Ecological Modelling 345 (2017) 10–20

Contents lists available at ScienceDirect

Ecological Modelling

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

Comparison of climate envelope models developed using expert-selected variables versus statistical selection

Laura A. Brandt[a,*], Allison M. Benscoter[b,1], Rebecca Harvey[b], Carolina Speroterra[b], David Bucklin[b], Stephanie S. Romañach[c], James I. Watling[b,2], Frank J. Mazzotti[b]

[a] U.S. Fish and Wildlife Service, 3205 College Avenue, Ft Lauderdale, FL 33314, USA
[b] University of Florida, Department of Wildlife Ecology and Conservation, Ft Lauderdale Research and Education Center, 3205 College Avenue, Ft Lauderdale, FL 33314, USA
[c] U.S. Geological Survey, 3321 College Avenue, Ft Lauderdale, FL 33314, USA

ARTICLE INFO

Article history:
Received 21 July 2016
Received in revised form 3 November 2016
Accepted 30 November 2016
Available online 14 December 2016

Keywords:
Climate adaptation
Conservation planning
Expert opinion
Florida
Threatened and endangered species

ABSTRACT

Climate envelope models are widely used to describe potential future distribution of species under different climate change scenarios. It is broadly recognized that there are both strengths and limitations to using climate envelope models and that outcomes are sensitive to initial assumptions, inputs, and modeling methods Selection of predictor variables, a central step in modeling, is one of the areas where different techniques can yield varying results. Selection of climate variables to use as predictors is often done using statistical approaches that develop correlations between occurrences and climate data. These approaches have received criticism in that they rely on the statistical properties of the data rather than directly incorporating biological information about species responses to temperature and precipitation. We evaluated and compared models and prediction maps for 15 threatened or endangered species in Florida based on two variable selection techniques: expert opinion and a statistical method. We compared model performance between these two approaches for contemporary predictions, and the spatial correlation, spatial overlap and area predicted for contemporary and future climate predictions. In general, experts identified more variables as being important than the statistical method and there was low overlap in the variable sets (<40%) between the two methods Despite these differences in variable sets (expert versus statistical), models had high performance metrics (>0.9 for area under the curve (AUC) and >0.7 for true skill statistic (TSS). Spatial overlap, which compares the spatial configuration between maps constructed using the different variable selection techniques, was only moderate overall (about 60%), with a great deal of variability across species. Difference in spatial overlap was even greater under future climate projections, indicating additional divergence of model outputs from different variable selection techniques. Our work is in agreement with other studies which have found that for broad-scale species distribution modeling, using statistical methods of variable selection is a useful first step, especially when there is a need to model a large number of species or expert knowledge of the species is limited. Expert input can then be used to refine models that seem unrealistic or for species that experts believe are particularly sensitive to change. It also emphasizes the importance of using multiple models to reduce uncertainty and improve map outputs for conservation planning. Where outputs overlap or show the same direction of change there is greater certainty in the predictions. Areas of disagreement can be used for learning by asking why the models do not agree, and may highlight areas where additional on-the-ground data collection could improve the models.

Published by Elsevier B.V.

1. Introduction

Climate change is creating new challenges for conservation. Within the next century it is expected to become one of the primary drivers of global biodiversity loss (Sala et al., 2000; Thomas et al., 2004; Urban, 2015). There are documented cases of species range shifts linked to changing climate (Chen et al., 2011; Parmesan and Yohe, 2003; Root et al., 2003) and climate change may have already

* Corresponding author.
E-mail addresses: Laura_brandt@fws.gov (L.A. Brandt), sromanach@usgs.gov (S.S. Romañach).

[1] Current address: U.S. Geological Survey, 3321 College Avenue, Ft Lauderdale, FL 33314 USA

[2] Current address: John Carroll University, University Heights, OH 44118 USA.

http://dx.doi.org/10.1016/j.ecolmodel.2016.11.016
0304-3800/Published by Elsevier B.V.

resulted in species extinction (Cahill et al., 2013; McLaughlin et al., 2002; Pounds et al., 2006). The conservation community has recognized that existing strategies for landscape and species protection (including existing protected area networks) may not be effective in the future because of shifting species distributions (Heller and Zavaleta, 2009; Kostyack et al., 2011). Efforts are underway to develop adaptation strategies (Glick et al., 2009) that will help to assess and respond to conservation challenges associated with climate change. Development of successful adaptation plans in the face of uncertainty requires tools for assessing climate change impacts and vulnerabilities of species and habitats. Models are one way to do this, and climate envelope models, a subset of species distribution models, are becoming more widely used in vulnerability assessments and adaptation planning (Franklin, 2013).

Climate envelopes for species are developed by correlating species occurrences with selected climate variables (Beaumont and Hughes, 2002; Berry et al., 2002; Huntley et al., 2010; Pearson and Dawson, 2003; Thuiller, 2003). They can be used to describe historical, current, and future potential climate space for species. The resulting maps of potential climate space are based on two assumptions: 1) climate variables play an important role in defining a species geographic range (Lomolino et al., 2005), and 2) empirical relationships between contemporary distributions of species and climate can be used to forecast species distributions under future climate change scenarios (Araújo and Peterson, 2012; Franklin, 2010). Climate envelope models require relatively little data on the specific biology of a species (they rely on correlating occurrence data with climate variables such as monthly temperatures and precipitation) and therefore can be developed for many species over broad geographic areas fairly rapidly (Lawler et al., 2006). Although climate envelope models can provide useful information, they also have received substantial criticism (Araújo and Guisan, 2006; Araujo and Peterson, 2012; Beale et al., 2008), in part because they do not incorporate specific biological information or consider all relevant factors that determine a species range (Real et al., 2013). There is increasing recognition that different modeling techniques and different inputs can yield different results (Baker et al., 2016; Elith and Graham, 2009; Synes and Osborne, 2011; Watling et al., 2012b), and that there is a need for a better understanding of the strengths, weaknesses, and sensitivity of resulting maps to initial assumptions and inputs (Araujo and Guisan, 2006; Whittaker et al., 2005).

Selection of climate predictor variables is a central step in climate envelope modeling (Austin and Van Niel, 2011; Harris et al., 2013). There are a number of ways to select variables, including automated statistical techniques or *a priori* selection of variables based on expert knowledge, where experts are individuals who have documented extensive knowledge about the subject. An advantage of automated approaches is that many species can be evaluated quickly. A disadvantage is that the resulting variables may be biologically implausible or irrelevant (Heikkinen et al., 2006). Selection of variables using expert knowledge also has advantages and disadvantages. Natural resource managers may be more comfortable with models developed through expert input than with those developed only by statistical methods because they have established relationships with trusted experts or are unfamiliar with statistical techniques (Addison et al., 2013). In addition, expert-selected variables may be more closely tied to empirical biophysical tolerances of species. However, there are shortcomings of using experts, probably the most important ones being bias and functional fixedness (Chi, 2006). Experts may be biased when the species occurs outside of the geographical area with which they are familiar resulting in variables only reflecting a subset of environmental conditions experienced by a species. Experts may become fixated on ideas that are familiar and find it challenging to accept new ideas or consider novel species environmental relationships that might occur with climate change. In addition, consultation with experts takes considerable time, experts may not agree on which variables are important, or experts may not exist for all species that need to be modeled, limiting the number of species that can be modeled in a given timeframe.

Performance metrics such as area under the receiver operating curve (ROC), known as AUC, Cohen's kappa, and TSS (True Skill Statistic) are used to quantify prediction accuracy of model outputs and numerous studies have examined how accuracy is affected by different inputs (see for example, Elith et al., 2006; Guisan et al., 2007; Hernandez et al., 2006; Segurado and Araújo, 2004). However, fewer studies (Bagchi et al., 2013; Baker et al., 2015; Baker et al., 2016; Bucklin et al., 2015; Syphard and Franklin, 2009; Watling et al., 2012b) have examined specifically how spatial configuration (spatial correlation and spatial overlap) of prediction maps varies even when performance metrics are high. Map outputs from models using different inputs might have equally high performance metrics but look different. Because in the context of natural resource management it is often the resulting map that is used for planning and management decisions, variations in spatial characteristics of prediction maps (which areas appear as high suitability) have the potential to result in different planning and management decisions.

We initiated this project to address concerns that models created via statistical variable selection would not be useful because of the belief that they would not reflect the ecology of the species. We did this by examining how climate envelope models and resulting prediction maps for 15 threatened or endangered species in Florida differed between two methods: 1) using predictor variables (temperature and precipitation) selected by experts as most important in describing the species climate envelopes, and 2) using variables selected by a statistical method. We evaluated model performance using traditional performance metrics of AUC and TSS. We projected model results to contemporary and future conditions and analyzed how spatial predictions (correlation of map suitability, spatial overlap, and area of map suitability) differed between the variable selection techniques. We compared similarity between expertly and statistically selected variable sets. We also used information gathered from experts on importance of temperature and precipitation for determining the range of the species and experts' level of confidence in variable selection, to examine if there were differences in confidence in variable selection or model outputs between models produced for species for which experts believed temperature or precipitation were very important compared to species for which they believed temperature or precipitation were not important.

2. Methods

Climate envelope models develop associations between species occurrence and environmental conditions described by climate data. Climate data are described by a set of predictor variables such as average temperature or average precipitation. Once these associations are developed an index of environmental suitability or probability of occurrence for the modeled species can be developed, tested, and projected spatially in the form of a map. Maps can be developed for both contemporary and future conditions. Below we describe how we applied these steps.

2.1. Input data for climate envelope models

2.1.1. Species' occurrence data

Our study species consisted of 15 terrestrial vertebrates classified as federally threatened or endangered species in the United States (Table 1), and contain either all or some of their distribu-