

## 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>).

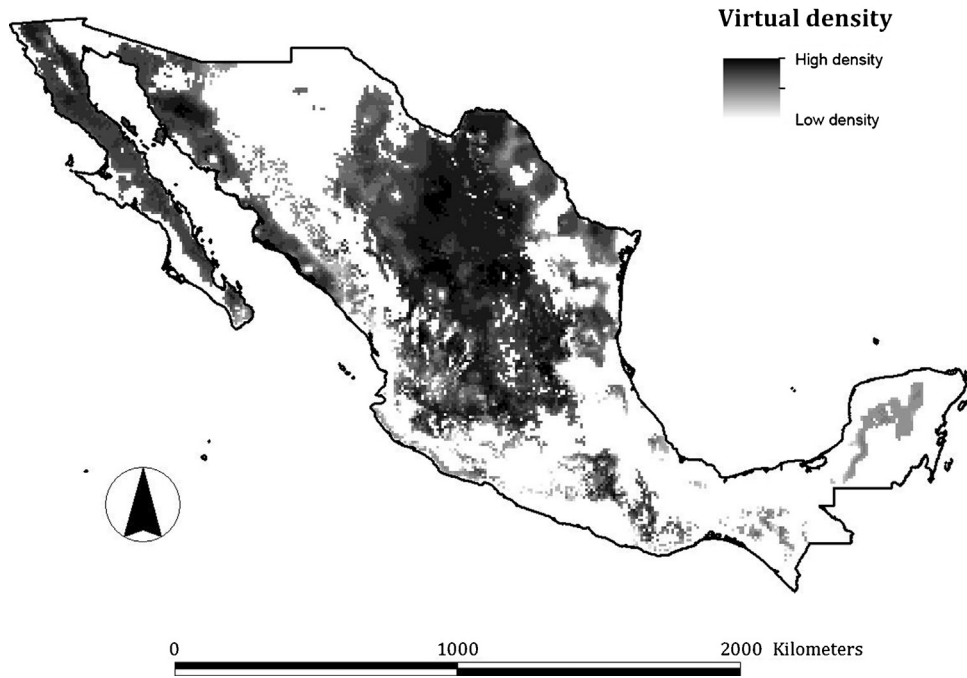
To populate the distribution range of the virtual species we followed a three-step protocol. First, we calculated the Mahalanobis environmental distance to the multivariate centroid of the four variables across all of the 24,000 cells of Mexico. Subsequently, we selected as the distribution area of the 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.

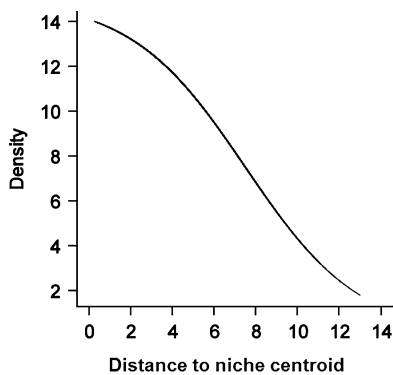


**Fig. 1.** Distribution area and density of a virtual species based on the Mahalanobis distance to the environmental centroid of four bioclimatic variables across Mexico.

Finally, we sought to replicate and evaluate the effects of different data intensities, thus for each sampling scenario we established three levels of sampling effort: 0.01%, 1%, and 5% of all cells within the range of the virtual species, corresponding to 14, 134, and 664 occurrence localities, respectively. We ran 100 replicates for each combination of sampling scenario and intensity, resulting in 900 replicate models.

### 2.3. The DNC method

Since our purpose was to measure the deviation of ‘biased’ centroids estimated under different sampling scenarios of the niche from the ‘true’ centroid, we followed the DNC approach (Martínez-Meyer et al., 2013 Yañez-Arenas et al., 2012) calculating the niche centroid based exclusively on sampled cells (i.e., the sample centroid), to subsequently estimate the Mahalanobis distance (as opposed to the Euclidian distance as originally proposed by the authors, to account for collinearity of variables) for all cells across the distribution of the virtual species.



**Fig. 2.** Logistic model used to transform distance to the environmental centroid of each cell to population density.

Mahalanobis distance values were then used as the predictor variable and the ‘true’ density values as the response variable.

### 2.4. Statistical analysis

We examined the performance of DNC results for each replicate of the three sampling scenarios and intensities by means of generalized additive models (GAMs) relating density values and distance to niche centroid distances. We preferred this model over an ordinary regression model in light of its flexibility in fitting trends, since relationships varied from linear to markedly non-linear. Given that the same four bioclimatic variables were subjacent to all models, we used the percentage of explained variation as the criterion for comparing model performance among sampling scenarios and intensities. All simulations and statistical tests were automated in R 2.15 (R\_Development\_Core\_Team, 2012) via the following libraries: boot (Canty and Ripley, 2006) for logistic back transformation, mgcv (Wood, 2010) for generalized additive models, and raster (Hijmans and Van Etten, 2010) for management of geospatial data.

We used Kruskal–Wallis post hoc multiple comparisons of mean ranks of groups (because of non-normality in data) to compare DNC performance between sampling scenarios and intensities. These tests were used on the collection of values of proportional explained deviance in each of the 100 models run under all sampling scenarios.

## 3. Results

Explained deviance increased significantly with sampling intensity in both “strict random” and “weighted by density” scenarios (Table 1; Fig. 3). However, under the “randomly biased” scenario, estimated densities for 5% sample size were not better than those with 1%. When sample size was very low (0.01%), only the “randomly biased” scenario accounted for slightly but statistically significant less variation in density than the “strict random” scenario. However, with a higher sampling effort, the three scenarios differed significantly in their capacity to predict

**Table 1**

Summary of post hoc multiple comparisons of mean ranks for all groups, comparing percentage of deviance explained by the relationship between actual population densities and predicted values by the Distance to Niche Centroid (DNC) method under three sampling scenarios and three data intensities. Sr = Strict random, Wd = Weighted by density, Rb = Randomly biased.

	Sr	Wd	Rb
0.01% vs 1%	143.17***	146.88***	42.24***
0.01% vs 5%	149.25***	149.25***	62.06***
1% vs 5%	138.59***	56.16***	2.00
0.01%	0.01%	1.0%	5.0%
Sr vs Wd	3.57	11.57***	107.68***
Sr vs Rb	7.20**	138.41***	149.25***
Wd vs Rb	0.87	124.80***	147.77***

\*\*  $P < 0.01$ .

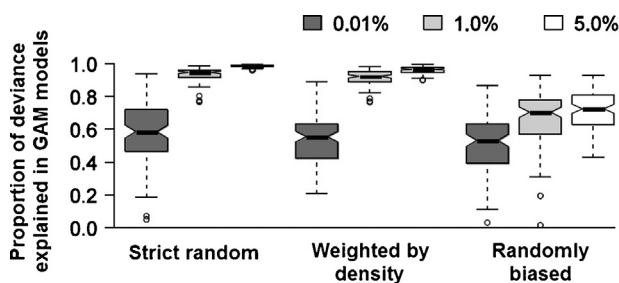
\*\*\*  $P < 0.001$ .

population density (Table 1; Fig. 3): strict random > weighted by density > randomly biased.

Under the best sampling scenarios (i.e., “strict random” and 5% sample size), DNC density estimations described an inverse, curvilinear relationship to actual density, with only minimal scatter (Fig. 4A). DNC predictions were more scattered with smaller sample intensities (Fig 4A), mainly because less suitable pixels (i.e., higher DNC values) varied more in their estimates. In general, increasing sampling effort reduced variability and increased estimation accuracy. This pattern was also observed for the “weighted by density” scenario (Fig 4B), but not for the “randomly biased” scenario (Fig 4C), where DNC prediction variability increased dramatically, showing triangular relationships even for the largest sample size (Fig. 4C).

#### 4. Discussion

Our goal was to evaluate the capacity of the Distance to the Niche Centroid (DNC) approach to estimate population density from presence-only data under different sampling schemes, using a simulated species to fix the most important sources of uncertainty that frequently influence the results of these modelling exercises. Our assessment was carried out under conditions quite distinct from those experienced by modellers with real species: the density of our virtual species is completely controlled by known environmental factors, models incorporated all of the required predictors, and the “true” geographic density pattern is known in advance. In addition, abundance held a multivariate normal distribution, as was observed empirically by Austin et al. (1984) and Martínez-Meyer et al. (2013), and expected from ecological niche theory (Brown, 1984; Hutchinson, 1957; Maguire, 1973). We tested effects of sample intensity and bias on



**Fig. 3.** Summary of the overall fit of Generalized Additive Models (GAMs) used to predict density via the Distance to the Niche Centroid (DNC) approach under three sampling scenarios of randomness or bias and three sampling intensities (0.01%, 1.0%, 5.0% of total cells). Box plots extend from the first quartile (Q1) to the third quartile (Q3); the median (Q2) is also represented.

performance of the DNC approach, examining whether model outputs have any potential to predict abundance patterns when deviating from these idealized conditions.

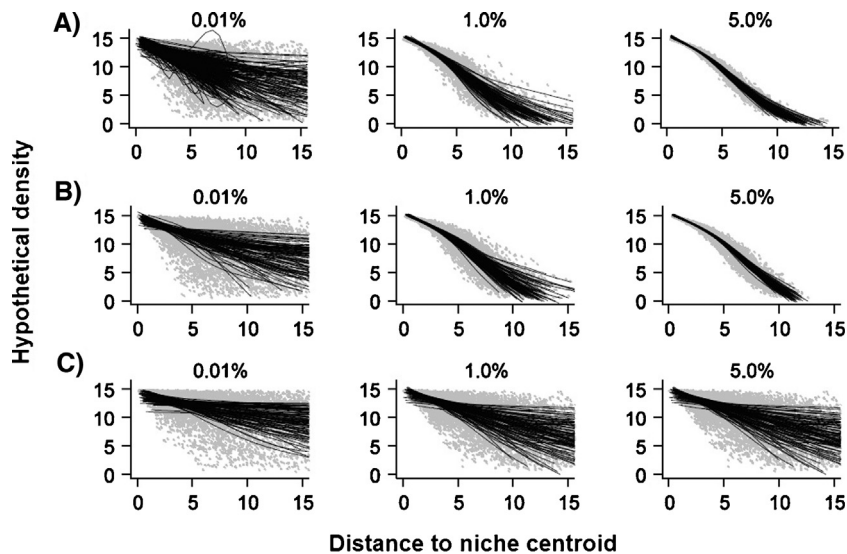
Our results indicated that DNC density estimates were best when presence information for training models came from unbiased sampling scenarios with high sampling intensities. A “strict random” sampling protocol produced the best density predictions, closely followed by the “weighted by density” scenario. That is, random surveys with complete detectability allow better characterization of species’ multidimensional niches. In the “weighted by density” scenario –probably the most common case for vagile species– although detectability was biased towards localities with higher densities, we obtained relatively reliable predictions perhaps because under this scheme most of the niche structure is still captured. Poorest DNC performance occurred when biases were generated by factors unrelated to density, and thus the structure is lost in the sample. In the “randomly biased” scenario, cells used to train models came disproportionately from attractor sites not related to the density of the species (Rocchini et al., 2011). This sampling scenario mimics frequent overconcentration of records that plague many biodiversity datasets (Hortal et al., 2008).

Regarding sampling effort, model performance is well known to be poor with small sample sizes of calibration data (Bojórquez-Tapia et al., 1995; Jiménez-Valverde, 2011; McPherson et al., 2004; Stockwell and Peterson, 2002; Wisz et al., 2008). Similarly, DNC performance increased with sample size; reliable density estimates were obtained with as low as 1% sampling density ( $N = 134$ ) for the “strict random” and “weighted by density” scenarios. The principal effect of reducing sampling effort was to increase uncertainty at low and moderate densities (see Fig 4A).

GAM results indicated that DNC outputs accounted for an average of 52–99% of explained deviance in density depending of the quality of data used. Previous studies aiming to estimate abundance by means of modelling procedures comparing the environmental conditions in presence localities against a set of background data randomly distributed in the territory (Gutiérrez et al., 2013; VanDerWal et al., 2009) show that the so obtained suitability values have a triangular relationship with densities (but see Bean et al., 2014). That is, low densities frequently manifest in both low- and high-suitability localities, and, in the best cases, models can account only for the upper limit of abundance. We also submitted our data to the most popular presence-background niche modelling technique, namely MaxEnt (Phillips et al., 2006), confirming that inaccuracies tend to appear in localities predicted as highly suitable (see Supplementary Material). On the contrary, DNC can generate high distance values for sites that may nonetheless present at least moderate densities, suggesting that inaccuracies do not seem to occur at high-density sites when optima can be established correctly, which can be considered a significant advantage of the DNC approach. In the best cases, with large sample sizes and a random sampling scheme, relationships between density and DNC were not triangular but slightly curvilinear, where different DNC values correspond to similar densities mainly in the far right portion of the relationship (Fig. 4). This has been observed before (Martínez-Meyer et al., 2013) and suggests that optimal niche conditions – where population size should theoretically be maximal – are present in an area around the niche centroid, beyond which populations decline steeply, as a consequence of the normal multivariate distribution of abundance; but probably this shape of the relationship is also due to the nature of the input data, since presence-only records inevitably lose information.

Our experiments were developed under ideal conditions. The use of occurrence data to identify optimal conditions and predict density patterns represents a significant step forward in niche modelling, but misleading interpretations of DNC results can emerge when optimal conditions do not present a unimodal,





**Fig. 4.** Relationships between “true” densities and predicted values according to the Distance to the Niche Centroid (DNC) approach under three sampling scenarios and three data intensities. Dark lines represent General Additive Model (GAM) best-fit curves and grey points show actual data in relationships. A = Strict random, B = Weighted by density, C = Randomly biased.

normal distribution (Austin, 2002), as can occur when representation of ecological niches is only partial under real-world circumstances (Browning et al., 2005; Pulliam, 2000; Rotenberry et al., 2006; Soberón, 2007). Problems may also occur if the derived niche optimum is not linked to the environmental predictors used in model calibration, but rather to other environmental dimensions, or to biotic, historical, dispersal constraints, chaotic dynamics, demographic or metapopulational factors (Krasnov et al., 2008). In such cases, distribution of species would be in non-equilibrium with environmental conditions and delimiting the extent of the study area may help in minimizing the role played by these factors (Acevedo et al., 2012).

## 5. Conclusions

Our results lead us to some basic recommendations when the aim is to make density or abundance estimates from presence information. First, we must be aware that local abundance is the result of the interplay of scenopoetic and dynamically linked variables of both the fundamental and realized niches; as well, the niche fitted from observed occurrences may be only a partial representation of them. In any case, as these contingent variables can have an environmental correspondence, the key issue is if we can obtain enough accurate occurrences reflecting population frequencies. Second, DNC can be a reasonable approach for estimating abundances when only presence data is available but sample size and bias should be taken into account before analyses are performed and conclusions drawn. Third, it is necessary to obtain presence data across the full range of environmental conditions under which the species occurs to have an adequate representation of the realized niche of the species. Full correction for detectability is advisable (Chen et al., 2012; Royle et al., 2005) or at least avoid localities with insufficient survey effort within the area of analysis. Finally, sample size must be large enough to ensure an adequate representation of the species' niche because insufficient and/or biased records have important negative effects on density estimations. As obvious as this may seem, these caveats represent a worldwide problem since occurrence data for most species are scarce and probably biased. Therefore our results highlight the need to continue collecting and assembling biodiversity information especially in ways that minimize biases.

## Acknowledgments

We thank A. Townsend Peterson for his valuable comments, suggestions and review of the manuscript. C.Y.-A. was supported by a doctoral fellowship (CONACYT: CB-2009-01 No. 130702) and a grant (CONACYT becas-mixtas 2011–2012: 212,502) for a research stay at the Museo de Ciencias Naturales de Madrid.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ecolmo-del.2014.09.014>.

## References

- Acevedo, P., Jiménez-Valverde, A., Lobo, J.M., Real, R., 2012. Delimiting the geographical background in species distribution modeling. *J. Biogeogr.* 39, 1383–1390.
- Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Model.* 157, 101–118.
- Austin, M.P., Cunningham, R.B., Fleming, P.M., 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* 55, 11–27.
- Bean, W., Prugh, L., Stafford, R., Butterfield, H., Westphal, M., Brashares, J., 2014. Species distribution models of an endangered rodent offer conflicting measures of habitat quality at multiple scales. *J. Appl. Ecol.* 51, 1116–1125.
- Blach-Overgaard, A., Svenning, J.C., Dransfield, J., Greve, M., Balslev, H., 2010. Determinants of palm species distributions across Africa: the relative roles of climate, non-climatic environmental factors, and spatial constraints. *Ecography* 33, 380–391.
- Bojórquez-Tapia, L.A., Azuara, I., Ezcurra, E., Flores-Villela, O., 1995. Identifying conservation priorities in Mexico through geographic information systems and modeling. *Ecol. Appl.* 5, 215–231.
- Brown, J.H., 1984. On the relationship between abundance and distribution of species. *Am. Nat.* 124, 255–270.
- Browning, D.M., Beaupre, S.J., Duncan, L., 2005. Using partitioned Mahalanobis D2 (k) to formulate a GIS-based model of timber rattlesnake hibernacula. *J. Wildl. Manage.* 69, 33–44.
- Calenge, C., Darmon, G., Basille, M., Loison, A., Jullien, J.-M., 2008. The factorial decomposition of the Mahalanobis distances in habitat selection studies. *Ecology* 89, 555–566.
- Canty, A., Ripley, B., 2006. *boot: bootstrap R (S-Plus) Functions* (Canty) – R Package version 1.2–27. <http://www.r-project.org/cran>
- Costa, G.C., Wolfe, C., Shepard, D.B., Caldwell, J.P., Vitt, L.J., 2008. Detecting the influence of climatic variables on species distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient. *J. Biogeogr.* 35, 637–646.

- Chen, G., Kéry, M., Plattner, M.K., Gardner, M.B., 2012. Imperfect detection is the rule rather than the exception in plant distribution studies. *J. Ecol.* 101, 183–191.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151.
- Escalante, T., Martínez-Meyer, E., 2013. Ecological niche modeling and wildlife management units (UMAs): an application to deer in Campeche, Mexico. *Trop. Subtrop. Agroecosyst.* 16, 183–191.
- Estrada, A., Arroyo, B., 2012. Occurrence vs abundance models: differences between species with varying aggregation patterns. *Biol. Conserv.* 152, 37–45.
- Franklin, J., 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge.
- Funk, V.A., Richardson, K.S., 2002. Systematic data in biodiversity studies: use it or lose it. *Syst. Biol.* 51, 303–316.
- Gormley, A.M., Forsyth, D.M., Griffioen, P., Lindeman, M., Ramsey, D.S.L., Scroggie, M. P., Woodford, L., 2011. Using presence only and presence-absence data to estimate the current and potential distributions of established invasive species. *J. Appl. Ecol.* 48, 25–34.
- Graham, C.H., Hijmans, R.J., 2006. A comparison of methods for mapping species ranges and species richness. *Global Ecol. Biogeogr.* 15, 578–587.
- Gutiérrez, D., Harcourt, J., Diéz, S.B., Gutiérrez-Illán, J., Wilson, R.J., 2013. Models of presence-absence estimate abundance as well as (or even better than) models of abundance: the case of the butterfly *Parnassius apollo*. *Landscape Ecol.* 28, 401–413.
- Hernández, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29, 773–785.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978.
- Hijmans, R.J., Garrett, K.A., Huaman, Z., Zhang, D.P., Schreuder, M., Bonierbale, M., 2000. Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. *Conserv. Biol.* 14, 1755–1765.
- Hijmans, R.J., Van Etten, J., 2010. *Raster: geographic analysis and modeling with raster data – R package version 1.3–11*. <http://www.r-project.org/cran>
- Hirzel, A., Guisan, A., 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecol. Model.* 157, 331–341.
- Hirzel, A., Hausser, J., Chessel, D., Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83, 2027–2036.
- Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M., Baselga, A., 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* 117, 847–858.
- Hutchinson, G.E., 1957. Concluding remarks. *Cold Spring Harbor Symp. Quant. Biol.* 22, 415–427.
- Jiménez-Valverde, A., 2011. Opinion: relationship between local population density and environmental suitability estimated from occurrence data. *Front. Biogeogr.* 3, 59–61.
- Kadmon, R., Farber, O., Danin, A., 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecol. Appl.* 14, 401–413.
- Krasnov, B.R., Shenbrot, G.I., Khokhlova, I.S., Vinarski, M., Korralo-Vinarskaya, N., Poulin, R., 2008. Geographical patterns of abundance: testing expectations of the 'abundance optimum' model in two taxa of ectoparasitic arthropods. *J. Biogeogr.* 35, 2187–2194.
- Loiselle, B.A., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G., Montiel, O.M., 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *J. Biogeogr.* 35, 105–116.
- Maguire, B., 1973. Niche response structure and the analytical potentials of its relationship to the habitat. *Am. Nat.* 107, 213–246.
- Martínez-Meyer, E., Díaz-Porras, D.F., Peterson, A.T., Yañez-Arenas, C., 2013. Ecological niche structure and rangewide abundance patterns of species. *Biol. Lett.* 9, 20120637.
- Martínez-Meyer, E., Peterson, A.T., Hargrove, W.W., 2004. Ecological niches as stable distributional constraints on mammal species: with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecol. Biogeogr.* 13, 305–314.
- McPherson, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *J. Appl. Ecol.* 41, 811–823.
- Nielsen, S.E., Johnson, C.J., Heard, D.C., Boyce, M.S., 2005. Can models of presence-absence be used to scale abundance? Two case studies considering extremes in life history. *Ecography* 28, 197–208.
- Owens, H.L., Campbell, L.P., Dornak, L.L., Saupe, E.E., Barve, N., Soberón, J., Peterson, A.T., 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecol. Model.* 263, 10–18.
- Pearce, J., Ferrier, S., 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biol. Conserv.* 98, 33–43.
- Peterson, A.T., 2003. Projected climate change effects on Rocky Mountain and Great Plains birds: generalities of biodiversity consequences. *Global Change Biol.* 9, 647–655.
- Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sánchez-Cordero, V., Soberón, J., Buddemeier, R.H., Stockwell, D., 2002. Future projections for Mexican faunas under global climate change scenarios. *Nature* 416, 626–629.
- Peterson, A.T., Papes, M., Eaton, M., 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and MaxEnt. *Ecography* 30, 550–560.
- Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araujo, M.B., 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259.
- Pulliam, H.R., 2000. On the relationship between niche and distribution. *Ecol. Lett.* 3, 349–361.
- R\_Development\_Core\_Team, 2012. *R: a language and environment for statistical computing*. Version 2.15.1. <http://cran.R-project.org> R Foundation for Statistical Computing, Vienna.
- Real, R., Barbosa, A.M., Rodríguez, A., García, F.J., Vargas, J.M., Palomo, L.J., Delibes, M., 2009. Conservation biogeography of ecologically interacting species: the case of the Iberian lynx and the European rabbit. *Divers. Distrib.* 15, 390–400.
- Reese, G.C., Wilson, K.R., Hoeting, J.A., Flather, C.H., 2005. Factors affecting species distribution predictions: a simulation modeling experiment. *Ecol. Appl.* 15, 554–564.
- Robertson, M., Caithness, N., Villet, M., 2001. A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. *Divers. Distrib.* 7, 15–27.
- Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jiménez-Valverde, A., Ricotta, C., Chiarucci, A., 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Prog. Phys. Geogr.* 35, 211–226.
- Rotenberry, J.T., Preston, K.L., Knick, S.T., 2006. GIS-based niche modeling for mapping species' habitat. *Ecology* 87, 1458–1464.
- Royle, J.A., Nichols, J.D., Kéry, M., 2005. Modelling occurrence and abundance of species when detection is imperfect. *Oikos* 110, 353–359.
- Soberón, J., 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecol. Lett.* 10, 1115–1123.
- Stockwell, D., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecol. Model.* 148, 1–13.
- Tórres, N.M., De Marco, P., Santos, T., Silveira, L., de Almeida Jácomo, A.T., Diniz-Filho, J.A.F., 2012. Can species distribution modelling provide estimates of population densities? A case study with jaguars in the Neotropics. *Divers. Distrib.* 18, 615–627.
- Van Couwenberghe, R., Collet, C., Pierrat, J.C., Verheyen, K., Gégout, J.C., 2013. Can species distribution models be used to describe plant abundance patterns? *Ecography* 36, 665–674.
- VanDerWal, J., Shoo, L.P., Johnson, C.N., Williams, S.E., 2009. Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. *Am. Nat.* 174, 282–291.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., 2008. Effects of sample size on the performance of species distribution models. *Divers. Distrib.* 14, 763–773.
- Wood, S.N., 2010. *mgcv: mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation – R package version 1.6–2*. <http://www.r-project.org/cran>
- Yañez-Arenas, C., Martínez-Meyer, E., Mandujano, S., Rojas-Soto, O., 2012. Modelling geographic patterns of population density of the white-tailed deer in central Mexico by implementing ecological niche theory. *Oikos* 121, 2081–2089.