Journal for Nature Conservation 28 (2015) 26–34

Contents lists available at ScienceDirect

Journal for Nature Conservation

journal homepage: www.elsevier.de/jnc

Niche modeling for management-ready information in little-studied, threatened frog species assemblages

Iris Holmes*, Kurt McLaren, Byron Wilson

The Department of Life Sciences, The University of the West Indies, Mona, Kingston 7, Jamaica

ARTICLE INFO

Article history:
Received 14 September 2014
Received in revised form 10 August 2015
Accepted 25 August 2015

Keywords:
Anurans
Climate change
Niche modeling
Frogs
Chytrid fungus
Conservation
Biodiversity

ABSTRACT

We use species distribution modeling to create easily testable hypotheses about the current and future distributions of Jamaican frogs, a little studied but highly endangered group. Our models simultaneously represent the best possible current estimate of the frogs' ranges and provide clear guidelines for future survey work and habitat preservation efforts. We identify areas that contain the highest frog biodiversity, the highest per-unit area frog conservation benefit, and areas that are putative climatic refuges from outbreaks of the frog disease chytridiomycosis. In addition, we use the distribution models to create a set of easily falsifiable predictions about frog presence or absence. Testing these predictions using presence/absence surveys will provide management-ready information about model quality, population trajectories, changes in realized climate tolerance, and disease presence. We present a method of generating targeted conservation recommendations that will be applicable to many little-studied, cryptic taxa worldwide.

1. Introduction

Effective conservation management requires both money and time. These resources are finite (Ehlers Smith, 2014), so managers must prioritize their application. For many taxa, we have little data on which to base our priority setting. An estimated 11% of fish, 33% of amphibians, 13% of reptile species have yet to be discovered (Scheffers, Joppa, Pimm, & Laurance, 2012). Of described species, 24.5% of amphibians, 15.2% of mammals, 18.3% of reptiles, 19.4% of ray-finned fishes are listed as Data Deficient by the IUCN (IUCN, 2013). The future of many little-studied taxonomic groups is uncertain, particularly considering the threats of global climate change (Keith et al., 2014; Gerick, Munshaw, Palen, Combes, & O'Regan, 2014), habitat destruction (Yahaya, Attuquayefie, Owusu, Holbech, & Ofori, 2013), and emerging infectious diseases (Fürst, McMahon, Osborne, Paxton, & Brown, 2014; Heard et al., 2013). Thus, we must get the greatest possible conservation value out of any available data, in the understanding that it may not be possible to collect more. In many species, collection locations from museum records make up the bulk of the available data (McPherson, 2014). Here, we use this presence-only data to direct future data-gathering surveys and prioritize habitat for preservation.

Species distribution modeling algorithms take the collection locations for a species and identify the values of a set of explanatory environmental variables occurring at those sites (McPherson, 2014). They return projected probabilities of presence for that species across a predefined study area, dependent on known explanatory variables. Provided that geographical distribution of the species is well represented by the sampled sites, the model should identify a hypervolume along the explanatory variable axes that accurately represents the realized niche of the species of interest (Hutchinson, 1991). The model will identify locations with variable values that fall within the hypervolume as having a higher probability of presence than those with values outside the hypervolume. The model returns an estimate for probability of presence of the species of interest across the area of interest.

In practice, many confounding factors can affect the results of an SDM. If unmeasured variables or historical events constrain the range of the species, the projection may extend beyond the true range. Species may be absent from areas predicted by niche modeling due to extirpations, biotic interactions, or because their site of origin is separated from another piece of viable habitat by a barrier (Robinson et al., 2011). Due to the large number of factors that can influence the results of species distribution models, in this paper we refer to their results as returning a projected geographi-

---

* Corresponding author. Permanent address: University of Michigan, Ann Arbor, University of Michigan Museum of Natural History, Ruthven Museums Building, 1109 Geddes Ave., Ann Arbor, MI 48109-1097, United States. Fax: +1 734 763 4080.

E-mail addresses:
iholmes@umich.edu, iah6@cornell.edu (I. Holmes), kurt.mclaren@uwimona.edu.jm (K. McLaren), byron.wilson@uwimona.edu.jm (B. Wilson).

http://dx.doi.org/10.1016/j.jnc.2015.08.005

cal distribution, rather than a projected realized niche. Further, we use the phrase "observed climate span" to describe the maximum and minimum values of a given explanatory variable found at the known collection locations. We use this vocabulary to clarify that our results are hypotheses about observable distributions and the management best practices that can be inferred from them, rather theoretical niche volume.

Despite these limitations, we investigate these methods because they allow data analysis even in little-known species. In low-information species, the majority of the available data may come from locality data attached to museum collections. The challenges of using museum data are well documented, but for many species there are few other sources of ecological information (Naimi, Hamm, Groen, Skidmore, & Toxopeus, 2014; Moudrý and Šímová, 2012). Species distribution modeling algorithms assume that the data they are provided are a random sample of the complete niche of the focal species (Phillips et al., 2009). This assumption may not be met if the data are preferentially collected in one portion of the species' range over another. Due to the logistical constraints of fieldwork, occurrence data are often collected along roads, or in a radius around a central camp. While the spatial autocorrelation in collection locations can be accounted for to an extent, it may still affect the predictions (Moudrý and Šímová, 2012). Despite these limitations, common to many museum collections, our species distribution models can make explicit, management-ready ecological predictions for little-studied groups of organisms.

In this paper, we acknowledge the limits of the available data. Our aim is to design a framework for using the data specifically for conservation managers. We propose a first test of a framework for data use loosely based on the philosophy of Bayesian statistics—we begin with a prior assumption, compare it to data, and update that prior to reflect the information in the data. In the absence of data, managers have no prior knowledge about which habitat patches have high conservation value. We use the data to alter our preconceptions to weight certain patches as more valuable. The goal of data analysis is specifically to improve the manager's knowledge and potential to act in a beneficial manner, rather than to approach a perfect geographical or climatic model for a given species

We chose our three species-distribution modeling algorithms to have complementary assumptions, such that if the assumptions of one model are violated, those of the others may not be. MaxEnt uses a maximum-entropy learning algorithm to create niche models from presence-only data (Phillips et al., 2004). It takes from the mean and variance of the sampled climate variables, and builds functions describing the relationship between the values of those climate variables and the probability of presence of the species of interest. The mean and variance of these functions must match the sampled means and variances of the variable, and the functions must be as uniform as possible (have maximum entropy). Probability of presence of the species at a given locality is calculated based on the functions described for each climate variable (Phillips, Dudík, & Schapire, 2004). The method is widely used for presence-only data, making our results comparable to many other studies. Random Forest (Breiman, 2001) and Boosted Regression Trees (Elith, Leathwick, & Hastie, 2008) function by splitting the data into maximally uniform groups according to explanatory climate variables. Random Forests draw random samples of data and explanatory variables to build trees, then average predictions over those trees. This approach avoids over fitting but can miss fine details. Boosted Regression Trees use a similar tree building approach but focus on placing difficult to fit data points, offering an opportunity to use patchily collected records to best possible effect. In sum, we believe that these three algorithms complement each other well, and provide the best possible opportunity for producing immediately useful modeling results.

We focus on endemic Jamaican frogs, which are at risk from all three types of threat (Wilson, 2011). These threats are disease, specifically the fungal pathogen *Batrachochytrium dendrobatidis*, invasive species, and habitat loss (Wilson, 2011). Many other taxa are also known largely from museum collections and face similar conservation problems as the Jamaican frogs (IUCN, 2013). Our methods can be generalized to these taxa. Jamaican frogs are an excellent example of low-information group, as the majority of our ecological knowledge about Jamaican frogs is drawn from locality data from museum collections (Habel, Husemann, Finger, Danley, & Zachos, 2014).

Jamaican frogs are an important component of their forest ecosystems (Beard, Eschtruth, Vogt, Vogt, & Scatena, 2003). Their extinction or severe reduction could have disruptive consequences for many other species (Connelly et al., 2014). Jamaican frogs represent two endemic radiations within Jamaica, of seventeen species of *Eleutherodactylus* and four of *Osteopilus* (Hedges 1989; Moen and Wiens, 2009). Over its geological history, Jamaica has at some points been comprised of two separate islands, which now exist as mountains on the eastern and western sides of the island, with low-lying limestone karst forest separating them (Haq, Hardenbol, & Vail, 1987). Evolutionarily, the Jamaican *Eleutherodactylus* comprise two old radiations, one on each former island, and a newer one in the intervening low elevation forest (Hedges, 1989). Like the better-studied *Anolis* radiations in the Caribbean, this natural experiment could yield considerable insight about the process of evolution and community assembly in island radiations. In order to be of use as an evolutionary model system, the Jamaican frog radiation must be protected.

Native species across the Caribbean are threatened by habitat loss, including Jamaican frogs. Few species are able to survive outside closed-canopy tropical forest. Climate change, both in temperature and precipitation, is a stressor for many species worldwide. Island species may be particularly at risk, as they have limited room to shift their range to match changing temperatures. Finally, the amphibian chytrid fungus, which has caused worldwide amphibian extinctions and declines (Eskew and Todd, 2013), is present in Jamaica and will be an ongoing conservation concern. With the goal of improving conservation outcomes, we present a set of testable hypotheses that will allow managers to maximize the use of available resources that can be devoted to habitat preservation or further data gathering. We designed these hypotheses such that they provide immediately actionable information whether they fail or are supported.

1.1. *Testable hypotheses*

Hypothesis one: adding new sampling localities will not significantly change the predictions of the geographical distribution or observable climate span for any Jamaican frog.

The predictive success of our models, and thus their utility for further analyses, depends on the samples being as representative as possible of the climate variables present full geographic range of the target species. We can test this hypothesis by establishing the presence or absence of the target species in areas that are at the edge of the predicted range of climate variation, in areas that have one or more climate variables outside the sampled range for the species. We encourage best practices for accounting for false negatives, including use of occupancy models (Zipkin, Grant, & Fagan, 2012) and better understanding of each species' thermal tolerance in a laboratory setting (Ruiz-Aravena et al., 2014), when funds and time allow. If we find the species in this marginal area, we can re-calculate the distribution model with the new, more extreme presence data and iterate the hypothesis test until a predicted area is free of the focal species, indicating that we have reached the limits of model improvement using new data.