Labeling Ecological Niche Models

Thiago Fernando Rangel & Rafael Dias Loyola*

Department of Ecology, Federal University of Goiás – UFG, Goiânia, GO, Brasil

Abstract

The ongoing biodiversity crisis is pushing ecologists and conservation biologists to develop models to foretell the effects of human-induced transformation of natural resources on the distribution of species, although ecology and biogeography still lacks a paradigmatic body of theory to fully understand the drivers of biodiversity patterns. Two decades of research on ecological niche models and species distributions have been characterized by technical development and discussions on a plethora of methods or algorithms to infer and predict species distributions. Here we suggest a metaphorical classification scheme for some of the most popular models based on their complexity, interpretability and suitability for specific applications in ecology and conservation biology. Our purpose is not to compare methods by their capacity to accurately predict the observed distribution of species, nor to criticize how they are commonly used in applied studies. Instead, we believe that a simple classification scheme can potentially highlight how some methods are more suited for specific applications in ecology and conservation biology. Envelope and distance-based models are grouped into the "fish bowl" category, for their transparency and simplicity. Statistical models are classified as "turbine" models, because of their hidden complexity and general applicability. Finally, machine-learning models are classified as "vault" models, for their high complexity and lack of interpretability of fit parameters. We conclude that the diversity of species distribution models used today is expected for a young research field, but the choice of modeling strategy depends on the purpose of the study. We provide some general guidelines for choosing models for studies of conservation planning and climate change mitigation and suggest models of intermediate complexity for conservation planning and forecast of climate change effects on biodiversity as they provide a good balance between interpretability, predictive power and robustness to model over-fit.

Key words: Bioclimatic Envelope Models, Biodiversity, Climate Change, Conservation Planning, MaxEnt, Species Distribution Models.

Introduction

"There is no part of natural history more interesting or instructive, than the study of the geographical distribution of animals." Alfred Russell Wallace (1823-1913)

Since immemorial times, from Wallace to modern scientists, the study of the geographical distribution of species has fascinated humans. Many studies have focused on the evolution, dynamics, and structure of geographic ranges (Brown *et al*. 1996; Gaston 2003; Diniz Filho *et al*. 2010), but, albeit their theoretical relevance, estimating geographic ranges and species distribution is still a challenging issue for ecologists and biogeographers. Modern scientists have been developing statistical and mathematical models to infer and predict geographic distribution of species by coupling data on species occurrences at different spatial scales with environmental (bioclimatic) data (Pearson & Dawson 2003).

According to the ecological niche theory, species are constrained by their tolerance to environmental factors. As so, these models define an environmental space describing the ecological niche of a species (at least its abiotic component, see Soberón 2007), which can be then projected into geographic space to estimate species' geographic distribution (Colwell & Rangel 2009) Models like these are now called "ecological niche models", although given the confusing terminology in the field, they are also referred to as "species distribution models" (Franklin 2009) or "bioclimatic envelope models" (Araújo & Peterson 2012). Araújo & Peterson (2012) and Peterson & Soberón (2012) provide recent clarification on this topic.

The ongoing biodiversity crisis is pushing ecologists and conservation biologists to develop models that are able to foretell the effects of human-induced transformation of natural resources on the distribution of species. Of course, forecasting is a statement about events that have yet to be observed, and require models with reliable level of predictive power, as well as minimum risk and low uncertainty. Ecological niche models (henceforth ENMs) associate known species occurrences across the geographic

Send correspondence to: Rafael Loyola
Department of Ecology, Federal University of Goiás – UFG,
CP 131, CEP 74001-970, Goiânia, GO, Brazil
E-mail: rdiasloyola@gmail.com

space to environmental conditions of the sites where species occurrences have been recorded. The association is used to define locations where viable populations of the species are likely to be maintained (Peterson *et al.* 2011). These models are an important tool for studies in ecology, biogeography, and conservation biology. Still, not all ecological processes that drive species distribution are known and understood by ecologists, and therefore phenomenological correlative models like ENMs are currently the most used and only general resort available. ENMs are now a hot topic, ranking among the most reviewed topics in the ecological literature (Araújo & Peterson 2012), and focus on recently published text books (Franklin 2009; Peterson *et al.* 2011). ENMs have been applied to a variety of problems ranging from biodiversity discovery, prediction of species invasion, to forecasting the effects of climate change on species distribution (Franklin 2009). Though widely used, their usefulness has been criticized by authors claiming that ENMs are built upon implausible assumptions and contradict empirical evidence (Araújo & Guisan 2006; Araújo & Peterson 2012).

The Jungle of Methods for Modeling Species' Ecological Niches and Geographic Distributions

"*Led by a new paradigm, scientists adopt new instruments and look in new places...*" Thomas S. Kuhn (1922-1996)

As expected for a new field, research on ENMs has been characterized by technical development and discussions on the best method or algorithm used for modeling ecological niches and inferring species' distributions. As Araújo & Peterson (2012) pointed out, the field publishes a rapidly increasing number of primary research papers and many syntheses. Still, it is surprising how it lacks conceptual and methodological unification.

Currently, there are several methods for modeling species' ecological niche as a function of environmental variables. As we will discuss later, techniques for generating ENMs range from very simple bioclimatic envelope models up to complex machine learning-based methods (Franklin 2009, Peterson *et al.* 2011). Table 1 summarizes the most common methods used for modeling species' ecological niches, a brief description of them, along with their best type of response and some available software in which they could be generated. These methods have been described and mathematically defined elsewhere (for a recent revision see Franklin 2009). However, independent of the methods used to produce ENMs, uncertainties in model predictions arise from many sources (Araújo & New 2007). For example, Diniz-Filho *et al.* (2009) found that 66% of variation in predictions of how climate change will affect species' distribution is due to uncertainty in model structure, whereas 14% is due to future climate scenario. Of course, such uncertainty in model structure affects ability to extrapolate

predictions in space and time (i.e. model transferability), and adds complexity to measure model fit (Lobo *et al.* 2008).

Ecological niche models, as most correlative models in ecology, are designed to identify and describe statistical patterns in species occurrence data by fitting values to model parameters (Franklin 2009; Peterson *et al.* 2011). The predictive power of model can be measured by some goodness-of-fit statistics, which quantifies the degree of similarity between a pattern estimated by the model using the adjusted parameters and an observed dataset (e.g. the input data). However, the ongoing debate on model performance, statistical fit and transferability indicates that it is currently difficult to determine what is the best method for modeling species' distribution, because their outcome is strongly dependent on data availability and the geographic scale used, as well as the logic and objective when building the models (Elith & Leathwick 2009). To cope with these problems, a combination of different projections built upon different conditions and methods – the ensemble forecasting approach – has been suggested as more conservative than single model analysis (Araújo & New 2007). Ensembles of forecasts should be used when it is impossible, or quite difficult, to determine which type of model should produce the most accurate predictions in a complex set of situations. Predictions from multiple models or from multiple input data sets are usually averaged and weighted by model accuracy, although estimating model accuracy is also challenging. Thus, by combining different model projections the final consensus may benefit from accurate models: although depending on how model predictions are combined, poor model predictions may cancel accurate models (Lawler *et al.* 2011). In the ensemble forecasting approach the final solution is an unique consensus, weighted by the overall statistical fit (e.g. TSS statistics) of combined models – from which is also possible to quantify and map uncertainties associated with ENMs (Diniz-Filho *et al.* 2009). In the next sections will discuss the applications of ENMs in ecology and conservation biology, propose a metaphorical classification for ENMS, and call attention to the advantages and pitfalls of using ENMs for conservation planning.

A Metaphorical Classification of ENMs

"*The metaphor is an origin, the origin of an image which acts directly, immediately.*" Gaston Bachelard (1884-1962)

As discussed above, ENM methods commonly applied in ecology and conservation biology vary widely in their underlying assumptions (hypothesized relationship between species' occurrence and environmental factors) and complexity (number of parameters used to fit the hypothesized relationship). ENMs, as any statistical model, are an attempt to reduce complex reality into a simplified model of the observed phenomena through a set of hypothesized relationships among relevant factors. However, deciding on what should be simplified in the

Table 1. Modeling methods commonly used for modeling ecological niches, their brief description, and indication of best type of response variables and some available software for building models.

Modeling method	Method description	Best response variable	Some available software
Envelope methods			
BIOCLIM	Simple bioclimatic envelope defining the environmental tolerance of a species for multiple predictors	Categorical presence-only	BIOCLIM, OpenModeller
Mahalanobis distance	Distance method that finds similarity to conditions where the species occurs. It does not estimate the importance of predictors	Categorical presence-only	DOMAIN, R, OpenModeller
ENFA	Compares the distribution of the localities where the focal species was observed to a reference set describing the whole study area.	Categorical	BIOMAPPER
Statistical methods			
GLM	Parametric regression methods with several response functions like linear, polynomial, piecewise, and interaction terms. GLM are extensions of linear models that can cope with non-normal distributions of the response variable	Quantitative, categorical, binomial	BIOMOD, Statistica, SYSTAT, GRASP, R
GAM	Non-parametric regression methods with smoothing function, estimated using local regression, splines or other method. GAMs are a non-parametric extension of GLMs	Quantitative, categorical	GRASP, R
MARS	Adaptive piecewise linear regression. MARS are a generalization of stepwise linear regression, also related to regression trees and GAMs	Quantitative, categorical	MARS, R
Machine-learning methods			
GARP	Complex method that combines decision rules using a genetic algorithm. In this method a population of classification rules is generated and then the rules "evolve" by a process analogous to natural selection until an optimal solution is reached	Categorical binomial	Desktop Garp
MaxEnt	Method based on a principle from statistical mechanics and information theory. It states that the most spread out and close to uniform probability distribution, subject to known constraints, is the best approximation for an unknown distribution. In MaxEnt, non-linear response functions can be described	Categorical binomial	MaxEnt
ANN	Method that hierarchically partition the data into subareas, focusing on the informative portions and ignoring the empty parts. ANNs response functions are non-linear decision boundaries in covariate space	Quantitative, categorical	NNETW, R

ENFA = ecological niche factor analysis, GLM = generalized linear models, GAM = generalized additive models, MARS = multivariate adaptive regression splines, GARP = genetic algorithm for rule set production, MaxEnt = maximum entropy, ANN = artificial neutral networks.

model structure depends not only on the reality to be described, but also on the purpose of the model and the state of science.

No model design is free of a trade-off between three desirable features – generality, realism and precision (Levins 1966) – and it is the final balance between these features that define the most proper application for a given model. Thus, it is not the choice of ENMs or dataset that draws the boundary between the inference and prediction of the distribution of species, but how the results of model are used and interpreted.

Scientific inference involves generalization about an ecological pattern or process, leaping from a known truth (observed data) to a logical conclusion that is presumed to be true with a degree of certainty (model structure). Usually inference is associated with an attempt to understand how a system behaves under certain conditions, and what are the processes involved. On the other hand,

prediction uses observed facts (empirical data) and scientific reasoning (model structure) to foretell a phenomenon under unobserved conditions. Predictions drawn for conditions that are similar or within the range of those that were observed are called interpolation, whereas these predictions will be called extrapolations if they are drawn for conditions that were never observed by the scientist.

We illustrate our argument by suggesting a classification scheme for some of the most popular ENMs models (see Table 2), based on their complexity, interpretability and suitability for specific applications in ecology and conservation biology. Our purpose is not to compare ENM methods by their ability to accurately predict the observed distribution of species, nor to criticize how they are

Table 2. Metaphorical classification of the most used ecological niche models and some general characteristics of models within each class.

Metaphor	"Fish bowl" models	"Turbine" models	"Vault" models
Common classification	Envelope models, distance models, classification models	Statistical models	Machine-learning models
Example of models under this class	BIOCLIM, Euclidian distance, Gower distance, Mahalanobis distance, Ecological Niche Factor Analysis (ENFA)	Generalized Linear Models (GLM), Generalized Additive Models (GAM), Multivariate Adaptive Regression Splines (MARS)	Random Forests, Genetic Algorithm for Rule Set Production (GARP), Maximum Entropy (MaxEnt), Artificial Neural Networks (ANN), Generalized Boosting Regression Models (GBM), Support Vector Machines (SVM)
Explanation for the metaphor	Geographic distributions of species are constrained by their tolerance to environmental factors; fish cannot survive outside the fish bowl, which is transparent and simple	Models work like turbines: data come in, predictions go out. Leverages under the control are used to regulate the amount of energy used to move the turbine, and therefore its rotation speed; parameters could be regulated to improve model predictions	As in vaults, the inner contents of these models are usually hidden from the scientist. Even if scientists were allowed inside the vault, it would take a great effort to make sense of the organization of all items
Complexity	Low	Intermediate	High
Transparency	High	From intermediate to low	Low
Strengths	Easy to understand, easy interpretation of parameters, intuitive, high generality, deeply rooted in ecological and physiological theory	Relatively easy to understand, predictive power higher than in fish bowl models, high precision, somewhat rooted in ecological theory	High statistical fit, high precision
Weaknesses	Low statistical fit, low precision and realism	Require addition scientific effort to extract information and interpret the ecological meaning of model parameter, not necessarily rooted in ecological theory, realism and generality often sacrificed	Difficult to understand, usually no constraint to limit the maximum number of parameters, over fitting, low generality, predictions may be highly contingent on the geographic position of input data
Best-practice application	More appropriate in studies aiming to understand and infer the form and direction of relationships between species occurrence and environmental conditions; good to infer species potential distribution	Best results will take place if species occurrence respond linearly to changes in environmental factors or incorporate the consequences of the interaction between factors for the distribution of species	Effective for predicting the distribution of species after appropriate validation of their predictive power by robust statistical methods; work well especially for describing actual (not potential) geographic distributions when all or almost all populations are known
Precaution with application in conservation planning	High; tend to overestimate species occurrences increasing commission errors	Intermediate; could have high model fit, but low biological meaning	From intermediate to high; should follow a good reasoning for their application given difficulties in parameter interpretation; tend to increase omission errors given their dependence on the quality of input data.

commonly used in specific studies. Instead, we believe that this simple classification scheme can potentially highlight how some methods are more suited for applications in ecology and conservation biology.

Fish bowl models

It is arguable that the simplest ENMs are those known as "envelope models", a category that includes, for example, BIOCLIM (Busby 1991), Euclidian and Gower distances (Carpenter *et al.* 1993), and Ecological-Niche Factor Analysis (ENFA) (Hirzel *et al.* 2002). These models assume that the geographic distributions of species are constrained by their tolerance to environmental factors. Such assumption is deeply rooted in several ecological theories, dating back to Gauss, Grinnell and Hutchinson. Not only these models rely on assumptions that are intuitive to any ecologist, but they also fit parameters with clear-cut interpretation. Fish bowl models are transparent – any ecologist can clearly see the ecological concepts behind the model, and easily interpret the parameters fit by the model. However, the substantial advantage in inference and interpretation power made possible by simplistic assumptions behind these models usually lead to a sacrifice in predictive power, especially when compared to models in the other two categories. Thus, fish bowl models tend to sacrifice precision and realism for the sake of generality. They are especially appropriate in studies that aim to understand and infer the form and direction of relationships between species occurrence and environmental conditions (see Table 2).

Turbine models

Regression-based methods, such as Generalized Linear Models (GLM) (Guisan *et al.* 2002), Generalized Additive Mode (GAM) since t (Hastie & Tibshirani 1986) and Multivariate Adaptive Regression Splines (MARS) (Friedman 1991) are more complex than "envelope models", both in terms of number of parameters they fit and the accepted relationship between species occurrence and environmental factors. Notice that within this category models may range widely in complexity (e.g. MARS is substantially more complex than GLM). Not surprisingly, more complex models usually have better explanatory power. It is arguable that these models are also somewhat rooted in ecological theory, especially those that assume that species occurrence respond linearly to change in environmental factors, or incorporate the consequences of the interaction between factors for the distribution of species. Thus, these models fit parameters that are potentially very useful to the understanding of species distribution, as they may be able to capture valuable ecological and evolutionary information about species (Franklin 2009). However, as opposed to the fish bowl models, the turbine models require addition scientific effort to extract information and interpret the ecological meaning

of model parameter. Thus, realism, and to some extent generality, are sacrificed for the sake of precision (Table 2).

We suggest the turbine analogy because these models usually work on their "sealed" state: data come in, predictions go out. Leverages under the control of the scientist are used to regulate the amount of energy used to move the turbine, and therefore its rotation speed. If the turbine is disassembled into small pieces, and its mechanisms are carefully studied, it is easier to understand that blades control fluid flow inside the turbine. Of course, fluid dynamics may not be always trivial, especially when the interaction between fluid and blades creates turbulence. Still, the rotation of the shaft creates useful work, which can be used to propel knowledge about the distribution of species.

Vault models

The most complex models are those that employ machine-learning algorithms to search (mine) for the parameters values that maximize the relationship between observed species occurrence and environmental since t (Franklin 2009). The most popular models that employ such technique are Genetic Algorithm for Rule Set Production (GARP) (Stockwell & Noble 1992), Random Forests (Breiman 2001), Artificial Neural Networks (ANN) (Manel *et al.* 1999), Generalized Boosting Regression Modelsince t (Friedman 2001), and Maximum Entropy (MaxEnt) (Phillips *et al.* 2006). The searching procedure involves maximizing model predictive power while minimizing the number of parameters, but there is usually no constraint to limit the maximum number of parameters or complexity of the relationship between species occurrence and environmental factors. Of course, when contrasted against models under the other categories, these models usually provide "predictions" that best fit the original data (Elith *et al.* 2006), but the complexity of the model prevents any interpretation of the parameters (Jiménez-Valverde *et al.* 2008)

Vault models tend to sacrifice realism and generality for the sake of precision. These models may be suited for predicting the distribution of species after appropriate validation of their predictive power by robust statistical methods (Peterson *et al.* 2011). However, over-fitting these models is a real concern, especially when they are used for extrapolations (Table 2).

We suggest the "vault" analogy because the inner contents of these models are usually hidden from the scientist. In addition, even if these scientists were allowed inside the vault, it would take a great effort to make sense of the organization of all items. In fact, very few personnel are allowed within the vault, while the scientist only interacts with the front-desk. When new items come in they are processed and recorded by automated algorithms, which decide the most economical way to store and secure them inside the vault. In addition, the same algorithm is required to find the exact location of each item, or to decipher the meaning of the label of

any given item inside the vault. However, it is guaranteed that if the vault is maintained under careful organization and constantly fed with additional items, it may eventually store enough resources to the scientist.

The Metaphors and their Consequences for Application of ENMs in Conservation Planning

"Go wisely and slowly. Those who rush stumble and fall." Friar Laurance, William Shakespeare's Romeo and Juliet, Act 2, Scene 3

Conservation scientists, practitioners and resource managers often work with limited data to answer critical questions, such as the allocation of scarce resources for conservation (Moilanen 2012). Species geographic extent is the raw material of conservation planning and science-based management (Lawler *et al*. 2011). It is also indispensable for assessing the risks of climate and land-use changes, and other human activities to natural populations. However, in many regions of the world, even this basic information is lacking (Whittaker *et al*. 2005; Lawler *et al*. 2011).

Ecological niche models have been applied in conservation biology mainly for three purposes: (1) discovering biodiversity, i.e. the use of ENMs for detecting new species or to increase effectiveness in sampling populations of a target species, (2) conservation planning and climate change effects, i.e. the use of ENMs for forecasting species distribution and evaluation of adaptation strategies, and (3) studying species' invasions, i.e. using ENMs to forecast invasions of introduced species (Peterson *et al*. 2011). Here we will focus on the limitations to the use of ENMs in conservation planning and climate change mitigation.

Conservation planning occurs at broad spatial extents and detailed data on species distribution are usually not available. As a shortcut, many studies use data on the extent of species' occurrences, typically depicted from taxonomic atlases or digital range maps. These maps, however, have a coarse resolution and tend to overestimate species occurrences (Lemes *et al*. 2011; Rodrigues 2011; Rondinini *et al*. 2011). To overcome these limitations the ENMs have been used to provide finer resolution estimates of species occurrences that feed spatial planning analyses (Lawler *et al*. 2011).

In this case, we must pay special attention to ENMs commission and omission errors (Table 1). If a site is indicated as priority because it complements the current established network of protected areas, although model predictions are wrong, we may face a serious problem of misallocating valuable conservation resources. What is the worst-case scenario then? Well, that depends on the specific situation (Lawler *et al*. 2011). For example, consider a given endangered species that occurs only at five localities, but only two sites will be set aside for protecting that species. If ENMs predicted the occurrence of the species at fifty sites it is likely that the two sites out of fifty will not protect,

therefore jeopardizing the efforts to protect the endangered species. In this case, commission errors favor resource misallocation. On the other hand, in the face of unlimited funds to protect a given species, omission errors may result in a subset of the actual habitat being protected (Lawler *et al*. 2011; Lemes *et al*. 2011). When predicting climate change effects on species distributions, commission errors lead to overestimation of range expansions whereas omission errors produce overestimates of range contractions. If one is employing such models to predict regions of climatic stability (Garcia *et al*. 2012; Loyola *et al*. 2012; Terribile *et al*. 2012), and to guide conservation actions, both omission and commission errors are of particular interest, as these errors are likely to produce huge bias in the results of gap-analysis as well (Marini *et al*. 2009).

Fish bowl models tend to inflate commission errors and individual species distribution, which are usually larger than predicted by other models. Turbine models provide intermediate predictions of range contraction/expansion, whereas vault models usually result in smaller distributions when compared to other models (Faleiro *et al*. 2013). Thus, if conservation practitioners want to reduce commission errors, vault models are perhaps the best option for the task, although in this case one will probably face the inconvenience of lack of clarity and difficulty of explaining results to stake holders and decision makers involved in the conservation process (Table 2).

How about the use of ensemble of forecasts? Arguably, the use of different ENMs to project species distributions may produce very distinct results increasing the uncertainties among the predictions and their applicability to conservation planning (Pearson *et al*. 2006; Araújo & New 2007). Diniz-Filho *et al*. (2009) were the first to provide a robust methodology to partition and map variation among ensemble of forecasts. Additional authors have also highlighted the need to measure and explicitly include ENMs uncertainty in conservation plans to increase its accountability and effectiveness (Diniz-Filho *et al*. 2009; Thuiller *et al*. 2009; Faleiro *et al*. 2013). It has been recently suggested that the use of ensemble of forecasts is preferred as oppose to model species distribution based on only one ENM (e.g. MaxEnt) (Loyola *et al*. 2012). Ensembles of model projections keep only the consensus-projected areas, minimizing variation among all projections (Araújo & New 2007). This is especially important for conservation purposes given that model uncertainty may mislead conservation efforts ending up being cost-ineffective.

Concluding Remarks

The famous quote on model building settles that "[…] essentially, all models are wrong, but some are useful [...]" (Box & Draper 1987). Thus, on the one hand, defining a specific model application is essential for any fair attempt of model comparison, whereas, on the other hand, a given model may be more useful and appropriate for specific

applications. For conservation planning and forecast of climate change effects on biodiversity we suggest models of intermediate complexity, which provides a good balance between interpretability, predictive power and robustness to model over-fit. Further, future scientific efforts should focus on the development of flexible and accurate mechanistic models of species distribution (Buckely *et al.* 2010). In any case, because ENMs vary widely in complexity and interpretability, a general classification scheme as proposed here is useful to highlight the differences in suitability of these models for inference of ecological processes, predictive power of species distribution, and trustful applications for conservation planning.

Acknowledgements

References

Araújo MB & Guisan A, 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33()(10):1677-1688. http://dx.doi.org/10.1111/j.1365-2699.2006.01584.x

Araújo MB & New M, 2007. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, 22(1):42-47. http://dx.doi.org/10.1016/j.tree.2006.09.010

Araújo MB & Peterson AT, 2012. Uses and misuses of bioclimatic envelope modeling. *Ecology*, 93(7):1527-1539. http://dx.doi.org/10.1890/11-1930.1

Box GEP & Draper NR, 1987. *Empirical model-building and response surfaces.* John Wiley & Sons.

Buckely LB *et al.*, 2010. Can mechanism inform species' distribution models. *Ecology Letters*, 13:1041-1054. http://dx.doi.org/10.3410/f.6630956.6777054

Breiman L, 2001. Random Forests. *Machine Learning*, 45(1):5-32. http://dx.doi.org/10.1023/A:1010933404324

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

Busby JR, 1991. BIOCLIM - a bioclimatic analysis and prediction system. In: Margules CR & Austin MP. *Nature Conservation*: cost effective biological surveys and data analysis. Intl Specialized Book Service Inc. *p.* 64-68.

Carpenter G, Gillison AN & Winter J, 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, 2(6):667-680. http://dx.doi.org/10.1007/BF00051966

Colwell RK & Rangel TF, 2009. Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences of the United States of America*, 106(Suppl. 2):19651-19658. http://dx.doi.org/10.1073/pnas.0901650106

Diniz Filho JAF *et al.*, 2010. The three phases of the ensemble forecasting of niche models: geographic range and shifts in climatically suitable areas of *Utetheisa ornatrix* (Lepidoptera, Arctiidae). *Revista Brasileira de Entomologia*, 54(3):339-349. http://dx.doi.org/10.1590/S0085-56262010000300001

Diniz-Filho JAF *et al.*, 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography*, 32(6):897-906. http://dx.doi.org/10.1111/j.1600-0587.2009.06196.x

Elith J *et al.*, 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2):129-151. http://dx.doi.org/10.1111/j.2006.0906-7590.04596.x

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

Faleiro FV *et al.*, 2013. Defining spatial conservation priorities in the face of land-use and climate change. *Biological Conservation*, 158:248-257.

Franklin J, 2009. *Mapping species distributions*: spatial inference and prediction. Cambridge: Cambridge University Press. p. 340.

Friedman JH, 1991. Multivariate adaptive regression splines. *Annals of Statistics*, 19(1):1-141. http://dx.doi.org/10.1214/aos/1176347963

Friedman JH, 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5):1189-1232. http://dx.doi.org/10.1214/aos/1013203451

Garcia RA *et al.*, 2012. Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Global Change Biology*, 18(4):1253-1269. http://dx.doi.org/10.1111/j.1365-2486.2011.02605.x

Gaston KJ, 2003. *The structure and dynamics of geographic ranges.* Oxford University Press.

Guisan A *et al.*, 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157(2-3):89-100. http://dx.doi.org/10.1016/S0304-3800(02)00204-1

Hastie T & Tibshirani R, 1986. Generalized Additive Models. *Statistical Science*, 1(3):297-310. http://dx.doi.org/10.1214/ss/1177013604

Hirzel AH *et al.*, 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, 83(7):2027-2036. http://dx.doi.org/10.1890/0012-9658(2002)083[2027:ENFAHT]2.0.CO;2

Jiménez-Valverde A *et al.*, 2008. Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, 14(6):885-890. http://dx.doi.org/10.1111/j.1472-4642.2008.00496.x

Lawler JJ *et al.*, 2011. Using species distribution models for conservation planning and ecological forecasting. In: Drew Yolanda A et al. (eds.). *Predictive Modeling in Landscape Ecology*. New York: Springer. p. 271-290. http://dx.doi.org/10.1007/978-1-4419-7390-0_14

Lemes P *et al.*, 2011. Refinando Dados Espaciais para a Conservação da Biodiversidade. *Natureza & Conservação*, 9(2):240-243. http://dx.doi.org/10.4322/natcon.2011.032

Levins R, 1966. The strategy of model building in population biology. *American Scientist*, 54:421-431.

Lobo JM *et al.*, 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17(2):145-151. http://dx.doi.org/10.1111/j.1466-8238.2007.00358.x

Loyola RD *et al.*, 2012. Severe loss of suitable climatic conditions for marsupial species in Brazil: challenges and opportunities for conservation. *PLoS ONE*, 7:e46257. http://dx.doi.org/10.1371/journal.pone.0046257

Manel S *et al.*, 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, 36(5):734-747. http://dx.doi.org/10.1046/j.1365-2664.1999.00440.x

Marini MA *et al.*, 2009. Major current and future gaps of Brazilian reserves to protect Neotropical savanna birds. *Biological Conservation*, 142(12):3039-3050. http://dx.doi.org/10.1016/j.biocon.2009.08.002

Moilanen A, 2012. Spatial Conservation Prioritization in Data-Poor Areas of the World. *Natureza & Conservação*, 10(1):12-19. http://dx.doi.org/10.4322/natcon.2012.003

Pearson RG & Dawson TP, 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, 12(5):361-371. http://dx.doi.org/10.1046/j.1466-822X.2003.00042.x

Pearson RG *et al.*, 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography*, 33(10):1704-1711. http://dx.doi.org/10.1111/j.1365-2699.2006.01460.x

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

Peterson AT *et al.*, 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press. v. 2011, *p.* 336.

Phillips S *et al.*, 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4):231-259. http://dx.doi.org/10.1016/j.ecolmodel.2005.03.026

Rodrigues ASL, 2011. Improving coarse species distribution data for conservation planning in biodiversity-rich, data-poor, regions: no easy shortcuts. *Animal Conservation*, 14(3):1-3. http://dx.doi.org/10.1111/j.1469-1795.2011.00451.x

Rondinini C *et al.*, 2011. Global habitat suitability models of terrestrial mammals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 366(1578):2633-41. http://dx.doi.org/10.1098/rstb.2011.0113

Soberón J, 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology letters*, 10(12):1115-23. http://dx.doi.org/10.1111/j.1461-0248.2007.01107.x

Stockwell DRB & Noble IR, 1992. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation*, 33(5-6):385-390. http://dx.doi.org/10.1016/0378-4754(92)90126-2

Terribile LC *et al.*, 2012. Areas of Climate Stability of Species Ranges in the Brazilian Cerrado: Disentangling Uncertainties Through Time. *Natureza & Conservação*, 10(2): 152-159. http://dx.doi.org/10.4322/natcon.2012.025

Thuiller W *et al.*, 2009. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, 32(3):369-373. http://dx.doi.org/10.1111/j.1600-0587.2008.05742.x

Whittaker RJ *et al.*, 2005. Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, 11(1):3-23. http://dx.doi.org/10.1111/j.1366-9516.2005.00143.x