Ecological Informatics 36 (2016) 42–49

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Impacts of the spatial scale of climate data on the modeled distribution probabilities of invasive tree species throughout the world

Ji-Zhong Wan, Chun-Jing Wang, Fei-Hai Yu *

School of Nature Conservation, Beijing Forestry University, Beijing 100083, China

ARTICLE INFO

Article history:
Received 15 July 2016
Received in revised form 1 October 2016
Accepted 4 October 2016
Available online 5 October 2016

Keywords:
Climate data scale
Distribution probability
Globe
Species distribution model
Tree invaders
Worldclim

ABSTRACT

Species distribution models (SDMs) are powerful tools to predict species distributions, and thus support invasion risk assessments for tree species at the global scale. However, SDMs may produce different species distribution probabilities depending on the spatial scale of climate data included in the model. Hence, we must understand impacts of the climate data scale on the modeled distribution probabilities of invasive tree species (ITS) throughout the world. We used nine ITS from the list of "The 100 of the World's Worst Invasive Alien Species" as our study species, and applied Maxent modeling based on presence and background points to model the distribution probabilities of these ITS across the globe using three climate data scales: 2.5, 5.0 and 10.0′. The average distribution probabilities of presence and background points across the nine focal ITS increased significantly from the 2.5 to the 10.0′ resolution, indicating that coarse climate data scales may increase the distribution probabilities of presence and background points for these focal species. The large gap between different climate data scales resulted in high prediction uncertainty for the distribution probabilities of ITS. We offer two suggestions for decreasing the prediction uncertainty of the distribution probabilities of ITS at the global scale due to the effects of the climate data scale when using SDMs: 1) use 5.0′ resolution as the input to SDMs when using GBIF or other specimen databases; and 2) decrease the gap between 2.5, 5.0 and 10.0′ in the number of presence points of ITS.

1. Introduction

Invasive tree species (ITS) have been suggested as a model group in plant invasion ecology at the global scale (Rejmánek, 2014; Rejmánek and Richardson, 2013). Previous studies have shown that climatic variables are the main driving factors shaping the global distribution patterns of ITS and may facilitate invasion of ITS via strong adaptation and rapid spread into areas of high protection value (Aguirre-Gutiérrez et al., 2015; Hof, 2015; Monahan et al., 2013; Nunez and Medley, 2011). The invasion of ITS can impact invaded systems in several ways: 1) ITS can occupy the suitable habitat of native species so that those native species may not survive; 2) ITS can change the ecological landscape and result in habitat fragmentation; 3) ITS can break the structure of communities and ecosystems (Donaldson et al., 2014; Nunez and Medley, 2011; Rejmánek, 2014; Rejmánek and Richardson, 2013; Rundel et al., 2014). Species distribution models (SDMs) are widely used to predict the global distributions of invasive plant species based on climatic variables (Donaldson et al., 2014; Mainali et al., 2015; Thuiller et al., 2005). The outputs of such modeling are used, for instance, to put forth feasibility suggestions for biological conservation and invasion risk control (Liang et al., 2014; Thuiller et al., 2005). Despite these important uses, there are still many technical challenges associated with SDMs, and solving such problems will greatly increase the prediction precision of the models and thus bolster environmental management or policymaking (Convertino et al., 2014; Mainali et al., 2015). For example, ecological transferability limits the application of SDMs for prediction of ITS distributions (Donaldson et al., 2014; Ray et al., 2016). To address this limit, ecologists have used SDMs to project the distributions of ITS based on climate data for native and invaded ranges at the global scale (Mainali et al., 2015; Shabani and Kumar, 2015).

Species distribution patterns and determinants are known to vary with the spatial scale of climate data (Rahbek and Graves, 2001; Wang et al., 2012). Reasons for this include: 1) a scale mismatch between large-scale ecological effects of climate change and species distributions with small scales of resolution (Rahbek and Graves, 2001); and 2) with the expansion of geographical extent, the explanatory power of climate variables such as environmental energy, water availability and climatic seasonality increase, while the explanatory power of habitat heterogeneity and human activities decrease (Wang et al., 2009, 2012). Therefore, projections of species distributions using SDMs may vary based on the climate data scales selected for models. Previous studies have shown that higher model performance was observed at finer data scales (Franklin et al., 2013; Gottschalk et al., 2011; Guisan et al., 2007). In comparison with the fine scale, SDMs at coarse scales may result in large prediction uncertainties for potential species distributions (Franklin et al., 2013). However, coarse-grained occurrence records, for example, from the Global Biodiversity Information Facility (GBIF),

* Corresponding author.
E-mail address: feihaiyu@bjfu.edu.cn (F.-H. Yu).

http://dx.doi.org/10.1016/j.ecoinf.2016.10.001
1574-9541/

are unable to accurately predict species' distributions at fine scales of climate data (Beck et al., 2014; Gottschalk et al., 2011; Song et al., 2013). These findings suggest that the response of species occurrence probability to different climate data scales is an important consideration for modelers estimating species distribution models at the global scale.

Here, we address two questions: 1) How do different climate data scales affect projections of distributions of ITS at the global scale? and 2) How can we reduce prediction uncertainty resulting from the impacts of the climate data scale on projections of distribution probability of ITS throughout the world? To address these questions, we selected nine ITS from the list of "The 100 of the World's Worst Invasive Alien Species" compiled by the Invasive Species Specialist Group (www.issg.org; Luque et al., 2014) as our focal study species, and used Maxent modeling, a common SDM method, to project the distributions of ITS throughout the world using three climate data scales: 2.5, 5.0 and 10.0′.

2. Methods and materials

2.1. Species data and climate data

Species with >100 occurrence records were chosen for this study to maximize the reliability of logistic SDMs (Wisz et al., 2008; Fig. 1 and Table 1). Occurrence records for the nine ITS, especially specimens or recorded sightings, were compiled from GBIF (www.gbif.org; Fig. 1).

Eight bioclimatic variables for input into the SDMs were downloaded from the WorldClim database (averaged from 1950 to 2000; www.worldclim.org; Table 2). These eight bioclimatic variables represent general trends (means), variation (seasonality), and limits (i.e. minimum and maximum) which are likely to influence the distribution and physiological performance of ITS (Hijmans and Graham, 2006). We used three spatial scales of resolutions of bioclimatic variables (2.5, 5.0 and 10.0′) because these resolutions are commonly used in SDMs (www.worldclim.org).

2.2. Species distribution modeling

We used Maxent (ver.3.3.3 k; http://www.cs.princeton.edu/~schapire/maxent/) to model the distribution of the nine ITS across the three scales of climate data based on maximum entropy (Phillips et al., 2006). Maxent modeling has the following advantages: 1) Maxent typically outperforms other methods in predictive accuracy based on the presence points (Merow et al., 2013); 2) Maxent is nonlinear, nonparametric, and not sensitive to multi-collinearity (Evangelista et al., 2011); 3) Maxent can estimate the importance of environmental variables to species distributions based on the jackknife method (Elith et al., 2011); 4) Maxent can have good prediction performance when the number of input species occurrence localities is low (Pearson et al., 2007; Wisz et al., 2008). Maxent produces a prediction map based on a logistic output format wherein cells with a value of 1 have the highest possibility of distribution, and those with a value of 0 the lowest. Species distribution areas were predicted based on similarity in climatic conditions between the study region and sites where occurrence localities have already been recorded (Merow et al., 2013). Maxent modeling may have possible applications in biological conservation, biological invasion and ecological restoration (Denoël and Ficetola, 2015; Donaldson et al., 2014; Gelviz-Gelvez et al., 2015; Thuiller et al., 2005).

When running the Maxent modeling, we removed the duplicated presence records in the same grid cell across the different scales (Elith et al., 2011; Phillips et al., 2006). The replicated run types were cross-validated to determine estimates of uncertainty for the response curves and predictions (Merow et al., 2013). We used a five-fold cross-validation approach to divide the presence dataset into five approximately equal partitions with four of the partitions used to train the model and the fifth to generate the SDM estimate (Merow et al., 2013). We set the regularization multiplier (beta) to 2.0 to produce a smooth and general response (Radosavljevic and Anderson, 2014). The convergence threshold was set to 0.0001. The maximum number of background points was 10,000, and default features were used in the model output. Other values were kept at default (after Elith et al., 2011).

We assessed the performance of the models using the area under the ROC curve (AUC). This statistic regards each value of the estimate as a possible threshold based on the corresponding sensitivity and specificity when randomly selected background points are removed from the dataset (Phillips et al., 2006). To ensure the high precision of SDM at the three spatial scales, we used SDMs with AUC values above 0.7 (Elith et al., 2011). The omission rate is the proportion of the sample units within grid cells that are predicted to be species absence within the occurrence localities (Phillips et al., 2006). These are 1-sided p-values for the null hypothesis that test points are predicted no better than a random prediction with the same fractional predicted area. The binomial probabilities were based on five common thresholds in Maxent modeling (10th percentile training presence; equal training

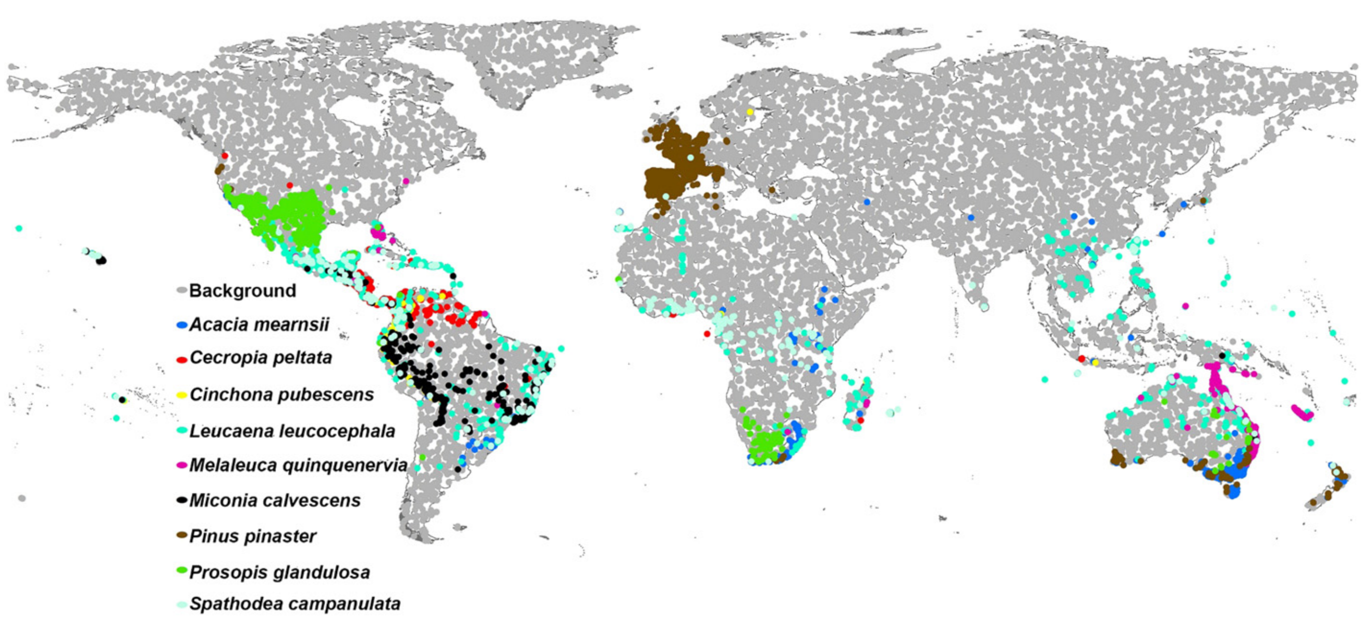

Fig. 1. Occurrence records of the nine focal invasive plant species (ITS), as well as background points.