Ecological Modelling 248 (2013) 57–70

Contents lists available at SciVerse ScienceDirect

Ecological Modelling

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

Comparing the relative contributions of biotic and abiotic factors as mediators of species' distributions

Constantino González-Salazar[a,c,*], Christopher R. Stephens[b,c,1], Pablo A. Marquet[d,e,f,2]

[a] Instituto de Biología, Universidad Nacional Autónoma de México, 04510 México, D.F., Mexico
[b] Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, 04510 México, D.F., Mexico
[c] C3 – Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, 04510 México, D.F., Mexico
[d] Center for Advanced Studies in Ecology and Biodiversity (CASEB) and Departamento de Ecología, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile
[e] Instituto de Ecología y Biodiversidad (IEB) Casilla 653, Santiago, Chile
[f] The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501 USA

ARTICLE INFO

Article history:
Received 6 February 2012
Received in revised form 1 September 2012
Accepted 7 October 2012
Available online 9 November 2012

Keywords:
Biotic interaction
Data mining
Environmental data
Ecological niche
Ecological modeling
Species distribution

ABSTRACT

Models to predict species' ranges have chiefly been limited to abiotic variables. However, the full ecological niche depends on a myriad of factors, both biotic and abiotic, that often correspond to completely different data types. We applied a methodology based on data mining techniques to construct ecological niche models composed of biotic as well as abiotic variables using three quite different sets of variables: climatic layers, maps of land cover and point collections of Mexican mammals. We show how potential ecological interactions can be inferred from geographic data using co-occurrences as proxies, and generate corresponding distribution models. We consider two case studies: an insect genus (*Lutzomyia* sp.) and a mammal species (*Lynx rufus*). We show that for both examples model predictability is higher using biotic versus abiotic variables, but even higher when both variable types are integrated together. Also, by identifying those variables that are most relevant in describing the suitable (niche) and unsuitable (anti-niche) areas we can establish an ecological profile for any geographic location and quantify the relative influence of each location and its impact on species. In conclusion, we show that including both abiotic and biotic factors not only leads to a fuller more comprehensive understanding of the niche, but also leads to more accurate prediction models.

1. Introduction

One of the most important goals in ecology and biogeography studies is to identify the principal factors that constrain the range of species. An understanding of these factors and their relative impact will permit us to better understand and model both current and future distributions (Lomolino et al., 2005; Araujo and Guisan, 2006; Araujo and Luoto, 2007). Unfortunately, the number of potential factors that can affect species' distributions is enormous. However, a great deal of information associated with many of these factors is now available in online databases. Thus, the most advanced techniques for describing and predicting species' distributions are nowadays based on data mining, where large volumes of observational data are systematically explored using different mathematical models, from standard regression type models to sophisticated artificial intelligence techniques. Such methods have recently been used in various ecological applications, such as biodiversity studies, modeling biological invasions and species distributions (Stockwell and Peters, 1999; Guralnick and Pearman, 2009).

Current niche modeling has chiefly been limited to abiotic variables, due to the difficulties of incorporating information associated with biotic interactions (Guisan and Thuiller, 2005; Araujo and Guisan, 2006). Recently, however, a methodology based on data mining techniques has been developed and applied in ecogeographic studies (Sánchez-Cordero et al., 2008; Stephens et al., 2009) that naturally facilitate the incorporation of biotic factors associated with point collection data. Although the methodology is general, what was not explicitly considered in those papers was the question of how to integrate together different data types, including both abiotic and biotic factors, thus permitting a deeper insight into the relative importance of the different types of factor in determining species' distributions.

Although modeling the range of a species without further insight into what factors potentially affect the range can be a useful goal, much greater insight can be gained by determining and

---

* Corresponding author at: Instituto de Biología, Universidad Nacional Autónoma de México, 04510 México, D.F., Mexico. Tel.: +52 555 623 0222x47880.
E-mail addresses: cgs@ibiologia.unam.mx (C. González-Salazar), stephens@nucleares.unam.mx (C.R. Stephens), pmarquet@bio.puc.cl (P.A. Marquet).
[1] Tel.: +52 555 622 4692; fax: +52 555 622 4693.
[2] Tel.: +56 2 6862639.

0304-3800/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2012.10.007

understanding which factors, barriers or biotic interactions are important for a particular species in a particular geographical location (Brown, 1995; Brown et al., 1996; Arntzen and Themudo, 2008). In other words, to better understand the relation between the geographical distribution of the species and its niche. Such considerations, for instance, could help to cover the current gap between local ecological observations and other regional processes (Guisan and Zimmermann, 2000; Pearson and Dawson, 2003; Guisan et al., 2006).

The main contribution of this article is to present a modeling framework in which both abiotic and biotic factors associated with different data types can be integrated together into models that lead to more accurate predicted distributions and to a fuller understanding of the corresponding niches of a given taxon. The methodology is an extension of that proposed in Stephens et al. (2009), where only biotic variables were included. We will show how to use such models to achieve the above stated goals of predicting taxa' distributions using a more complete set of potential niche variables, as well as quantifying the relative importance not only of biotic versus abiotic factors in explaining the presence of a taxon in a particular geographic location, but also of specifying exactly which factors are the most important.

Although our approach is quite general, specific results, such as the distribution of particular taxon and the associated niche variables, depend on the taxon being studied. We will therefore illustrate the method by determining the predictive power of regional variables in predicting the distribution and quantifying and characterizing the ecological niches of two very different taxa, the genus *Lutzomyia* (sandfly) and the species *Lynx rufus* (bobcat), describing their ecological profiles and the role played by different biotic and abiotic factors as constraints on their distributions. Note that, instead of choosing two distinct species here as our examples, we consider a species and a genus, thereby showing that the modeling framework we present is applicable to different taxonomic levels. One might, of course, ask whether or not it is appropriate to talk of predicting the distribution and the associated niche of a taxon other than a species. For instance, if the niches of different species within a genus are quite distinct, it may well be that the genus niche is so smeared out it has no characterizing features and subsequently the corresponding distribution model is only weakly predictive. The very fact that our results are positive – leading to an accurate predictive model and a characteristic niche for the genus Lutzomyia – show that, in this case at least, it makes perfect ecological sense to model at the genus level. This does not, of course, guarantee that this will always be the case.

So, in the first case study we consider data associated with actual and potential reservoirs of an important emerging disease, Leishmaniasis, a disease widely distributed in tropical regions that is transmitted by sandflies. As Leishmaniasis is a zoonotic tropical disease, sylvan reservoirs are crucial to the maintenance of the parasite in ecological communities and, further, are intimately associated with human transmission (Wolfe et al., 2007). For our second case study, we consider the effect of abiotic and biotic factors on the distribution of the carnivore *Lynx rufus*, given that it has not been collected south of the Isthmus of Tehuantepec in Mexico, in spite of the fact that there are geographic regions below the Isthmus that coincide with the ecological requirements associated with its fundamental niche (Sánchez-Cordero et al., 2008).

By using two such contrasting taxa (an insect and a mammal) we can illustrate the scope of this methodology in inferring different types of inter-specific interactions (e.g., competition, mutualism, commensalism), and their relative importance for taxa distributions. Additionally, we can directly compare and contrast the role of climatic factors to that of biotic interactions to determine which variables most affect the presence or absence of a species and in what way the prediction of species' ranges could be improved (Guisan et al., 2006; Heikkinen et al., 2007).

2. Materials and methods

In applying our methodology we used a class of biotic variables – collection data for Mexican mammals – as well as abiotic (climatic) variables (e.g. temperature, precipitation) and land cover. Thus, models were built using four sets of explanatory variables: (1) abiotic variables only; (2) mammals species only; (3) land cover only; and (4) abiotic variables, mammals and land cover together.

2.1. Data types

2.1.1. Abiotic variables

2.1.1.1. Climatic data. Nineteen bioclimatic variables were used as environmental layers (Table 1) obtained from WorldClim with a spatial resolution of 30″ (http://www.worldclim.org/current.htm; Hijmans et al., 2005). They represent annual trends (e.g., annual mean temperature and precipitation), seasonality (e.g. annual temperature and precipitation ranges), environmental extremes (e.g., highest and lowest values of temperature for the warmest and coolest months) of temperature and precipitation, and characterize the dimensions of climate considered particularly relevant in determining species distributions (Waltari and Guralnick, 2009).

2.1.2. Biotic variables

2.1.2.1. Land cover. We used the Inventario Nacional Forestal (INF) 2000 (Palacio et al., 2000) as a base for current land use and vegetation types in México. INF 2000 is based on both LandSat satellite imagery interpretation and ground field validation of the main vegetation types and land use in Mexico and scaled at 1:250,000. It is jointly produced by the Instituto de Geografía of the Universidad Nacional Autónoma de México (www.igeograf.unam.mx), and the Instituto Nacional de Estadística y Geografía (www.inegi.gob.mx). This layer included 77 types of vegetation in México.

2.1.2.2. Species occurrence data. The data set consisted of point collection data associated with one Class, Mammalia, including the species – *Lynx rufus* – and one genus – *Lutzomyia* – of the class Insecta. The mammal data set contains 37,297 unique point collections from georeferenced localities for 427 terrestrial mammals occurring in Mexico. Data are based on museum voucher specimens from national and international collections, public electronic databases (GBIF; www.gbif.org, and CONABIO; www.conabio.gob.mx), and published records (Hall, 1981; Guevara-Chumacero et al., 2001). For *L. rufus* there were 220 collections points. For *Lutzomyia*, there were 270 collections points belonging to 11 species (see supplementary material) taken from published literature and from national collections: Instituto de Diagnóstico y Referencia Epidemiológica (InDRE, Mexico City), the Colección Entomológica Regional Universidad Autónoma de Yucatán (UADY, Mérida) and the Laboratorio de Medicina Tropical at the Universidad Nacional Autónoma de México (UNAM, Mexico City). For all data sets, each locality was georeferenced to the nearest 0.01 degrees of latitude and longitude using 1:250,000 topographic maps (INEGI; www.inegi.gob.mx; Instituto de Geografía, Universidad Nacional Autónoma de México; www.igeograf.unam.mx). Point collection data was, of course, not collected in order to provide and unbiased sampling of underlying species abundance and therefore must be considered carefully to understand potential statistical biases that might be present. The utility and limitations of point collection data have been amply discussed in (Ponder et al., 2001; Graham et al., 2004).

With respect to the data set for Mexican mammals, this data has been collected over a period of more than 100 years with a