Ecological Modelling 269 (2013) 9–17

Contents lists available at ScienceDirect

Ecological Modelling

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

Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample sizes

Mariya Shcheglovitova[a,*,1], Robert P. Anderson[a,b,c]

[a] Department of Biology, City College of the City University of New York, 160 Convent Avenue, New York, NY 10031, USA
[b] Graduate Center, City University of New York, 365 5th Avenue, New York, NY 10016, USA
[c] Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA

ARTICLE INFO

Article history:
Received 16 May 2013
Received in revised form 19 August 2013
Accepted 20 August 2013
Available online 19 September 2013

Keywords:
Ecological niche model
Feature class
Jackknife
Maxent
Species distribution model
Tuning

ABSTRACT

Algorithms for producing ecological niche models and species distribution models are widely applied in biogeography and conservation biology. However, in some cases models produced by these algorithms may not represent optimal levels of complexity and, hence, likely either overestimate or underestimate the species' ecological tolerances. Here, we evaluate a delete-one jackknife approach for tuning model settings to approximate optimal model complexity and enhance predictions for datasets with few (here, <10) occurrence records. We apply this approach to tune two settings that regulate model complexity (feature class and regularization multiplier) in the presence-background modeling program Maxent for two species of spiny pocket mice in Ecuador and southwestern Colombia. For these datasets, we identified an optimal feature class parameter that is more complex than the default. Highly complex features are not typically recommended for use with small sample sizes in Maxent. However, when coupled with higher regularization, complex features (that allow more flexible responses to environmental variables) can obtain models that out-perform those built using default settings (employing less complex feature classes). Although small sample sizes remain a serious limitation to model building, this jackknife optimization approach can be used for species with few localities (<approximately 20–25) to produce models that maximize the utility of the little information available.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Ecological niche models (ENMs) and species distribution models (SDMs) based on presence-only occurrence data constitute widely used tools for many areas of biogeographic research, as well as for conservation planning (Papeş and Gaubert, 2007; Wilting et al., 2010; Lawler et al., 2011; Anderson, 2013). Here, we follow the paradigm of ecological niche modeling of the conditions suitable for the species in model calibration, evaluation, and interpretation (Peterson et al., 2011; Anderson, 2012). However, the methodological advances we apply are equally applicable to models aimed at characterizing the species' occupied distribution (SDMs, *sensu stricto*). ENMs examine associations between known occurrences of a species and abiotic environmental (often climatic) data in the geographic region of interest. The resulting model approximates the environmental conditions that the species can inhabit (the species' existing fundamental niche, subject to clear assumptions); that model then can be applied to geography, yielding estimates of the corresponding areas with suitable environmental conditions (its abiotically suitable distribution; see Peterson et al. (2011) for terminology and assumptions regarding the characteristics of occurrence and environmental data).

Despite their broad appeal, ENMs may be especially problematic when implemented with species for which few occurrence records exist; nevertheless, such situations often correspond to precisely the species most in need of predictive models for conservation-based initiatives (Gaubert et al., 2006). Specifically, model accuracy decreases and model variability increases with decreasing sample size (Wisz et al., 2008). If possible, the paucity of occurrence data should be rectified by increasing efforts put into field surveys and data sharing (Cayuela et al., 2009). However, this seldom is feasible in the time frame within which conservation decisions need to be made. As an alternative, optimizing or tuning model settings (sometimes called "smoothing") to estimate optimal model complexity can result in higher-quality output than employing default settings (Elith et al., 2010; Anderson and Gonzalez, 2011; Warren and Seifert, 2011; Radosavljevic and Anderson, in press). Furthermore, optimal settings likely vary among species as well as for different combinations of the occurrence localities, study region, and environmental data at hand. Therefore, we explored model tuning as a

* Corresponding author Tel.: +1 347 410 3228; fax: +1 202 994 6100.
E-mail address: mshcheg@gwu.edu (M. Shcheglovitova).
[1] Present address: Department of Biological Sciences, George Washington University, 2023 G St. NW, Washington, DC 20052, USA.

0304-3800/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2013.08.011

way of improving ENMs for datasets with few occurrence records. In particular, we used a delete-one jackknife approach suggested for model evaluation recently (a form of *k*-fold cross validation where *k* is equal to the number of occurrence localities in the original dataset; Peterson et al., 2011; see also Pearson et al., 2007). Although this approach may also be useful for higher sample sizes (e.g., up to ca. 25 records), we here employ it for species with very few records (<10).

As an assessment of this approach, we used the presence-background modeling software Maxent (Phillips et al., 2006) to generate ENMs for two species of spiny pocket mice across a range of program settings (Supplementary Fig. 3). We compared the performance of default settings to a variety of user-specified settings. Maxent identifies geographic areas of suitable conditions for a species, based on known occurrence records, by applying a maximum entropy model to estimate the species' response given a set of constraints (environmental variables). We chose Maxent because it: (1) is in common use; and (2) has been found to perform well for small sample sizes in previous studies (Wisz et al., 2008); yet, (3) is sensitive to model settings that affect model complexity (Elith et al., 2010; Anderson and Gonzalez, 2011; Warren and Seifert, 2011; Syfert et al., 2013). In the tuning experiments that led to the current default settings, Phillips and Dudík (2008) stated that for datasets unlike those used in that study, it may be necessary to use further tuning to optimize the program's performance. Even though we tested our approach using Maxent, this jackknife approach for model tuning with small sample sizes is general and can be extended to other modeling methods. We assessed models based on quantitative evaluations of performance, and compared optimal to default models using measures of similarity. Independently, we evaluated model output qualitatively.

2. Materials and methods

2.1. Study species and region

We used two species of spiny pocket mice, *Heteromys australis* and *Heteromys teleus* (Rodentia: *Heteromyidae*), to conduct our tuning experiments. These species represent suitable entities for the current study for several reasons. Recent taxonomic research provides high-quality (although limited) occurrence data, as well as general natural-history information regarding the habitats occupied by the species. Furthermore, strong climatic gradients exist in the regions occupied by these species, facilitating both model calibration and interpretation.

Both species inhabit western Ecuador. In addition, the range of *H. australis* extends into Colombia, eastern Panama, and western Venezuela (Anderson and Jarrín-V, 2002 Fig. 1). In northwestern Ecuador and southwestern Colombia, *H. australis* can be found in very wet and unseasonal evergreen forests, while *H. teleus* inhabits slightly drier and markedly seasonal, but still evergreen forests in central–western Ecuador (Anderson and Jarrín-V, 2002). Both species occur in a wide range of altitudes on the Pacific coastal lowlands and western slopes of the Andes (from up to ca. 2000 m; Anderson and Jarrín-V, 2002). Preliminary conservation assessments were undertaken several years ago for these species in Ecuador using cruder climatic data and a different modeling method (Anderson and Martínez-Meyer, 2004). Our aim here is to explore model complexity with Maxent, leaving conservation-related questions for these species to other ongoing studies (Burneo, pers. comm.).

We modeled the environmental requirements of *H. teleus* in its full known distribution and those for *H. australis* in part of its range (Ecuador and southwestern Colombia). We did so for *H. australis* because high-quality occurrence data exist for it in this region. We acknowledge that the present dataset may underestimate the species' full environmental tolerances. For each species, we delimited a rectangle that surrounds the occurrence records, specifically one whose borders were the nearest half degree from the most-peripheral occurrence record after filtering (0.5°S to 2°N, 77.5–80°W for *H. australis*; 2.5°S to 1°N, 78.5–81.5° W for *H. teleus*; see Section 2.2 for a description of filtering).

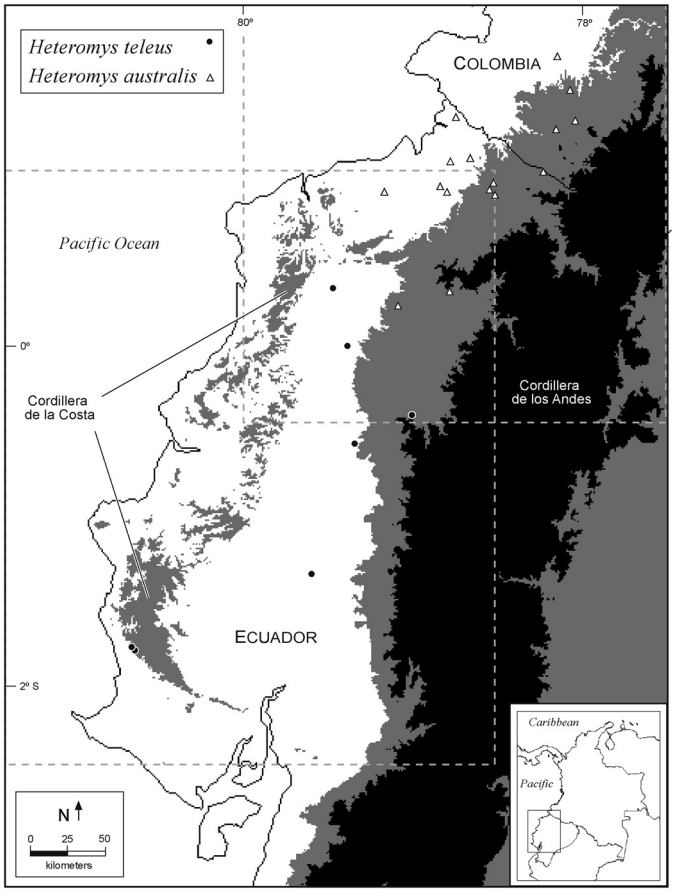

Fig. 1. Map showing all (unfiltered) occurrence records for *Heteromys australis* (triangles) and *H. teleus* (circles) in Ecuador and southwestern Colombia (data from Anderson and Jarrín-V, 2002). Regions above 300 m are shown in gray, and areas above 2000 m appear in black. The dashed boxes indicate the study regions used here in modeling the abiotically suitable areas for each species using spatially filtered localities.

2.2. Occurrence and environmental data

We created jackknife sets for each species after filtering occurrence records to reduce the likely effects of spatial autocorrelation due to biased sampling typical of museum and herbarium data (Peterson et al., 2011; Anderson, 2012). The probable environmental bias introduced by spatially autocorrelated occurrence records has been observed to affect model complexity (Anderson and Gonzalez, 2011). Additionally, Maxent tends to produce overfit predictions when used with biased occurrence records (Peterson et al., 2007). An overfit model is more complex than the true relationships between the species' niche and the examined environmental variables (Peterson et al., 2011). Had we not filtered occurrence records, geographically proximate records with similar environmental characteristics may have led to inflated estimates of performance (Veloz, 2009) and, therefore, to selection of overly complex models as optimal.

To filter occurrence records, we only retained those with a linear distance more than 30 km to neighboring records, such that

Download English Version:

https://daneshyari.com/en/article/4376039

Download Persian Version:

https://daneshyari.com/article/4376039

Daneshyari.com