Ecological Modelling 388 (2018) 1–9

Contents lists available at ScienceDirect

Ecological Modelling

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

The effect of pseudo-absence selection method on transferability of species distribution models in the context of non-adaptive niche shift

Wanwan Liang[a,*], Monica Papeş[b,c], Liem Tran[d], Jerome Grant[a], Robert Washington-Allen[e], Scott Stewart[a], Gregory Wiggins[c]

[a] Department of Entomology and Plant Pathology, University of Tennessee, 2505 E. J. Chapman Drive, Knoxville, TN, 37996, United States
[b] Department of Ecology and Evolutionary Biology, University of Tennessee, 569 Dabney Hall Knoxville, TN 37996, United States
[c] National Institute for Mathematical and Biological Synthesis, 1122 Volunteer Boulevard, University of Tennessee, Knoxville, TN 37996, United States
[d] Department of Geography, University of Tennessee, Burchfiel Geography Bldg. 1000 Phillip Fulmer Way, Knoxville, TN 37996, United States
[e] Department of Agriculture, Nutrition and Veterinary Sciences, University of Nevada, Reno, 1664 North Virginia Street, Reno, NV 89557, United States

ARTICLE INFO

Keywords:
Ensemble models
Ecological niche modeling
Kudzu bug
Megacopta cribraria
Invasive species
Non-adaptive niche shift

ABSTRACT

Transferability of species distribution models (SDMs) is key to predicting invasion patterns and can be challenged if niche shift occurs in the invaded range. When using native occurrences to estimate potential invasions with presence-only modeling methods, it is important to constrain the pseudo-absence (PA) sampling to the species' native range. However, some studies including highly cited ones, do not follow this approach to selecting PA samples. In this research, we addressed two questions using an invasive species in the United States (U.S.), kudzu bug (*Megacopta cribraria*): 1) is model transferability challenged by a non-adaptive niche shift? and 2) is model performance affected by use of PA samples from outside the native range of the species? Kudzu bug is native to Asia, with recently observed non-adaptive niche shift in the U.S. To answer the first question, we quantified the environmental space anisotropy and non-adaptive niche change, and then evaluated the performances of seven SDMs. To answer the second question, we further compared the interpolation and transferability of seven SDMs trained with PAs from the native range and from both native and invaded ranges. We confirmed that the environmental space anisotropy ($P = 0.01$) and non-adaptive niche change ($P = 0.01$) are both statistically significant. Of the seven SDMs used, four models had transferability indices higher than 0.9. Boosted regression tree and random forests both had good interpolation and transferability (AUC > 0.80 and kappa > 0.60), whereas three other models showed good interpolation and fair transferability (AUC > 0.70 and kappa > 0.40). Inclusion of pseudo-absences from the invaded range significantly increased the interpolation ($P < 0.001$) but decreased the transferability ($P < 0.01$) of almost all models. Our findings suggest that SDMs can show good transferability with non-adaptive niche shift, thus native occurrence information should be used in similar situation. We confirmed that it is crucial to constrain the PAs to the same spatial range as presences to accurately model potential invasions.

1. Introduction

Species distribution models (SDMs) or ecological niche models have been commonly used to predict the potential distribution of species for various purposes, including biological conservation, invasion prediction, paleobiology, spatial epidemiology, and impacts of climate change on biodiversity (Franklin, 2013; Guisan and Thuiller, 2005; Mainali et al., 2015; Svenning et al., 2011). SDMs identify relationships between observed occurrences and environmental variables by using statistical models or theoretically derived response curves (Guisan and Thuiller, 2005; Elith and Franklin, 2013). The ecological niche, more specifically the fundamental niche (FN), or a subset of it, has been increasingly used to estimate species' geographic extent (Guisan and Zimmermann, 2000; Peterson and Soberón, 2012). The ecological niche is described as a multi-dimensional space, including both abiotic and biotic factors, that permits positive growth of a given species, whereas the FN delineates only the abiotic environmental conditions (Hutchinson, 1957; Pearson and Dawson, 2003). Estimating the potential distribution of a species across different spatial and temporal scales has become an increasing practical use of SDMs (Elith and Leathwick, 2009; Mainali et al., 2015). In this practice, SDMs are usually trained with occurrence information and environmental

* Corresponding author.
E-mail address: wliang@vols.utk.edu (W. Liang).

https://doi.org/10.1016/j.ecolmodel.2018.09.018
Received 24 March 2018; Received in revised form 16 September 2018; Accepted 18 September 2018
0304-3800/

predictors from one spatio-temporal range, and then projected to a different range to identify potential distributional areas of a given species (Peterson, 2003; Verbruggen et al., 2013). One underlying assumption of this practice is that the given species conserves its niche across different spatial and temporal scales (Wiens and Graham, 2005). Another assumption is that the given species is in equilibrium with the environment in the spatial range from where the occurrence information is extracted for model training (Elith and Leathwick, 2009; Gallien et al., 2012).

Estimating potential biological invasions has become an important tool to manage non-native species. To predict the potential invasion pattern of a given species in a new range, native occurrence information is generally used in SDMs. Thus, successfully predicting invasion patterns closely relies on the transferability of models, which is defined as the ability of SDMs to predict occurrence in a largely unsampled spatial range or time period (Heikkinen et al., 2012; Randin et al., 2006). However, several researchers have suggested that species could shift their niche in new spatial ranges (Broennimann et al., 2007; Early and Sax, 2014; Gallagher et al., 2010; Medley, 2010; Parravicini et al., 2015), which challenges the transferability of SDMs as the same species may survive under different environmental conditions in the invaded ranges (Broennimann et al., 2007; Early and Sax, 2014; Parravicini et al., 2015). As a result, the use of only native occurrence information to model potential invasion is debatable. Broennimann and Guisan (2008) and Jiménez-Valverde et al. (2011) suggested that SDMs for estimating potential invasions could be developed using occurrence information from both native and invaded ranges. However, the drawback of using occurrence information from both ranges is that the equilibrium assumption is likely violated if the invasion is still ongoing (Early and Sax, 2014; Elith and Leathwick, 2009; Jiménez-Valverde et al., 2011).

A study of 50 terrestrial plant invaders, however, suggested that substantial niche shifts are rare (Petitpierre et al., 2012), while the same conclusion was also made for birds and other taxa (Peterson, 2011; Strubbe et al., 2013). Soberón and Peterson (2011) suggested that, in some cases, the “niche shift” is more likely a result of differences in environmental conditions, or environmental space anisotropy, between two spatial ranges than a true, adaptive niche shift of species. The concern that the non-adaptive niche shift would affect the transferability of SDMs has not been fully addressed. Specifically, it is unclear whether both native and invaded range occurrence datasets are needed to estimate invasion patterns when a non-adaptive niche shift exists, given the drawback of invaded range data violating the equilibrium assumption underlying SDMs.

When estimating potential invasion patterns with presence-only data, the spatial range from which pseudo-absences (PAs, also called background data) are extracted certainly impacts the transferability of SDMs (Phillips, 2008; Barbet-Massin et al., 2012). For studies that use only occurrences from the native range of a species to train models, the PAs also should be restricted to the native range (Peterson, 2003; Phillips, 2008). However, in several highly cited studies, the PAs were extracted from both the native and invaded ranges when only native occurrences were used to construct the SDMs (Broennimann and Guisan, 2008; De Meyer et al., 2010; Gallardo et al., 2013; Verbruggen et al., 2013). Additionally, in some studies, the methods used to extract the PAs were not mentioned (Fitzpatrick et al., 2007; Loo et al., 2007; Mau-Crimmins et al., 2006). Evaluation of the impact of PA data on model performance is not rare (Anderson and Raza, 2010; Barbet-Massin et al., 2012; Zhu et al., 2014). However, to the best of our knowledge, what is missing is a systematic and quantitative assessment of the impact of selecting PAs from both native and invaded ranges, while only native occurrences are used, on both interpolations and transferability of multiple commonly used models. This information is important to consider, especially when niche shift is observed between the two ranges.

Our study had two objectives: 1) evaluate the interpolative accuracy and, more importantly, transferability (also called extrapolative accuracy) of SDMs obtained with seven commonly used techniques, under non-adaptive niche shift between native and invaded ranges, and 2) examine how inclusion of PAs from the invaded range, when only native occurrences are used for model training, impacts model interpolation and transferability. We conducted our research with kudzu bug (*Megacopta cribraria*; Hemiptera: Plataspidae) for several reasons. First, kudzu bug is a newly invasive species in the United States (U.S.), but is a well studied species in Asia, its native range. Kudzu bug spread fast, such that it had been reported in more than 650 counties in the U.S. by the end of 2017, ensuring enough occurrence data in the invaded range for SDMs. Additionally, a non-adaptive niche shift was observed between native and invaded populations of kudzu bug in the U.S. (Liang et al., 2018). Modeling the potential distribution of kudzu bug is also important for practical applications. Results of this research can provide valuable information for management of kudzu bug, as it has become a pest in both agricultural and urban areas (Eger et al., 2010). To fulfill the first objective, we quantified the differences of both available and occupied environmental spaces between native and invaded ranges of kudzu bug. As empirical research, these findings can be used to guide selection of SDMs when a non-adaptive niche shift is observed and to select PAs for accurate modeling of species’ distribution.

2. Methodology

2.1. Presence and pseudo-absence data

2.1.1. Presence

To compare the interpolation and transferability of different models, we collected point presence data of kudzu bug in both native and invaded ranges. Readers are referred to Liang et al. (2018) and Zhu et al. (2012) for detailed information on data acquisition. We deleted redundant observations to ensure only one observation per 1000-m grid. In total, we generated three datasets: Dataset I included 164 presence records in the native range (Asia), Dataset II contained 152 presence records in invaded range of kudzu bug (U.S.), and Dataset III combined occurrences from both the native and invaded ranges and therefore had 316 presence records (Fig. 1). All occurrence data are available to readers upon request.

2.1.2. Pseudo-absence (PA)

To address the lack of absence data, we generated 10,000 Pa for model training and intrinsic evaluation. To quantify and compare the accuracy of model interpolation and transferability, we extracted PAs in the same spatial range as the presences used for model training. However, to fulfill the second objective, we generated Dataset IV, which included the presences from Asia but 10,000 Pa extracted from both Asian and U.S. ranges.

2.2. Environmental variables

To characterize the environmental space in both ranges (native and invaded), we used an elevation variable from the Hydro-1k digital elevation model dataset (USGS, 1996) and nine climatic variables from the “WorldClim” dataset (Hijmans et al., 2005). Instead of using all 19 climatic variables available in WorldClim, we selected only nine variables that were not highly correlated ($r < 0.8$) based on the climatic conditions of 10,000 randomly selected points. The nine climatic variables are mean diurnal temperature range (BIO2), maximum temperature of the warmest month (BIO5), mean temperature of the wettest quarter (BIO8), mean temperature of the coldest quarter (BIO11), annual mean precipitation (BIO12), precipitation seasonality (BIO15), precipitation of the driest quarter (BIO17), precipitation of the warmest quarter (BIO18), and precipitation of the coldest quarter (BIO19). All 10 variables (elevation and climate) used had a resolution of 30 arc-seconds (approximately 1 km).