



## Predicting species' abundances from occurrence data: Effects of sample size and bias



Carlos Yañez-Arenas<sup>a</sup>, Roger Guevara<sup>b</sup>, Enrique Martínez-Meyer<sup>c</sup>,  
Salvador Mandujano<sup>d</sup>, Jorge M. Lobo<sup>e,\*</sup>

<sup>a</sup> División de Posgrado, Instituto de Ecología A.C., Carretera Antigua a Coatepec, No. 351, El Haya, Xalapa 91070, Veracruz, Mexico

<sup>b</sup> Red de Biología Evolutiva, Instituto de Ecología A.C., Carretera Antigua a Coatepec, No. 351, El Haya, Xalapa 91070, Veracruz, Mexico

<sup>c</sup> Departamento de Zoología, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad Universitaria, Mexico City 04510, Mexico

<sup>d</sup> Red de Biología y Conservación de Vertebrados, Carretera Antigua a Coatepec, No. 351, El Haya, Xalapa 91070, Veracruz, Mexico

<sup>e</sup> Departamento de Biogeografía y Cambio Global, MNCN-CSIC, c/José Gutiérrez Abascal 2, 28006 Madrid, Spain

### ARTICLE INFO

#### Article history:

Received 12 June 2014

Received in revised form 19 September 2014

Accepted 19 September 2014

Available online xxx

#### Keywords:

Distance to the niche centroid  
Ecological niche modelling  
Sample intensity  
Population density  
Abundance  
Sampling scenarios

### ABSTRACT

Modelling geographic patterns of abundance/density of species is an important step forward in ecological niche modelling, with implications for theoretical and applied ecology. The distance to the niche centroid approach (DNC) is a methodological development toward better understanding how the internal structure of species' ecological niches is related to geographic patterns of abundance. We evaluated this approach under combinations of three sampling scenarios and three sampling intensities for a hypothetical species for which abundance patterns were ideal and strictly controlled. Our results indicate that predictive ability of the DNC approach increased with sample intensity, particularly under a strict random sampling scheme. Model performance under a sampling scenario biased by species' density fell slightly, but was importantly reduced when the source of the biases were attractor sites unrelated with species' traits. We conclude that the DNC approach is only suitable to model species' abundances/densities under particular conditions. First because it is necessary fulfill some assumptions (discussed in this paper), and second because its performance strongly depends on sampling characteristics that are unusual in most biodiversity data.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

Ecological niche modelling (ENM) has become a popular approach to estimate environmental conditions under which species can maintain viable populations (i.e., the ecological niche) and the associated potential geographic distribution, particularly when the presence of species is the only information available (Peterson et al., 2011). Presence data, represented as a collection of localities where the species' has occurred, are related to a set of environmental variables via various computational algorithms to predict potential or realized distributions (Franklin, 2009; Peterson et al., 2011).

For years, the field has focused on analysing the robustness of different methods to predict both realized and potential geographic distributions of species (Elith et al., 2006; Graham and Hijmans, 2006; Peterson et al., 2007), transfer models in space and time (Martínez-Meyer et al., 2004; Peterson, 2003; Peterson et al.,

2002), or study effects of environmental variables on model performance (Blach-Overgaard et al., 2010; Costa et al., 2008; Gormley et al., 2011). Some studies have explored the capacity of some of these models to inform about abundance (or population density) patterns of species (Bean et al., 2014; Estrada and Arroyo, 2012; Gutiérrez et al., 2013; Jiménez-Valverde, 2011; Nielsen et al., 2005; Pearce and Ferrier, 2001; Real et al., 2009; Tôrres et al., 2012; Van Couwenbergh et al., 2013; VanDerWal et al., 2009). Results of these studies are inconclusive, but in general show that the continuous suitability values provided by models based on presence-only or presence-absence, pseudo-absence or background data rarely are able to reflect abundance patterns adequately.

Recently, Martínez-Meyer et al. (2013) documented that the internal structure of ecological niches derived from distributional information correlates with abundance, and implemented the so-called Distance to the Niche Centroid approach (DNC) to predict geographic patterns of abundance/density; a procedure similar to those applied formerly based on orthogonal variables (Calenge et al., 2008; Hirzel et al., 2002; Robertson et al., 2001). The rationale behind this procedure is that optimal conditions for a

\* Corresponding author. Tel.: +34 91 4111328x1145.  
E-mail address: [mcnj117@mncn.csic.es](mailto:mcnj117@mncn.csic.es) (J.M. Lobo).

species are found towards the centroid of the ecological niche in a multidimensional space (Brown, 1984; Hutchinson, 1957; Maguire, 1973). According to this assumption, a locality close to the centroid of the  $n$ -dimensional niche harbor better conditions for the species and as a consequence higher abundances/densities would be expected at such localities. This idea of a niche optimum towards the center of the hypervolume has been previously suggested (Hutchinson, 1957) or assumed (Brown, 1984; Maguire, 1973), but not empirically tested until recently (Martínez-Meyer et al., 2013; Van Couwenberghe et al., 2013; Yañez-Arenas et al., 2012; but see Austin et al., 1984).

However, it is reasonable to think that abundance is not only determined by scenopoetic, physiologically limiting variables of the fundamental niche, but also by dynamically linked variables (*sensu* Peterson et al., 2011) of the realized niche, such as availability of resources or the effect of biotic interactions like predation or parasitism. Moreover, the direct link between DNC and the local density/abundance of a species may be roiled by factors intrinsic to the species (e.g., density-dependence, dispersal) and even because in most instances niche estimations are incomplete in their key scenopoetic and dynamically linked axes or may be based on biased samples unable to represent accurately the niche of the species. Thus, under real-life conditions the realized and fundamental niches (*sensu* Peterson et al., 2011) estimated from field observations can be a subset of “real” ones (Browning et al., 2005; Pulliam, 2000; Rotenberry et al., 2006; Soberón, 2007).

The success of any spatial model depends upon the quality of input data, and model performance depends strongly on sample size and bias (Hernández et al., 2006; Kadmon et al., 2004; Loiselle et al., 2008; Owens et al., 2013; Reese et al., 2005; Stockwell and Peterson, 2002; Wisz et al., 2008). For ecological niche models, minimum sample sizes seem to depend on the algorithm used for building models. For some methods, reliable models can supposedly be developed with as few as 50–100 unique occurrence points (Elith et al., 2006; Kadmon et al., 2004; Loiselle et al., 2008; Stockwell and Peterson, 2002; Wisz et al., 2008). On the other hand, biodiversity data frequently have sampling bias problems, both geographic and environmental, that pose significant challenges for the success of such analyses (Funk and Richardson, 2002; Hijmans et al., 2000; Hirzel and Guisan, 2002; Hortal et al., 2008; Kadmon et al., 2004; Rocchini et al., 2011).

In this study we used simulated data under ideal conditions to evaluate the effects of different sampling intensities and scenarios on geographic predictions of population density provided by the DNC approach. Previous studies support the use of this method to predict abundance/density variations from real occurrence data (Escalante and Martínez-Meyer, 2013; Martínez-Meyer et al., 2013; Yañez-Arenas et al., 2012). However, these results veil the detection of caveats and uncertainties due to biases of input data because these sources of uncertainty are unknown. Thus, a population density map with an abundant-centre pattern was developed for a hypothetical species for which its niche and driving environmental variables were known in advance, in order to test the performance of the DNC approach under three different sampling intensities and three sampling scenarios representative of biases common in real occurrence data. The aim was to learn the reach and limitations of this procedure in estimating geographic patterns of abundances under controlled conditions.

## 2. Material and methods

### 2.1. Virtual species population density map

A density map for a hypothetical species was generated across Mexico as the study area. We selected four bioclimatic variables

from the WorldClim dataset (Hijmans et al., 2005) as the niche dimensions: maximum temperature of the warmest month, minimum temperature of the coldest month, precipitation of the wettest month, and precipitation of the driest month. All variables were in geographic coordinates (datum WGS84), z-standardized (mean=0; SD=1), and resampled to a spatial resolution of 1' (~2 km<sup>2</sup>).

To populate the distribution range of the virtual species we followed a three-step protocol. First, we calculated the Mahalanobis environmental distance to the multivariate centroid of the four variables across all of the 24,000 cells of Mexico. Subsequently, we selected as the distribution area of the virtual species the 50% of the cells closest to the environmental centroid (i.e., 12,000; prevalence=0.5), in order to better estimate both commission and omission errors (Fig. 1). Finally, distance values within this distribution area were rescaled via a generalized logit regression model (density = inverse logit  $[2.729 - 0.364 \times D] \times 16$ , where  $D$  is the Mahalanobis distance) to obtain variation of the population density directly related with the used environmental variables. This procedure resulted in maximum density values of 15 individuals/km<sup>2</sup> when distance to the centroid was zero, and density values approached asymptotically to zero when progressively increasing distances from the centroid (Fig. 2). Under this scheme, we assumed in effect a multivariate normal distribution of density, wherein optimal conditions and maximum abundances of the virtual species occur at sites closest to its ecological niche centroid, as expected from both theoretical arguments (Brown, 1984; Hutchinson, 1957; Maguire, 1973) and empirical evidences (Austin et al., 1984; Martínez-Meyer et al., 2013). Note that these are ideal conditions for the DNC approach, since it was formulated based on the same ecological theory, thus a good performance of DNC was expected under a representative and unbiased sampling scheme. Therefore, our aim was to evaluate the way and magnitude in which DNC performance was affected when sampling departed from ideal, simulating real-life data sampling problems.

### 2.2. Sampling scenarios

We generated three data collection scenarios, an ideal sampling method (“strict random”) and two biased sampling schemes (“weighted by density” and “randomly biased”), aiming to simulate scenarios of data collection typical of real-world field surveys and museum specimens. For the “strict random” scenario we selected a given number of cells at random within the distribution area. This random survey design assumed that all presence cells are equally detectable, independent of their densities (total detectability); this sampling scenario is equivalent to a protocol in which a field biologist gathers data randomly across the entire range of a species.

The second sampling scenario was “weighted by density.” Here, we set sampling effort to be proportional to local densities of the species. Cells with high density (>10 individuals/km<sup>2</sup>) were sampled with a probability of 0.64; cells with intermediate densities (5–10 individuals/km<sup>2</sup>) were sampled with a probability of 0.21; and low-density cells (<5 individuals/km<sup>2</sup>) were sampled with a probability of 0.14. This sampling scenario thus emulates a situation in which the probability of recording a presence of a species at a locality is directly proportional to the local population density; localities with very low population densities will often be recorded as absences under this sampling scenario.

The third sampling scenario was “randomly biased,” in which data were collected randomly around particular ‘attractor’ areas. This method simulates reserves, universities, or other intensively-surveyed sites, around which biologist tend to concentrate sampling efforts. We distributed sampling effort randomly around 10 attractors within a radius of ~100 km. The positions of attractors were randomly shifted in each run.

Download English Version:

<https://daneshyari.com/en/article/6296695>

Download Persian Version:

<https://daneshyari.com/article/6296695>

[Daneshyari.com](https://daneshyari.com)