

# ECOGRAPHY

## Forum

### On population abundance and niche structure

Luis Osorio-Olvera, Jorge Soberón and Manuel Falconi

*L. Osorio-Olvera (https://orcid.org/0000-0003-0701-5398) and M. Falconi (https://orcid.org/0000-0002-4296-7258), Univ. Nacional Autónoma de México, Facultad de Ciencias, México. LO-O also at: Centro del Cambio Global y la Sustentabilidad A.C., Villahermosa, Tabasco, Mexico. – J. Soberón (https://orcid.org/0000-0003-2160-4148) ✉ (jsoberon@ku.edu), Biodiversity Inst. and Dept of Ecology and Evolutionary Biology, Univ. of Kansas, Lawrence, KS, USA.*

#### Ecography

42: 1415–1425, 2019

doi: 10.1111/ecog.04442

Subject Editor and

Editor-in-Chief: Miguel Araújo

Accepted 25 February 2019

Recent published evidence indicates a negative correlation between density of populations and the distance of their environments to a suitably defined ‘niche centroid’. This empirical observation lacks theoretical grounds. We provide a theoretical underpinning for the empirical relationship between population density and position in niche space, and use this framework to understand the circumstances under which the relationship will fail. We propose a metapopulation model for the area of distribution, as a system of ordinary differential equations coupled with a dispersal kernel. We present an analytical approximation to the solution of the system as well as R code to solve the full model numerically. We use this tool to analyze various scenarios and assumptions. General and realistic demographic assumptions imply a good correlation between position in niche space and population abundance. Factors that modify this correlation are: transitory states, a heterogeneous spatial structure of suitability, and Allee effects. We also explain why the raw output of the niche modeling algorithm MaxEnt is not a good predictor of environmental suitability. Our results elucidate the empirical results for spatial patterns of population size in niche terms, and provide a theoretical basis for a structured theory of the niche.

Keywords: Allee effects, metapopulations, niche structure, population density

#### Introduction

Population density was for some time thought to be related to the position of populations in their geographic range, with larger populations close to the center of the range (Udvardy 1969, Rapoport 1975). This so called ‘abundant centre’ hypothesis has been criticized on empirical grounds (Sagarin and Gaines 2002, Sagarin et al. 2006, Tuya et al. 2008). Another attempt to establish relationships between population density and distributional range used the outputs of ecological niche models (ENMs) as predictors of density (Nielsen et al. 2005, Van der Wal et al. 2009, Törres et al. 2012, Van Couwenberghe et al. 2013, Thuiller et al. 2014, Acevedo et al. 2017, Dallas and Hastings 2018). However, these authors have found inconsistent or weak relationships between density and the output of ENMs. In this vein, on the basis of ideas suggested by Maguire (1973) and Brown (1984), it has been proposed that abundance in a site should be related to the position of the environment of the site in niche



space, with the highest population densities in those sites with environments closest to the center of the fundamental niche (Yáñez-Arenas et al. 2012, Martínez-Meyer et al. 2013, Ureña-Aranda et al. 2015), where fitness is expected to be highest. This would be the niche distance–abundance (NDA) relationship (Dallas and Hastings 2018). Niche distance relationships also exist for the genetic structure of populations (Lira-Noriega and Manthey 2014).

Although there is indeed empirical evidence of a NDA (Yáñez-Arenas et al. 2012, Martínez-Meyer et al. 2013), and theoretically the relationship is to be expected (Maguire 1973), recent papers (Dallas et al. 2017, Dallas and Hastings 2018, Santini et al. 2018) reported consistent near zero correlation values for these variables, for several thousand species. The explanations given for the lack of a consistent relationship include methodological and ecological factors (Dallas et al. 2017, Dallas and Hastings 2018, Santini et al. 2018), the lack of correlation between geographic location and niche suitability, interactions with other species, geographic barriers, and non-linearity in the relationship (Pironon et al. 2017, Dallas and Hastings 2018, Santini et al. 2018). The studies addressing the NDA relationship have been mostly correlative, but to understand a body of empirical results (even negative) a theoretical formulation of the problem is often very useful. This can be done by using an explicit mathematical model of the relationship, as recommended by Dallas and Hastings (2018). Here, we aim to provide a conceptual framework for the problem by modeling the area of distribution of a virtual species as a set of populations linked by dispersal (a metapopulation), on a heterogeneous environment, with the habitat suitability of local populations determined by their fundamental niche and localities linked by migratory rates. The local population dynamics include Allee effects, which may create barriers and areas of zero abundance in marginally favorable habitats, and therefore are very important to understand the distribution of a species (Keitt et al. 2001). Moreover, our numerical simulations emphasize the transient nature of the relationship that may run from impossible to detect, to very good, at different times.

The geographical expression of the fundamental niche is a key factor to understand how abundance changes (Santini et al. 2018). However, a comprehensive treatment of the problem requires rigorously defining the niche variables (interacting, non-interacting), and what shape it is assumed to have (for instance, a convex hull, or an ellipsoid). In this work we follow the conventions of the Grinnellian niche theory (Hutchinson 1978, Soberón 2010, Peterson et al. 2011), which is based on a multidimensional space of non-interactive environmental variables (like climate), and on the relationship between niche space and geographic space called ‘Hutchinson’s duality’ (Colwell and Rangel 2009). The fundamental niche, which is the crucial object being modeled, is defined in terms of population fitness as a function of environmental variables, in the absence of biotic interactions and dispersal limitations. For a given

fitness measure (the intrinsic growth rate, for example), a suitable threshold can be found, above which populations could grow. This creates a level curve for the niche fitness function and establishes the set of all environmental combinations with positive intrinsic growth rate. This set is the fundamental niche.

The metapopulation structure is achieved by allowing dispersal of individuals among the cells of the grid (Smolik et al. 2010, Nenzén et al. 2012, Schurr et al. 2012, Osorio-Olvera et al. 2016). Such a model will generate population abundances (including empty cells, and sink populations) over a landscape with known environmental features and explicit inter-cell movements. This establishes a direct relationship between the niche position of a site, and the abundance of the species in the site. The major complicating factor is movement, which is modeled explicitly, and thus relaxes the steady-state assumption of much species-distribution’s modeling (Svenning and Skov 2004).

The potential area of distribution of a species may be defined simply as the collection of localities where the intrinsic growth rate (without immigration) is greater than zero (Holt et al. 1996, Pulliam 2000), i.e. source populations. However, sink populations sometimes cannot be distinguished from sources using simple presence data, and thus the geographic distribution of a species may be defined as the collection of sinks and sources. A spatially-explicit metapopulation model provides a natural first-principles model for species distributions because it allows to model source and sink dynamics between populations in patches that experiment different migration rates. This kind of models have been studied among others, by Vandermeer (1972), Holt et al. (1997), Kirkpatrick and Barton (1997), Pulliam (2000) and Keitt et al. (2001).

Although simplified in the sense that we assume a deterministic single-species ecology at coarse resolutions (i.e. no habitat heterogeneity is included), with no age-structure and no evolution, our model is, on the one hand, general and robust, and on the other, quite challenging numerically and analytically because the population growth is modeled by non-linear equations, over a grid of potentially many thousands of cells, coupled by dispersal. We present the results of: 1) analytic approximations, and a general theorem on the existence of a solution, as well as 2), numerical results of the relationship between the position of a population in niche space and its size, under a number of scenarios. Although a relationship between population density and distance to a niche centroid is built into the structure of our model, several complications can upset the basic relationship. We explore four major themes: 1) the effect to the relation between abundance and niche, of choosing different metrics of distance in environmental space. 2) The effect to the relation between abundance and niche of the spatial structure of suitability in relation to initial conditions and Allee effects. 3) The effect of transient times in the population dynamics. 4) Why direct outputs of ENMs often fail to correlate with abundance.

## Material and methods

To create a grid for the virtual species we use the region from the eastern side of the Isthmus of Tehuantepec (Mexico) to the border of Panama with Colombia. This was arbitrarily selected as the extent of a grid with  $i = 1, 2, \dots, 638$  cells or patches. The resolution of the grid's cells is half a degree. This region was selected only for purposes of illustration. Each cell is characterized by environmental features. We used the first three principal components (PC) performed on the 19 variables of WorldClim (Hijmans et al. 2005), at 10' resolution, for the entire world.

The virtual species is inspired in *Cryptotis nigrescens*, a shrew complex which has a distributional range in Mesoamerica (Ceballos and Oliva 2005). The virtual species consists of: 1) an ellipsoidal fundamental niche that determines the intrinsic population growth rate, and 2) the parameters for density dependence, the dispersal kernel, initial conditions, and Allee effects. In order to have a semi-realistic fundamental niche geographic coordinates ( $n=32$ ) for recorded populations of the species were obtained from the global biodiversity informatics facility (GBIF) and checked for geographic consistency. These points were used to fit an ellipsoidal shape, containing 95% of the points. This is regarded as a hypothesis for a fundamental niche in the three PCs described above. Strictly speaking fundamental niches cannot be estimated without physiological or biophysical data (Kearney and Porter 2009), and here we simply provide a virtual species with a not entirely arbitrary fundamental niche. We assume the following metapopulation model:

$$\dot{x}_i = (r_i x_i - a x_i^2)(x_i - M) + \sum_b \delta_{b,i} x_b - \sum_b \delta_{i,b} x_i \quad (1)$$

where  $\dot{x}$  is the rate of change of a population, growing in the cell  $i$  at time  $t$ .  $r_i$  is the intrinsic growth rate in cell  $i$ ,  $a$  is a constant density-dependent parameter,  $M$  is the threshold of the Allee effect, and  $\delta_{i,b}$  is the migration rate from cell  $i$  to cell  $b$ . Without migration and Allee effects, Eq. 1 are simple logistics with steady-state values of  $\hat{x}_i = r_i / a$ , or zero. Notice that although the density-dependent parameter  $a$  is constant over cells, the steady-state population size changes as a function of the environment, via the intrinsic growth-rate. This is the link between population density and the position of the local environment in niche space.

The intrinsic growth rate  $r_i$  in each cell is a decreasing function of the Mahalanobis distance between the centroid of the hypothetical fundamental niche ( $\mathbf{N}_F$ ) and the environmental values in the cell. If  $\mu$  is the centroid of  $\mathbf{N}_F$ ,  $\Sigma$  its shape matrix,  $\bar{e}_i$  is the vector of environmental variables in cell  $i$ , and  $r_m$  is the maximum possible value of the intrinsic growth rate, then

$$r_i(\mu, \bar{e}_i) = r_m \exp\left(-\frac{1}{2}(\mu - \bar{e}_i)^T \Sigma (\mu - \bar{e}_i)\right) \quad (2)$$

If the distance between the actual environment in a cell and the centroid of the niche is smaller than a given threshold, the local intrinsic growth rate would be positive. The above is a simple model of population growth expressed as a function of niche distance, and by design creates the effect of correlating suitability with population density. It is very important to notice that this mechanism creates a positive relationship between the fitness function of the fundamental niche and population density, and not between the realized niche and population density. This is a major difference with methods that attempt to correlate outputs of ENMs (by definition, models of the realized niche) to population density.

In our simulation, the maximum value of the intrinsic growth rate is that of *Cryptotis nigrescens*. A value for the growth rate ( $r_m = 0.115 \text{ yr}^{-1}$ ) was obtained from the database of life-history parameters COMADRE (Salguero-Gómez et al. 2016). This value is modulated by the Mahalanobis distance, in such a way that when the distance of environmental values to the centroid of the niche is zero, the intrinsic growth rate is  $0.115 \text{ yr}^{-1}$ . The value of the density-dependence parameter was set to  $a = 0.001$ , to yield population densities (in units of individuals  $\text{km}^{-2}$ ) of about 100 in the most suitable cells. This value was selected to produce reasonable simulation times.

To define the fundamental niche we use the centroid and covariance matrix of the GBIF data points:  $\mu = (5.258301, 3.135807, 1.820532)$  and

$$\Sigma = \begin{pmatrix} 2.74 & -1.44 & -1.50 \\ -1.44 & 1.01 & 0.76 \\ -1.50 & 0.76 & 2.03 \end{pmatrix}$$

### Dispersal kernel

To model the dispersal between cells we used an exponential kernel (Nathan et al. 2012, Nenzén et al. 2012) given by

$$\delta_{i,b}(w_{i,b}) = \begin{cases} k \exp[-w_{i,b}^b] & w_{i,b} \leq D_{\max} \\ 0 & w_{i,b} > D_{\max} \end{cases} \quad (3a)$$

$$(3b)$$

where  $D_{\max}$  is the maximum distance that an individual can travel per year, set to 2 cells,  $w_{i,b}$  is the distance between patches,  $\delta_{i,b}(w_{i,b})$  is the migration rate of individuals going from cell  $i$  to cell  $b$ ,  $k$  is the maximum capacity of migration, that we set to  $0.00575 \text{ degrees d}^{-1}$ , and  $b$  is a positive constant that modulates the traveling capacity. Modeling dispersal is notoriously difficult, and we use the values mostly as an illustration. The distance between cells is simply the Euclidean distance using the coordinates of cells  $i$  and  $b$ . Note that barriers that prevent the migration from patch  $i$  to patch  $b$  can be simulated by taking  $\delta_{i,b} = 0$ . A dispersal matrix that can be partitioned in subsets inaccessible from other subsets is called reducible; on the other hand, a dispersal matrix is called irreducible when every element of the grid can be reached from any other cell. Whether subsets of cells can be reached or not from another subset depends on the dispersal

capabilities of the species in question, and on whether there are Allee effects and cells of low suitability. Using the above dispersal kernel, and ignoring Allee effects, we have an irreducible dispersal. In order to explore the effects of Allee-induced barriers, for each simulation we seeded the species in every one of the cells of grid.

### Distance metrics

To correlate population abundance with position in niche space we use first, the Euclidean distance between the local environments and the fundamental niche centroid. This is straightforward but multivariate niche space variables may be correlated and have different units (for instance, mean temperature in °C and precipitation in mm). Therefore it is important to use a distance metric that will not be affected by those complications. Standardizing variables (Martínez-Meyer et al. 2013) or using PCA-transformed data only partially solves these problems.

The second metric we use is then the Mahalanobis distance measure, which corrects those problems (Legendre and Legendre 1998). The shape matrix of the Mahalanobis distance is the inverse of the covariance matrix for the occurrence points.

### Ecological niche modelling

Several authors have directly used the outputs of the algorithm Maxent (Phillips et al. 2006), as predictors of population size (Van der Wal et al. 2009, Törres et al. 2012, Weber and Grelle 2012). Since Maxent is the most popular ENM algorithm (Joppa et al. 2013), we use it as a third metric. Maxent was calibrated using the GBIF points, using the linear, quadratic, and product features, to avoid over-fitted models. We disabled clamping and left the other parameters with their default values, and used the raw output because this is the output with the most straightforward probabilistic interpretation. Since Maxent is sensitive to background, we fitted models for three different backgrounds (Supplementary material Appendix 1 Fig. A1); the first was located in Mesoamerica; the second one corresponds to a region between eastern Mexico and northern Colombia; finally, the third extends from the United States to Brazil. See Merow et al. (2013) for a discussion on parameterizing a Maxent analysis. We report the AUC reported by the Maxent software as indices of model performance of the Maxent algorithm.

We compared the performance of these three metrics by obtaining Spearman correlations between the value of the metric and population abundance at different time scenarios: 10, 500 and 2500 yr after invasion of an initial cell. The relationships are non-linear, and this is the reason to prefer an order statistic.

### Allee effects

We modeled populations with and without Allee effects (threshold of 20 individuals). We used as initial conditions

all of the single-cell possibilities. We initiated a population in every possible cell, and report the abundances at three different times: 10, 500 and 2500 yr. After 2500 yr, changes in abundance are minor, and a steady state has been achieved. We know by the theoretical result proved in the Supplementary material Appendix 1, 2 that without an Allee effect there is a single global equilibrium point for the system (and all our simulations reach one).

There are dozens of possible ways to model an Allee effect (Boukal and Berec 2002); as a pilot we use a cubic equation (Keitt et al. 2001) yielding our model (1). This point still requires further theoretical analysis, but our simulations consume much computer time making it difficult to do a comprehensive exploration.

### Analytic approximation

Because an analytic solution for Eq. 1 is generally impossible, we developed a simple analytic approximation in order to enable preliminary explorations into the behavior of Eq. 1. First, in the Supplementary material Appendix 1, 2 it is demonstrated that Eq. 1 have a unique positive equilibrium point; however, analytical expressions for this point are intractable. To approximate the solution of the system in the steady state state, we first consider a global emigration rate for cell  $i$ . This is simply the sum

$$\sum_b \delta_{ib} = \delta_i \quad (4)$$

Now we make two assumptions. First, we assume that there is an average, constant total number of migrants approximated by the sum of the migration-independent equilibrium (in every cell) with no Allee effects:

$$R = \frac{\sum_b r_b}{n} \quad (5)$$

Therefore, noticing that the dispersal rates are symmetric, the term describing immigration reduces to:

$$R \sum_b \delta_{ib} = R \delta_i \quad (6)$$

Now the model becomes:

$$\frac{dx_i}{dt} = (r_i - \delta_i)x_i - ax_i^2 + \delta_i R \quad (7)$$

which has an analytic solution for the steady state (denoted with an asterisk) in every cell:

$$x_i^* = \frac{r_i - \delta_i + \sqrt{(r_i - \delta_i)^2 + 4\delta_i a R}}{2a} \quad (8)$$

Eq. 8 are expressions of the steady-state equilibrium point for every cell in the grid under the assumption that there

was a suitable initial population (that is, we ignore initial conditions and Allee effects).

## Numerical solutions

Our analytic approximation only estimates the steady-state and ignores the transitory period and Allee effects; to obtain solutions over time intervals we resort to numerical simulations (developed as an R package deposited in the GitHub archive). This package permits the user to create and parameterize the set of ordinary differential equations of the model (1) to solve them numerically. A compiled C++ routine for ordinary differential equations (from the odeint library (Ahnert and Mulansky 2011)) was used. The computation time grows exponentially with the number of cells; thus, if the grid contains more than a few hundred cells, computation time becomes unmanageable. To illustrate different aspects of the time dynamics of the system, we take three time slices: 10, 500 and 2500 yr, at which the population is in a steady state. Our main results are presented as correlations between observed population density (at each time slice), and the predictors (suitability, Mahalanobis distance to centroid, Euclidean distance to centroid, and the three MaxEnt raw outputs). The relationships between abundance and the predictors tends to be non-linear; thus, we use the Spearman rank correlation, which is not affected by monotonic nonlinearity.

## Data deposition

Data available from the Dryad Digital Repository: <<https://doi.org/10.5061/dryad.33k4gp8>> (Osorio-Olvera et al. 2019).

## Results

### Analytic results

Without Allee effects, as long as the environments in cell  $i$  are inside the fundamental niche, and thus  $r_i > 0$ , a cell can be invaded. With no migration, the steady-state abundance in an invaded cell would be simply the steady-state of the logistic equation for that cell, and 0 if  $r_i < 0$ .

$$\hat{x}_i = \frac{r_i}{a} \quad (9)$$

Because of the way the model is built, with the intrinsic growth rate as a function of the environmental parameters of each cell  $[r(\mu, \vec{e}_i)]$ , at the steady-state and ignoring movements, there can be a direct relationship between  $r$  and abundance, contrary to what Dallas and Hastings (2018) claim. Therefore an inverse relationship between the Mahalanobis distance of the environments in cell  $i$  and the centroid of  $\mathbf{N}_F$  should be expected, as shown in Eq. 2. However, when dispersal is allowed, the analytical approximation reveals an immediate

complication. As shown by Eq. 8, the steady-state abundance is a function of the dispersal rates in ways that may be unrelated to local suitability (unless the parameter determining the immigration rate  $\delta_i$  would be directly correlated with  $r(\mu, \vec{e}_i)$ ). This is a classic source–sink dynamics. Therefore, in general, movements may interfere with the direct correlation between abundance and suitability. In Fig. 1 it is shown how well the numerical data (at the steady state, for a single randomly-chosen locality) correlates with the simplest logistic output, and with the analytic approximation. The simple set of logistic models without movements under-predict abundance, and the analytic approximation slightly over-predict abundance.

Thus, we have a null result: without the complicating effects of dispersal, barriers, transitory states, and Allee effects, equilibrium population density should be inversely related to the distance between the local environments and the centroid of the fundamental niche, at least in an ordinal sense.

## Numerical results

In Fig. 2, we take one randomly selected starting point of the simulation to illustrate how the steady-state abundance is related to various predictors of density, for every point in the grid and with no Allee effects. For most initial points, there is a positive relationship between suitability and population density, monotonously negative between the environmental distances and suitability, and for most Maxent plots, there is a peak, with positive relationship to the left and negative relationships to the right of the peak.

To display the effect of the time dynamics we show boxplots of the Spearman rank correlation ( $\rho$ ) between population density (for times 10 (A), 500 (B) and 2500 (C)), and suitability, inverse Mahalanobis distance to the centroid, inverse Euclidean distance to the centroid (in both distances we take the Maximum – observed values), and raw scores of Maxent for the three backgrounds used (Fig. 3, 4).

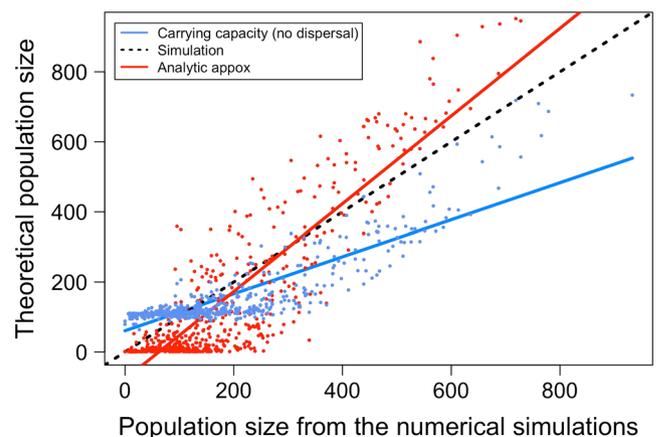


Figure 1. Relationship between population size, calculated numerically, from the 638 initial conditions, without Allee effect, and: the analytical approximation (red); the identity function (dashed line in black); and the simple uncoupled logistic steady-state values (blue).

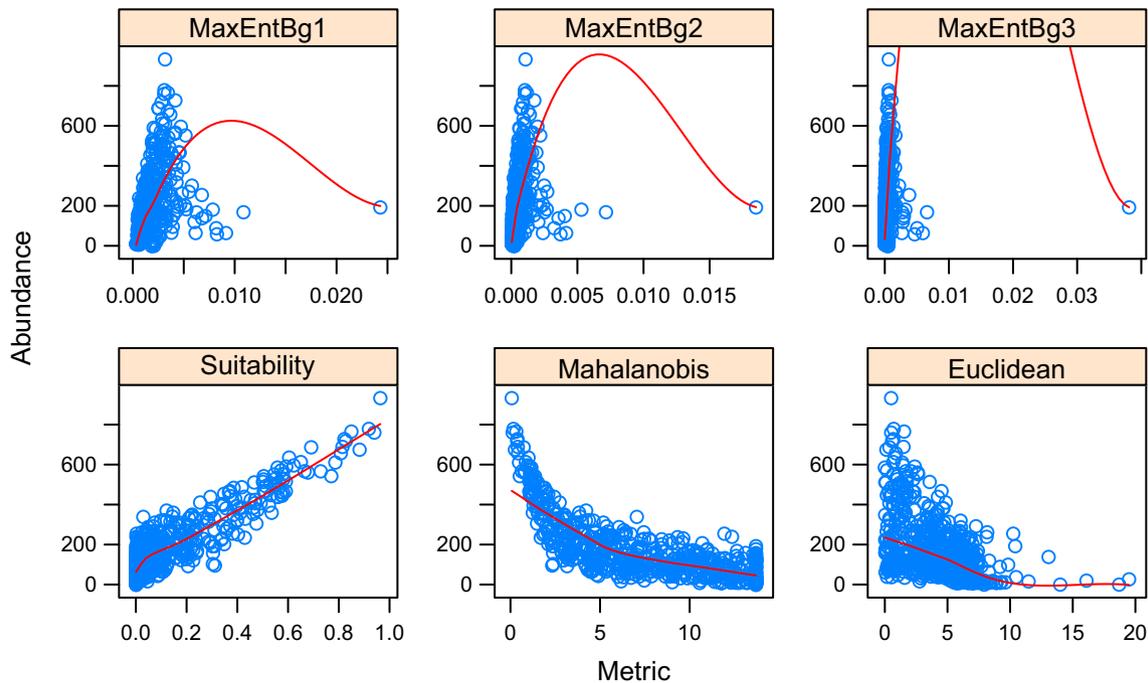


Figure 2. Scatterplots of the relationship between population size, in every grid-cell, as a function of the different predictors, for two randomly-selected initial points in the simulations.

We show the results for population sizes reached after initiating growth in each one of the 638 cells, against the suitability value, without Allee effects.

The results clearly show that during the transitory period correlations between population density and suitability of habitat can be bad or mediocre. However, towards the steady state (Fig. 3c), the  $\rho$  value becomes high, and with low variance for all predictors. This is to be expected with suitability, due to the way the model was constructed, and the value is also as high for a Mahalanobis distance to the niche centroid (as the suitability function is the rescaled version of suitability), but lower for the Euclidean distance. The Spearman correlation is high for the three MaxEnt raw outputs. At the steady state, practically there is no variance in the results. This is explained because without an Allee effect, every point in the grid is accessible from every other point and this erases difference due to initial conditions (5A).

The inclusion of Allee effects greatly complicates this. In Fig. 4 we show boxplots of  $\rho$  values between the same variables, for three times along the trajectory (10, 500 and 2500 yr). The Allee effect acts primarily by reducing the median value of the rank correlation  $\rho$ , and also increasing its variance, because there are now many localities from which a starting population cannot reach suitable regions (this is a result of the Allee effect creating ‘barriers’. This effect has a strong geographical component, as shown in Fig. 5b, where it is shown how the barrier effect is highly structured geographically: the Allee effect creates regions in the map where the correlations are high, and others where the correlations are low.

## Discussion

If niche position and fitness correlate, as it is expected, (Maguire 1973) all things being equal, position in niche space would correlate with population density. As long as the fitness function defining the niche has some regular structure with maximum fitness in its center, as Maguire (1973) suggested, an appropriate measure of distance to the center should correlate with abundance. However, this expectation is complicated by a number of factors, as our results show.

### Movements and Allee effects

The main complication that we discuss is the effect, shown in Eq. 8, and confirmed by the simulations, that dispersal for a population with a spatial structure (i.e. a set of metapopulations), can substantially interfere with the expected theoretical correlation between local environmental suitability and population density. In our model, metapopulation dynamics interferes with the imposed relationship between abundance and suitability. We suspect this is the main reason why recent studies have failed to report a correlation between population density and position in niche space. Equation 8 suggest that the confounding effect of immigration is mediated by the rates of arrival to patches. If there is no covariance between suitability and patch attractiveness, as determined by  $\delta_p$ , or the covariance is negative, then the expected null relationship between abundance and suitability (distance to the centroid) may be seriously upset (Soberon 1986).

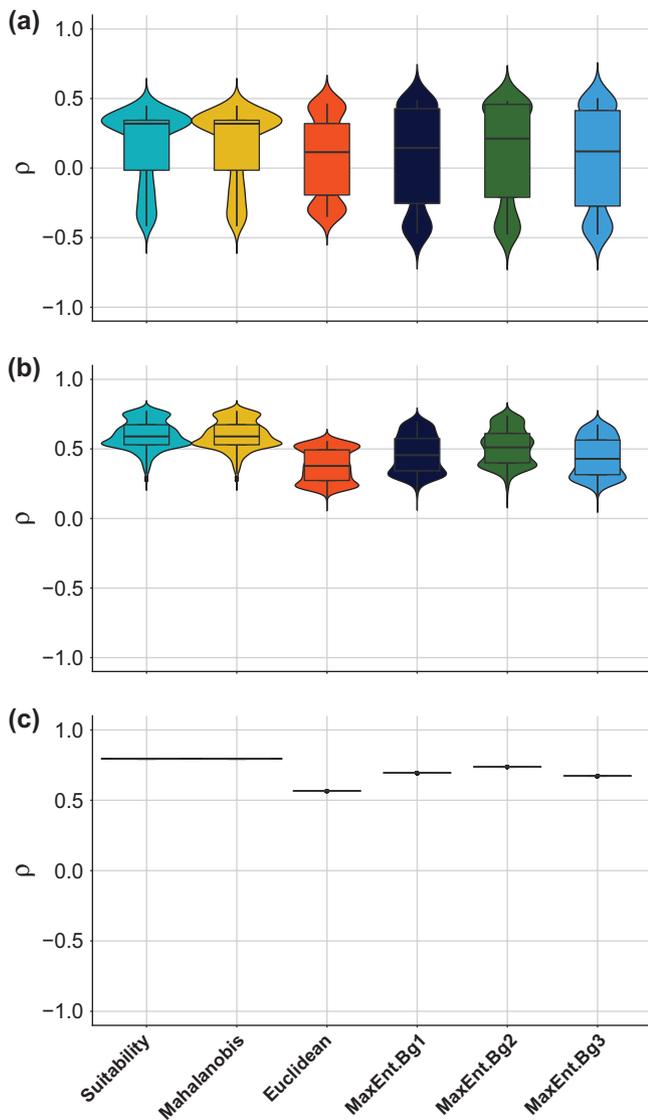


Figure 3. Distribution of the correlation value between niche metric and population abundance for different times in the model without Allee effect. (a) Ten years after dispersal; (b) 500 yr period; (c) 2500 yr period. A Spearman rank correlation was used for variables with a monotonic relationship with abundance (suitability, Mahalanobis and Euclidean distances), and a Spearman correlation for the Maxent outputs.

Another subtle effect of dispersal is expressed via the Allee effect. Having an Allee effect implies that there may be barriers due to regions of positive but relatively low suitability, because migration fails to start a population under an Allee threshold. This is an interaction between the process of actually moving, and the process of establishing subpopulations that may act as stepping-stones. The Allee effects amplify the role of low suitability regions to act as barriers (Keitt et al. 2001, Pironon et al. 2017), and make information about the initial conditions crucial to determine the time trajectory of the spread of an invasion, or the expansion of a population. This effect has been studied theoretically by Kirkpatrick

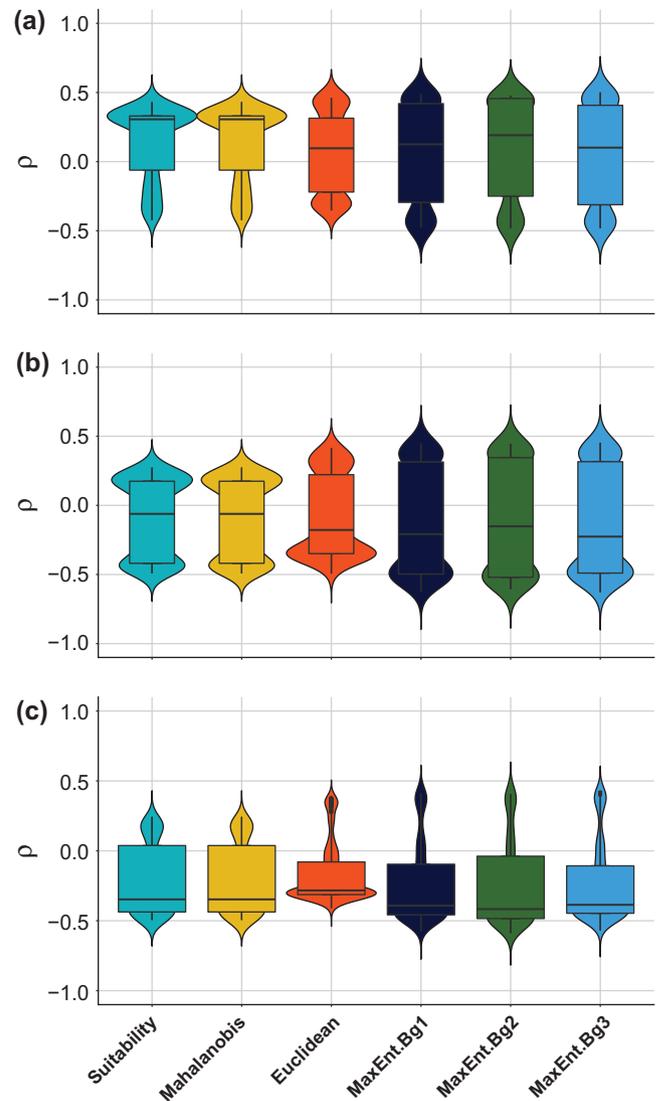


Figure 4.  $\rho$  distribution of the relationship between niche metric and population abundance for different times in the model with Allee effect. (a) Ten years after dispersal; (b) 500 yr period; (c) 2500 yr period. A Spearman rank correlation was used for all variables.

and Barton (1997) and Holt (2009a), but our method and software permits realistic simulations (using actual geographic regions and parameters obtained from data), as we show with our semirealistic virtual species.

The main lesson we extract, in the context of the problem under study, is that if a geographic region contains regions of low suitability, even if movements are allowed among all patches, then low suitability and Allee effects may create barriers isolating empty but perfectly suitable subregions, depending on the initial conditions. Clearly, such suitable but empty localities would upset the correlation between abundance and distance to the centroid, or abundance and correlative niche models. This is one of the mechanisms that Dallas et al. (2017) use to explain the low correlations they found in their review of

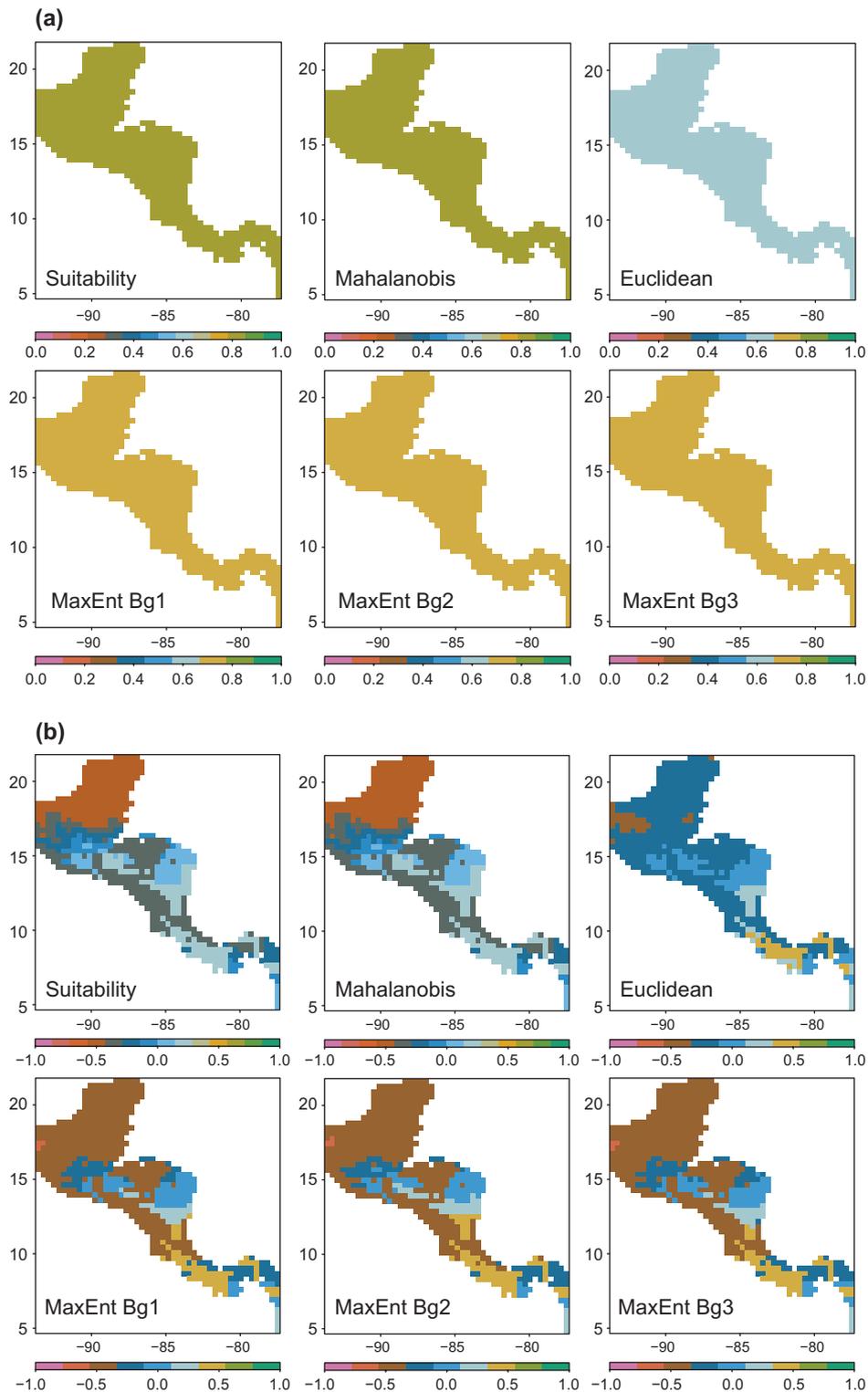


Figure 5.  $\rho$  maps of the relationship between niche metric and population abundance after 2500 yr of invasion. (a)  $\rho$  map for each initial condition (grid-cell) in the model without Allee effect. (b)  $\rho$  map for each initial condition (pixel) in the model with an Allee effect.

the relationship between population density and distance to niche centroid. Although our software permits realistic simulations, exploring this problem is made difficult by the

very demanding numerical challenges. This is one of the reasons why we did not perform explorations of a wider range of parameters or Allee models.

## Maxent

As stated previously, empirical studies have sought a correlation between the mean abundance and the direct output of presence-background ENM algorithms. The most commonly used method is the machine-learning algorithm called MaxEnt (Phillips et al. 2004). Maxent scores have been used to correlate with species density (Van der Wal et al. 2009, Tórres et al. 2012, Weber et al. 2017). The results have been inconsistent or weak. However, the same outputs organized as niche distances can be predictive of abundance (Martínez-Meyer et al. 2013). We hypothesize that this is a consequence of two things. First, the raw output of the Maxent program is not a measure of fitness or suitability, but simply a measure of how similar the environments in the background cells are to the environments of those cells in which the species has been found (Merow et al. 2013). Maxent finds a solution for the probability  $P(\bar{z}_i | Y = 1)$ , i.e. the probability of having the environmental variables  $\bar{z}_i$  given the fact that the species has been observed ( $Y = 1$ ) in cell  $i$ . This is not an index of suitability. This is an important and almost universally disregarded point: the outputs of ENM algorithms may be used to classify cells in a spatial grid in terms of similarity to a subset where a species has been detected; and whether this would coincide with a classification based on a generally unavailable fitness function is an open question. Simply assuming that the coincidence would exist is unwarranted.

Second, a correlative ENM algorithm predicts a potential distribution, not an actual one (Peterson 2011), and therefore many regions with high suitability value may zero densities. ENMs find similarities, not fitness.

However, the assumption that the fitness function defines a regular shape in niche space allows points to be ordered in relation to such a shape. In their empirical explorations, (Martínez-Meyer et al. 2013) use not the value of the raw output of Maxent, but first selected the cells with observed presences, ordered them per value of the raw output, retained the upper 90%, and used these to estimate a mean value per variable in the model. This set of mean values constitutes their approximation to the centroid of a niche with an implicitly assumed structure expressed in fitness terms; and it is from this centroid that distances (Euclidean, in their case) can be calculated to each point in the region. If it is true that the fitness function indeed describes an ellipsoidal structure in niche space, then the distance between the environments  $\bar{z}_i$  and the hypothesized niche centroid  $\mu$  is at least an ordinal measure of suitability. The empirical association between the distance to the centroid and abundance found by Martínez-Meyer et al. (2013) is therefore consistent with the hypothesis of a structure of the niche in terms of fitness.

When true absence data is available, it may be possible to estimate a true probability of occurrence. Thuiller et al. (2014) used presence-absence data to study the correlation between probability of occurrence of species in plots (using a GLM regression), and estimations of three demographic parameters (the intrinsic growth rate  $r$ , a 'carrying capacity'  $K$ , and the initial population size  $N$ ). These were obtained

from locally-defined Ricker models for population growth, without metapopulation structure. Their results show that there is some tendency for probability of occurrence to predict the value of  $N$ ; however, as they remark, their results are inconclusive, among other issues, because without dispersal effects simple species distribution models ignore critical information required to understand range dynamics. This is precisely the main result of our work: spatial dynamics is essential to understand the relationship between niche and population abundance.

## Effects due to choice of distance metric

Since we defined fitness as a function of a Mahalanobis distance to a niche centroid, it is to be expected that using another distance will yield poorer correlations, which is what we observe. However, the lesson from our results is more than an artifact of an ellipsoidal choice for the fundamental niche. Environmental space is intrinsically non Euclidean (the variables have non-zero correlations). Therefore using an Euclidean distance probably is not appropriate in general, as we saw in Fig. 3. Even if the space has been transformed by principal components, as it is the case in the study of Dallas et al. (2017), the specific set of occurrence points may be not well described by the principal components transformation. The Mahalanobis distance automatically corrects for different units and it can help to correct the non-orthogonality of axes for the occurrence points, and for these reasons we would expect that in general it would be a better metric to calculate distance to a niche centroid.

## Effects due to transient phases

Non-equilibrium behavior is mentioned in the literature as one possible explanation of poor correlations between population density and niche position (Dallas et al. 2017, Santini et al. 2018), and our results confirm this. Unsurprisingly, when the system is far from its steady-state, the correlation between abundance and any of its predictors is not very good. Although this is an intuitive result that has been reported in the context of distribution modeling (Svenning and Skov 2004), a more general conclusion can be extracted: systems suspected to be far from a steady-state (for instance, an invasive species at an early phase of its invasion) should not be expected to have a good correlation between presence and measures of niche suitability (Vaclavik and Meentemeyer 2012); thus, a good correlation with abundance is much less likely.

## Conclusion

For theoretical reasons related to basic population dynamics one would expect that the relationship between population abundances in a locality, and the distance to some 'niche centroid' estimated using ENM, should be negative: fitness, and hence abundance (under general assumptions) should be related to environmental suitability. This is in some sense the

null hypothesis. However, by analytically and numerically examining a spatially-explicit population model, we show how the use of an inappropriate measure of distance, transitory (non-equilibrium) dynamics (Svenning and Skov 2004), and metapopulation dynamics (Pulliam 2000) coupled with Allee effects (Holt 2009b) can interfere with the simple pattern expected by a population dynamics argument.

These points have practical implications for testing the NDA. First, since vagile species are expected to have poorer NDA relationships than sedentary species, special care should be taken when using such species as tests of the NDA. Second, the spatial structure of niche suitability in relation to rates of movement into patches (i.e. the relationship between  $r_i$  and  $\delta_i$ ) is also important. Non correlated  $r_i$  and  $\delta_i$  may create a net of populations with size determined more by movements than by intrinsic environmental suitability. This can obscure the NDA relationship. Third, species in the process of invading new regions are out of equilibrium and should not be used to test the NDA. Finally, correlative species distribution modeling predicts potential ranges of distribution, rather than actual distributions. Moreover, correlative ENM methods do not measure fitness directly. Therefore using the outputs of correlative methods like Maxent as predictors for population density adds further assumptions to testing the NDA. Although we did not find a major difference between the raw output of Maxent and niche distances, it is important to keep in mind always the fact that a distribution is due to more than just suitable environments, and thus testing the NDA on the basis of correlative models should be done with great care. Keeping in mind these complications will help to interpret correctly the growing body of data on the relationship between population dynamics and the structure of niche space.

*Acknowledgements* – Blitzi Soberon provided moral encouragement to our work. Town Peterson gave us many useful suggestions and we are grateful for his help.

*Funding* – JS acknowledges partial support from NSF grant ABI 1458640, and LOO from a CONACyT (339233) scholarship, the Posgrado en Ciencias Biológicas, UNAM and CONACyT-FORDECyT 273646. MF and LOO acknowledges to PAPIIT-UNAM IN116018 (2018).

## References

Acevedo, P. et al. 2017. Population dynamics affect the capacity of species distribution models to predict species abundance on a local scale. – *Divers. Distrib.* 23: 1008–1017.

Ahnert, K. and Mulansky, M. 2011. Odeint – solving ordinary differential equations in C++. – *AIP Conf. Proc.* 1389: 1586–1589.

Boukal, D. S. and Bercé, L. 2002. Single-species models of the Allee effect: extinction boundaries, sex ratios and mate encounters. – *J. Theor. Biol.* 218: 375–394.

Brown, J. H. 1984. On the relationship between abundance and distribution of a species. – *Am. Nat.* 124: 255–279.

Ceballos, G. and Oliva, G. 2005. Los mamíferos silvestres de México. – Fondo de Cultura Económica y CONABIO.

Colwell, R. K. and Rangel, T. F. 2009. Hutchinson's duality: the once and future niche. – *Proc. Natl Acad. Sci. USA* 106: 19651–19658.

Dallas, T. A. and Hastings, A. 2018. Habitat suitability estimated by niche models is largely unrelated to species abundance. – *Global Ecol. Biogeogr.* 27: 1448–1456.

Dallas, T. et al. 2017. Species are not most abundant in the centre of their geographic range or climatic niche. – *Ecol. Lett.* 20: 1526–1533.

Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – *Int. J. Climatol.* 25: 1965–1978.

Holt, R. D. 2009a. Predation and community organization. – In: Levin, S. A. (ed.), *The Princeton guide to ecology*. Princeton Univ. Press, pp. 274–281.

Holt, R. D. 2009b. Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. – *Proc. Natl Acad. Sci. USA* 106: 19659–19665.

Holt, R. D. et al. 1996. Food webs in space: an island biogeographic perspectives. – In: Polis, G. A. and Winemiller, K. (eds), *Food webs. Integration and patterns and dynamics*. Chapman and Hall, pp. 313–323.

Holt, R. et al. 1997. On the relationship between range size and local abundances: back to the basics. – *Oikos* 78: 183–190.

Hutchinson, G. E. 1978. *An introduction to population ecology*. – Yale Univ. Press.

Joppa, L. N. et al. 2013. Troubling trends in scientific software use. – *Science* 340: 814–815.

Kearney, M. and Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. – *Ecol. Lett.* 12: 334–350.

Keitt, T. H. et al. 2001. Allee effects, invasion pinning, and species' borders. – *Am. Nat.* 157: 203–216.

Kirkpatrick, M. and Barton, N. H. 1997. Evolution of a species range. – *Am. Nat.* 150: 1–23.

Legendre, P. and Legendre, L. 1998. *Numerical ecology*. – Elsevier Scientific Publishing Company.

Lira-Noriega, A. and Manthey, J. D. 2014. Relationship of genetic diversity and niche centrality: a survey and analysis. – *Evolution* 68: 1082–1093.

Maguire, B. 1973. Niche response structure and the analytical potentials of its relationship to the habitat. – *Am. Nat.* 107: 213–246.

Martínez-Meyer, E. et al. 2013. Ecological niche structure and rangewide abundance patterns of species. – *Biol. Lett.* 9: 20120637.

Merow, C. et al. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. – *Ecography* 36: 1058–1069.

Nathan, R. et al. 2012. Dispersal kernels: review. – In: Clobert, J. et al. (eds), *Dispersal ecology and evolution*. Oxford Univ. Press, pp. 186–210.

Nenzén, H. K. et al. 2012. demoniche – an R-package for simulating spatially-explicit population dynamics. – *Ecography* 35: 577–580.

Nielsen, S. E. et al. 2005. Can models of presence-absence be used to life abundance? Two case studies considering extremes in life history. – *Ecography* 28: 197–208.

Osorio-Olvera, L. A. et al. 2016. Sobre la relación entre idoneidad del hábitat y la abundancia poblacional bajo diferentes escenarios de dispersión. – *Rev. Mex. Biodivers.* 87: 1080–1088.

- Osorio-Olvera, L. et al. 2019. Data from: On population abundance and niche structure. – Dryad Digital Repository, <<https://doi.org/10.5061/dryad.33k4gp8>>.
- Peterson, A. T. 2011. Ecological niche conservatism: a time-structured review of evidence. – *J. Biogeogr.* 38: 817–827.
- Peterson, A. T. et al. 2011. Ecological niches and geographic distributions. – Princeton Univ. Press.
- Phillips, S. et al. 2004. A maximum entropy approach to species distribution modeling. – 21st Int. Conf. Mach. Learn.
- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Pirone, S. et al. 2017. The ‘Hutchinsonian niche’ as an assemblage of demographic niches: implications for species geographic ranges. – *Ecography* 41: 1103–1113.
- Pulliam, H. R. 2000. On the relationship between niche and distribution. – *Ecol. Lett.* 3: 349–361.
- Rapoport, E. H. 1975. Areografía: estrategias geográficas de las especies. – Fondo de Cultura Económica.
- Sagarin, R. D. and Gaines, S. D. 2002. Geographical abundance distributions of coastal invertebrates: using one-dimensional ranges to test biogeographic hypotheses. – *J. Biogeogr.* 29: 985–997.
- Sagarin, R. D. et al. 2006. Moving beyond assumptions to understand abundance distributions across the ranges of species. – *Trends Ecol. Evol.* 21: 524–530.
- Salguero-Gómez, R. et al. 2016. COMADRE: a global data base of animal demography. – *J. Anim. Ecol.* 85: 371–384.
- Santini, L. et al. 2018. Addressing common pitfalls does not provide more support to geographical and ecological abundant-centre hypotheses. – *Ecography* doi:10.1111/ecog.04027
- Schurr, F. M. et al. 2012. How to understand species’ niches and range dynamics: a demographic research agenda for biogeography. – *J. Biogeogr.* 39: 2146–2162.
- Smolik, M. G. et al. 2010. Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. – *J. Biogeogr.* 37: 411–422.
- Soberón, J. 1986. The relationship between use and suitability of resources and its consequences to insect population size. – *Am. Nat.* 127: 338–357.
- Soberón, J. M. 2010. Niche and area of distribution modeling: a population ecology perspective. – *Ecography* 33: 159–167.
- Svenning, J.-C. and Skov, F. 2004. Limited filling of the potential range in European tree species. – *Ecol. Lett.* 7: 565–573.
- Thuiller, W. et al. 2014. Does probability of occurrence relate to population dynamics? – *Ecography* 37: 1155–1166.
- Tórres, N. M. et al. 2012. Can species distribution modelling provide estimates of population densities? A case study with jaguars in the Neotropics. – *Divers. Distrib.* 18: 615–627.
- Tuya, F. et al. 2008. Testing the ‘abundant centre’ hypothesis on endemic reef fishes in south-western Australia. – *Mar. Ecol. Prog. Ser.* 372: 225–230.
- Udvardy, M. 1969. Dynamic zoogeography: with special reference to land animals. – van Nostrand Reinhold.
- Ureña-Aranda, C. et al. 2015. Using range-wide abundance modeling to identify key conservation areas for the micro-endemic Bolson tortoise (*Gopherus flavomarginatus*). – *PLoS One* 10: 1–14.
- Vaclavik, T. and Meentemeyer, R. K. 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. – *Divers. Distrib.* 18: 73–83.
- Van Couwenberghe, R. et al. 2013. Can species distribution models be used to describe plant abundance patterns? – *Ecography* 36: 665–674.
- Van der Wal, J. et al. 2009. Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. – *Am. Nat.* 174: 282–291.
- Vandermeer, J. H. 1972. Niche theory. – *Annu. Rev. Ecol. Syst.* 3: 107–132.
- Weber, M. M. and Grelle, C. E. 2012. Does environmental suitability explain the relative abundance of the tailed tailless bat, *Anoura caudifer*? – *Nat. Conserv.* 10: 221–227.
- Weber, M. M. et al. 2017. Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. – *Ecography* 40: 817–828.
- Yáñez-Arenas, C. et al. 2012. Modelling geographic patterns of population density of the white-tailed deer in central Mexico by implementing ecological niche theory. – *Oikos* 121: 2081–2089.

Supplementary material (available online as Appendix ecog-04442 at <[www.ecography.org/appendix/ecog-04442](http://www.ecography.org/appendix/ecog-04442)>). Appendix 1–2.