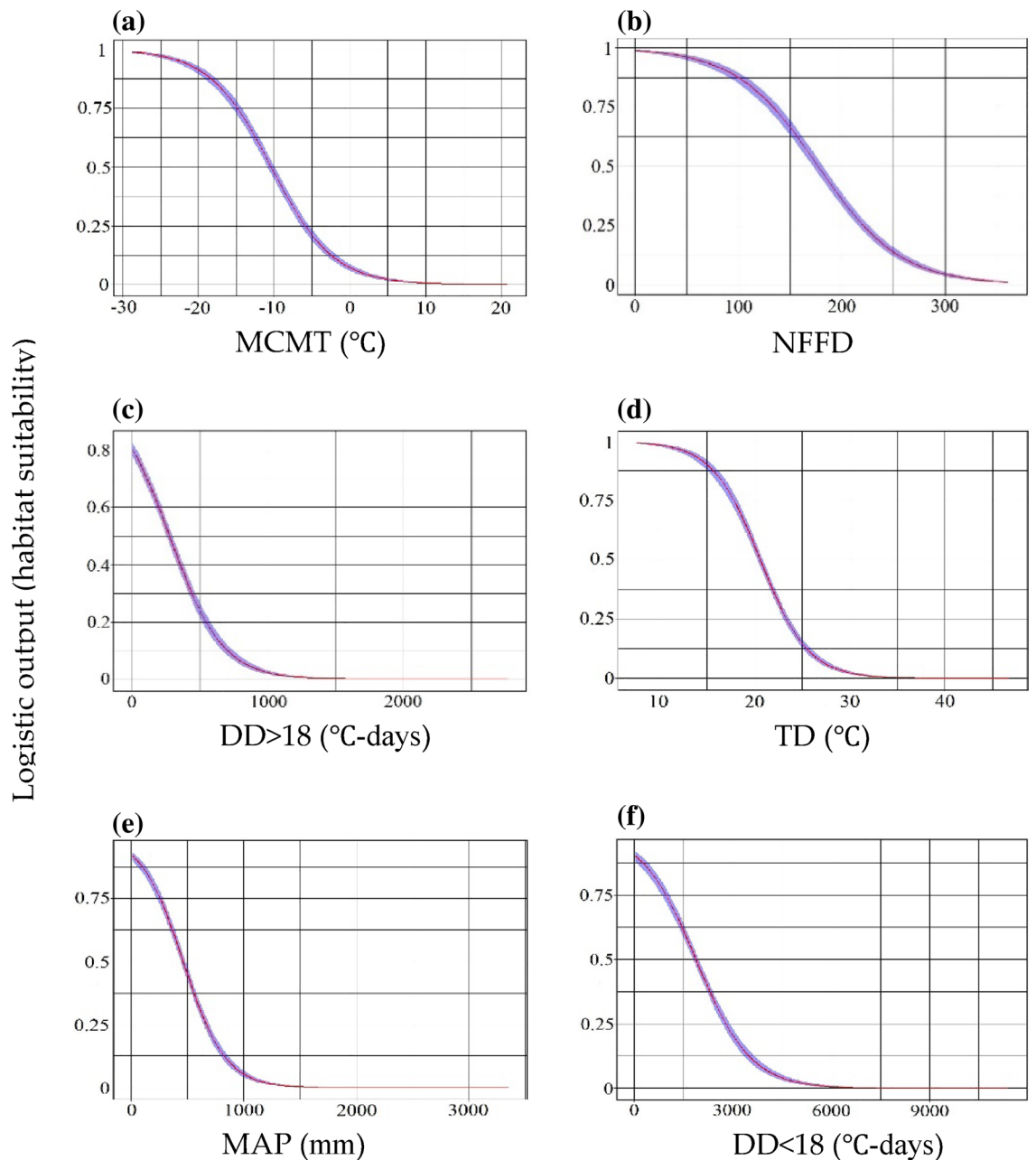
Variable <sup>1</sup>	Units	Overall contribution
MCMT	°C	189.24
NFFD	day	180.69
DD > 18	°C-days	104.77
TD	°C	72.82
MAP	mm	69.43
DD < 18	°C-days	64.12
DD > 5	°C-days	56.27
AHM		54.88
DD < 0	°C-days	44.01
PAS	mm	28.54

**Table 1.** Contributions of the most influencing climate variables to the *M. azedarach* random forest (RF) model. <sup>1</sup>See Table S1 for variables abbreviations.

(Fig. 4a), between 0 and 175 days for NFFD (Fig. 4b), between 0 and 250 for DD > 18 (Fig. 4c), between 5 and 21 °C for TD (Fig. 4d), between 0 and 480 mm for MAP (Fig. 4e), and between 0 and 1750 for DD < 18 (Fig. 4f).

**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 Fig. 5. The overall suitable habitat was mainly distributed between 18 and 40° N (Fig. 5a). These habitats were classified as: (1) high-suitable habitats (mainly scattered in Shandong (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%; Fig. 5b); (2) medium-suitable habitats (scattered around the high-suitable habitats, covering  $6.8 \times 10^5$  km<sup>2</sup> (7%; Fig. 5b) 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%; Fig. 5b)), 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)).

**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 (Fig. 6). The projected range increase was greatest under RCP 8.5 as compared to RCP 4.5 (Fig. 6). More specifically, the expanded area would increase by  $562.6 \times 10^3$  km<sup>2</sup> and  $584.5 \times 10^3$  km<sup>2</sup> by 2020s,  $807.4 \times 10^3$  km<sup>2</sup> and  $930.3 \times 10^3$  km<sup>2</sup> by 2050s, and  $906.1 \times 10^3$  km<sup>2</sup> and  $1486.3 \times 10^3$  km<sup>2</sup> by 2080s under the RCP4.5 and RCP8.5 scenarios, respectively (Fig. 6g). 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 (Fig. 6f). Additionally, the species stable range area showed the same change pattern as that of the expanded area (Fig. 6g). 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

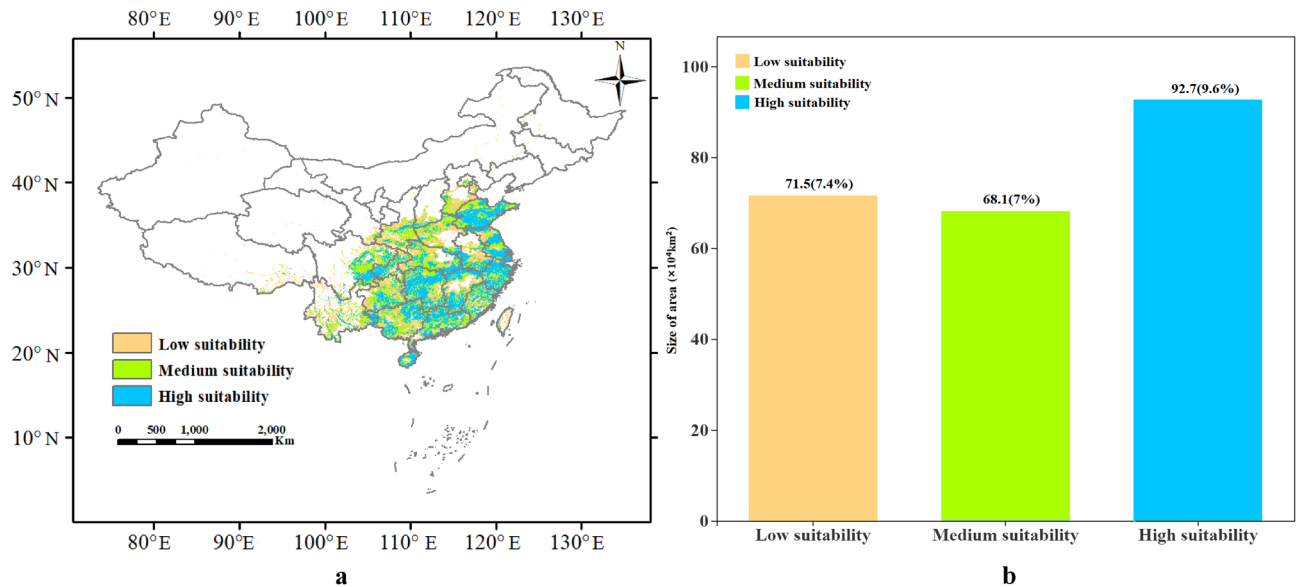


**Figure 4.** Response curves of the top six important climate variables (a–f) in the RF model. When the logical output  $> 0.5$ , the probability of species presence under this condition is higher than that under a typical condition, indicating that the condition is suitable for tree species.

(SC) (Fig. 6a–f). Furthermore, the species area loss exhibited an opposite trend to that of expansion and stable range areas (Fig. 6f) and most of the loss area was mainly distributed in eastern coastal provinces near  $30\text{--}38^\circ\text{N}$  (e.g., Shandong (SD)) (Fig. 6).

## Discussion

**Model performance.** Here, we used the AUC, Kappa, and OA to evaluate the performance of seven species range prediction models (GLM, GBM, MaxEnt, SVM, XGBoost, NBM, and RF) to predict *M. azedarach* contemporary and future ranges under two climate scenarios (RCP 8.5 and RCP 4.5). The results showed that RF and XGBoost were the top-performing models with RF being the best, while NBM and GLM were the low-performing with the NBM being the worst. Similarly, multiple lines of evidence support the superiority of the RF algorithm<sup>26</sup>. In a study in northern California, the GLM, ANN, RF and ME models were used to predict new occurrences for rare plants, and the results showed that RF provided the best prediction<sup>27</sup>. Akpoti et al. used BRT, GLM, MAXNT and RF algorithms to predict rice production suitability and the results showed that RF has better generalizability<sup>28</sup>. Silva et al. found the highest model quality for the RF and GAM algorithms when assessing the limitations of different species distribution models using the Azorean Forest as an example<sup>29</sup>. The RF is an

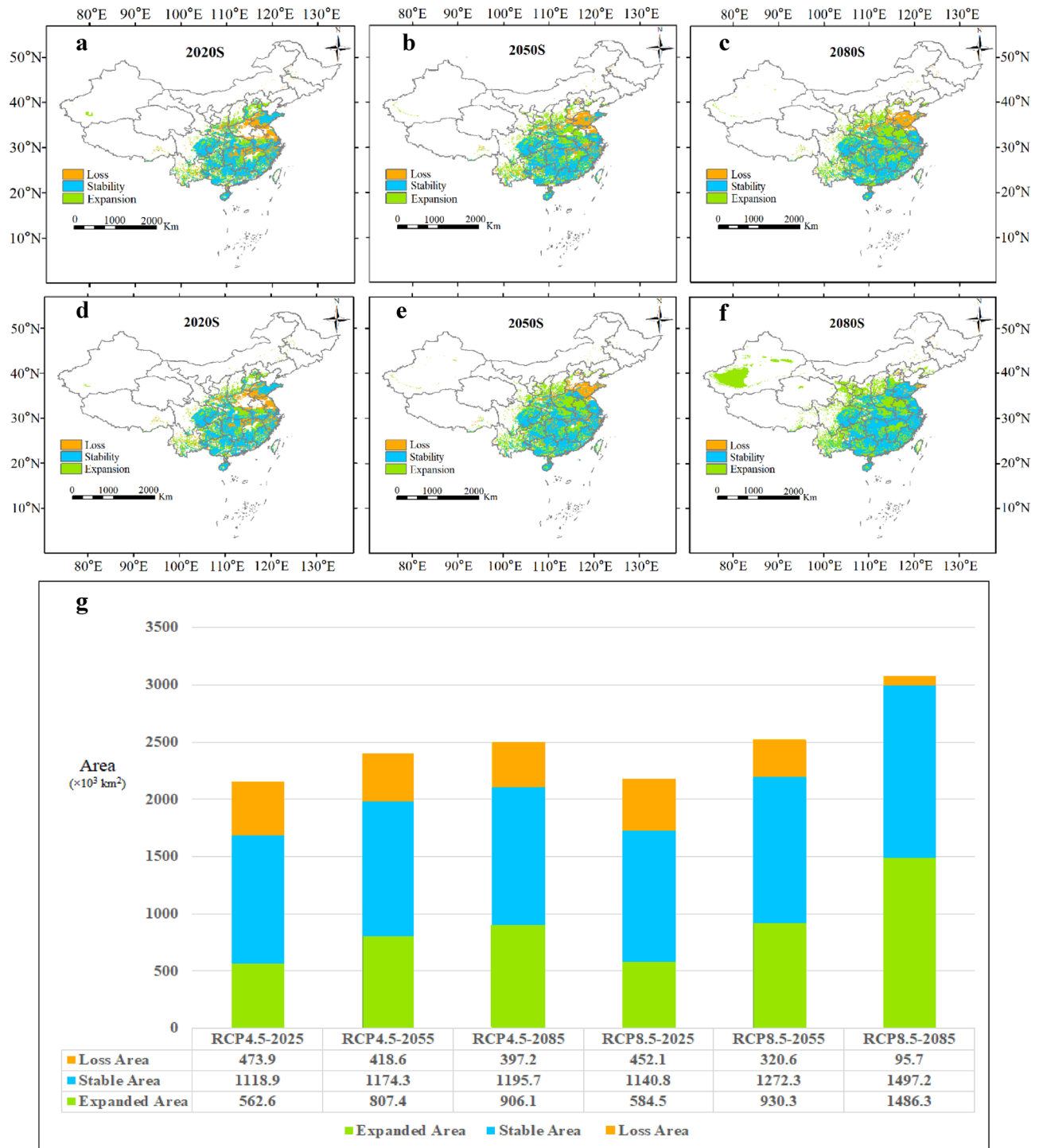


**Figure 5.** (a) *M. azedarach* contemporary suitable habitats distributions (1960–1990) and (b) their percentage representations. The map was created in ArcMap 10.2 of the Environmental System Resource Institute, Inc. (<https://www.esri.com/zh-cn/arcgis/products/arcgis-desktop/resources>).

ensemble machine-learning model that could handle data with multi-dimensional, non-linear relationships, high-order correlations, and missing values<sup>30</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>31</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>32</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>33</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>34</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 XGBoost was very close to that of RF as the XGBoost has good generalization performance<sup>35</sup>. Although, previous studies have shown that MaxEnt, SVM, and GBM models performed well in simulating species suitability distribution<sup>36,37</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>38</sup>.

**The importance of climate variables.** Our study along with several others<sup>39–41</sup> were based on the assumption that species distribution is mainly determined by climate<sup>42,43</sup>. It is well documented that climatic factors are key elements for most species' population regeneration<sup>44</sup>. Here, our results indicated that temperature-associated climate factors have greater influence on *M. azedarach* suitable habitats than precipitation factors. Specifically, five of the top six contributing climatic variables were related to low temperature (MCMT, NFFD, and DD < 18) and continentality (TD), with MCMT contributing the most. This shows that low temperature was the main climatic factor that restricted *M. azedarach* suitable habitat, 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>45</sup>. The extension of the number of frost-free days (NFFD) was beneficial to increasing *M. azedarach* seed size and quality, thereby improving the survival rate<sup>46</sup>. In addition, MAP also influences the distribution of *M. azedarach* under certain TD conditions, as a warm and humid climate favors the growth and biomass accumulation of *M. azedarach*<sup>47</sup>. Xu et al. also confirmed that the ground diameter of *M. azedarach* tended to increase with increasing precipitation<sup>48</sup>.

**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 (Fig. 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>46</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 (Fig. 5), including the Xinjiang Tarim Basin (RCP8.5) (Fig. 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



**Figure 6.** 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). The map created in ArcMap 10.2 of the Environmental System Resource Institute, Inc. (<https://www.esri.com/zh-cn/arcgis/products/arcgis-desktop/resources>).

warmer, resulting in a contemplated rise of mountains tree line, which would ultimately provide the species with a potential of geographic range expansion<sup>49</sup>. At the same time, we noted that the suitable habitat in the Shandong region would experience substantial range loss (Fig. 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>50</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>51</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>52</sup>.

**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>53</sup>. Some tree species adapted to the new climatic conditions by migrating to the same environmental gradient or evolving<sup>54</sup>; however, other tree species would benefit from climate change<sup>55</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>56</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>57</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>58</sup>. Assisted migration initiatives should apply 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>59</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>60,61</sup>, these factors collectively affect the contemporary and future distribution of *M. azedarach*.

## 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.

## Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 21 October 2021; Accepted: 12 July 2022

Published online: 23 July 2022

## References

- Chen, L. *et al.* Geographic variation in traits of fruit stones and seeds of *Melia azedarach*. *J. Beijing For. Univ.* **36**, 15–20 (2014).
- 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).
- 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).
- Khoshraftar, Z., Safekordi, A., Shamel, A. & Zaeifzadeh, M. Evaluation of insecticidal activity of nanoformulation of *Melia azedarach* (leaf) extract as a safe environmental insecticide. *Int. J. Environ. Sci. Technol.* **17**, 1159–1170. <https://doi.org/10.1007/s13762-019-02448-7> (2020).
- Sivaraj, I., Nithanial, S., Bhooma, V., Senthilkumar, U. & Parani, M. Species delimitation of *Melia dubia* Cav. from *Melia azedarach* L. complex based on DNA barcoding. *Botany* **96**, 329–336. <https://doi.org/10.1139/cjb-2017-0148> (2018).
- Liao, B. *et al.* Population structure and genetic relationships of *Melia* Taxa in China assayed with sequence-related amplified polymorphism (SRAP) markers. *Forests* **7**, 81. <https://doi.org/10.3390/f7040081> (2016).
- 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).
- Hoegh-Guldberg, O. *et al.* The human imperative of stabilizing global climate change at 1.5 C. *Science* **365**, eaaw6974. <https://doi.org/10.1126/science.aaw6974> (2019).
- 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).
- 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).
- 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).
- Dyderski, M. K., Paź, S., Frelich, L. E. & Jagodziński, A. M. How much does climate change threaten European forest tree species distributions?. *Glob. Change Biol.* **24**, 1150–1163. <https://doi.org/10.1111/gcb.13925> (2018).
- Zhong, Y. *et al.* A generalized linear mixed model approach to assess emerald ash Borer diffusion. *ISPRS Int. J. Geo Inf.* **9**, 414. <https://doi.org/10.3390/ijgi9070414> (2020).



14. Chang, Z., Meng, J., Shi, Y. & Mo, F. Lnc RNA recognition by fusing multiple features and its function prediction. *CAAI Trans. Intell. Syst.* **13**, 928–934. <https://doi.org/10.11992/tis.201806008> (2018).
15. Shiferaw, H., Bewket, W. & Eckert, S. Performances of machine learning algorithms for mapping fractional cover of an invasive plant species in a dryland ecosystem. *Ecol. Evol.* **9**, 2562–2574. <https://doi.org/10.1002/ece3.4919> (2019).
16. Tang, X., Yuan, Y., Li, X. & Zhang, J. Maximum entropy modeling to predict the impact of climate change on pine wilt disease in China. *Front. Plant Sci.* **12**, 764. <https://doi.org/10.3389/fpls.2021.652500> (2021).
17. 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. *J. Agric. Sci.* **158**, 25–37. <https://doi.org/10.1017/S0021859620000350> (2020).
18. 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).
19. Hijmans, R. J., Huaccho, L. & Zhang, D. In *I International Conference on Sweetpotato. Food and Health for the Future* **583**, 41–49.
20. Luo, M., Wang, H. & Lyu, Z. Evaluating the performance of species distribution models Biomod2 and MaxEnt using the giant panda distribution data. *J. Appl. Ecol.* **28**, 4001–4006. <https://doi.org/10.13287/j.1001-9332.201712.011> (2017).
21. 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. *Front. Agric. Sci. Eng.* **4**, 448–458. <https://doi.org/10.15302/J-FASE-2017172> (2017).
22. Yang, X.-Q., Kushwaha, S., Saran, S., Xu, J. & Roy, P. Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L Lesser Himalayan foothills. *Ecol. Eng.* **51**, 83–87. <https://doi.org/10.1016/j.ecoleng.2012.12.004> (2013).
23. Pepe, M. S., Cai, T. & Longton, G. Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometrics* **62**, 221–229. <https://doi.org/10.1111/j.1541-0420.2005.00420.x> (2006).
24. McHugh, M. L. Interrater reliability: The kappa statistic. *Biochem. Med.* **22**, 276–282. <https://hrcaj.srce.hr/89395> (2012).
25. 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).
26. 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. <https://doi.org/10.5846/stxb201306031292> (2014).
27. Williams, J. N. *et al.* Using species distribution models to predict new occurrences for rare plants. *Divers. Distrib.* **15**, 565–576. <https://doi.org/10.1111/j.1472-4642.2009.00567.x> (2009).
28. Akpoti, K., Kabo-Bah, A. T., Dossou-Yovo, E. R., Groen, T. A. & Zwart, S. J. Mapping suitability for rice production in inland valley landscapes in Benin and Togo using environmental niche modeling. *Sci. Total Environ.* **709**, 136165. <https://doi.org/10.1016/j.scitotenv.2019.136165> (2020).
29. Dutra Silva, L., de Brito, A. E., Vieira Reis, F., Bento Elias, R. & Silva, L. Limitations of species distribution models based on available climate change data: a case study in the Azorean forest. *Forests* **10**, 575. <https://doi.org/10.3390/f10070575> (2019).
30. 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).
31. 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).
32. Zhang, H. The optimality of naive Bayes. *Am. Assoc. Artif. Intell.* **1**, 3 (2004).
33. Wang, Q., Nguyen, T.-T., Huang, J. Z. & Nguyen, T. T. An efficient random forests algorithm for high dimensional data classification. *Adv. Data Anal. Classif.* **12**, 953–972. <https://doi.org/10.1007/s11634-018-0318-1> (2018).
34. Zheng-tao, Y., Bin, D., Bo, H., Lu, H. & Jian-yi, G. Word sense disambiguation based on bayes model and information gain. *Proc. Int. J. Sci. Technol.* **2**, 153–157. <https://doi.org/10.1109/FGCN.2008.188> (2009).
35. Yu, B. *et al.* SubMito-XGBoost: Predicting protein submitochondrial localization by fusing multiple feature information and eXtreme gradient boosting. *Bioinformatics* **36**, 1074–1081. <https://doi.org/10.1093/bioinformatics/btz734> (2020).
36. 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. *Ecol. Inf.* **42**, 79–89. <https://doi.org/10.1016/j.ecoinf.2017.10.001> (2017).
37. Ramirez-Reyes, C. *et al.* Embracing ensemble species distribution models to inform at-risk species status assessments. *J. Fish Wildl. Manag.* **12**, 98–111. <https://doi.org/10.3996/JFWM-20-072> (2021).
38. 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).
39. 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).
40. 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).
41. Wang, T., Hamann, A., Spittlehouse, D. L. & Murdock, T. Q. ClimateWNA—high-resolution spatial climate data for western North America. *J. Appl. Meteorol. Climatol.* **51**, 16–29. <https://doi.org/10.1175/JAMC-D-11-043.1> (2012).
42. Feng, L. *et al.* Predicting suitable habitats of *ginkgo biloba* L. fruit forests in China. *Clim. Risk Manag.* **34**, 100364. <https://doi.org/10.1016/j.crm.2021.100364> (2021).
43. 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).
44. Guo, Y. *et al.* Spatial prediction and delineation of Ginkgo biloba production areas under current and future climatic conditions. *Ind. Crops Prod.* **166**, 113444. <https://doi.org/10.1016/j.indcrop.2021.113444> (2021).
45. 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**, 2. <https://doi.org/10.1007/s00344-020-10096-2> (2021).
46. Thakur, S., Thakur, I. & Sankanur, M. Assessment of genetic diversity in drek (*Melia azedarach*) using molecular markers. *J. Tree Sci.* **36**, 78–85. <https://doi.org/10.5958/2455-7129.2017.00011.5> (2017).
47. Sivasubramaniam, K. *et al.* Seed priming: Triumphs and tribulations. *The Madras Agricultural Journal* **98**, 197–209. <https://www.researchgate.net/publication/267298497> (2011).
48. Xu, L. *et al.* Effect of salt stress on growth and physiology in *Melia azedarach* seedlings of six provenances. *Int. J. Agric. Biol.* **20**, 471–480. <https://doi.org/10.17957/IJAB/15.0618> (2018).
49. 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* **320**, 1768–1771. <https://doi.org/10.1126/science.1156831> (2008).
50. Ou-Yang, C.-F. *et al.* Impact of equatorial and continental airflow on primary greenhouse gases in the northern South China Sea. *Environ. Res. Lett.* **10**, 065005. <https://doi.org/10.1088/1748-9326/10/6/065005> (2015).
51. Liu, B., Zhu, C., Su, J., Ma, S. & Xu, K. Record-breaking northward shift of the western North Pacific subtropical high in July 2018. *J. Meteorol. Soc. Japan* **97**, 913–925. <https://doi.org/10.2151/jmsj.2019-047> (2019).
52. Huang, J. *et al.* Dryland climate change: Recent progress and challenges. *Rev. Geophys.* **55**, 719–778. <https://doi.org/10.1002/2016RG000550> (2017).

53. Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T. & Prentice, I. C. Climate change threats to plant diversity in Europe. *Proc. Natl. Acad. Sci.* **102**, 8245–8250. <https://doi.org/10.1073/pnas.0409902102> (2005).
54. Waldvogel, A. M. *et al.* Evolutionary genomics can improve prediction of species' responses to climate change. *Evol. Lett.* **4**, 4–18. <https://doi.org/10.1002/evl3.154> (2020).
55. 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).
56. He, X., Li, J., Wang, F., Zhang, J. & Chen, X. Variation and selection of *Melia azedarach* provenances and families. *J. Northeast For. Univ.* **47**, 1–7. <https://doi.org/10.13332/j.1000-1522.20170321> (2019).
57. 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).
58. 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. *Evol. Appl.* **10**, 965–977. <https://doi.org/10.1111/eva.12521> (2017).
59. 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.* **42**, 101–114. <https://doi.org/10.1016/j.chnaes.2021.09.012> (2021).
60. 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).
61. 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).

### 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), the Scholarship for Studying Abroad by China Scholarship Council (No. 202008320472), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX22\_1118).

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-16571-y>.

**Correspondence** and requests for materials should be addressed to G.W. or T.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022