

## The crucial role of the accessible area in ecological niche modeling and species distribution modeling

Narayani Barve<sup>a</sup>, Vijay Barve<sup>a,1</sup>, Alberto Jiménez-Valverde<sup>a,1</sup>, Andrés Lira-Noriega<sup>a</sup>, Sean P. Maher<sup>a,2</sup>, A. Townsend Peterson<sup>a,\*</sup>, Jorge Soberón<sup>a</sup>, Fabricio Villalobos<sup>b</sup>

<sup>a</sup> Biodiversity Institute, University of Kansas, Lawrence, KS 66045, USA

<sup>b</sup> Posgrado en Ciencias Biológicas and Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70-275, CP 04510 México City, Mexico

### ARTICLE INFO

#### Article history:

Received 3 July 2010

Received in revised form 4 December 2010

Accepted 13 February 2011

Available online 7 April 2011

#### Keywords:

Accessible area

Area of distribution

Ecological niche

Geographic extent

Pleistocene

### ABSTRACT

Using known occurrences of species and correlational modeling approaches has become a common paradigm in broad-scale ecology and biogeography, yet important aspects of the methodology remain little-explored in terms of conceptual basis. Here, we explore the conceptual and empirical reasons behind choice of extent of study area in such analyses, and offer practical, but conceptually justified, reasoning for such decisions. We assert that the area that has been accessible to the species of interest over relevant time periods represents the ideal area for model development, testing, and comparison.

© 2011 Elsevier B.V. All rights reserved.

The area of distribution of a species is a complex expression of its ecological and evolutionary history (Brown et al., 1996; Gaston, 2003). Among factors determining distributional areas, the most important are the limits of the species' tolerances and needs for certain abiotic conditions, the suite of other species with which it interacts, and the potential for dispersal and colonization within a given time period (Grinnell, 1917; Udvardy, 1969; MacArthur, 1972; Pulliam, 2000; Soberón and Peterson, 2005). These factors are expressed across a geographic matrix that is in itself dynamic: climatic conditions, coastlines, locations of rivers, mountain ranges, and other possible barriers change over time (Jackson and Overpeck, 2000; Zachos et al., 2001). A species can respond in diverse ways to its physical and biotic environments: ecological niches may evolve, or may remain conserved; in the latter case, shifting conditions can cause geographic isolation and eventual speciation (Wiens, 2004; Byrne et al., 2008; Stigall, 2008).

The processes affecting species' geographic distributions have been summarized in mathematical models that only take into account ecological factors (Vandermeer, 1972; Pulliam, 2000; Holt,

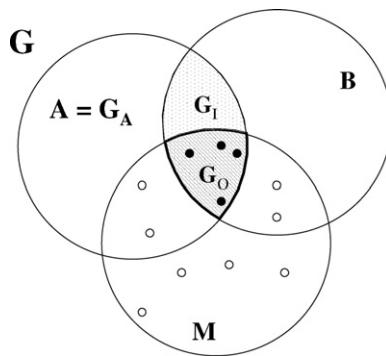
2003; Soberón, 2010); those that include microevolutionary (Holt, 2003) and macroevolutionary processes (Rangel et al., 2007; Roy and Goldberg, 2007) tend to be very complex (Gotelli et al., 2009). In formal terms, if interest is in changes in the area of distribution, particularly in terms of the mechanistic factors that influence it, one would need to solve ecological and evolutionary equations simultaneously, which is normally not tractable analytically. For this reason, Soberón and Peterson (2005) proposed a much-simplified approach based on a static approximation to describe three classes of factors, as follows: (1) Environmental factors with values not dependent on the dynamics of the species' population. These are mostly abiotic dimensions such as climate, topography, and solar radiation. These variables have their own intrinsic dynamics, but may be regarded as independent of the presence or abundance of the species in question. The geographic region presenting favorable conditions of this type is called **A**, the geographic expression of what Jackson and Overpeck (2000) call the "potential niche" of a species and Peterson et al. (2012) the "existing fundamental niche." This term actually describes the intersection of the fundamental ecological niche (defined physiologically as the ranges or values or environmental variables where a species can survive; cf. Hutchinson, 1957) with the set of environments that actually exist on the landscape in question (Soberón and Nakamura, 2009). (2) Sets of variables that are dynamically linked to the population numbers of the species of interest include limited food resources; presence and influence of competitors, predators, and mutualists;

\* Corresponding author. Tel.: +1 785 864 3926.

E-mail address: [town@ku.edu](mailto:town@ku.edu) (A.T. Peterson).

<sup>1</sup> Present address: Departamento de Biología Animal, Facultad de Ciencias, Universidad de Málaga, 29071 Málaga, Spain.

<sup>2</sup> Present address: Odum School of Ecology, University of Georgia, 140 E. Green St., Athens, GA 30602.



**Fig. 1.** A **BAM** diagram (Soberón and Peterson, 2005) to illustrate the three interacting factors that determine a first-order view of species' geographic distributions: **Biotic**, **Abiotic**, and **Movement**. Closed circles = source populations, open circles = sink populations.

and other factors, mostly biotic in nature. This region is denoted by **B**, although it is difficult to estimate owing to the fine spatiotemporal resolution and potentially complex nature of biotic dimensions (Engler and Guisan, 2009; Anderson and Raza, 2010); indeed, in the simulations to be developed in this paper, we neglect **B** entirely (see discussion in Soberón, 2010). (3) The parts of the world that have been accessible to the species via dispersal over relevant periods of time are symbolized by **M**. As a first approximation, the intersection of these three sets (i.e.,  $B \cap A \cap M$ ) determines the region in which one can find a species (Fig. 1), while  $B \cap A$  defines the distributional potential of the species were dispersal barriers to be removed (e.g., in the case of invasive species); this simple heuristic framework is termed the **BAM** diagram (Soberón and Peterson, 2005).

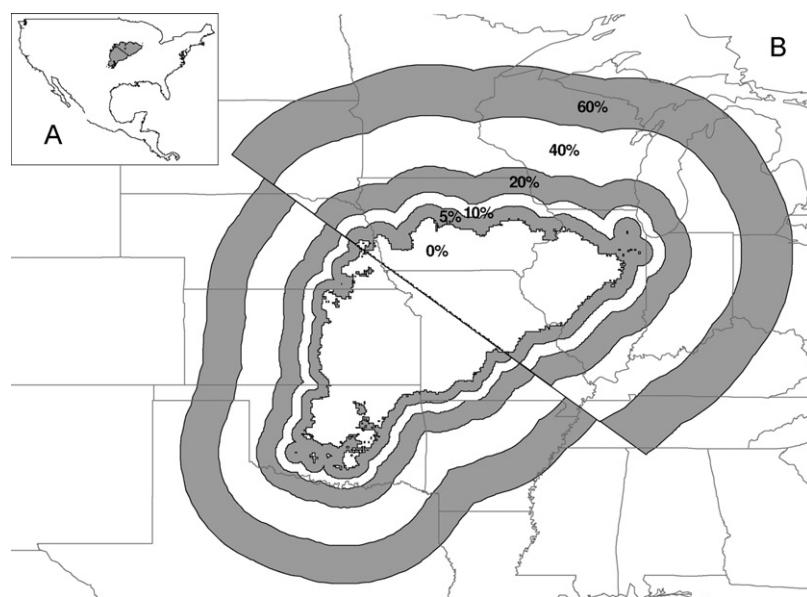
Ecological niche modeling (ENM) is used to estimate the realized (as opposed to the fundamental) coarse-resolution environmental requirements of species, which can be projected onto real-world landscapes to identify regions in which the requirements of species are manifested; given the coarse-resolution nature of the variables employed, and the practical problems involved in measuring variables related to **B**, this region is defined mostly by variables related to **A**. The region **M** depends on opportunities for and constraints on movements of the species, factors not often included in modeling efforts, although some exceptions exist (Kot et al., 1996; Engler and Guisan, 2009; Cabral and Schurr, 2010; Smolik et al., 2010). However, over the period of the species' existence, its populations may experience a suite of environmental conditions, only some of which are within its fundamental niche. Because estimation of ecological niches using correlative methods relies upon comparisons of environmental characteristics of sites of known occurrence with those associated with sites of the background of conditions presented across the study area (Stockwell and Peters, 1999; Phillips et al., 2006), it is crucial to understand and outline the area that the species has potentially visited (**M**), as *the species will be absent from outside of this area for reasons unrelated to A* (Anderson and Raza, 2010).

In this paper, we present a framework for thinking about and estimating **M** in ecological niche modeling (and the related species distribution modeling). We present a simulation of the next step of the **BAM** scheme, in which we model **M** as well as **A** using simple and straightforward rules. To make this scheme operational, we take advantage of heretofore unavailable data on past climates that affected species' distributions. Our scheme is simple conceptually, but computationally intensive, which is becoming a distinct modality of science (Hey et al., 2009). Although our emphasis is on estimating the set of areas that a species has "sampled" over its history to inform a niche model, the ideas we explore have additional implications for fields such as biogeography, macroecology, and phylogeography.

## 1. Importance of estimating the region **M**

The extent used during the niche modeling process has pervasive influences on the outcome of the model. Specifically, if the extent under consideration is too limited to represent **M** entirely, the importance of coarse-resolution factors such as climate in delimiting species' distributions may be underestimated. A rather dramatic example of this limitation was the recent conclusion that climate has negligible influences on species' distributions (Beale et al., 2008), a conclusion that turned out to be heavily extent-dependent (Jiménez-Valverde et al., 2010). More generally, **M** has important implications in all aspects of ecological niche modeling studies, including model parameterization, model validation, and model comparisons, as follows:

- (1) *Effects on model training:* In model training, **M** determines the arena of comparison – that is, the area within which presences may exist and within which absences are meaningful, in that they represent sites with the broader background landscape actually likely to have been "tested" by the species for suitability, but not occupied. VanDerWal et al. (2009) illustrated the effect of varying the extent on parameterization of ENM models: using 12 vertebrate species in Australia, they fitted maximum entropy models (Phillips et al., 2006; Phillips and Dudík, 2008) within 10 different buffers around known occurrence points. Their results showed that increasing the extent reduced the number of variables included in the models, which in turn changed predicted geographic patterns, such that suitable areas were broader when models were calibrated at greater extents. Anderson and Raza (2010) present parallel analyses that also show significant effects of training area on model results, although in their case modeled suitable areas were smaller when models were calibrated at broader extents.
- (2) *Effects on model validation:* Lobo et al. (2008) highlighted effects of extent (which should be equivalent to **M**) on results of model validation tests and, in particular, on the results of receiver operating characteristic (ROC AUC) tests. During validation exercises, areas outside **M** (where the species cannot occur) will generally be predicted at lower suitability levels; as a result, inclusion of these areas (which hold no presence data, but owing to restrictions resulting from **M** and not **A**) makes the model look better than it actually is. Increasing the extent also often includes absences that are more distant environmentally from the presences, so the ROC AUC (as well as other discrimination measures, such as sensitivity or specificity) values increase, because it is easy to parameterize models with good discrimination capacity but that are low in useful information (e.g., it is easy to see that no polar bears live in the Tropics, but this information is scarcely useful; VanDerWal et al., 2009). In consequence, high and statistically significant AUC values can be obtained simply by increasing the area of study (Jiménez-Valverde et al., 2008); here, careless use of these statistics procedures gives results that are of dubious merit (Lobo et al., 2008). Indeed, such careless applications may cast doubt on the robustness of the entire field, given lack of means of objective model evaluation (Peterson, 2005).
- (3) *Effects on model comparisons:* Warren et al. (2008) highlighted the importance of correct specification of null hypotheses in comparisons of modeled niches: some previous studies have tested hypotheses of niche identity (Graham et al., 2004), while others have tested hypotheses of relative niche similarity (Peterson et al., 1999). Warren et al. (2008) developed and presented software tools enabling both types of comparisons (ENMTools, <http://enmtools.com>). Importantly, only the background similarity tests include specification of an "area of interest" – such that niches are "similar" only relative to some



**Fig. 2.** Distributions of two virtual species generated from the set of environments most similar to the climate of Lawrence, Kansas (B; details in text). The two species share the same fundamental niche, but are separated by a hypothetical barrier to dispersal (line running northwest to southeast). The distributional area of each virtual species is shown in relation to the 6 extents across which examples of model calibration, evaluation, and comparison are developed (0%, 5%, 10%, 20%, 40%, and 60%). Inset (A) shows the location of the ranges of the two “species” relative to North America.

background area, which we would argue should be equivalent to **M**. Specification of this area of interest has critical and direct effects on the relative similarity of two niches sampled from that space (see example developed below).

In sum, **M** has pervasive effects throughout the realm of ecological niche modeling and species distribution modeling. Essentially no aspect of this modeling approach can be developed appropriately without a carefully considered hypothesis of **M**, as this region is the only appropriate arena for such analyses. In the following section, we illustrate these effects by means of exploration of the niche and distribution of a virtual species.

## 2. Worked example: a virtual species

A virtual ecological niche was generated by means of visualizing climatic variation across North America with respect to the conditions presented at Lawrence, Kansas (7 of the so-called “bioclimatic variables” at a resolution of  $0.17^\circ$  from [Hijmans et al., 2005](#): annual mean temperature, mean diurnal temperature range, maximum temperature of warmest month, minimum temperature of coldest month, annual precipitation, and precipitation of the wettest and driest months). We transformed this 7-dimensional space by means of principal components analysis (PCA) to create 4 orthogonal axes summarizing climatic variation across North America. We chose the 5% of the overall distribution of environments associated with pixels that were most similar climatically to our reference site, and identified the geographic area matching those environments. This area, curiously, is in large part coincident with the original range of tall-grass prairie in the central part of the continent.

We then created a pair of virtual “sister” species sharing this same fundamental ecological niche by bisecting its spatial footprint with a barrier that might be a large river crossing the region from northwest to southeast. Our two “species” are found in the suitable regions fitting the same virtual niche on either side of this barrier (Fig. 2a). We emphasize that the two species share a single fundamental ecological niche; the only difference between them is that they inhabit the spatial footprints of that niche on either side of a barrier, which may manifest different portions of the envi-

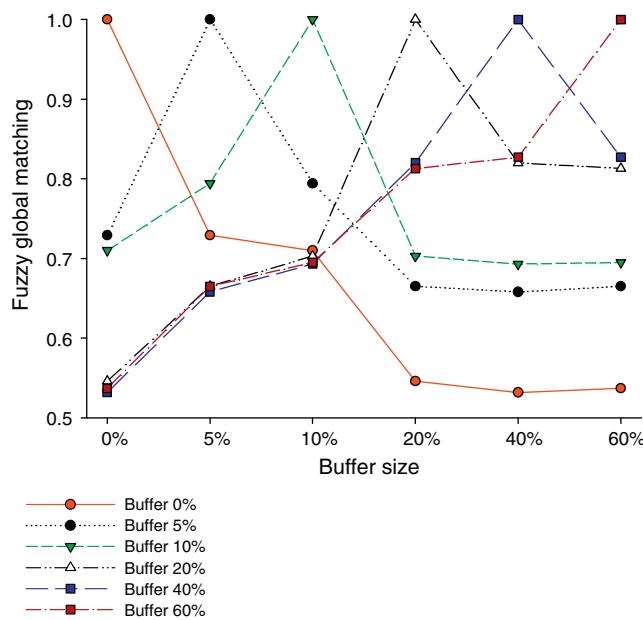
ronmental space within the niche, thus differing in their existing fundamental ecological niche.

To illustrate the effects of assumptions regarding **M** on model calibration, we plotted 14 points (number chosen for the purpose of illustration only) from within the distribution of the northeastern species, and built models based on calibration areas of different sizes. Specifically, we buffered the habitable area of the species by 0%, 5%, 10%, 20%, 40%, and 60% of additional area (Fig. 2b); note that, in the case of the 0% buffer, all pseudoabsence or background data are selected from suitable areas, so the model is in the end fit to noise. Then, we calibrated models using Maxent version 3.3.1 ([Phillips et al., 2006](#)) on default parameter settings using the same input occurrence data, but over the six different extents, and with a random testing percentage of 50%.

To illustrate the effects of the different areas on model calibration, we used global fuzzy pattern comparisons ([Power et al., 2001](#)) among all predictions using the freely available software Map Comparison Kit ([Visser and de Nijs, 2006](#)). To depict the area of presence, we thresholded model outputs based on the lowest suitability value assigned to any training presence data point ([Pearson et al., 2007](#)). Comparisons of the predictions among the six calibration extents show no two results to be the same, and a tendency towards a higher similarity as extent of analysis increases (Fig. 3).

To illustrate effects of assumptions regarding **M** on model evaluation, we used the non-thresholded logistic output from the niche model and the six background areas (0–60%) for the northeastern species described above. This raster grid presents values between 0 and 1, based on the logistic output of the Maxent algorithm ([Phillips and Dudík, 2008](#)). We calculated traditional receiver operating characteristic area under the curve (ROC AUC) statistics using ROC bootstrapping software developed for alternative ROC calculations ([Peterson et al., 2008](#)). Thus, we used the same “independent” testing points in relation to the same model output, and changed only the extent across which the model was tested.

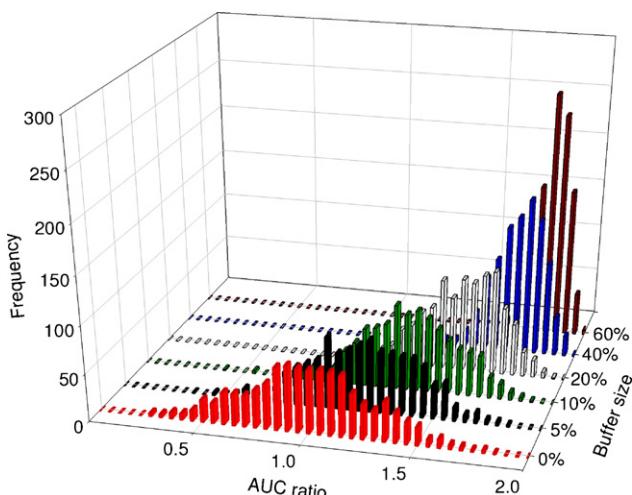
When evaluation was limited to areas close to the species’ range (buffers of 0% and 5% in Fig. 2b), the AUC scores did not indicate predictive ability significantly better than random expectations (both  $P \gg 0.05$ ). However, if the distributional area was



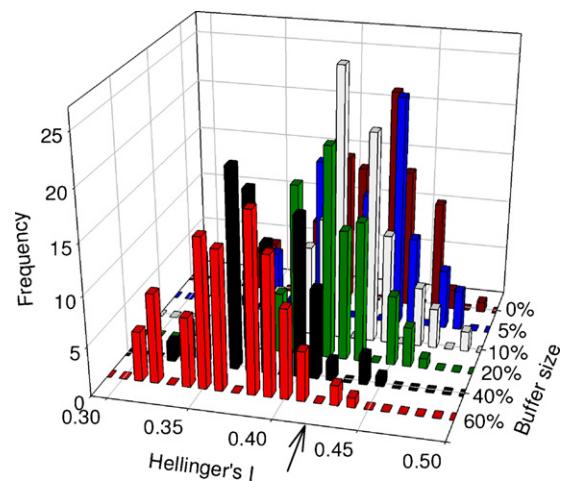
**Fig. 3.** Fuzzy global matching among the Maxent models generated for the first six extents of analysis.

buffered by 10–20%, the evaluation exercise indicated significant predictive ability (both  $P < 0.05$ ). Statistical significance was exaggerated still more as broader areas around the distributional area were included within the testing area (40% and 60% buffers, both  $P \ll 0.05$ ; Fig. 4). A researcher interested in evaluating this model might choose any of these different areas of analysis as extents of analyses, but we can see that the “answer” that she obtains can range from non-significant to highly significant. This example thus illustrates how model evaluation exercises can be subjective, if careful and well-founded assumptions regarding  $\mathbf{M}$  are not stated *a priori*.

Model comparisons are also affected by choice of  $\mathbf{M}$ , although the ideas and tools for such tests are much younger than model calibration and model evaluation. We illustrate this point using the randomization tests proposed by Warren et al. (2008). In particular, we evaluated effects of the same set of buffered extents on the results of background niche similarity tests comparing our two virtual species (see Fig. 5). Briefly, the background similarity test



**Fig. 4.** Effect of the extent of study area in model validation. AUC values tend to increase, meaning a higher model significance, as model testing area increases.



**Fig. 5.** Comparison of the similarity of the two virtual species depending on the background area used to select pseudoabsences (i.e., different buffer sizes around the distribution of the southwestern species) for the  $I$  index from Warren et al. (2008) implemented in ENMTools. Note that the buffer size axis is reversed from that of Fig. 4, but that again the likelihood of the observed  $I$  value being a consequence of random differences decreases markedly with buffer size.

of Warren et al. (2008) uses two similarity indices, Hellinger's  $I$  and Schoener's  $D$ , to quantify the level of overlap between models, and compares this observed overlap to a null distribution of values generated from models associated with random points from within the “area of interest.” More specifically, we used each of the six buffer extents (0–60%) as a distinct “area of interest” to select 14 points at random 99 times and generate niche models with Maxent for each random draw. The parameters in Maxent were kept at default settings and models were reduced from continuous probability maps to presence/absence maps using the minimum presence threshold setting (Pearson et al., 2007) to avoid complications notable at higher levels of prediction in Maxent output (Peterson et al., 2007). The observed  $I$  and  $D$  values were 0.42 and 0.32; similarity tests compared the northeastern species' occurrence points against the background of the other species at each buffer size. As summarized in Fig. 5, observed values were in the non-significant region of the null distribution ( $0.05 < P < 0.95$ ) when buffer sizes were small (<10%). Warren et al. (2008) suggested that such observations are equivocal, and that no inference about similarity can be made. At 10–20%, the  $I$  metric was non-significant, but comparisons using the  $D$  metric indicated that the two models were statistically similar ( $P > 0.95$ ). When buffers were >20%, we observed significant similarity ( $P > 0.95$ ) in both metrics.

In sum, then, our virtual species example illustrates how choice of  $\mathbf{M}$  has important implications for three aspects of ecological niche modeling – model calibration, model evaluation, and model comparison. One or more of these functionalities is used in all niche modeling applications. As a consequence, the influence of  $\mathbf{M}$  in the field is simultaneously pervasive and ignored – we now proceed to outline means by which  $\mathbf{M}$  can be estimated or otherwise approximated in real-world applications.

### 3. Approaches to estimating $\mathbf{M}$

Although effects of spatial scale (i.e., extent) on resource selection and ENM studies are well-known (Boyce, 2006; Meyer and Thuiller, 2006), selection of appropriate regions within which to develop analyses is not straightforward. Most frequently, researchers take this decision without any biologically meaningful basis (Meyer and Thuiller, 2006). For instance, in most studies, a geopolitical unit is used to delimit the area of analysis without

justification – in the best and most fortunate cases, these geopolitical units coincide with significant biotic regions (e.g., Madagascar). Too frequently, however, geopolitical units include significant biogeographic heterogeneity (e.g., areas east and west of the Isthmus of Tehuantepec in Mexico, or areas east and west of the Isthmus of Panama in Panama). We recommend strongly that niche models be calibrated, evaluated, and compared in light of explicit, *a priori* hypotheses regarding the extent of  $\mathbf{M}$  as the delimitation of the study area (Anderson and Raza, 2010). This assertion, of course, begs the question of how to estimate  $\mathbf{M}$  in real-world situations.

We have identified and explored three potential approaches to estimating  $\mathbf{M}$ , which should be carried out *prior* to initiation of analyses. These approaches are as follows:

**1. Biotic regions:** A favorite challenge of biogeographers in past decades was to delineate “biotic regions” as areas sharing sets of species that are distinct from those of other regions (Wallace, 1860; Herbertson, 1905; Shelford, 1963; Dasmann, 1972; Bailey, 1996; Lomolino et al., 2005). In general, boundaries of these regions correspond to shared sets of distributional limits of species across landscapes, which may be informative about barriers that have repeatedly constrained the distributional potential of species. In this sense, it may be reasonable to take the set of biotic regions within which a species is known to occur as a hypothesis of the areas that have been available to it over relevant time periods. This approach is quite simple, and may prove the most operational (Soberón, 2010) – others, as will be clear from discussions below, are more intricate.

**2. Niche-model-based reconstructions:** Previous studies have explored the possibility of reconstructing historical distributions of species from models based on their present-day ecological niche characteristics (Martínez-Meyer et al., 2004; Peterson et al., 2004; Martínez-Meyer and Peterson, 2006; Waltari et al., 2007; Nogués-Bravo et al., 2008; Pearman et al., 2008; Jakob et al., 2009; Nogués-Bravo, 2009), with the result that such ‘hindcasting’ offers useful predictions regarding past distributional potential. It is then feasible to use an initial round of niche model calibration to estimate the basic dimensions of a species’ distributional potential through time, in which a present-day niche model would be back-projected onto historical conditions for relevant time periods (e.g., Pleistocene Last Glacial Maximum, Last Interglacial). These potential distributional areas back through time can then be buffered by some estimate of long-term dispersal potential, and used as an estimate of  $\mathbf{M}$  in a second round of model calibration, now with a more restricted estimate of the arena of distributional possibilities. This approach runs some risk of circularity, in that the initial round of modeling is conducted without reference to a hypothesis of  $\mathbf{M}$ , and yet determines the dimensions of  $\mathbf{M}$  for the final round of model development. However, this approach is operational, and could be implemented readily.

**3. Full dynamic dispersal model:** Finally, in theory, it is feasible to join estimates of the niche with scenarios of dispersal potential to develop detailed simulations of distributions extending back into the past through periods of environmental change (Engler and Guisan, 2009; Cabral and Schurr, 2010). This approach would be more realistic than the first two proposed, as it takes into account explicitly the spatially path-dependent nature of effects of environmental change on species’ dispersal reach and consequent distributional potential, but it poses formidable computational challenges. As a first illustration of this general framework within which such a simulation would be erected, we present a first simulation, albeit preliminary and quite simplified.

## 4. The simulation

### 4.1. The geographic setting

We used a simple  $12 \times 12$  cell grid to denote the geographic domain of interest in this simulation,  $\mathbf{G}$ . For geographic reality, we used environmental conditions corresponding to the region  $103\text{--}109^\circ\text{W}$ ,  $34.5\text{--}40.5^\circ\text{N}$ . The occupied area (i.e., the actual area of distribution/area of occupancy) of a species at time  $t$  can be represented by the symbol  $\mathbf{G}_0(t)$ , a vector of ones and zeros corresponding to cells where the species is present (1) or absent (0). The coordinates of each cell are maintained in another vector with a one-to-one correspondence. We define symmetric adjacency matrices  $\mathbf{D}$  with elements  $d_{ij}$  as follows: if cell  $j$  can be accessed from cell  $i$  in one time step, then  $d_{ij}=1$ ; if it cannot be reached from  $i$ , then  $d_{ij}=0$ .  $\mathbf{D}$  therefore represents the connectivity of different cells within the geographic domain. We note that future implementations of these ideas might include more complex dispersal kernels, for example, a probabilistic view of adjacency, as well as a time-dependent version of  $\mathbf{D}$ , which would take into account changing opportunities for dispersal among cells.

Connectivity of areas, and its representation by adjacency matrices as just described, is strongly dependent on the natural history of the species and the geographic structure of the world at time  $t$  (e.g., positions of barriers like mountain ranges, rivers, deserts, and oceans). This point will be explored in greater depth later, but we will take the adjacency matrices as a given for the moment. Assume first that the entire world is favorable to the species (i.e., the region  $\mathbf{G} \subseteq \mathbf{A}$ ). Let  $\mathbf{G}_0(0)$  be the occupied distributional area of the species of interest at some initial time  $t=0$ . At  $t=1$ , the following multiplication produces a vector  $\mathbf{J}(1)=\mathbf{D}\mathbf{G}_0(0)$ . The elements of  $\mathbf{J}(1)$  contain the number of ways in which every cell of the world can be reached, in one step, from  $\mathbf{G}_0(0)$ . The occupied distributional area after one step,  $\mathbf{G}_0(1)$ , then, is simply a vector of zeros and ones corresponding to those cells where  $\mathbf{J}(1)>0$ . Repeated application of the adjacency matrix produces a vector  $\mathbf{J}(t)=(\prod_{\tau=0}^{t-1}\mathbf{D})\mathbf{G}_0(0)$  representing the number of ways in which every cell can be reached in  $t$  steps from the initial distribution  $\mathbf{G}_0(0)$ , and the species’ occupied distributional area at time  $t$  is simply the vector  $\mathbf{G}_0(t)$  of ones where  $\mathbf{J}(t)>0$ , and zeroes otherwise.

If the connectivity matrix  $\mathbf{D}$  is constant (as assumed herein), the above equation reduces to  $\mathbf{J}(t)=\mathbf{D}^t\mathbf{J}(0)$ ; assuming that  $\mathbf{D}$  has some standard properties (i.e., irreducibility, non-periodicity), the elements of the first eigenvector of  $\mathbf{D}$  are proportional to the “accessibility” of the elements. We stress that the elements of  $\mathbf{J}(t)$  are integers, in principle  $>1$ , so it is necessary to convert  $\mathbf{J}(t)$  to a binary vector by recoding all elements with value  $>0$  as 1. This step yields a final vector  $\mathbf{G}_0(t)$ , which summarizes the corresponding occupied distributional area after  $t$  time steps, which is, in effect, the total area reached by the species during the entire simulation process.

### 4.2. The fundamental ecological niche

In reality, of course, a cell in  $\mathbf{G}$  is not occupied simply by reaching it. Rather, the correct suite of environmental and biotic conditions ( $\mathbf{A}$  and  $\mathbf{B}$  respectively) must also be fulfilled. We will simply ignore the biotic factors in  $\mathbf{B}$ , given their very complex nature and lack of available data. Following Soberón and Nakamura (2009), and for the sake of illustration, we assume that some convex envelope in a space of  $v$  environmental variables that constitute the environment can represent the fundamental niche of a species. For example, the fundamental niche might be represented as an ellipsoid of the form  $\mathbf{N}_F=(x-\mu)\Sigma(x-\mu)^T-1$ , wherein  $\Sigma$  is a symmetric positive definite matrix that defines the semi axes of the ellipsoid, and  $\mu$  represents its centroid.

This ellipsoid in our example represents the fundamental niche; its intersection with  $\mathbf{E}(t)$ , the available environmental space at time  $t$ , which can be expressed in symbols as  $\mathbf{N}_F \cap \mathbf{E}(t)$ , gives the regions of  $\mathbf{G}$  that are suitable for settling at time  $t$  because the abiotic environment is favorable. This intersection, which was first identified by [Jackson and Overpeck \(2000\)](#) as of extreme importance in niche theory, we term the “existing fundamental niche” ([Peterson et al., 2012](#)).

A simple and natural assumption is that individuals of the species will be able to settle in a cell  $i$  that has been reached, as long as the  $v$ -dimensional environment in cell  $i$  [denoted by  $\bar{\mathbf{e}}_i(t)$ ] is contained within the envelope that represents the fundamental niche of the species. In simple words, in our simulation, a new cell is added to the occupied distributional area in a time step if (1) it can be reached from an occupied cell in time  $t - 1$ , and (2) its environment is favorable. To formalize this idea, we use a diagonal matrix  $\mathbf{S}(t)$ , with 1s or 0s along the diagonal, depending on whether the environment is suitable or not, respectively, at time  $t$ . A cell  $i$  will have a 1 if  $\bar{\mathbf{e}}_i(t) \in \mathbf{N}_F \cap \mathbf{E}(t)$ , or in other words, if  $\bar{\mathbf{e}}_i(t)$  is within the existing fundamental niche at time  $t$ . Notice that this step assumes strict conservatism of the fundamental niche, but a dynamic environment.

Multiplication of  $\mathbf{S}(t)$  by  $\mathbf{D}$  defines, for one time step, which cells can be both reached and colonized. If we substitute the new vector after each multiplication, we get

$$\mathbf{G}_0(t+1) \leftarrow \mathbf{J}(t+1) = \mathbf{S}(t)\mathbf{D}\mathbf{G}_0(t) \quad (1)$$

where the arrow indicates that positive values in  $\mathbf{J}(t+1)$  are transformed to 1.

Region  $\mathbf{M}$ , by definition above, is that part of  $\mathbf{G}$  that has been accessible to a species in a given, relevant time interval. It is the region that the species has “sampled” over the period from 0 to  $t$  (expressed hereafter as  $[0,t]$ ). Several ways of defining  $\mathbf{M}$  are thus possible: a very natural one is simply to consider the union of all the vectors of cells that have been accessed, with or without successful establishment. This set is

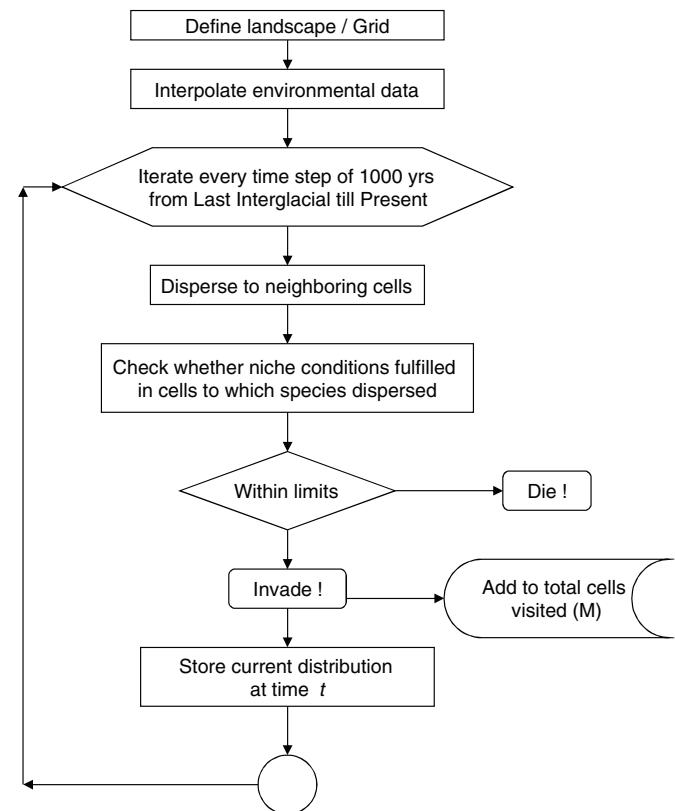
$$\mathbf{M}(t) = \bigcup_{\tau=0}^t \mathbf{G}_0(\tau) \quad (2)$$

where  $\cup$  indicates the union of the succeeding terms over  $[0,t]$ . This definition is that which we propose for  $\mathbf{M}$  for the time interval  $[0,t]$ .

#### 4.3. Calculating $\mathbf{M}$

To illustrate how the preceding scheme would work, within the simulation, we set the species as initially present in 6 contiguous pixels near the center of the region ([Fig. 6](#)). We defined the relevant time period as ranging from the Last Interglacial (LIG; 135,000 BP), through the globally cold period of the Last Glacial Maximum (LGM; 20,000 BP), and up to the present. Climate data for LIG, LGM, and present are available in the form of “bioclimatic” variables developed from monthly temperature and precipitation data ([Hijmans et al., 2005](#); [Waltari et al., 2007](#)); environments across our simple  $\mathbf{G}$  were interpolated between these points in time at 1000-year intervals using R scripts available upon request from the authors. Hence, we used three climatic data matrices:  $\mathbf{S}(t = -135,000)$ ,  $\mathbf{S}(t = -20,000)$  and  $\mathbf{S}(t = 0)$ . These matrices were then interpolated to create 132 intermediate matrices by using the following weighting formula:

$$\begin{aligned} \mathbf{S}(t) &= \mathbf{S}(-135,000)[e^{-6e^{-0.00005(t+20,000)}} - e^{-6}] \\ &\quad + \mathbf{S}(-20,000)[1 - e^{-6e^{-0.00005(t+20,000)}}] \quad t < -20,000 \\ \mathbf{S}(t) &= \mathbf{S}(-20000)e^{-0.001e^{0.0005(t+20000)}} \\ &\quad + \mathbf{S}(0)[1 - e^{-0.001e^{0.0005(t+20000)}}] \quad t \geq -20,000 \end{aligned}$$



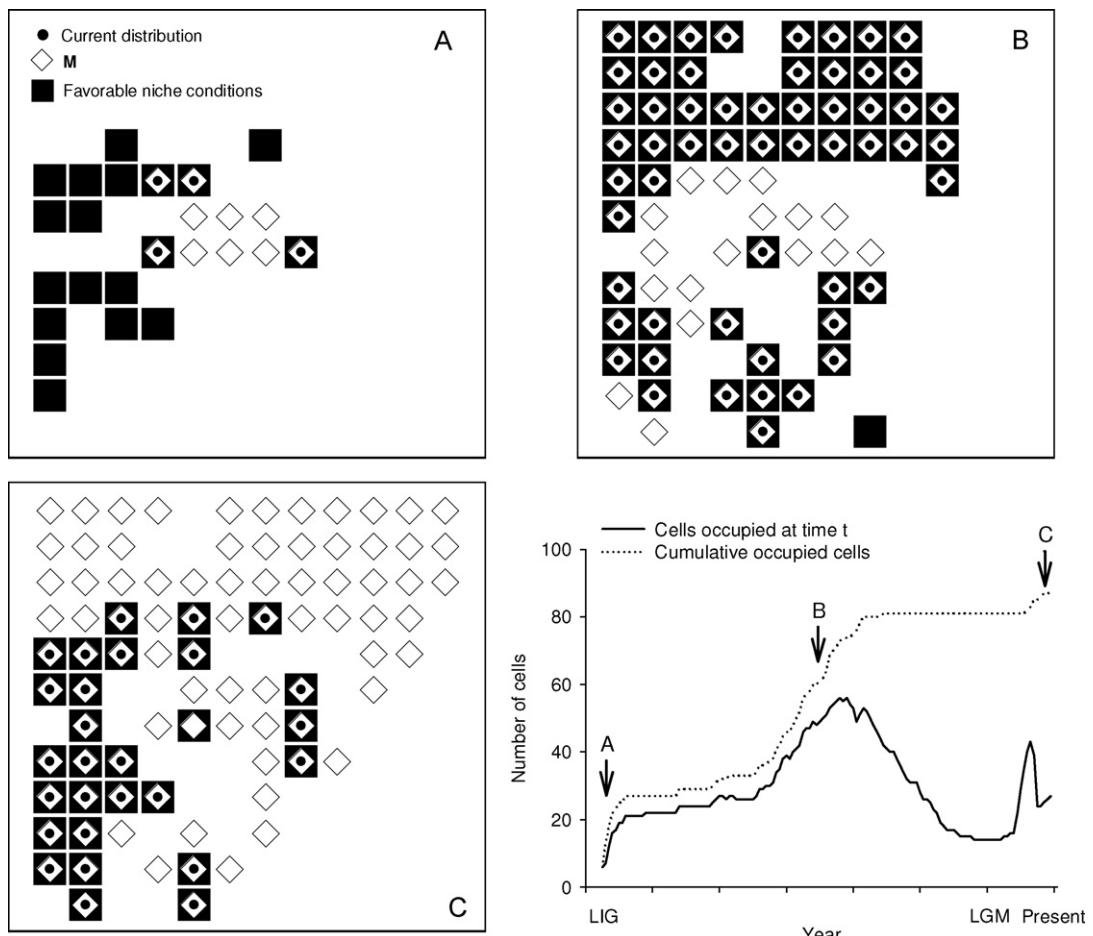
**Fig. 6.** Flow chart of simulation used to estimate  $\mathbf{M}$ .

We defined the fundamental niche of our hypothetical species  $\mathbf{N}_F$  as an ellipsoid with centroid  $\mu = (10, 500)$  and axes matrix  $\mathbf{S} = \begin{bmatrix} 0.001 & 0.00031 \\ 0.00031 & 0.00015 \end{bmatrix}$ , which yields the following expression for the fundamental niche in our simulation:

$$\begin{aligned} 39.8113 - 0.333671x + 0.00100925x^2 - 0.152572y \\ + 0.000626971xy + 0.000146302y^2 \end{aligned}$$

From the initial conditions (i.e., presences in the 6 central cells), we simulated dispersal and colonization by the species according to the scheme of Eq. (1). First, the species had to be able to actually reach the cell physically via dispersal: we assumed that the species would disperse to all first-order neighboring cells in a given 1000-year time period, a process summarized in the corresponding adjacency matrix  $\mathbf{D}$ , which we held constant through time in this example. Second, the matrices  $\mathbf{S}(t)$  (as described and defined above) were used to check whether a given cell is suitable or unsuitable at time  $t$  (i.e., whether its environments are within  $\mathbf{N}_F$ ). Cells that are both accessible and suitable are assumed in this simple simulation to be settled successfully.

At each time step, each cell on the landscape is checked as to whether it presents favorable conditions (i.e., whether its environment is within the fundamental niche), which is summarized as a “niche matrix.” The distribution of the species in the next time step is then calculated as the product of the adjacency matrix,  $\mathbf{D}$ , the distribution vector,  $\mathbf{G}_0$ , and the niche matrix,  $\mathbf{N}_F$ , and then iterated through the entire period of the simulation. The union of all cells visited (whether or not successfully colonized) through the entire course of the simulation constitutes the total area explored by the species over time, producing in effect a direct estimate of  $\mathbf{M}$ . The temporal sequence of the evolution of the size of the distributional area [denoted  $|\mathbf{G}_0(t)|$ ], and of our estimate of  $\mathbf{M}$ , are presented in [Fig. 7](#).



**Fig. 7.** Simulation results, showing the number of cells occupied by the species at a particular time and the cumulative tracking of cells that have been occupied by it at any point previously in the simulation. Shown are the continuous traces of cells occupied at any point in time and cumulatively, as well as three time “snapshots” showing actual distributions.

#### 4.4. Simulation results

The number of occupied cells at any particular time (from 135,000 BP to present) during the simulation varied depending on their suitability and reachability in relation to the favorable conditions for, and the dispersal of, the species. The area of distribution,  $G_0$ , expanded initially until the LGM, when the extent of suitable environments was reduced. The total number of cells explored increased until the LGM as well, when it stabilized until  $\sim 5000$  BP, after which it increased again, but only slightly (Fig. 7). The sum of all cells explored is, in turn, a hypothesis of the complete set of sites that have been available (i.e., within the dispersal distance of an occupied cell at least once) to the species during the entire simulation, and may be used as the best estimate for  $M$ , where both historical (e.g., environmental shifts) and ecological processes (e.g., dispersal and establishment) have been taken into consideration.

### 5. Discussion

#### 5.1. General scenarios

The arguments and examples presented in the first section of this paper should – we believe – suffice to convince the reader that  $M$  is an important consideration in studies of distributional ecology. Basically, it is the realm within which the species has sampled the landscape in question, and so it is the appropriate arena for training, validating, and comparing ecological niche models. Because our simulation was little more than a caricature of the processes

involved in producing a complex and more realistic  $M$ , we do not expect that such simulation approaches will soon see extensive use – as discussed below, the complexity of such approaches is daunting, at least for the time being. As a consequence, the niche modeler who is convinced that  $M$  is important will (obviously) ask how can one estimate it in a very practical and real-world sense.

Hence, we explore possible approaches to estimating  $M$  that may or may not be appropriate in a particular situation. Specifically, the time span involved in a particular question becomes fundamental – some species may only have been “exploring” a landscape for a few years (e.g., invasive species), whereas other species may have been present on landscapes for millions of years and have responded distributionally to diverse climatic changes and geological or environmental shifts. As a result, we can consider three sets of strategies as bases for hypothesizing the spatial footprint of  $M$ , as a function of the time that the species has been present on the landscape:

- *History nil:* In the case of a species that is just arrived on a landscape, the area that is sampled by it is more or less equivalent to its generation-to-generation maximum dispersal distance. That is, one can consider the present-day distribution of the species, buffered by an estimate of that maximum dispersal distance, as an estimate of  $M$ . This simple process will identify areas likely sampled by the species in determining its distributional potential on the landscape of interest. Examples of this sort of situation and corresponding approach are relatively few, as we generally focus on species that have a longer history of presence on

a landscape – nonetheless, see [Anderson et al. \(2006\)](#) for a partial example.

- *History included but no environmental change:* When a species has had a longer-term presence on a landscape, a single generation's maximum dispersal distance will underestimate its maximum dispersal "reach," in the sense that the species may have had the possibility to explore out several-fold times more than the basic, individual dispersal distance, and so  $\mathbf{M}$  will be underestimated. Under such situations, dispersal simulations can be iterated by a number of generations that corresponds approximately to its time of presence on that landscape (e.g., [Smolik et al., 2010](#)). It should be noted, however, that such simple, iterative consideration of dispersal assumes stable environmental conditions, and therefore must still be over time spans that are reasonably short.
- *History included, environmental change:* The most challenging – and unfortunately perhaps most generally relevant – of these situations are those in which the species has had a long period of presence on the landscape of interest, as would be the case of most or all resident/native species. "Long" is defined as sufficient time such that environments represented in  $\mathbf{G}$  have not been constant over the time period of interest. This case presents both advantages and disadvantages to the niche modeling exercise. On the positive side, the species is much more likely to have achieved distributional equilibrium: that is, it has probably explored via dispersal out in all directions, and has more likely filled in all of the areas that are suitable (i.e., areas within  $\mathbf{A} \cap \mathbf{B}$ ) and that are also within  $\mathbf{M}$ , and thus that are within  $\mathbf{G}_0$  – in such cases, apparent commission error is much reduced and areas from which the species is absent are more likely to be lacking the species *for reasons of environmental suitability*, rather than because of dispersal limitation. On the negative side, however, such species will have experienced a much greater variety of environmental shifts, and may have had much more diverse opportunities to colonize suitable areas that might otherwise appear to be inaccessible – consider, for example, the "sky islands" situation in the southwestern United States, and the many isolated conspecific populations of montane organisms that likely were founded in the region under quite-different climatic conditions ([Patterson, 1982](#)).

Managing situations such as the latter is correspondingly much more complicated than the previous two situations. In essence, we have the challenge of taking into account historical environmental changes that may have broadened the overall dispersal reach of the species. Above we outlined three more or less practical approaches to this situation – while none of the three is completely satisfactory, and all have limitations, they represent explicit assumptions regarding  $\mathbf{M}$  and its extent, and may thus take the niche modeling process one or more steps closer to being fully reasonable and realistic by linking it explicitly to ecological and biogeographic theory.

## 5.2. The eventual challenge regarding $\mathbf{M}$

The process of estimating ecological niches is complex, and this paper aims to clarify one aspect of this process that has heretofore not been explicit – that of how to delineate the area of analysis appropriately. To date, this area has been generally chosen as a convenient rectangle enclosing the area of interest ([Peterson, 2001](#)), the outline of the country in which the species of interest is distributed ([Peterson et al., 2002](#)), or some other area that is convenient. No good rationale has been used as a basis for this decision, and in most cases the study area is delimited without any statement of reasoning as to why ([Anderson and Raza, 2010](#)). In this paper, we argue that  $\mathbf{M}$  from the BAM framework ([Soberón and Peterson, 2005](#)) provides an appropriate conceptual framework for this decision (see [Anderson and Raza, 2010](#) for similar conclusion).

However, estimating  $\mathbf{M}$  is only rarely easy and straightforward (e.g., when history is a negligible component of the situation). More frequently, multiple options will be available, and perhaps several such scenarios should be explored, to make the niche estimation exercise as realistic as possible. Explicit statement of the rationale for these hypothesized scenarios in the Methods sections of papers presenting these analyses should constitute a critical step in the niche modeling process.

$\mathbf{M}$  depends simultaneously on the natural history and dispersal characteristics of the species in question, the geography of the landscape of interest, the spatial configuration of suitable habitats, and the time span relevant to the species' presence on the landscape and any environmental changes that occurred during that time period. Each of these points must be weighed carefully in outlining  $\mathbf{M}$ , or the estimate may be incomplete, biased, or misleading. The following list outlines key considerations:

1. *Estimate dispersal characteristics:* The importance of dispersal-driven processes in range dynamics has only been recently incorporated into ENM studies ([Cabral and Schurr, 2010](#); [Smolik et al., 2010](#)) in an effort to link ecological theory with the modeling of species' niches and distributions, in spite of the great potential importance of these considerations ([Engler and Guisan, 2009](#)). A first challenge in incorporating these processes is that of estimating the dispersal characteristics of species over the time period that makes sense for the problem, or, in other words, estimating the parameters of the matrix  $\mathbf{D}$  or any equivalent dispersal kernels. On ecological time scales, dispersal has been estimated by measuring seed shadows ([Carey et al., 1995](#)), pollen dispersal ([Okubo and Levin, 2001](#)), and results of mark-recapture studies ([Dobzhansky and Wright, 1943](#); [Dobzhansky et al., 1979](#)). In this latter case, the separation of natal and breeding sites in space estimates the dispersal characteristics of the individual in question. The problem with mark-recapture approaches is that individuals that are not recaptured may have died, but also may have dispersed much-longer distances, which ends up determining much of the dynamics of the system ([Trakhtenbrot et al., 2005](#)). As a result, such individual-tracking studies will frequently fail to characterize the dispersal characteristics of the species fully. Other options include use of population-genetic profiles, which, though indirect, have the potential to integrate over the full dispersal profile of the species ([Avise, 2000](#)).
2. *Estimate the niche:* As discussed above, several approaches to understanding  $\mathbf{M}$  require initial estimates of the *fundamental* niche of the species. Indeed, our caricature of the dynamic approach to estimation of  $\mathbf{M}$  required us to postulate a fundamental niche as part of the design of the simulation – in real-life examples, this niche is unknown, and so would have to be estimated either from first-principles approaches ([Kearney and Porter, 2004](#)), experimental approaches ([Etterson and Shaw, 2001](#)), or a first round of correlational niche modeling. Although this latter approach runs the risk of circular reasoning, in which biases and complications introduced in initial iterations propagate through the process and cause downstream problems, some initial estimate of the niche will be needed, and correlational approaches frequently offer the only or the most practical recourse ([Peterson et al., 2012](#)).
3. *Estimate relevant time span:* This point is quite critical: witness the implications of different time spans discussed above. When a species is newly arrived, or relatively newly arrived, on a landscape, many aspects of its associated  $\mathbf{M}$  can be estimated from dispersal considerations only. When, however, the species has a longer history on a landscape, the process becomes considerably more complex, and environmental changes must be incorporated into the estimation process. Phylogeographic studies may be particularly relevant here: population history can be

estimated using approaches like the coalescent (Edwards and Beerli, 2000; Knowles, 2004), and these estimates can then at least provide an order-of-magnitude view of how far back into time environmental history must be considered (Carstens and Richards, 2007). However, we note that time-resolution information regarding paleoclimates is scarce (particularly prior to the Last Glacial Maximum), which may present an obstacle to full implementation of these approaches.

**4. Identify relevant environmental changes:** Once the time span of the species' distributional history on a landscape has been determined, the environmental changes relevant to its distribution can be summarized. Frequently, these changes will be those related to the dramatic climatic fluctuations of the Pleistocene (Davis and Shaw, 2001), with cold glacial maxima contrasting with warm interglacials. Recent niche-model-based explorations have taken advantage of global estimates of Last Glacial Maximum and Last Interglacial climate conditions to explore Pleistocene distributional patterns. These estimates, derived from general circulation models tuned to relevant landscape characteristics and then-current atmospheric composition and conditions, are not without error, but succeed in reconstructing many relevant features of paleoclimates (Otto-Bliesner et al., 2006), and have been quite successful in applications to distributional ecology (Nogués-Bravo, 2009). In the example presented above, we used a simple regularly distributed interpolation, but more realistic interpolations can be developed via reference to detailed global temperature profiles on more realistic time scales (Dansgaard et al., 1993).

It is important to note that, although we favor a spatially explicit and dynamic mechanistic approach as the best way to estimate **M**, taking into account dispersal and habitat-driven processes, other important factors that depend on the time frame considered can have a crucial role in determining **M**. A first consideration is of the effects of **B**, the set of areas that are suitable in terms of biotic considerations – although neglected in our simulation presented here, **B** and **M** may interact in important ways to affect estimates of ecological niches (Anderson and Raza, 2010). Finally, evolutionary adaptive processes can shift the fundamental niche of a species (Holt and Gaines, 1992). Nonetheless, in our example above (and likely also in other studies), the time frame (135,000 BP) used may not be large enough for evolutionary processes to take place, at least for vertebrates and other long-lived species (Rodríguez-Trelles and Rodríguez, 1998; Peterson, 2011).

Computational challenges enter rather dramatically into the methodologies that we would ideally recommend. Realistic simulation of spatially explicit population processes is always numerically challenging (Bolker and Pacala, 1997). The example presented above was a 12 × 12 matrix of cells in a caricature of a real-world landscape. Adjacency matrices grow in size as the square of the number of cells. A space-of-configurations representation grows exponentially with the number of cells in the grid. Since real landscapes will frequently be represented by arrays of 10<sup>5</sup>–10<sup>6</sup> pixels, these simulations will be computationally expensive, likely out of reach of the computing capacities currently available. A final step that will augment computational demands considerably is that of making these simulations probabilistic and not deterministic, which will require numerous replicate simulations to be able to characterize both the behavior of a system and the variation in that behavior.

The challenge outlined in this paper boils down to that of understanding the distributional potential of species over time across a changing environmental landscape, plus the species' dispersal potential over that time, in a more mechanistic manner than the traditional phenomenological approach (Guisan and Thuiller, 2005). This exercise in and of itself would be of considerable interest

to biogeographers, who focus on understanding the distributional potential of lineages through time (Liow and Stenseth, 2007) – as such, it presents considerable challenges, and we in no sense purport to have solved the challenge in this paper. What we *have* done in this paper, nonetheless, is to raise the issue, for the first time explicitly and in detail in the niche modeling literature (Elith et al., 2010), such that researchers can incorporate these ideas as part of the process of their investigations. As we have demonstrated, this explicit presentation of assumptions has critical implications for the robustness and validity of models that have been previously developed and those yet to come.

## Acknowledgements

Microsoft Research provided funding for much of this effort. ALN (fellowship #189216) and FV were supported by graduate fellowships from Consejo Nacional de Ciencia y Tecnología, Mexico, and FV by the Posgrado en Ciencias Biológicas, UNAM. AJV was supported by a postdoctoral fellowship from the Ministerio de Educación y Ciencia, Spain (Ref.: EX-2007-0381) and the Juan de la Cierva Program.

## References

- Anderson, R.P., Peterson, A.T., Egbert, S.L., 2006. Vegetation-index models predict areas vulnerable to purple loosestrife (*Lythrum salicaria*) in Kansas. *Southwestern Naturalist* 51, 471–480.
- Anderson, R.P., Raza, A., 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography* 37, 1378–1393.
- Avise, J.C., 2000. *Phylogeography: The History and Formation of Species*, 3rd ed. Harvard University Press, Cambridge, Mass.
- Bailey, R.G., 1996. *Ecosystem Geography*. Springer, New York.
- Beale, C.M., Lennon, J.I., Gimona, A., 2008. Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences of the United States of America* 105, 14908–14912.
- Bolker, B., Pacala, S.W., 1997. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology* 52, 179–197.
- Boyce, M.S., 2006. Scale for resource selection functions. *Diversity and Distributions* 12, 269–276.
- Brown, J.H., Stevens, G.C., Kaufman, D.M., 1996. The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics* 27, 597–623.
- Byrne, M., Yeates, D.K., Joseph, L., Kearney, M., Bowler, J., Williams, M.A.M., Cooper, S., Donnellan, S.C., Keogh, I.S., Ley, R., Melville, J., Murphy, D.J., Porch, N., Wyrnwall, K.H., 2008. Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology* 17, 4398–4417.
- Cabral, J.S., Schurr, F.M., 2010. Estimating demographic models for the range dynamics of plant species. *Global Ecology and Biogeography* 19, 85–97.
- Carey, P.D., Watkinson, A.R., Gerard, F.O., 1995. The determinants of the distribution and abundance of the winter annual grass *Vulpia ciliata* ssp. *ambigua*. *Journal of Ecology* 83, 177–187.
- Carstens, B.C., Richards, C.L., 2007. Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution* 61, 1439–1454.
- Dansgaard, W., Johnsen, S.J., Clausen, H.B., Dahl-Jensen, D., Gundestrup, N.S., Hammer, C.U., Hvidberg, C.S., Steffensen, J.P., Sveinbjörnsdóttir, A.E., Jouzel, J., Bond, G.C., 1993. Evidence for general instability of past climate from a 250 kyr ice-core record. *Nature* 264, 218–220.
- Dasmann, R.F., 1972. Towards a system of classifying natural regions of the world and their representation by national parks and reserves. *Biological Conservation* 4, 247–255.
- Davis, M.B., Shaw, R.G., 2001. Range shifts and adaptive responses to quaternary climate change. *Science* 292, 673–679.
- Dobzhansky, T., Powell, J.R., Taylor, C.E., Andregg, M., 1979. Ecological variables affecting the dispersal behavior of *Drosophila pseudoobscura* and its relatives. *American Naturalist* 132, 325–334.
- Dobzhansky, T., Wright, S., 1943. Genetics of natural populations. X. Dispersion rates in *Drosophila pseudoobscura*. *Genetics* 28, 304–340.
- Edwards, S.V., Beerli, P., 2000. Gene divergence, population divergence, and the variance in the coalescence time in phylogeographic studies. *Evolution* 54, 1839–1854.
- Elith, J., Kearney, M., Phillips, S., 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution* 1, 330–342.
- Engler, R., Guisan, A., 2009. MigClim: predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions* 15, 590–601.
- Ettersson, J.R., Shaw, R.G., 2001. Constraint to adaptive evolution in response to global warming. *Science* 294, 151–153.

- Gaston, K.J., 2003. The Structure and Dynamics of Geographic Ranges. Oxford University Press, Oxford.
- Gotelli, N.J., Anderson, M.J., Arita, H.T., Chao, A., Colwell, R.K., Connolly, S.R., Currie, D.J., Dunn, R.R., Graves, G.R., Green, J.L., Arvid, J., Grytnes, Y.-H.J., Jetz, W., Lyons, S.K., McCain, C.M., Magurran, A.E., Rahbek, C., Rangel, T.F.L.V.B., Soberón, J., Webb, C.O., Willig, M.R., 2009. Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters* 12, 873–886.
- Graham, C.H., Ron, S.R., Santos, J.C., Schneider, C.J., Moritz, C., 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* 58, 1781–1793.
- Grinnell, J., 1917. Field tests of theories concerning distributional control. *American Naturalist* 51, 115–128.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8, 993–1009.
- Herbertson, A.J., 1905. The major natural regions, an essay in systematic geography. *Geographical Journal* 25, 300–312.
- Hey, T., Tansley, S., Tolle, K. (Eds.), 2009. The Fourth Paradigm: Data-intensive Scientific Discovery. Microsoft Research, Redmond, Washington.
- 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. *International Journal of Climatology* 25, 1965–1978.
- Holt, R.D., 2003. On the evolutionary ecology of species' ranges. *Evolutionary Ecology Research* 5, 159–178.
- Holt, R.D., Gaines, M.S., 1992. Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. *Evolutionary Ecology* 6, 433–447.
- Hutchinson, G.E., 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22, 415–427.
- Jackson, S.T., Overpeck, J.T., 2000. Responses of plant populations and communities to environmental changes of the Late Quaternary. *Paleobiology* 26 (supplement), 194–220.
- Jakob, S.S., Martínez-Meyer, E., Blattner, F.R., 2009. Phylogeographic analyses and paleodistribution modeling indicate Pleistocene *in situ* survival of *Hordeum* species (Poaceae) in southern Patagonia without genetic or spatial restriction. *Molecular Biology and Evolution* 26, 907–923.
- Jiménez-Valverde, A., Barve, N., Lira-Noriega, A., Maher, S.P., Nakazawa, Y., Papeş, M., Soberón, J., Sukumaran, J., Peterson, A.T., 2010. Dominant climate influences on North American bird distributions. *Global Ecology and Biogeography*.
- Jiménez-Valverde, A., Lobo, J.M., Hortal, J., 2008. Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions* 14, 885–890.
- Kearney, M., Porter, W.P., 2004. Mapping the fundamental niche: physiology, climate, and the distribution of a nocturnal lizard. *Ecology* 85, 3119–3131.
- Knowles, L.L., 2004. The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology* 17, 1–10.
- Kot, M., Lewis, M.A., Driessche, P.v.d., 1996. Dispersal data and the spread of invading organisms. *Ecology* 77, 2027–2042.
- Liow, L.H., Stenseth, N.C., 2007. The rise and fall of species: implications for macroevolutionary and macroecological studies. *Proceedings of the Royal Society B* 274, 2745–2752.
- Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17, 145–151.
- Lomolino, M.V., Riddle, B.R., Brown, J.H., 2005. Biogeography, 3rd ed. Sinauer Associates, Sunderland.
- MacArthur, R., 1972. Geographical Ecology. Princeton University Press, Princeton, NJ.
- Martínez-Meyer, E., Peterson, A.T., 2006. Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-recent transition. *Journal of Biogeography* 33, 1779–1789.
- 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 Ecology and Biogeography* 13, 305–314.
- Meyer, C.B., Thuiller, W., 2006. Accuracy of resource selection functions across spatial scales. *Diversity and Distributions* 12, 288–297.
- Nogués-Bravo, D., 2009. Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography* 18, 521–531.
- Nogués-Bravo, D., Rodríguez, J., Hortal, J., Batra, P., Araújo, M.B., 2008. Climate change, humans, and the extinction of the Woolly Mammoth. *PLoS Biology* 6, e79.
- Okubo, A., Levin, S.A., 2001. Diffusion and Ecological Problems: Modern Perspectives, 2nd ed. Springer-Verlag, New York.
- Otto-Biesner, B.L., Marshall, S.J., Overpeck, J.T., Miller, G.H., Hu, A., 2006. CAPE Last Interglacial Project members Simulating Arctic climate warmth and icefield retreat in the Last Interglaciation. *Science* 311, 1751–1753.
- Patterson, B.D., 1982. Pleistocene vicariance, montane islands, and the evolutionary divergence of some chipmunks (Genus *Eutamias*). *Journal of Mammalogy* 63, 387–398.
- Pearman, P.B., Guisan, A., Broennimann, O., Randin, C.F., 2008. Niche dynamics in space and time. *Trends in Ecology and Evolution* 23, 149–158.
- Pearson, R.G., Raxworthy, C., Nakamura, M., Peterson, A.T., 2007. Predicting species' distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34, 102–117.
- Peterson, A., 2001. Predicting species' geographic distributions based on ecological niche modeling. *Condor* 103, 599–605.
- Peterson, A.T., 2005. Kansas gap analysis: the importance of validating distributional models before using them. *Southwestern Naturalist* 50, 230–236.
- Peterson, A.T., 2011. Ecological niche conservatism: a time-structured review of evidence. *Journal of Biogeography*.
- Peterson, A.T., Ball, L.G., Cohoon, K.C., 2002. Predicting distributions of Mexican birds using ecological niche modelling methods. *Ibis* 144, e27–e32.
- Peterson, A.T., Martínez-Meyer, E., González-Salazar, C., 2004. Reconstructing the Pleistocene geography of the *Aphelocoma* jays (Corvidae). *Diversity and Distributions* 10, 237–246.
- 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., Papeş, M., Soberón, J., 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modelling. *Ecological Modelling* 213, 63–72.
- Peterson, A.T., Soberón, J., Anderson, R.P., Pearson, R.G., Martínez-Meyer, E., Nakamura, M., Araújo, M.B., 2012. Ecological Niches and Geographic Distributions: A Modeling Perspective. Princeton University Press, Princeton.
- Peterson, A.T., Soberón, J., Sánchez-Cordero, V., 1999. Conservatism of ecological niches in evolutionary time. *Science* 285, 1265–1267.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190, 231–259.
- Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 21, 161–175.
- Power, C., Simms, A., White, R., 2001. Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science* 15, 77–100.
- Pulliam, H.R., 2000. On the relationship between niche and distribution. *Ecology Letters* 3, 349–361.
- Rangel, T.F., Diniz-Filho, J.A., Colwell, R., 2007. Species-richness and evolutionary niche dynamics: a spatial pattern-oriented simulation experiment. *American Naturalist* 170, 602–616.
- Rodríguez-Trelles, F., Rodríguez, M.A., 1998. Rapid micro-evolution and loss of chromosomal diversity in *Drosophila* in response to climate warming. *Evolutionary Ecology* 12, 829–838.
- Roy, K., Goldberg, E.E., 2007. Origination, extinction, and dispersal: integrative models for understanding present-day diversity gradients. *American Naturalist* 170, S71–S85.
- Shelford, V.E., 1963. The Ecology of North America. University of Illinois Press, Urbana.
- Smolik, M.G., Dullinger, S., Essl, F., Kleinbauer, I., Leitner, M., Petersen, J., Stadler, L.M., Vogl, G., 2010. Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *Journal of Biogeography* 37, 411–422.
- Soberón, J., 2010. Niche and area of distribution modeling: a population ecology perspective. *Ecography* 33, 159–167.
- Soberón, J., Nakamura, M., 2009. Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America* 106, 19644–19650.
- Soberón, J., Peterson, A.T., 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2, 1–10.
- Stigall, A.L., 2008. Tracking species in space and time: assessing the relationships between paleobiogeography, paleoecology and macroevolution. In: Kelley, P.H., Bambach, R.K. (Eds.), From Evolution to Geobiology: Research Questions Driving Paleontology at the Start of a New Century, Paleontological Society Short Course. The Paleontological Society Papers, pp. 227–242.
- Stockwell, D.R.B., Peters, D.P., 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* 13, 143–158.
- Trakhtenbrot, A., Nathan, R., Perry, G., Richardson, D.M., 2005. The importance of long-distance dispersal in biodiversity conservation. *Diversity and Distributions* 11, 173–181.
- Udvardy, M.D.F., 1969. Dynamic Zoogeography. Van Nostrand Reinhold Company, New York.
- Vandermeer, J., 1972. Niche theory. *Annual Review of Ecology and Systematics* 3, 107–132.
- 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. *American Naturalist* 174, 282–291.
- Visser, H., de Nijs, T., 2006. The map comparison kit. *Environmental Modeling and Software* 21, 346–358.
- Wallace, A.R., 1860. On the zoological geography of the Malay Archipelago. *Proceedings of the Linnean Society of London* 4, 172–184.
- Waltari, E., Perkins, S., Hijmans, R., Peterson, A.T., Nyári, Á., Guralnick, R., 2007. Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS ONE* 2, e563.
- Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62, 2868–2883.
- Wiens, J.J., 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution* 58, 193–197.
- Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms and aberrations in global climate 65 Ma to present. *Science* 292, 686–693.