Basic Questions in Biogeography and the (Lack of) Simplicity of Species Distributions: Putting Species Distribution Models in the Right Place**

Joaquín Hortal[1]*, Jorge M. Lobo[1] & Alberto Jiménez-Valverde[2]

[1] *Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales – CSIC, Madrid, Spain*

[2] *Departamento de Biología Animal, Universidad de Málaga, Spain*

**All authors contributed equally to this work.*

"In considering the distribution of organic beings over the face of the globe, the first great fact which strikes us is, that neither the similarity nor the dissimilarity of the inhabitants of various regions can be wholly accounted for by climatal and other physical conditions." Charles Darwin (1872, p. 316)

Abstract

Biogeographers have tried to explain the geographic patterns of biodiversity for more than two centuries. As early as the 18th Century differences in climate were the first cause invoked to explain why the composition of species and their forms vary in space. The large amount of evidence in favor of a covariation between climatic and biodiversity gradients supports the use of contemporary factors as the main predictor of species distribution ranges. However, although it is clear that current environmental conditions affect species distributions, this *ahistorical* view may neglect the role played by other processes. Species distributions are also the result of many other factors related with the evolutionary history of the species, their dispersal ability and/or the history of land and climate. The different aspects of this *historical* view have been also largely studied and acknowledged. Here we discuss how these historical factors can produce spatial patterns of diversity that correlate with contemporary climatic gradients. We review recent advances that synthesize a more complex conceptualization of the factors affecting species distributions, highlighting the differences between potential and realized distributions. We discuss how can non-contemporary factors be partly taken into account in Species Distribution Modelling (SDM) through (i) an informed choice of the extent of the analysis – that restricts it to the areas where no historical processes have limited the range of the species; and (ii) a good choice of the evaluation dataset – that allows to determine the success of the model in describing the particular aspect of the distribution of the species that is being studied. We recommend a thoughtful approach to SDMs, based on a careful design of the analyses that ensures that they are the most adequate to help answering the particular questions of each study.

Key words: Species Distribution Models, Environmental Niche Models, Correlation, Causality, Historical Effects, Extent, Model Evaluation and Validation.

Some Little History

The general purpose of biogeography is to understand why life varies through space. That is, where are the populations of species, their different forms and the communities they constitute located; how species, forms and communities vary in space; and which are the causes of these variations. These questions already appear in the foundations of biogeography as a body of science in the edge of the 18th and 19th Centuries (see e.g. Briggs & Humphries 2004 for an account). As early as 1808 Alexander von Humboldt was not only able to relate the differences in the geographical distributions of the families – and species – of plants with the climatic conditions of the regions where they are most abundant. He also related their forms – and functions – with the character (i.e. configuration) of the landscapes they inhabit, and even with the functioning of these landscapes (see e.g. von Humboldt 1850, p. 279). Implicit in Humboldt's work is the idea that the differences in the biotas of different continents are in part the consequence of changes through time. But it is Charles Darwin itself (with a large influence of Alfred Russel Wallace) who relates geography – both in terms of space and topographic barriers – with time as two of the main components of evolutionary changes (Darwin 1872). Since then, knowledge on distributional patterns cannot be dissociated from an evolutionary perspective. One of the main challenges of biogeography as a science is

Send correspondence to: Joaquín Hortal
Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales – CSIC, C/ José Gutiérrez Abascal 2, 28006, Madrid, Spain
E-mail: jhortal@mncn.csic.es

relating present-day distributional data with the historical variations of Earth's geography and environment. This poses questions such as to what point are observed biogeographical patterns the reflection of past speciation, extinction and range shift processes? Do evolutionary processes leave any geographical trace? Can we derive the evolutionary processes generating biodiversity from current geographical patterns? How old and detectable are these signals? This deeply evolutionary reasoning is in the very essence of the origin of biogeography as a scientific discipline.

Being space, time and form the three components of the evolutionary cradle, biogeographers are interested in the origin of form. That is, in the processes occurring over space and in time that are responsible that a certain *taxon* appears at a given point of the map (Croizat 1962, p. 6). From the very beginning species distributions have been correlated and/or explained by the role played by two main types of factors: contemporary and historical (Brown *et al.* 1996). When the first biogeographers (or the first naturalists interested in what we now call biogeography) started to think about why the species are where they are and not elsewhere, climate and other present-day environmental factors were arguably the first cause to be invoked. During the second half of the 18th Century both Carolus Linnaeus and George-Louis Leclerc, Compte de Buffon, maintained that similar climates in different continents host different species because they dispersed from a common center of origin – the Paradise or Mount Ararat – , eventually becoming separated in space; but always tracking the climatically most adequate climatic zones (see Briggs & Humphries 2004). In 1820 Augustin de Candolle included the "geography of plants" as a part of their physiology "[...] for it investigates the laws according to which climate, temperature, soil, elevation […]" and distance to the equator "[…] operate upon the production of plants [...]" (de Candolle 1821, p. 263). Although he conceded that plant distributions and forms were connected with their history, he minimized such effects, emphasizing the role of the physical environment as an "essential influence" on plant growth and distribution. Similarly, Forbes (1843, p. 152, cf. Briggs & Humphries 2004) argued that distributions were influenced by "climate, mineral structure, and elevation" and "climate, sea composition and depth" for terrestrial and marine animals, respectively. This pre-eminence of contemporary factors as the main determinant of the distribution of life continues to be supported by a number of recent studies (e.g. Currie & Paquin 1987; Wright *et al.* 1993; Hawkins *et al.* 2003a). Despite the major imprint of evolutionary thinking on biogeography, the evidences in favour of environmentally-driven biodiversity patterns sustain a point of view in which the historical trace on extant geographical patterns is of minor relevance. Here we will denominate this standpoint as *ahistorical* for simplicity, as part of a dichotomy that includes the *historical* perspective that relies on adaptations to former conditions, biotic interactions and co-evolutionary, dispersal and extinction processes as major determinants of species distributions (see below).

Modelling Species Distributions from Contemporary Factors

According to the ahistorical point of view, the information about present environmental conditions could be used to obtain reliable biogeographical estimations, not only for present-day distributions but also to transfer these predictions to past and future scenarios. Following this rationale, a number of statistical techniques are now widely used to produce hypotheses about species' distributions based on contemporary environmental predictors – mainly climatic – and the available data about species occurrences – almost exclusively presence information. The application of these Species Distribution Models (SDMs) increased exponentially since 1995 (Lobo *et al.* 2010), so that currently about one paper is published every day in scientific journals of international scope. SDMs are commonly used to predict both the location of species' range limits and the variability in the probability of occurrence within these limits. Here note that SDMs are also called "Environmental Niche Models", "Ecological Niche Models" (both called ENMs), "Habitat Suitability Models" (HSMs), "Bioclimatic Envelope Models" (BEMs) or climate envelopes. We deliberately use the more general – and less problematic – term Species Distribution Models to refer to them because they model known species occurrences (i.e. known distribution) to generate hypotheses about species distributions, rather than modelling their niche or the actual suitability of each place (see Figure 1 below) (but see Peterson & Soberón 2012). There are many recent reviews and books on why and how to use these techniques in different contexts (e.g. Franklin 2009; Elith & Leathwick 2009; Peterson *et al.* 2011; Svenning *et al.* 2011; Varela *et al.* 2011; Anderson 2012), so we will not discuss it here. We refer to these works for further details on the current common usage of SDMs. But we would like to highlight that most recent SDM applications typically follow the recommendations from a single study (Elith *et al.* 2006) about which techniques are best in generating accurate distributional predictions based on partial data, despite criticisms to the design and findings of that study (Lobo 2008; Hijmans 2012).

The particular physiology and morpho-anatomical constraints of each organism imposes a series of tolerance limits to environmental gradients that constrain the places where it can maintain successful populations. It is thus undeniable that current distribution patterns must retain a – more or less – strong environmental signal and, consequently, SDMs may offer some sort of approximation to the environmental requirements (Soberón 2007). However, uncritical and automatic practises may be implicitly promoting an ahistorical view, offering a flawed picture on species distributions, whose origin is more complex than mere responses to current environmental conditions. We believe that three factors have contributed

to the common thought that, in general, SDMs can provide reliable accounts of species distributions: (i) a persistent oblivion of the difference between correlation and causality; (ii) an oversimplified and narrow view of species distributions and the associated lack of understanding on the meaning of SDM results; and (iii) the use of invalid and arbitrary training and validation procedures. We provide a brief account of the first factor, provide an updated overview on the determinants of species distributions and their implications for the use of SDM, and discuss in some extension the importance of validation to determine our current conceptualization of SDMs.

Missing the Cause for the Correlation

In practice, in most occasions SDMs are used just to produce an accurate predictive function from observational data; that is, to forecast the species' distribution range (*sensu* Legendre & Legendre 1998, p. 482). This implies that SDMs can provide accurate predictions even without considering truly causal factors. In fact, being based on purely statistical fitting, there is no need to presume causality to build them. Of course, it would always be preferable to use predictors that are known to influence the geographical responses of the species or to consider the available ecological and physiological knowledge (Austin 2007). It is interesting to highlight that, at least in some occasions, the spatial coordinates of species' occurrences can act as accurate surrogates of the "true" causes of their distributions (see Bahn & McGill 2007). This is because the spatial structure in these occurrences may be associated with the spatial structure in the processes influencing species distributions, if these processes are spatially structured –as is often the case with processes occurring in the geographical space. On the down side, this also implies that the mere presence of spatial autocorrelation in both the response and explanatory variables may be the cause of their covariation. Under these – common – circumstances it is possible to obtain reliable predictions from spurious correlations with predictors that are not related with the response variable –in this case the presence and absence of the species.

It is (or it should be) well known that there is a significant degree of uncertainty when correlations are used to identify or discriminate causality. Any variable can be identified as being a good predictor due to its high discrimination power within a given dataset, but fail to describe the relationship with the response variable in the rest of the – statistical – population it describes if the sample is not representative enough, or if more adequate predictors have not been considered (see, e.g., Legendre & Legendre 1998). This impedes to *demonstrate* causality by correlating field data on species occurrences and environmental variables. In contrast, to be accurate the projections of SDM results in space and/or time (the most common uses of SDMs) need that the model parameters reflect true causality – and that such causal relationship holds true in new conditions and territories. At least, it is expected that the correlation
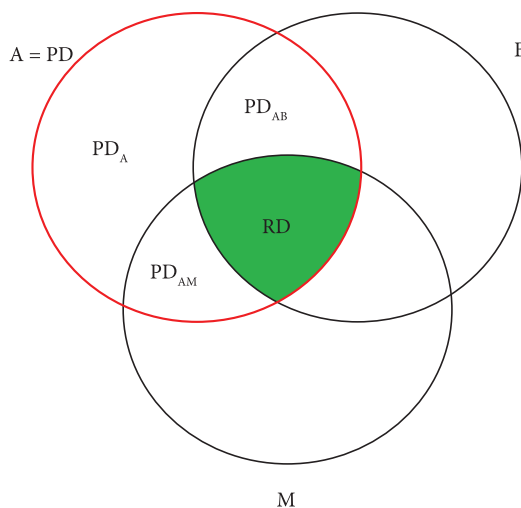

Figure 1. Venn diagram exemplifying the main factors affecting species distributions (i.e. the so-called BAM diagram from Soberón & Peterson 2005; Soberón 2007), modified according to Jiménez-Valverde *et al.* (2011a). *A* would be the geographic area in which the scenopoetic environmental variables – that are unaffected by the presence of the species – define places suitable for the survival of the species –where its intrinsic growth rate is positive. Climatic conditions can, in some degree, be conditioned by the organisms, thus compromising their scenopoetic character (Odling-Smee *et al.* 2003). Similar arguments apply to edaphic and other variables via niche construction (e.g., Boogert *et al.* 2006). For simplicity we assume that the majority of these variables, and in particular the climatic ones, are related with extrinsic forces generating variations in Earth's orbit (e.g., Dynesius & Jansson 2000). *B* would be the geographic area where biotic interactions are favorable for the persistence of populations of the species (e.g. because there are populations of its host species, or no competitive exclusion). *M* is the geographic region that has been accessible to the species over relevant periods of time (Barve *et al.* 2011). The extent of this region is determined by geographic and/or environmental barriers and some biological attributes of the species (e.g. dispersal capacity, niche specialization); but also by the location of the origin of the target species, the amount of time since such origin or the spatio-temporal pattern of past climatic changes. This combination of historical factors and dispersal limitations leaves a profound impact on current species distributions. The geographical representation of the realized niche (the realized distribution or *RD*; in green) is the accessible area with favorable biotic and environmental conditions, while the whole *A* circle represents the potential distribution (*PD*) composed by the realized distribution, the area with favorable biotic and environmental conditions that remains un-colonized (PD_{AB}), the accessible area with favorable environmental conditions but inappropriate due to biotic interactions (PD_{AM}), and the inaccessible area with favorable environmental conditions but unsuitable due to biotic interactions (PD_A). All these areas conforming *PD* (red circle) can be potentially invaded depending on the species' capacity to overcome current dispersal limitations and its plasticity to develop new biotic interactions that allow sustaining its populations.

structure among the predictors and between the true causal variables and the surrogates holds in all the studied territory (Jiménez-Valverde *et al.* 2009; Dormann *et al.* 2012). This makes transferring model results outside the training data a risky business, no matter whether the response variable

is modelled using its relationship with the environment or its spatial structure (see Hortal & Lobo 2011). It follows that although SDMs can be used to estimate realized or potential distribution ranges (see below), they present serious drawbacks that need to be taken into consideration when interpreting their results. Regardless of the technique and predictors used, the key aspect for obtaining fair SDM results is a consistent training sample that covers the environmental and spatial variability of the whole extent of the target territory, as well as a good validation dataset and a reliable validation procedure that allows estimating the prediction errors. We have reviewed and discussed in extent the quality, problems and biases of biodiversity data elsewhere (see, e.g., Hortal *et al.* 2007, 2008; Lobo *et al.* 2007; Rocchini *et al.* 2011), so hereafter we will focus on why relationships between species occurrences and environmental data are not enough, and how a well-designed validation procedure can help assessing SDM results with fairness and even refining their final performance. In any case, it is important to highlight that SDM predictions that are consistent with the evaluation data can reinforce and suggest hypotheses about the role of the predictors that end up being part of the final models (Kearney & Porter 2009; Dormann *et al.* 2012), but they can never be considered as prove of causality (Shipley 2000).

A More Complex View on Species Distributions

During the 19th Century the foundations of modern biogeography were built thanks to the work of many naturalists – from which the abovementioned work of Darwin and, in particular, Wallace outstand. During this period of scientific expeditions the importance of dispersal processes and the effect of geographical barriers in restricting species distribution and generating different regional biotas became evident (e.g. Wallace 1876). This reinforced the importance of dispersal or isolation and added the spatial configuration of the geographical matrix to the set of known determinants to species distributions, a view that gained a deeper temporal perspective after Alfred Wegener's works on continental drift (Briggs & Humpries 2004). But it is during the 20th Century when a greater variety of aspects of the geographic distributions of species start to be examined. The multi-faceted appearance of the concept of ecological niche leads to recognize that growth and persistence of populations is affected by both environmental conditions (e.g. Grinnell 1917) and ecological interactions (e.g. Elton 1927). The latter implies that the distributions of species can be restricted further than the areas imposed by their environmental requirements (and dispersal limitations). This makes describing the distribution of species based on the location of their populations a hard and complex task (Erickson 1945, cf. Soberón 2007). But also allowed George E. Hutchinson (1957, 1978) to unify both types of niches – scenopoetic (i.e. environmental) and biotic (i.e. determined by interactions) – into a single conceptual

framework, setting up the basis for an inclusive treatment of ecological and physiological processes in the study of species distributions (Soberón 2007, 2010; Soberón & Nakamura 2009; Colwell & Rangel 2009). The temporal perspective provided by paleoecologists also included into this picture the importance of past climate on current distributions (Jackson & Overpeck 2000). The development of metapopulation ecology during the second half of the last Century (see e.g. Hanski & Gilpin 1991) added a further dimension to the understanding of the location of populations both within and outside species distributions (Pulliam 1988, 2000). Species distributions are now understand as complex entities affected by a complex mixture of scenopoetic factors, biotic interactions, geographic and historical processes, and occupancy dynamics (Hortal *et al.* 2010). To understand such complexity it is useful to recourse to conceptual simplifications. Building upon the simplified account of Soberón & Peterson (2005), Soberón (2007) conceptualized some of the main factors influencing current species distributions into a single theoretical framework, which has been improved in successive revisions (Soberón 2007, 2010; Soberón & Nakamura 2009; Barve *et al.* 2011; Jiménez-Valverde *et al.* 2011a). This theoretical framework is described in Figure 1. Here note that Soberón's (2007) framework does not deal explicitly with the effect of historical processes. A complementary (and deeper) temporal perspective is in part provided by Jackson & Overpeck (2000) (see Figure 1).

A direct consequence of such diversity (and complexity) of factors is that studying species distributions is not straightforward, and needs to rely on both ahistorical and historical biogeographical perspectives. According to the – ahistorical – environmental control model, contemporary environmental gradients would be the main force controlling species distributions. It is impossible to deny the important role of productivity and several climatic variables causally related to it (evapotranspiration, temperature, water-energy balance, etc.) on limiting populations and diversity, both now and in the past (Hawkins 2004). Available energy and primary resources affect current diversity and have been doing so over evolutionary time (Fischer 1960). Therefore, the variables describing their gradients may hold a high power for explaining species distributions and/or the geographical variability in other diversity attributes. However, if we follow the steps of classic biogeographers, a more complex picture emerges. The shape and location of species distributions at global scale cannot be adequately explained without considering the singular and contingent history of clades, biomes, climatic conditions and land masses (Ricklefs & Schluter 1993; McGlone 1996; Dynesius & Jansson 2000; Hawkins *et al.* 2003b; Wiens & Donoghue 2004; Mutke & Barthlott 2005; Fine & Ree 2006; Ricklefs 2004, 2006; Hortal *et al.* 2011; Baselga *et al.* 2012a, b), together with the inability of organisms to occupy all the *a priori* environmentally favorable places (Svenning & Skov 2004). At last resort, even the distinction between contemporary environmental and historical factors is

naïve, because the role of current climate in delimiting species ranges is the consequence of the adaptations of the species developed throughout the evolutionary history of their ancestors. The inertia of such evolutionary process impedes the development of new adaptations that are not rooted into those of the ancestral lineages, thereby imposing a series of physiological and morpho-anatomical constraints (Ricklefs 2006).

The generally high predictive power of contemporary environmental determinants, and most importantly of climate, has led to the general thought that their use may be enough, or almost enough, to obtain accurate predictions of species distributions. And it is true that they are powerful predictors of the patterns of composite biodiversity variables such as species richness, perhaps because the individual determinants of each species are averaged across all of them. However, their ability to predict the geographic distributions of individual species is limited. The distribution of species is the consequence of an individualistic – or autoecologic – behavior where species track the frequent environmental changes by shifting their distributions (spatial adaptation; Hengeveld 1997). Due to this, dispersal limitations and historical factors are highly likely to exert a major influence on species' geographic ranges. Both historical and dispersal processes are geographically explicit and – in almost all cases – operate through spatial contagion. It follows that they can generate spatially structured patterns that are, in practice, indistinguishable from those generated by contemporary environmental conditions (Legendre & Legendre 1998). Building upon the considerations on correlation and causality discussed before (see section "*Missing the cause for the correlation*"), models of the potential distribution of species calibrated on their current geographic ranges are likely to fail (Soberón & Peterson 2011). When based solely on field data, determining which variables are influential for species distribution and estimating their model parameters may misrepresent the true abiotic niche (*A* in Figure 1), generating inaccurate projections when transferred or extrapolated to a different time or space.

To be influential, environmental factors must act on evolutionary and biogeographical processes such as speciation, extinction and dispersal (Wiens & Donoghue 2004). This implies that the likelihood that a given factor is a good predictor for the distribution of a species depends on its capacity of triggering evolutionary processes and the history of climate in the studied region. Interestingly, the origin of most modern species is older than the Pleistocene (Klicka & Zink 1997; Magallón & Sanderson 2001). This implies that deeper temporal scales need to be considered to detect the influence of speciation processes on current distributional patterns. At large spatial and temporal scales, fossil evidence shows that speciation rates are higher in the tropical areas, while extinctions and range shifts predominate under temperate conditions (Krug *et al.* 2009). As a consequence, basal clades accumulate species in the tropics while derivate

ones appear in temperate biomes, which in turn experience much higher extinction and immigration rates (Wiens & Donoghue 2004; Hawkins *et al.* 2006). This implies that the limits to the distribution of many temperate species will be difficult to predict from scenopoetic variables simply because other contingent factors have a higher probability of impeding the equilibrium of distribution ranges with current environmental conditions (Baselga *et al.* 2012a). On the other hand, the difficulties for predicting the distributions of many tropical species with narrow geographic ranges and high specializations will come from the need to consider fine-tuning microenvironmental variables and complex biotic interactions in the modelling process.

Modelling Species Distributions

The existence of the diverse array of historical and ahistorical processes discussed above has many crucial implications for the use of Species Distribution Models to describe, project and study geographic ranges. Most of the examples above are extreme examples on how different evolutionary and biogeographical histories may affect species distributions in different ways. But they serve to highlight that the actual performance of SDM applications will be dramatically improved if they are based on a good knowledge about the factors influencing species distributions, their responses at the studied scale and the questions that are to be answered (Kamino *et al.* 2012a, b). For example, the distribution of dung beetles within a landscape is highly reliant on the availability of their food source (Lobo *et al.* 2006a), but at the whole extent of Europe –where dung is widely available from cattle and sheep herding–temperature gradients determine the distribution of most widespread species as a direct effect of their evolutionary history (Hortal *et al.* 2011). Also, knowledge or estimates of the past dispersal processes may allow determining if a spatial structure not related with environmental variables should be expected. While widespread European tree species that have fewer limitations to dispersal are most affected by current climate, geographically restricted species are probably more related to the conditions during Last Glacial Maximum due to their difficulty to spread from their glacial refugia (Svenning & Skov 2007). To model the distribution of these latter species it is necessary to recourse to geographical constraints and/ or spatially-structured predictors. Also, a number of species' traits are known to affect SDM performance (McPherson & Jetz 2007; Syphard & Franklin 2010; Chefaoui *et al.* 2011; Swab *et al.* 2012), so a good knowledge on the biology of the species is critical for a correct assessment of the adequacy of SDM techniques.

A direct consequence of such complex interrelation of factors is that there are several aspects to species distributions. And all of these aspects need to be considered in order to understand the origin and dynamics of geographic distribution ranges. We first need to distinguish the potential distribution – all the places where the species could live – from the realized distribution – all the places where it

actually lives. This distinction is often neglected due to the reliance of ahistorical biogeography, and therefore SDM results are often interpreted as actual distributions while they model potential responses, or *vice versa*. However, a correct interpretation of which aspect is being actually modelled could be the basis for sound ecological research (Kamino *et al*. 2012a, b). The choice of SDM method – and the associated choices on the type of data used – determines the adequacy of the output to model different aspects of species distributions (Jiménez-Valverde *et al*. 2008a; Chefaoui & Lobo 2008). Models of potential distributions should aim to describe the whole Abiotic niche (*A* in the BAM diagram, Figure 1), and therefore be inclusive rather than exclusive in terms of data and predictions. Whereas models of realized distributions should constrain predictions to the areas where the species is currently present (*RD* in Figure 1). It follows that SDMs describing realized distributions may not be adequate for being projected to different times or regions. They can be used to assess current threats and conservation status (e.g. Platts *et al*. 2010). In contrast, SDMs describing potential distributions are not adequate for, say, the design of current areas for conservation (see, e.g., Smart *et al*. 2010). But they can be for identifying areas under risk of invasion (e.g. Roura-Pascual *et al*. 2004), assessing threats of disease contagion to wild species by the spread of their vectors (e.g. Acevedo *et al*. 2007), finding new populations of poorly-known species (e.g. Jiménez-Valverde *et al*. 2011b) or studying temporal changes in species distributions (e.g. Varela *et al*. 2011). And the differences between the estimates of potential and realized species distributions can be used to assess the conservation status of endangered species or the existence of barriers to dispersal (e.g. Lobo *et al*. 2006b; Chefaoui & Lobo 2007; Jiménez-Valverde *et al*. 2008b).

The dichotomy between historical and ahistorical factors also poses a basic decision problem at the very beginning of the modelling process: the selection of the sampling universe. The selection of the appropriate spatial scale has long been recognized as a central challenge in ecological research (Wiens 1989). However, the selection of the spatial extent has been largely ignored by the SDM community, despite being an elemental statistical question – the selection of the sample – which is also in the root of how the researchers interpret the distribution maps. Only recently a few studies have explicitly addressed how to choose the limits of the training area in SDM (which defines the calibration data) (Anderson & Raza 2010; Barve *et al*. 2011; Acevedo *et al*. 2012). Such choice should not be made at will; rather, it has to take into account that geographic ranges are more than the mere expression of the scenopoetic niche. According to Barve *et al*. (2011), the training sample has to be selected from the area that has been accessible for the species in a relevant period of time (M in the BAM diagram, Box 1). In other words, records of the absence of the species in sites that it has not had the opportunity to reach are entirely uninformative (to describe its potential distribution; see also Lobo *et al*. 2010). Although enlightening, this definition is rather non-operational. Subsequently, Acevedo *et al*. (2012)

specified that the training area should be that in which the species is actually interacting with the environment, and proposed to delimit this region using Trend Surface Analysis (a geostatistical technique, see e.g. Legendre & Legendre 1998). This way historical effects are already accounted for while delimiting the spatial extent, so the models obtained could be considered "purely" ecological (to a certain extent). The selection of the spatial extent is also important because it has dramatic effects on the parameterization of the models and on the statistics used to evaluate them (Lobo *et al*. 2008; Giovanelli *et al*. 2010; Barve *et al*. 2011; Acevedo *et al*. 2012). Given the little work conducted so far this topic will, without doubt, generate much discussion in the near future.

Evaluating the Results of Species Distribution Models

Model evaluation is a central part of Species Distribution Modelling, indicating whether the models are reliable and useful (Vaughan & Ormerod 2005) and, to a certain degree, if the theory behind them is correct or not (but see section "*Missing the cause for the correlation*" above). Once SDMs are parameterized, they can be transferred either in space or time. For example, one could be interested in forecasting the possible contractions or retractions of the geographic ranges due to climate change, or in identifying the suitable localities for a certain species in a new territory which has the risk of being invaded. In both cases there is usually no data set available for validations, since the focus species have not (yet) reached the new region or time. Validating the projections of potential distributions is quite difficult (or, rather, impossible) using empirical field data, since the object of the evaluation is the capacity of the models to estimate a potential, and thus non-realized, distribution (Jiménez-Valverde 2012). Historical processes (e.g. Varela *et al*. 2010), dispersal limitations (e.g. Lobo *et al*. 2006b) and the dynamics of natural systems (e.g. Zurell *et al*. 2009) impede most (if not all) species to occupy their whole potential distribution. Thus, the absence of the species from places identified as potentially suitable cannot serve as proof that the model is failing.

Despite this, under certain circumstances it may be possible to conduct some sort of assessment of the performance of SDM projections in describing potential distributions. For instance, in the rare occasions when distributional data exists for two time periods (e.g. Martínez-Meyer *et al*. 2004; Araújo *et al*. 2005; Araújo & Rahbek 2006; Nogués-Bravo *et al*. 2008); when an independent distributional hypothesis – not based on SDMs (e.g. phylogeographic, mechanistic) – is available for one time period (e.g. Waltari *et al*. 2007; Elith *et al*. 2010); or when the species has already invaded a new territory and is expanding its range throughout it (Peterson 2003). Albeit possible, evaluation in these occasions is yet tricky and risky. Usually, distributional data for one of the time periods is rather incomplete and biased, such as the cases of fossil records

(Varela *et al.* 2011) or invasive species whose distributions have not reached equilibrium (Jiménez-Valverde *et al.* 2011a). Under these circumstances, the measures that are often used to assess the discrimination power of models are useless, since there are no events to separate; that is, there are no cases of absence to separate from the cases of presence. The application of discrimination measures when no true absence data are available can lead to misleading results (Jiménez-Valverde 2012). The only statistic that could give an idea of the reliability of the models may be the rate of cases of presence that are incorrectly classified as absences. It makes sense that if the models have captured the environmental response of the species correctly – and assuming that it has not changed between time/spatial scenarios – most presences should be accurately predicted (Nogués-Bravo *et al.* 2008; Peterson *et al.* 2011).

Transferring SDM results or estimating the potential distribution of species are required for a number of applications, but researchers are often interested in estimating realized distributions (Peterson 2006). These estimates are necessary because the geographic range of the species is never fully characterized by the available biodiversity data (the Wallacean shortfall; Lomolino 2004; Rocchini *et al.* 2011). Modelling realized distributions requires reliable absence and presence data (ideally coming from a random and representative sampling). These distributional data can be randomly divided into a training data set and a validation data set; the results of SDM based in the former dataset are evaluated through their performance in the latter, using different measures of discrimination and calibration (Pearce & Ferrier 2000). Arguably, this assessment informs about how closely the estimated suitability scores reflect the observed probability of presence of the species. However, discrimination measures, such as the Area Under the Curve (AUC), depend on the statistical distribution of the estimated suitability values, a characteristic that is context-dependent (Jiménez-Valverde *et al.* 2012). This implies that they cannot actually inform about how "good" a model is. Well-calibrated models can yield low discrimination values if the species shows a generalist environmental response in the studied region. Thus, although discrimination provides some information about the usefulness of SDMs in a certain context, calibration offers a more general picture of model reliability, and should be reported together with discrimination (Jiménez-Valverde *et al.* 2012).

Nevertheless, the evaluation of realized distributions faces two important issues. First, as discussed before, any spatially-structured pattern can be modelled using spatially-structured variables (Bahn & McGill 2007; see above). This means that the fact that SDM results predict reasonably well a random subset of distributional data (i.e. cross-validation) does not necessarily imply that the model is biologically meaningful. Second, the (common) presence of spatial autocorrelation in the data usually results in SDM performance being largely overestimated (Hijmans 2012). This problem could be solved by resampling the evaluation data in order to reduce spatial proximity between data cases. Although this approach has been pursued a few times (e.g. Boyce *et al.* 2002; Segurado & Araújo 2004), how to reduce the degree of spatial contagion in the observations is a question that deserves further attention. An alternative to the use of a randomly splitted evaluation data set is to develop regionally constrained SDMs (i.e., calibrated with certain populations) and evaluate their performance using distributional data from other populations (Jiménez-Valverde *et al.* 2011a). This way, one can try to avoid the limitations caused by autocorrelation, subjecting SDMs to a more challenging test. Here it is important to note that distinct environmental adaptations between populations (Osborne & Suárez-Seoane 2002) can make good models to perform badly in the evaluation datasets. Although model evaluation is the step which could, in theory, tell us how much we can rely on SDMs, it is still an area under development. Most importantly, we are starting to see hitherto unrecognized problems that may have biased our perception of the actual performance of SDMs.

Concluding Remarks

Species Distribution Models are powerful statistical techniques that can be used to describe species' geographic ranges and project their results to different times and spaces. However, as we have discussed extensively, SDMs face a number of theoretical and practical problems that compromise the reliability of their results and their adequacy to answer certain questions. This by no means implies that SDMs cannot be part of the biogeographer's toolbox; rather, they should be handled with care and thoughtfulness. To decide which combination of distributional data, predictors and SDM technique is most useful in each case, it is necessary to pay attention to a number of aspects. First it is important to define what is your question, and which part of the BAM diagram are you interested in to answer it. Then, determine whether you have got data to represent it or not, and finally (only finally) pick the best SDM technique and modelling strategy according to the question that is intended to answer (Kamino *et al.* 2012b).

Currently there is an extensive literature on SDM, including books (Franklin 2009; Peterson *et al.* 2011) and many practical examples. We have provided a number of hints about how to model species distributions, but rather than providing a quick guide to the practical aspects of Species Distribution Modelling, we want to promote a thoughtful approach to this discipline. Using a particular SDM technique or modelling the distribution of a species are not true research or technical goals; they are means to an objective. And there are many of these objectives, so there are many ways to model the geographic distribution of a species in a meaningful way. And in many cases the available data, predictors and/or techniques will not be adequate to the intended objective. Thus, rather than applying SDM techniques indiscriminately, we enthusiastically recommend to first think about your question and which

aspect of species distributions you need to model, conduct new surveys if necessary, consider making field and/or lab experiments, and be prepared to face that some questions cannot be answered, at least not just yet.

Acknowledgements

References

Acevedo P *et al.*, 2007. Invasive exotic aoudad (*Ammotragus lervia*) as a major threat to native Iberian ibex (*Capra pyrenaica*): a habitat suitability model approach. *Diversity and Distributions,* 13:587-597. http://dx.doi.org/10.1111/j.1472-4642.2007.00374.x

Acevedo P *et al.*, 2012 Delimiting the geographical background in species distribution modelling. *Journal of Biogeography*, 39:1383-1390. http://dx.doi.org/10.1111/j.1365-2699.2012.02713.x

Anderson RP, 2012. Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences*, 1260:66-80. PMid:22352858. http://dx.doi.org/10.1111/j.1749-6632.2011.06440.x

Anderson RP & 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. http://dx.doi.org/10.1111/j.1365-2699.2010.02290.x

Austin M, 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, 200:1-19. http://dx.doi.org/10.1016/j.ecolmodel.2006.07.005

Araújo MB & Rahbek C, 2006. How does climate change affect biodiversity? *Science*, 313:1396-1397. PMid:16959994. http://dx.doi.org/10.1126/science.1131758

Araújo MB *et al.*, 2005. Validation of species-climate impact models under climate change. *Global Change Biology*, 11:1504-1513. http://dx.doi.org/10.1111/j.1365-2486.2005.01000.x

Bahn V & McGill BJ, 2007. Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, 16:733-742. http://dx.doi.org/10.1111/j.1466-8238.2007.00331.x

Barve N *et al.*, 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222:1810-1819. http://dx.doi.org/10.1016/j.ecolmodel.2011.02.011

Baselga, A *et al.*, 2012a. Historical legacies in world amphibian diversity revealed by the turnover and nestedness components of beta diversity. *PLoS ONE*, 7(2):e32341. http://dx.doi.org/10.1371/journal.pone.003234

Baselga A *et al.*, 2012b. Global patterns in the shape of species geographical ranges reveal range determinants. *Journal of Biogeography*, 39:760-771. http://dx.doi.org/10.1111/j.1365-2699.2011.02612.x

Boogert NJ *et al.*, 2006. The implications of niche construction and ecosystem engineering for conservation biology. *BioScience*, 56:570-578. http://dx.doi.org/10.1641/0006-3568(2006)56[570:TIONCA]2.0.CO;2

Boyce MS *et al.*, 2002. Evaluating resource selection functions. *Ecological Modelling*, 157:281-300. http://dx.doi.org/10.1016/S0304-3800(02)00200-4

Briggs JC & Humphries CJ, 2004. Early classics. In: Lomolino MV, Sax DF & Brown JH (eds.). *Foundations of biogeography Classic papers with commentaries*. Chicago; London: The University of Chicago Press. p. 5-13.

Brown JH *et al.*, 1996. The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics*, 27:597-623. http://dx.doi.org/10.1146/annurev.ecolsys.27.1.597

Chefaoui RM & Lobo JM, 2007. Assessing the conservation status of an Iberian moth using pseudo-absences. *Journal of Wildlife Management*, 71:2507-2516. http://dx.doi.org/10.2193/2006-312

Chefaoui RM & Lobo JM, 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, 210:478-486. http://dx.doi.org/10.1016/j.ecolmodel.2007.08.010

Chefaoui RM *et al.*, 2011. Effects of species' traits and data characteristics on distribution models of threatened invertebrates. *Animal Biodiversity and Conservation*, 34:229-247.

Colwell RK & Rangel TF, 2009. Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences USA*, 106:19651-19658. http://dx.doi.org/10.1073/pnas.0901650106

Currie DJ & Paquin V, 1987. Large-scale biogeographical patterns of species richness of trees. *Nature*, 329:326-327. http://dx.doi.org/10.1038/329326a0

Croizat L, 1962. *Space, Time, Form*: The biological synthesis. Caracas: Published by the author.

Darwin C, 1872. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. 6h ed. with additions and corrections. London: John Murray.

De Candolle AP, 1821. *Essai elementaire de geographie botanique*: Dictionnaire des sciences naturelles. Strasbourg: Flevrault. vol 18, p. 1-64.

Dormann CF *et al.*, 2012. Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography*, 39:2119-2131. http://dx.doi.org/10.1111/j.1365-2699.2011.02659.x

Dynesius M & Jansson R, 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National*

Academy of Sciences USA, 97:9115-9120. PMid:10922067 PMCid:16831. http://dx.doi.org/10.1073/pnas.97.16.9115

Elith J & Leathwick JR, 2009. Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution and Systematics*, 40:677-697. http://dx.doi.org/10.1146/annurev.ecolsys.110308.120159

Elith J *et al.*, 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29:129-151. http://dx.doi.org/10.1111/j.2006.0906-7590.04596.x

Elith J *et al.*, 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1:330-342. http://dx.doi.org/10.1111/j.2041-210X.2010.00036.x

Elton CS, 1927. *Animal Ecology*. London: Sidwick and Jackson.

Erickson RO, 1945. The *Clematis fremontii* Var. *Riehlii* population of the Ozarks. *Annals of the Missouri Botanical Garden*, 32: 413-460. http://dx.doi.org/10.2307/2394445

Fine PVA & Ree RH, 2006. Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. *American Naturalist*, 168:796-804. PMid:17109321. http://dx.doi.org/10.1086/508635

Fischer AG, 1960. Latitudinal variations in organic diversity. *Evolution*, 14:64-81. http://dx.doi.org/10.2307/2405923

Forbes E, 1843. Report on the Mollusca and Radiata of the Aegean Sea, and on their distribution considered as beating on geology. In: *Proceedings of the Thirteenth meeting of the British Association for the Advancement of Science*; 1983. vol. 13, p. 130-207

Franklin J, 2009. *Mapping species distributions. Spatial inference and prediction*. Cambridge University Press.

Giovanelli JGR *et al.*, 2010. Modeling a spatially restricted distribution in the Neotropics: How the size of calibration area affects the performance of five presence-only methods. *Ecological Modelling*, 221:215-224. http://dx.doi.org/10.1016/j.ecolmodel.2009.10.009

Grinnell J, 1917. The niche relationships of the California thrasher. *Auk*, 34:427-433. http://dx.doi.org/10.2307/4072271

Hanski I & Gilpin M, 1991. Metapopulation Dynamics - Brief-History and Conceptual Domain. *Biological Journal of the Linnean Society*, 42:3-16. http://dx.doi.org/10.1111/j.1095-8312.1991.tb00548.x

Hawkins B, 2004. Are we making progress toward understanding the global diversity gradient? *Basic and Applied Ecology*, 5:1-3. http://dx.doi.org/10.1078/1439-1791-00214

Hawkins BA *et al.*, 2003a. Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84:3105-3117 http://dx.doi.org/10.1890/03-8006

Hawkins BA *et al.*, 2003b. Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology*, 84:1608-1623. http://dx.doi.org/10.1890/0012-9658(2003)084[1608:PAHAPO]2.0.CO;2

Hawkins BA *et al.*, 2006. Post-Eocene climate change, niche conservatism, and the latitudinal diversity gradient of New World birds. *Journal of Biogeography*, 33:770-780. http://dx.doi.org/10.1111/j.1365-2699.2006.01452.x

Hengeveld R, 1997. Impact of biogeography on a population-paradigm shift. *Journal of Biogeography*, 24:541-547. http://dx.doi.org/10.1111/j.1365-2699.1997.tb00069.x

Hijmans RJ, 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, 93:679-688. http://dx.doi.org/10.1890/11-0826.1

Hortal J & Lobo JM, 2011. Can species richness patterns be interpolated from a limited number of well-known areas? Mapping diversity using GLM and kriging. *Natureça & Conservação*, 9:200-207. http://dx.doi.org/10.4322/natcon.2011.026

Hortal J *et al.*, 2007. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife (Canary Islands). *Conservation Biology*, 21:853-863. http://dx.doi.org/10.1111/j.1523-1739.2007.00686.x

Hortal J *et al.*, 2008. Historical bias in biodiversity inventories affects the observed realized niche of the species. *Oikos*, 117:847-858. http://dx.doi.org/10.1111/j.0030-1299.2008.16434.x

Hortal J *et al.*, 2010. Understanding (insect) species distributions across spatial scales. *Ecography*, 33:51-53. http://dx.doi.org/10.1111/j.1600-0587.2009.06428.x

Hortal J *et al.*, 2011. Ice age climate, evolutionary constraints and diversity patterns of European dung beetles. *Ecology Letters*, 14:741-748. http://dx.doi.org/10.1111/j.1461-0248.2011.01634.x

Jackson ST & Overpeck JT, 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology*, 26:194-220. http://dx.doi.org/10.1666/0094-8373(2000)26[194:ROPPAC]2.0.CO;2

Jiménez-Valverde A, 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, 21:498-507. http://dx.doi.org/10.1111/j.1466-8238.2011.00683.x

Jiménez-Valverde A *et al.*, 2008a. Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, 14:885-890. http://dx.doi.org/10.1111/j.1472-4642.2008.00496.x

Jiménez-Valverde A *et al.*, 2008b. Challenging species distribution models: the case of *Maculinea nausithous* in the Iberian Peninsula. *Annales Zoologici Fennici*, 45:200-210. http://dx.doi.org/10.5735/086.045.0305

Jiménez-Valverde A *et al.*, 2009. Environmental correlation structure and ecological niche model projections. *Biodiversity Informatics*, 6:28-35.

Jiménez-Valverde A *et al.*, 2011a. Use of niche models in invasive species risk assessments. *Biological Invasions*, 13:2785-2797. http://dx.doi.org/10.1007/s10530-011-9963-4

Jiménez-Valverde A *et al.*, 2011b. Environmental suitability of new reported localities of the funnelweb spider Macrothele calpeiana: an assessment using potential distribution modelling with presence-only techniques. *Journal of Biogeography*, 38:1213-1223. http://dx.doi.org/10.1111/j.1365-2699.2010.02465.x

Jiménez-Valverde A *et al.*, 2012. Discrimination capacity in species distribution models depends on the representativeness of the environmental domain. *Global Ecology and Biogeography*. In press. http://dx.doi.org/10.1111/geb.12007

Kamino LHY *et al.*, 2012a. Challenges and perspectives for species distribution modelling in the neotropics. *Biology Letters*, 8:324-326. http://dx.doi.org/10.1098/rsbl.2011.0942

Kamino LHY *et al.*, 2012b. The application of species distribution models in the megadiverse Neotropics poses a renewed set of research questions. *Frontiers of Biogeography*, 4:7-10. http://escholarship.org/uc/item/74v6t43m

Kearney M & Porter W, 2009. Mechanistic niche modeling: combining physiological and spatial data to predict species ranges. *Ecology Letters*, 12:334-350. PMid:19292794. http://dx.doi.org/10.1111/j.1461-0248.2008.01277.x

Klicka J & Zink RM, 1997. The importance of recent ice ages in speciation: a failed paradigm. *Science*, 277:1666-1669. http://dx.doi.org/10.1126/science.277.5332.1666

Krug AZ *et al.*, 2009. Generation of Earth's first-order biodiversity pattern. *Astrobiology*, 9:113-124. PMid:19215204. http://dx.doi.org/10.1089/ast.2008.0253

Legendre P & Legendre L, 1998. *Numerical Ecology*. 2nd ed. Amsterdam: Elsevier.

Lobo JM, 2008. More complex distribution models or more representative data? *Biodiversity Informatics*, 5:14-19.

Lobo JM *et al.*, 2006a. Regional and local influence of grazing activity on the diversity of a semiarid dung beetle community. *Diversity and Distributions*, 12:111-123. http://dx.doi.org/10.1111/j.1366-9516.2006.00194.x

Lobo JM *et al.*, 2006b. Environmental and geographical factors affecting the Iberian distribution of flightless *Jekelius* species (Coleoptera: Geotrupidae). *Diversity and Distributions*, 12:179-188. http://dx.doi.org/10.1111/j.1366-9516.2005.00214.x

Lobo JM *et al.*, 2007. How does the knowledge about the spatial distribution of Iberian dung beetle species accumulate over time? *Diversity and Distributions*, 13:772-780. http://dx.doi.org/10.1111/j.1472-4642.2007.00383.x

Lobo JM *et al.*, 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17:145-151. http://dx.doi.org/10.1111/j.1466-8238.2007.00358.x

Lobo JM *et al.*, 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, 33:103-114. http://dx.doi.org/10.1111/j.1600-0587.2009.06039.x

Lomolino MV, 2004. Conservation Biogeography. In: Lomolino MV & Heaney LR (eds.). *Frontiers of Biogeography: new directions in the geography of nature*. Sunderland: Sinauer Associates, Inc. *p.* 293-296.

Magallón S & Sanderson MJ, 2001. Absolute diversification rates in angiosperm clades. *Evolution*, 55:1762-1780. PMid:11681732.

Martínez-Meyer E *et al.*, 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. http://dx.doi.org/10.1111/j.1466-822X.2004.00107.x

McGlone MS, 1996. When history matters: scale, time, climate and tree diversity. *Global Ecology and Biogeography Letters*, 5:309-314. http://dx.doi.org/10.2307/2997586

McPherson JM & Jetz W, 2007. Effects of species' ecology on the accuracy of distribution models. *Ecography*, 30:135-151. http://dx.doi.org/doi:10.1111/j.2006.0906-7590.04823.x

Mutke J & Barthlott W, 2005. Patterns of vascular plant diversity at continental to global scales. *Biologiske Skrifter*, 55: 521-531.

Nogués-Bravo D *et al.*, 2008. Climate change, humans and the extinction of the woolly mammoth. *PLoS Biology*, 6:e79. http://dx.doi.org/10.1371/journal.pbio.0060079

Odling-Smee JF *et al.*, 2003. *Niche Construction*: The Neglected Process in Evolution. New Jersey: Princeton University Press.

Osborne PE & Suárez-Seoane S, 2002. Should data be partitioned spatially before building large-scale distribution models? *Ecological Modelling*, 157:249-259. http://dx.doi.org/10.1016/S0304-3800(02)00198-9

Pearce J & Ferrier S, 2000. An evaluation of alternative algorithm for fitting species distribution models using logistic regression. *Ecological Modelling*, 128:127-147. http://dx.doi.org/10.1016/S0304-3800(99)00227-6

Peterson AT, 2003. Predicting the geography of species´ invasions via ecological niche modeling. *Quarterly Review of Biology*, 78:419-433. PMid:14737826. http://dx.doi.org/10.1086/378926

Peterson AT, 2006. Uses and requirements of ecological niches models and related distributional models. *Biodiversity Informatics*, 3:59-72.

Peterson AT *et al.*, 2011. *Ecological niches and geographic distributions*. Princeton: Princeton University Press.

Peterson AT & Soberón J, 2012. Species Distribution Modeling and Ecological Niche Modeling: Getting the concepts right. *Natureza & Conservação*, 10(2): 102-107. http://dx.doi.org/10.4322/natcon.2012.019

Platts PJ *et al.*, 2010. Can distribution models help refine inventory-based estimates of conservation priority? A case study in the Eastern Arc forests of Tanzania and Kenya. *Diversity and Distributions*, 16:628-642. http://dx.doi.org/10.1111/j.1472-4642.2010.00668.x

Pulliam HR, 1988. Sources, sinks and population regulation. *American Naturalist*, 132:652-661. http://dx.doi.org/10.1086/284880

Pulliam HR, 2000. On the relationship between niche and distribution. *Ecology Letters*, 3:349-361. http://dx.doi.org/10.1046/j.1461-0248.2000.00143.x

Ricklefs RE, 2004. A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, 7:1-15. http://dx.doi.org/10.1046/j.1461-0248.2003.00554.x

Ricklefs RE, 2006. Evolutionary diversification and the origin of the diversity-environment relationship. *Ecology*, 87:S3-S13. http://dx.doi.org/10.1890/0012-9658(2006)87[3:EDATOO]2.0.CO;2

Ricklefs RE & Schluter D (eds.), 1993. *Species diversity in ecological communities*: historical and geographical perspectives. Chicago: The University of Chicago Press. PMid:17816896.

Rocchini D *et al.*, 2011. Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography*, 35:211-226. http://dx.doi.org/10.1177/0309133311399491

Roura-Pascual N *et al.*, 2004. Geographical potential of Argentine ants (Linepithema humile Mayr) in the face of global climate change. *Proceedings of the Royal Society of London B*, 271:2527-2534. http://dx.doi.org/10.1098/rspb.2004.2898

Segurado P & Araújo MB, 2004. An evaluation of methods for modelling species distributions. J*ournal of Biogeography*, 31:1555-1568. http://dx.doi.org/10.1111/j.1365-2699.2004.01076.x

Shipley B, 2000. *Cause and Correlation in Biology*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511605949

Smart SM *et al.*, 2010. Empirical realised niche models for British higher and lower plants - development and preliminary testing. *Journal of Vegetation Science*, 21:643-656

Soberón J, 2007. Grinnellian and Eltonian niches and geographic distribution of species. *Ecology Letters*, 10:1115-1123 PMid:17850335. http://dx.doi.org/10.1111/j.1461-0248.2007.01107.x

Soberón J & Peterson AT, 2005. Interpretation of models of fundamental ecological niches and species' distribution areas. *Biodiversity Informatics*, 2:1-10

Soberón J & Peterson AT, 2011. Ecological niche shifts and environmental space anisotropy: a cautionary note. *Revista Mexicana de Biodiversidad*, 82:1348-1355.

Soberón J & Nakamura M, 2009. Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences USA*, 106:19644-19650 PMid:19805041 PMCid:2780935. http://dx.doi.org/10.1073/pnas.0901637106

Soberón J, 2010. Niche and area of distribution modeling: A population ecology perspective. *Ecography*, 33:159-167 http://dx.doi.org/10.1111/j.1600-0587.2009.06074.x

Svenning J-C & Skov F, 2004. Limited filling of the potential range in European tree species. *Ecology Letters*, 7:565-573. http://dx.doi.org/10.1111/j.1461-0248.2004.00614.x

Svenning J-C & Skov F, 2007. Ice age legacies in the geographical distribution of tree species richness in Europe. *Global Ecology and Biogeography*, 16:234-245. http://dx.doi.org/doi:10.1111/j.1466-8238.2006.00280.x

Svenning J-C *et al.*, 2011. Applications of species distribution modeling to paleobiology. *Quaternary Science Reviews*, 30:2930-2947. http://dx.doi.org/10.1016/j.quascirev.2011.06.012

Swab RM *et al.*, 2012. Niche models tell half the story: spatial context and life-history traits influence species responses to global change. *Journal of Biogeography*, 39:1266-1277. http://dx.doi.org/10.1111/j.1365-2699.2012.02690.x

Syphard AD & Franklin J, 2010. Species traits affect the performance of species distribution models for plants in southern California. *Journal of Vegetation Science*, 21:177-189. http://dx.doi.org/10.1111/j.1654-1103.2009.01133.x

Varela S *et al.*, 2011. Using species distribution models in palaeobiogeography: A matter of data, predictors and concepts. *Palaeogeography*, *Palaeoclimatology*, *Palaeoecology*, 310:451-463. http://dx.doi.org/10.1016/j.palaeo.2011.07.021

Varela S *et al.*, 2010. Were the Late Pleistocene climatic changes responsible for the disappearance of the European spotted hyena populations? Hindcasting a species geographic distribution across time. *Quaternary Science Reviews*, 29:2027-2035. http://dx.doi.org/10.1016/j.quascirev.2010.04.017

Vaughan IP & Ormerod SJ, 2005. The continuing challenges of testing species distribution models. *Journal of Applied Ecology*, 42:720-730. http://dx.doi.org/10.1111/j.1365-2664.2005.01052.x

Von Humboldt A, 1850. *Views of nature*, *or*, *Contemplations on the sublime phenomena of creation*: with scientific illustrations. Translated from the German by E.C. Otté and Henry G. Bohn; with a frontispiece from a sketch by the author, a fac-simile of his handwriting, and a comprehensive index. London: H.G. Bohn. http://dx.doi.org/10.5962/bhl.title.4802

Wallace AR, 1876. *The geographical distributions of animals*. London: Macmillan. 2 vol.

Waltari E *et al.*, 2007. Locating Pleistocene refugia: Comparing phylogeographic and ecological niche model predictions. *PloS One*, 7: e563. PMid:17622339 PMCid:1905943. http://dx.doi.org/10.1371/journal.pone.0000563

Wiens JA, 1989. Spatial scaling in ecology. *Functional Ecology*, 3:385-397. http://dx.doi.org/10.2307/2389612

Wiens JJ & Donoghue MJ, 2004. Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution*, 19:639-644. PMid:16701326. http://dx.doi.org/10.1016/j.tree.2004.09.011

Wright DH *et al.*, 1993. Energy supply and patterns of species richness on local and regional scales. In: Ricklefs RE & Schluter D (eds.). *Species diversity in ecological communities*: historical and geographical perspectives. Chicago: The University of Chicago Press. *p.* 66-74.

Zurell D *et al.*, 2009. Static species distribution models in dynamically changing systems: how good can predictions really be? *Ecography*, 32:733-744. http://dx.doi.org/10.1111/j.1600-0587.2009.05810.x