

# The three phases of the ensemble forecasting of niche models: geographic range and shifts in climatically suitable areas of *Utetheisa ornatix* (Lepidoptera, Arctiidae)

José Alexandre Felizola Diniz Filho<sup>1</sup>, Viviane G. Ferro<sup>1</sup>, Thiago Santos<sup>2</sup>, João Carlos Nabout<sup>3</sup>, Ricardo Dobrovolski<sup>2</sup> & Paulo De Marco Jr.<sup>1</sup>

<sup>1</sup>Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Caixa Postal 131, 74001-970 Goiânia-GO, Brazil

<sup>2</sup>Programa de Pós-Graduação em Ecologia & Evolução, ICB, UFG

<sup>3</sup>Programa PNPD/CNPq

---

**ABSTRACT.** The three phases of the ensemble forecasting of niche models: geographic range and shifts in climatically suitable areas of *Utetheisa ornatix* (Lepidoptera, Arctiidae). Species' geographic ranges are usually considered as basic units in macroecology and biogeography, yet it is still difficult to measure them accurately for many reasons. About 20 years ago, researchers started using local data on species' occurrences to estimate broad scale ranges, thereby establishing the niche modeling approach. However, there are still many problems in model evaluation and application, and one of the solutions is to find a consensus solution among models derived from different mathematical and statistical models for niche modeling, climatic projections and variable combination, all of which are sources of uncertainty during niche modeling. In this paper, we discuss this approach of ensemble forecasting and propose that it can be divided into three phases with increasing levels of complexity. Phase I is the simple combination of maps to achieve a consensual and hopefully conservative solution. In Phase II, differences among the maps used are described by multivariate analyses, and Phase III consists of the quantitative evaluation of the relative magnitude of uncertainties from different sources and their mapping. To illustrate these developments, we analyzed the occurrence data of the tiger moth, *Utetheisa ornatix* (Lepidoptera, Arctiidae), a Neotropical moth species, and modeled its geographic range in current and future climates.

**KEYWORDS.** Arctiidae; climate change; geographic range; niche models; uncertainty.

**RESUMO.** As três fases da projeção combinada de modelos de nicho: distribuição geográfica e deslocamento de áreas climaticamente adequadas para *Utetheisa ornatix* (Lepidoptera, Arctiidae). A distribuição geográfica das espécies tem sido considerada como a unidade básica em macroecologia e biogeografia, mas ainda há dificuldades em mensurá-la de forma adequada, por diferentes razões. Há cerca de 20 anos atrás, os pesquisadores começaram a utilizar dados locais da ocorrência das espécies para estimar essas distribuições utilizando modelos de nicho ecológico. Entretanto, ainda há uma série de problemas na avaliação dos modelos e em suas aplicações, e uma das soluções é utilizar um consenso de diferentes modelos, projeções climáticas, cenários de emissão e combinação de variáveis, que são fontes de incerteza durante o processo de modelagem de nicho. Neste artigo nós discutimos essa abordagem de consenso e a dividimos em três fases, com níveis crescentes de complexidade. A Fase I é simplesmente a combinação de mapas e a obtenção e interpretação de um único mapa de consenso. A Fase II envolve a descrição das diferenças entre os mapas utilizando técnicas de análise multidimensional, enquanto que a Fase III consiste em analisar quantitativamente e mapear a magnitude relativa das diferentes fontes de incerteza. A fim de ilustrar essa abordagem, nós analisamos dados de ocorrência de *Utetheisa ornatix* (Lepidoptera, Arctiidae), uma mariposa distribuída na região Neotropical, modelando sua distribuição geográfica com base em dados climáticos atuais e projeções de mudança climática.

**PALAVRAS-CHAVE.** Arctiidae; distribuição geográfica; incerteza; Arctiidae; modelagem de nicho; mudanças climáticas.

---

Species' geographic ranges are usually considered as basic units in macroecology and biogeography, and a large number of studies over the last 20 years have focused on the geographic ranges' properties, including the geographic range size, internal structure, position and shape, as well as the ecological and evolutionary processes at their borders. The analysis of species' ranges overlap (i.e., richness), is also important because it determines the regional component of community structure and, when analyzed at broad geographic scales, usually generates latitudinal gradients (see Hawkins *et al.* 2003; Mittelbach *et al.* 2007 for a review of ecological and evolutionary processes driving range overlap). More recently, understanding patterns and processes associated with species' ranges and range limits has also become important in improving conservation plans because better knowledge

allows the adoption of strategies to increase population viability and even to predict where undiscovered species can be found (avoiding the "Wallacean shortfall", *e.g.*, Bini *et al.* 2006).

Despite the important theoretical and applied issues related to species' ranges, it is still difficult to measure them accurately for many reasons. First, the definition of a species range may be scale-dependent, and in many macroecological and biogeographical studies, researchers usually define species' ranges based on their extent of occurrence, which is a broad definition of the "external" limits of the range. The detailed structures within this extent are not known and are usually not considered as important in establishing patterns at broad scales (*e.g.*, latitudinal gradients). However, this is usually insufficient for more refined analyses, such as those

required for establishing effective conservation planning at local scales. In a more practical aspect, detailed data for species' occurrences, including both determining where the species is present and absent, are lacking for most of the world, especially for tropical, mega-diverse regions. A compromise between these two worldviews is not easy to achieve (Diniz-Filho *et al.* 2010a).

Approximately 20 years ago, some researchers started to use local data on species occurrences (and eventually their absences) to estimate broad scale ranges based on a modeling approach (see Pearson & Dawson 2003). The idea is to use occurrences to establish what is typically called the "bioclimatic envelope", which reflects, in part, the species' niche (or at least one component of the niche, the Grinnelian niche – *sensu* Soberon 2007). Once the Grinnelian niche is established, and considering that the climatic data used to define the envelope have a spatial component, it is possible to use this "duality" (Colwell & Rangel 2009) to project the niche into geographical space and thereby obtain an overall picture of the species' range(s). Of course, the projection of a niche model into geographical space will produce a generalization, and spatially explicit models directly related to the species' range(s) (i.e., dispersal and geographically structured components of population dynamics – see DeMarco *et al.* 2008) would not be considered at first.

This general approach actually created a new research field, and niche models are now widely used to reconstruct species' geographic ranges with different purposes (Guisan & Thuiller 2005; see Escalante *et al.* 2009 for a recent application in insect biogeography), although of course many improvements are still necessary (e.g., Araújo & Guisan 2006). Here we review the main aspects of this research field, which we call the niche modeling of species distribution. Many other reviews on this subject are already written, focusing on theoretical and conceptual aspects (Pearson & Dawson 2003; Soberon 2007; Pearman *et al.* 2008; Colwell & Rangel 2009) as well as methodological developments (e.g., Segurado & Araújo 2004; Elith *et al.* 2006; Elith & Graham 2009). We will focus on the new idea of ensemble forecasts (i.e., combining multiple models for better predictions) (Araújo & New 2007) and the evaluation of uncertainties during the process of generating the species niche (Diniz-Filho *et al.* 2009a). To illustrate these developments, we will analyze the occurrence data for the tiger moth *Utetheisa ornatrix* (Linnaeus, 1758) (Lepidoptera, Arctiidae). The larvae of this Neotropical moth feed mainly on non-ripened pods and developing seeds of Fabaceae shrubs of the genus *Crotalaria* (Ferro *et al.* 2006; Signoretti *et al.* 2008). The larvae sequester pyrrolizidine alkaloids from their host plants and use them as a defense mechanism and the precursors of sexual pheromones. This insect occurs in natural and disturbed areas throughout South and Central America and a portion of the United States (Pease 1968; see also Ferro & Teston 2009).

#### Niche Modeling and the Geographic Range of Species

Currently there are many methods available to reconstruct the Grinnelian component of species' niches, usually based on climatic data. Projecting the outcome of the niche models

into geographic space enables reconstruction of the potential species' range(s). These methods are all based on the overall idea of correlating the presence (and sometimes the absence) of the species in certain localities for which climatic or environmental data are available and building a "bioclimatic envelope." However, the various methods use these data in very different forms, ranging from very simple models that use the maximum and minimum values, or ellipsoidal fits, of the environmental data in localities where the species are found (i.e., the BIOCLIM and distance methods), to several more complex models based on optimization algorithms and artificial intelligence (i.e., maximum entropy, artificial neural networks and others) that search for irregular shapes in hyperspace defined by the presence of records for environmental variables. These more complex methods produce non-linear functions of the relationships between species' occurrences and the environment (Elith *et al.* 2006; see also Table I). All these different methods arose gradually over time, and their applications in the literature follow clear patterns. Some of the methods are in fact becoming very popular (i.e., MAXENT), whereas the applications of others are decreasing.

We used ten of these methods, described in Table I, and implemented in different software, to model the niche of *U. ornatrix* based on 87 records obtained from specimens deposited in the ten most important Brazilian entomological collections and from the literature (Biezanko & Freitas 1938; Biezanko & Seta 1939; Zikan & Zikan 1968; Ferreira *et al.* 1995; Marinoni & Dutra 1996; Teston & Corseuil 2004; Teston *et al.* 2006; Ferro & Teston 2009) (Fig. 1). Initially, we modeled these occurrences as a function of nine environmental variables (i.e., maximum temperature in the hottest month, minimum temperature in the coldest month, annual temperature variation, annual precipitation, precipitation for the driest month, precipitation for the wettest month, precipitation standard deviation, average maximum temperature, and average minimum temperature) and projected them into geographic space. Some examples of the maps using different types of models (optimization and distances) and algorithms produced using different models (MAXENT, Neural Networks, Random Forests and Mahalanobis distances) show that the predictions vary slightly, although, in general, they clearly show the distribution of *U. ornatrix* through the Neotropics, concentrated on the southeastern coast of Brazil (Atlantic forest) and parts of Central America (Fig. 2). Mahalanobis distances map shows high suitability throughout the region, despite a patch of high suitability in the southeastern South America, the region of highest suitability according to the other models.

One of the main current problems in the literature is how to evaluate the "quality" of these models, and several statistics are now available. Niche models usually generate a continuous prediction that tends to express the "centrality" of an environmental condition with respect to species' niche centers in multidimensional space. Thus, when this model is projected into geographical space, a continuous map appears. It is expected that localities with environmental conditions



Fig. 1. Geographical distribution of the 87 occurrences used for niche modeling of *Utheteisa ornatrix*.

close to the species range center in environmental space will have high “suitability,” indicating “Climatically Suitable Areas” (CSA) for the species (e.g., Diniz-Filho *et al.* 2010b). In practice, it is usually necessary to truncate these continuous predictions to establish where the species occurs and its range (although the range limits are, on a small scale, dynamic and not well defined by a specific boundary). Different criteria may also be used to establish this truncation threshold.

In general, the statistics for model evaluation are computed based on a comparison between the observed and expected values, obtained when the continuous distribution is truncated to generate a discrete prediction of the range. The idea is to build a confusion matrix that records the presence and absence of the species (or computer-generated absence in some cases) and the corresponding predictions of the model. The simplest metric, for example, is the omission rate, or false negative. This is the proportion of occurrences that are not correctly predicted by the model (or its counterpart, the commission error, which is the proportion of absences). Many statistics can be built using the confusion matrix (see Allouche *et al.* 2006), including the Receiving Operator Curve (ROC), which plots the “true positives” rate (sensitivity) against the “false positives” rate (1 - true negative rate or 1 - specificity) calculated for multiple confusion matrices based on increasing truncation thresholds. The area under the ROC curve, called

the AUC statistics, is one of the most popular metrics, despite some criticisms (see Lobo *et al.* 2008; Peterson *et al.* 2008).

As an illustration, we present the AUC values for the ten models built for *U. ornatrix*, which indeed show large variation (Table II), although all values are above the expected value of 0.5 under the null hypothesis of absence of relationship between environmental variables and occurrences. Values for AUC that are close to one express the maximum fit (in principle, a very good model), whereas values closer to 0.5 (half of the area under the ROC curve) indicate a prediction that is no better than a random prediction. In our models, AUC values range from 0.859 (for CTA) to 0.937 (for RF). Thus, all of the models have a very good fit, and one could simply choose the niche model with higher AUC values to project the niche into space and to build a geographic range map. However, there are problems with this reasoning that arise from many statistical and theoretical aspects (see Lobo *et al.* 2008). One of the most important criticisms to this reasoning is that using a complex empirical function, despite generating a good fit, does not necessarily produce a good model of the species’ range.

Underlying this main criticism is the assumption that species are in “equilibrium” with their environments (see Araújo & Pearson 2005). When sampling species’ occurrences, one assumes that those (and only those) are the

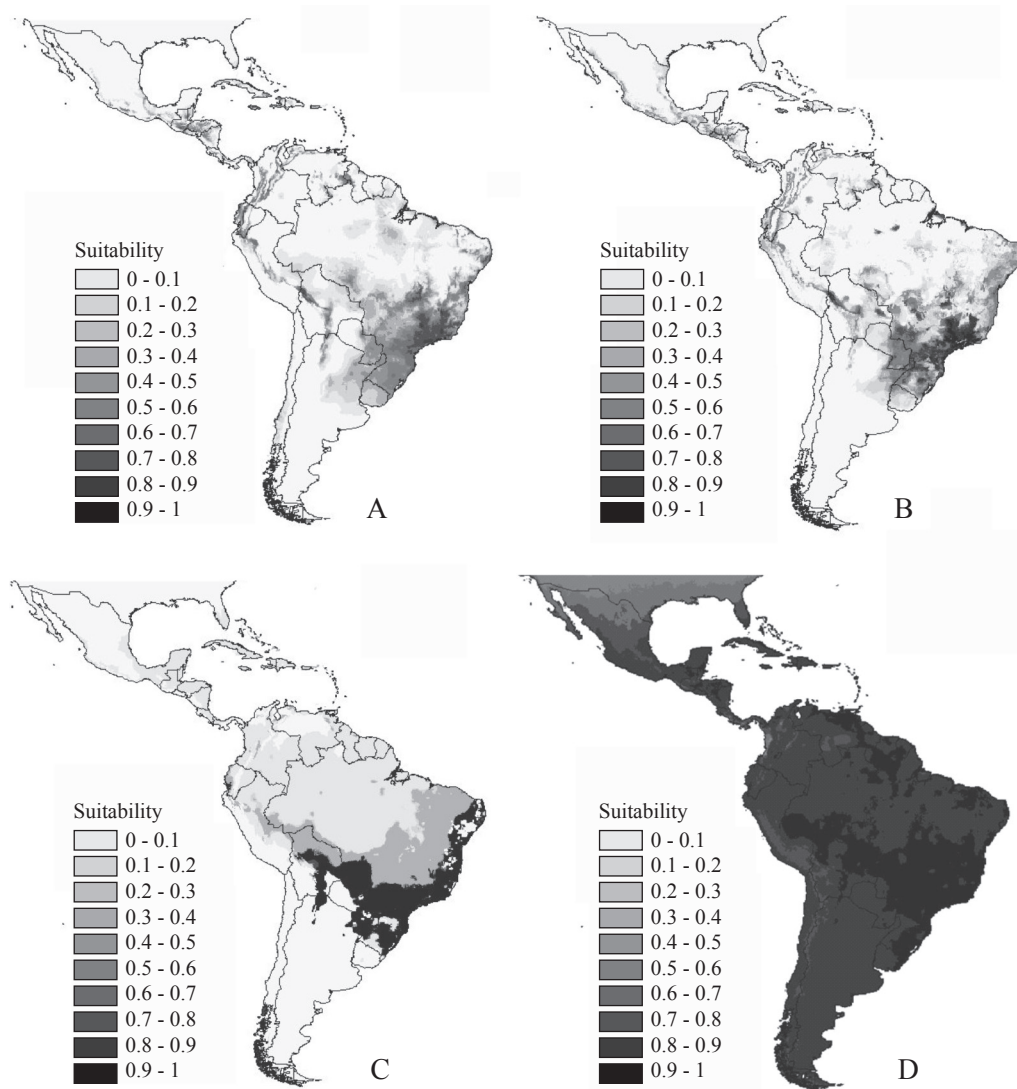


Fig. 2. Geographic ranges of *Uthetseia ornatrix* obtained by spatial projections of niche models, built with (A) Maximum Entropy; (B) Random Forests; (C) Artificial Neural Networks; (D) Mahalanobis distances.

environmental conditions in which the species can be found, as expected if the species has colonized all of the possible suitable habitats. However, many processes can disturb this equilibrium and, thereby, generate an underestimation of the environmentally suitable space. For example, historical effects can constrain the colonization of other habitats (see Nogués-Bravo 2009), or spatially irregular or low sampling efforts can bias occurrences towards particular sub-regions of the suitable space (e.g., Hortal *et al.* 2008). The consequence of violating the equilibrium assumption is that a model with high fit does not necessarily have a high “transferability,” which is a critical issue if one is interested not only in determining species’ ranges but also in investigating how the niche model can be used to predict species’ occurrences in other regions (i.e., after a biological invasion) or in future environments (i.e., responses to climate change) (see Peterson *et al.* 2007 for a discussion).

Another problem is that adding more parameters (i.e., increasing the number of environmental variables or producing complex truncations or polynomial expansions of them) to a model will increase the fit but will also produce an accurate model of any of the errors involved in the data. Without clear theoretical ecological reasoning—which is unfortunately lacking in most cases—hyper-complex models, despite possessing a high fit, will tend to produce bad predictions (Randin *et al.* 2006; see also Elith *et al.* 2006; Peterson *et al.* 2007 for a discussion).

Thus, if statistics (at least some of them) fail to indicate a good model, the main issue becomes how to select a model. Most users simply choose a model according to computational facility and availability. Others prefer simple methods that can produce overall descriptions of the niche, such as MAHAL or statistical models such as the GLM (when absence data are available). Others believe that fit statistics, such as AUC, can



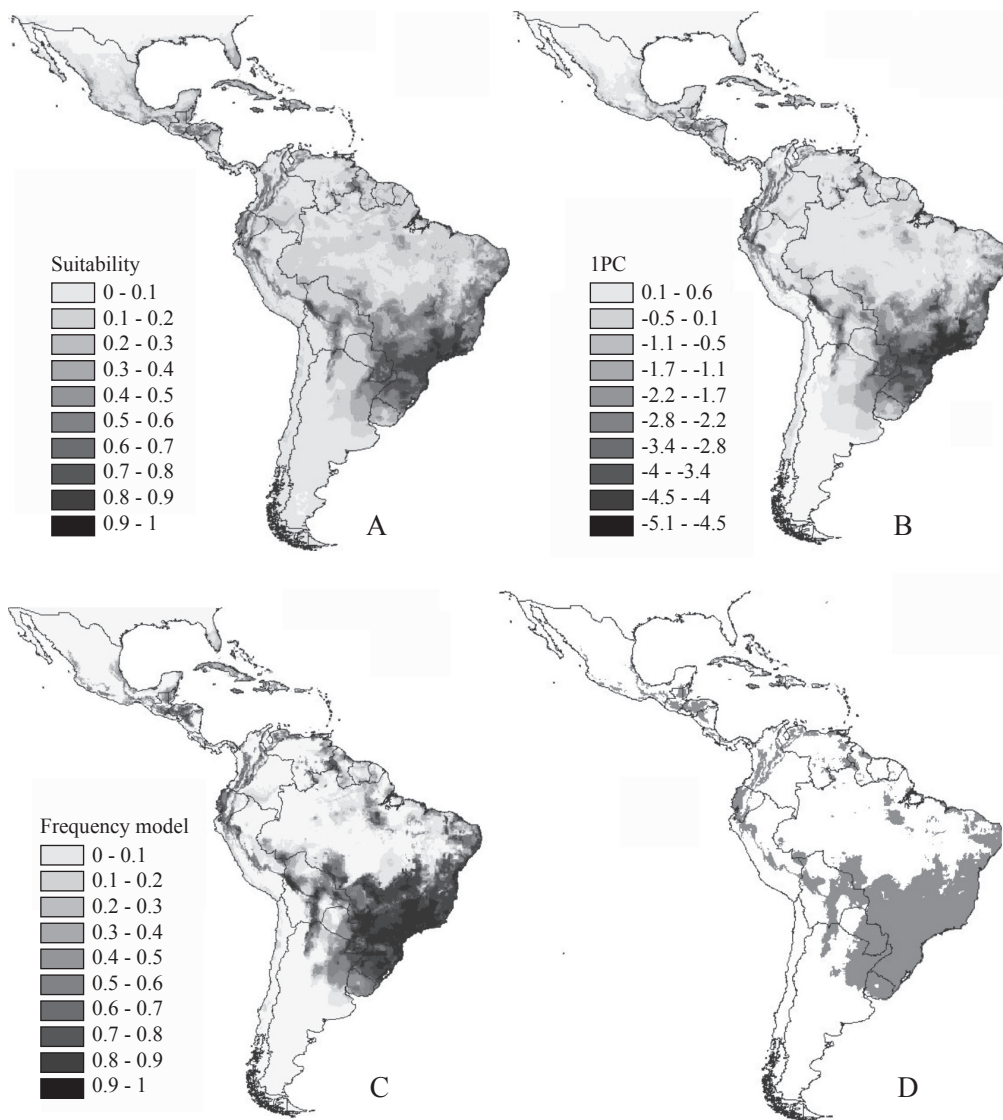


Fig. 3. Ensemble forecasting of 10 geographic ranges derived from distinct niche models for *Utheteisa ornatrix*. A) mean suitability (which define climatically suitable areas, or CSAs); B) First principal component of suitabilities; C) frequency of models that indicates species as present in each region and; D) majority consensus of frequencies (i.e., species is present in a region if 50% of the model suggest it is present).

be a reasonable guide for selecting the best model to use. A completely different approach, proposed by Araújo & New (2007), is to use as many models as possible to generate consistent and robust predictions; this approach is called “ensemble forecasting.”

#### Ensemble Forecasting: Phase I. Establishing Consensus Projections

The idea of ensemble forecasting involves the consideration that different sources of errors (sensitivity to data, lack of absence data, errors in environmental variables) will affect each niche model in different ways. It is expected that, by using a consensus of these models, errors will tend to cancel each other out and produce a trustworthy and more conservative solution.

Araújo & New (2007) considered multiple hierarchically-

arranged sources of variation among species' range predictions (including niche models, model parameterization, data sampling schemes, models and emission scenarios, and so on). The overall idea of ensemble forecasting is to combine different models using different approaches. Marmion *et al.* (2009) recently produced a comparative analysis of several strategies and showed that simple means seem to produce a more robust solution (see also Roura-Pascal *et al.* 2009; Coetzee *et al.* 2009; O' Haney 2009).

To illustrate the possibilities of using ensemble forecasting in geographical projections of niche models, we combined the ten niche models for *U. ornatrix* and generated different forms of ensembles. We used a simple average suitability model (Fig. 3A) and the principal component of the correlation matrix among the suitability values (Fig. 3B), which explained 77.9% of the correlation structure (see also

Thuiller 2004; Araújo *et al.* 2005, 2006). These two ensembles use the original suitability values of each model, and therefore, they are not truncated to produce discrete (0/1) predictions of where species can be found. However, it is also possible to truncate each model to produce a discrete prediction and then to calculate the frequency at which different models appear in a cell (Fig. 3C). Finally, if a discrete final consensus map is desired, one can use, for example, a majority consensus rule and state that only cells for which half of the models indicate the presence of species be considered as occupied by the species (Fig. 3D). In this case, all of the maps are quite similar and reinforce the high suitability in the Atlantic forest and in Central America.

Generating the maps shown in Fig. 3 and describing the above-mentioned method can be the “first phase” of model combination, in which the purpose is to obtain a unique map of the species’ distributions. However, in addition to producing a consensus map, the idea of ensemble forecasting also involves understanding the origins of uncertainty and why different models are produced for certain species. This would be the second methodological phase in the ensemble forecasting approach, i.e., understanding the structure of the differences among the niche models and their associated distribution maps.

#### Ensemble Forecasting: Phase II. Multidimensional Analysis of Uncertainty Patterns

The first way to disentangle these uncertainties is to use multivariate statistical analysis to compare the maps (Thuiller 2004). The principal component analysis (PCA) of the suitability maps is a way to evaluate the differences among the niche models. The scores of the PCA are the coordinates of each observation (in this case, the cells in the maps) in each principal component, and therefore, they can be mapped (see Fig. 3B). On the other hand, the loadings of the PCA, which reveal the contribution of each map to the principal component, are a useful way to assess the similarities among the maps.

For instance, in our analysis of *U. ornatrrix*, the first principal component explained 77.9% of the correlation structure among the suitability maps. The plot of loading on the two axes (Fig. 4A) clearly shows that all maps have similar directions along the first axis, and therefore, they tend to be quite similar. The only exception is the Mahalanobis distance map, which tends to depart from the other maps because, when using distance, the higher suitability produced by this method make it different from the others (although the overall “shape” of suitability is not too different).

The same patterns—grouping all similar models and contrasting the map generated by Mahalanobis distances—appears in a UPGMA cluster analysis based on Euclidian distances among these maps (Fig. 4B).

#### Ensemble Forecasting: Phase III. Measuring and Mapping the Variance Components of Uncertainty

Diniz-Filho *et al.* (2009a) and Buisson *et al.* (2010) have recently proposed that linear models can be used to assess

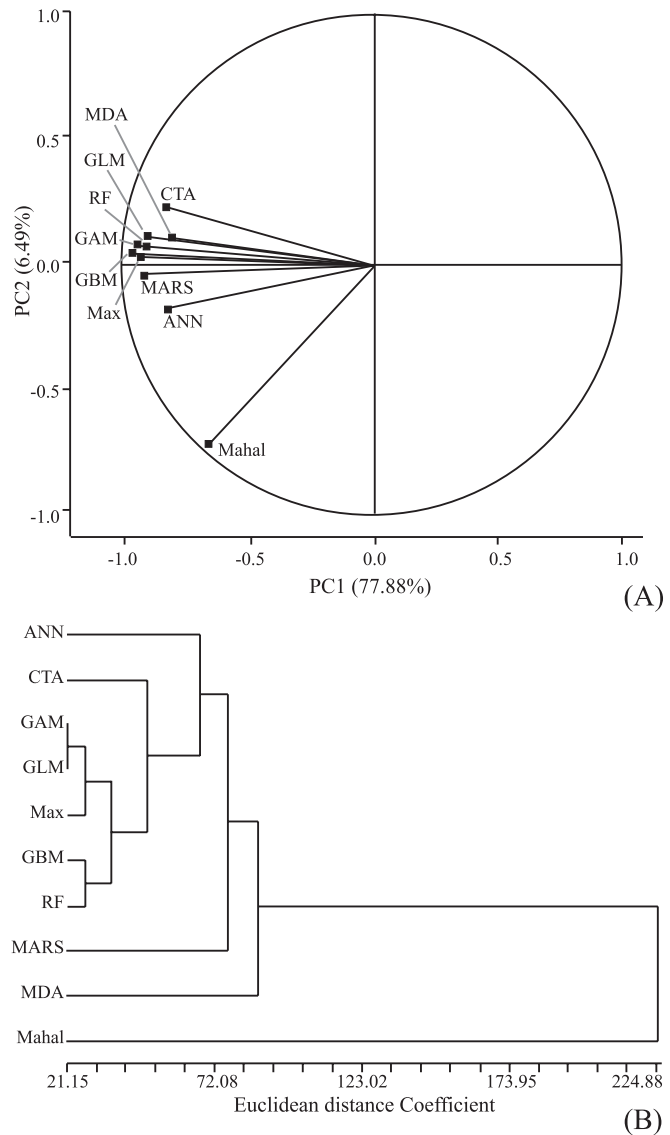


Fig. 4. Similarity of the 10 distribution maps of *Utheteisa ornatrrix*, based on Euclidian distances of suitabilities and analyzed by multivariate techniques, including A) Principal component loadings; B) UPGMA clustering. See Table 1 for codes.

the relative magnitude of different sources of variation in ensemble forecasting. Thus, this would be a third phase in the ensemble analysis, the purpose of which is to disentangle the different sources of uncertainty and to evaluate (and map, as shown below) their magnitude quantitatively.

To increase the number of sources of variation to investigate and to better illustrate our variance partitioning approach, we analyzed the projections of the ten niche models built for *U. ornatrrix* in the future using different models (coupled Atmospheric-Ocean Global Circulation Models - AOGCM) for climate change. These AOGCMs were the CCCma (Canadian Centre for Climate Modeling and Analysis), CSIRO (Australia’s Commonwealth Scientific and Industrial Research Organization), and the HadCM3 (Hadley Centre for Climate Prediction and Research’s General Circulation Model) (only the pessimistic scenarios were used), the

Table I. Short description of the models used in this paper niche modeling of *Uthetseia ornatrix*, using the software BIOMOD (see Thuiller 2003; Thuiller *et al.* 2009) or MAXENT (Phillips & Dudik 2007). Mahalanobis distance was implemented in the MATLAB platform.

Method	Short description	Software	Parameters used
Artificial Neural Network (ANN)	The ANN are a flexible way to generalize a linear model, selecting the best network through a N-fold cross-validation procedure. Criterion is based on the number of hidden units and weight decay (accuracy) of each possible result.	BIOMOD	5-fold cross-validation
Classification Tree Analysis (CTA)	The CTA is a method that splits the response in groups as homogenous as possible, based on the explanatory variables. The tree that minimizes the heterogeneity of each leaf with the smallest number of leafs is found by cross validation among possible trees.	BIOMOD	100-fold cross-validation
Generalized Additive Models (GAM)	GAMs are a generalization of the well-known GLMs (Generalized Linear models) that identify the appropriate transformation of the data, maximizing the relationship between the response and explanatory variables. To achieve this, GAMs use a series of smoothers to fit the data locally, so that these models cope for a wide range of response shapes.	BIOMOD	3 smoothers
Generalized Boosting Models (GBM)	GBMs are models that generate a large number of relative simple models (when compared to a GLM or a CTA) whose combined prediction is expected to produce a robust estimate of the response. Here we used a boosting regression tree (BRT) algorithm which combines the results of a given specified number of CTA's classifications trees.	BIOMOD	2000 classification trees
Generalized Linear Models (GLM)	GLMs are mathematical extensions of standard linear models developed to deal with non-normal error distribution and allow non-linear relationships between the response and the explanatory variables.	BIOMOD	Cubic polynomial
Mahalanobis distance	The distance model based on the Mahalanobis metrics determine the distance of each point in multivariate space to the centroid of this space (optimum of the species), taking into account the covariance structure among explanatory variables.	implemented in Matlab	No user defined parameters
Maximum Entropy (MaxEnt)	Maxent is a method for making predictions based on incomplete information, estimating a target probability distribution by finding the distribution of maximum entropy, which is the closest to uniform. This optimization is subjected to a set of constraints representing the incomplete information about the target distribution.	MaxEnt	1000 iterations
Mixture Discriminant Analysis (MDA)	MDAs are a flexible extension of the classical linear discriminant analysis, which searches for the linear combinations of variables that best discriminated between site groups defined by the presence or absence of the species. In MDA.	BIOMOD	No user defined parameters
Multivariate Adaptive Regression Spline (MARS)	MARS are an alternative regression for fitting non-linear responses using a method similar to a piecewise, non-stationary, regression, rather than smooth functions used in GAMs. It takes into account that the coefficients to be fitted have different optimal values across different levels of the explanatory variables.	BIOMOD	No user defined parameters
Random Forest (RF)	RF is a generalization of CTA, and "grows" a large number of CTAs based on a randomized subset of the predictors, so that each tree is a full-sized tree and the final result is obtained by averaging all the partial results.	BIOMOD	No user defined parameters

predictions for which were downloaded from Worldclim (<http://www.worldclim.org/futdown.htm>). Thus, it is possible to evaluate which source of variation (niche model or climatic model) is more important for the overall variation.

First, we can map the simple unweighted consensus map under current (Fig. 5A) and future (Fig. 5B) climates and evaluate the differences by mean suitability values (Fig. 5C). It is possible to see that a loss of climatically suitable areas will appear in the central part of the species' range and will tend to move in the southeastern direction. However, it is important to notice that a large variance around these estimates of climatically suitable areas exists in the future (Fig. 5D).

The variance partition proposed by Diniz-Filho *et al.* (2009a) allows us to explain why these variances arise in

geographical space; this involves performing a model II two-way Analysis of Variance (ANOVA) for each cell in the grid, comparing the suitability of the maps from different niche models (Factor I) and the climate models (Factor II). Because there is no replication, it is impossible to disentangle the error from the interaction term, but it is still possible to evaluate the main effects. On average, 95.17% of the variation among the maps is due to niche models (ranging from 23.71% to 99.9%), and 0.8% of the variation is due to the differences among the climate models (ranging from less than 0.0001% to 37.89%).

More importantly, because ANOVA is performed independently for each cell, it is possible to map these variance components (Fig. 6A). It is also possible to use spatial autocorrelation analyses based on Moran's I correlograms

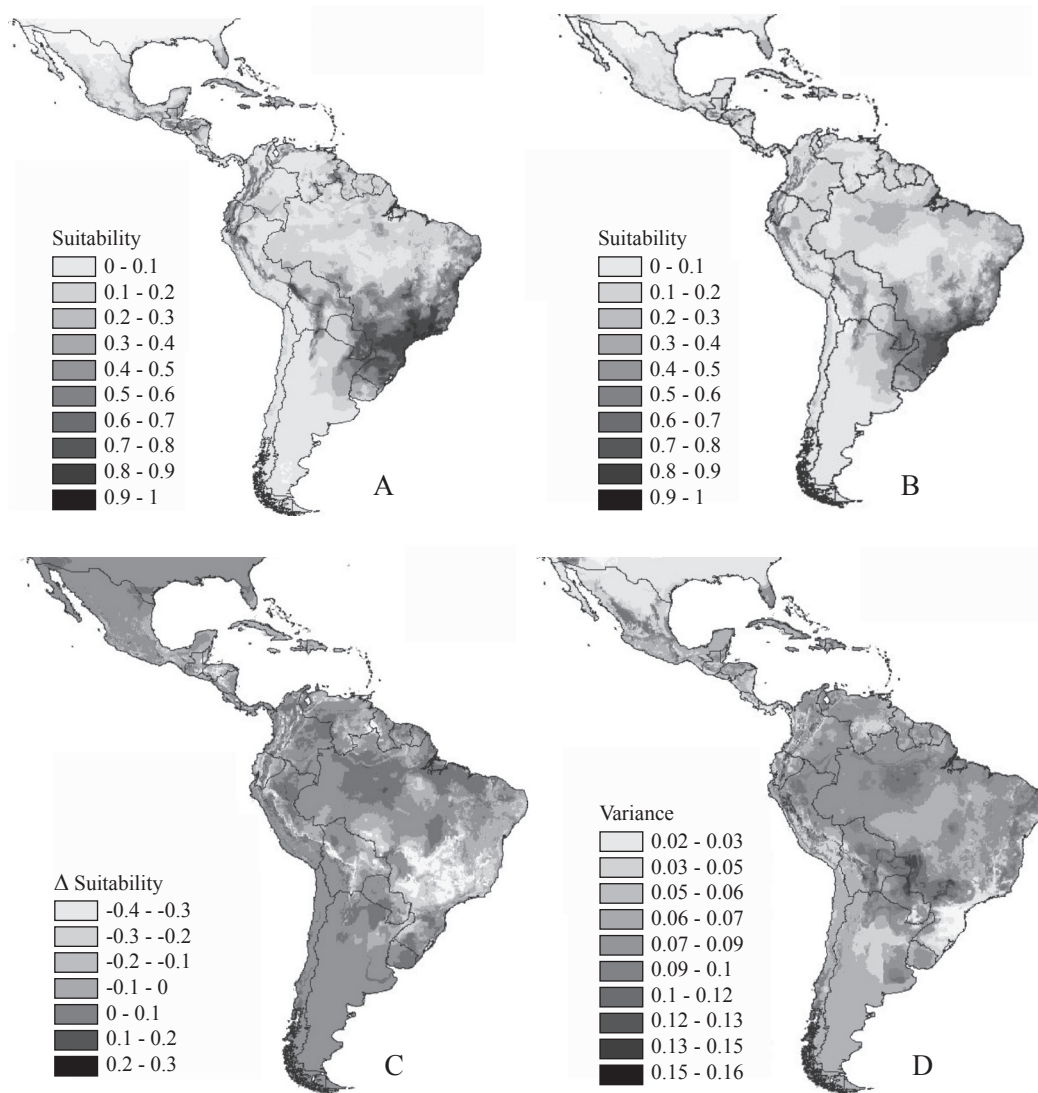


Fig. 5. Projecting geographic ranges into the future and the shifts in Climatically Suitable Areas (CSA) of *Utheteisa ornatrix*. A) mean current suitability; B) mean suitability in the future; C) shifts in CSAs between present and future and; D) variance of differences in CSA among niche models and climatic models.

(Diniz-Filho *et al.* 2003) to determine if spatial structures are present in this variance component. The variance component due to the effect of niche modeling shows a strong spatial pattern, with high Moran's I forming a clinal correlogram (Fig. 6B). As expected under the assumption of phase I of ensemble forecasting, the places in which the species are indicated as present have much less uncertainty due to the niche models than places where the species is clearly not found. The variance component due to the climate model does not reveal any geographic pattern of variation, because it has a non-significant Moran's I correlogram.

For instance, our results are similar to those obtained by Diniz-Filho *et al.* (2010b), who modeled the distribution of the grasshopper *Tropidacris cristata* in the Neotropics. They showed the large effect of niche modeling on the shifts in climatically suitable areas towards southeastern Brazil.

However, in their analysis, the effect of AOGCM and the interaction with niche modeling was much more pronounced. Diniz-Filho *et al.* (2009b) also showed the strong effect of niche modeling when evaluating the responses of Cerrado fauna (species turnover) to climate change.

Dormann *et al.* (2008) used a similar ANOVA design to evaluate uncertainty in niche modeling, but it is important to highlight that their response variables were the fit statistics from each model. Therefore, although their reasoning is similar to that proposed by Diniz-Filho *et al.* (2009a), the approach does not really show the explicit differences among the geographic range maps. Rather, it shows which effects are responsible for model fitting. This is because we can expect that, when using the same data, a very high model fit is generated by the same distribution map. However, when the model fit decreases (as commonly observed), similar values



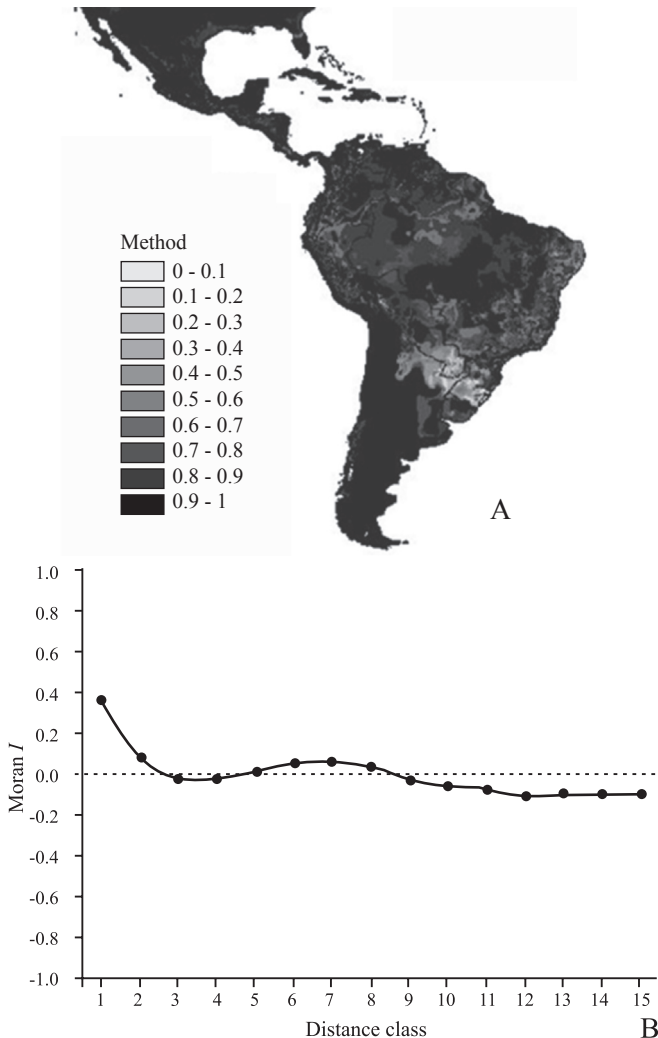


Fig. 6. A) Map of the variance component due to niche model, and B) spatial correlogram of variance component.

for the given fit statistics can be produced based on different maps.

#### Concluding Remarks

We used the niche modeling analysis of *U. ornatix*, a Neotropical tiger moth species, to illustrate how recent advances in niche modeling can be conservatively used, under the framework of ensemble forecasting, to generate distribution maps. Although this approach does not overcome the critical issue of species' equilibrium with the environment, it may represent a compromise solution if the methods deal differently with this assumption (e.g., are not equally robust to violations in the assumption). The maps from niche models allow both a description of the internal structure of geographic ranges, in terms of establishing climatically suitable areas for the species, and an understanding of how these areas will shift under climate change. We propose three phases for ensemble forecasting studies. Phase I includes the combination of maps for achieving a consensual—and hopefully conservative—solution, using different models, data and methods. Phase II describes the differences among the maps used, and Phase

Table II. Values of the area under the receiving operating curve (AUCs) of each technique used for modeling niche of *Utheteisa ornatix*.

Method	AUCs
ANN	0.880
CTA	0.859
GAM	0.932
GBM	0.931
GLM	0.926
Mahalanobis distance	0.911
MARS	0.908
MDA	0.898
MaxEnt	0.936
RF	0.937

III is the quantitative evaluation of the relative magnitude of uncertainties from different sources and their mappings.

Phase I was actually proposed by Araújo & New (2007) as an approach that allows researchers to keep working, considering the uncertainties and discussions regarding the validity and use of different models, both for describing species' niches and for predicting climate changes. However, at the same time, they recognize that we can use uncertainty levels to improve our knowledge and move the field forward. Phase II has been used for several years (e.g., Thuiller 2004; Araújo *et al.* 2005, 2006; Pearson *et al.* 2006) and can facilitate decisions concerning outlier models in multidimensional space. Diniz-Filho *et al.* (2009a; see also Buisson *et al.* 2010) proposed Phase III as a more complex way to evaluate the magnitude of uncertainty that can be attributed to each source.

Both Phase II and Phase III can be used to improve the modeling approach. Our results clearly show that the main source of uncertainty in generating geographic ranges is niche modeling, and therefore, an effort must be made to improve their evaluation. Reducing uncertainty requires a better evaluation of models and, perhaps, the development of new statistics that find a better compromise between model fit and transferability. This may take time, but it may also be a promising research solution. In the meantime, the three phases of ensemble forecasting can be an alternative for modeling species distribution.

Acknowledgements: We thank Miguel Araújo, David Nogués-Bravo, Thiago F. Rangel, Joaquin Hortal e Jorge Lobo for useful discussions on ensemble forecasting and niche modeling. Our work on niche modeling in the Neotropics have been continuously supported by different grants CNPq/MCT and CAPES, and by BIOIMPACT project coordinated by Miguel B. Araújo and supported by BBVA foundation.

#### REFERENCES

- Allouche, O.; A. Tsoar & R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**: 1223–1232.
- Araújo, M. B. & A. Guisan. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* **33**: 1677–1688.
- Araújo, M. B. & M. New. 2007. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* **22**: 42–47.
- Araújo, M. B. & R. G. Pearson. 2005. Equilibrium of species' distributions with climate. *Ecography* **28**: 693–695.

- Araújo, M. B.; W. Thuiller & R. G. Pearson. 2006. Climate warming and the decline of amphibians and reptiles in Europe. **Journal of Biogeography** **33**: 1712–1728.
- Araújo, M. B.; R. J. Whittaker; R. Ladle & M. Erhard. 2005. Reducing uncertainty in projections of extinction risk from climate change. **Global Ecology and Biogeography** **14**: 529–538.
- Biezanko, C. M. & R. G. de Freitas. 1938. Catálogo dos insetos encontrados na cidade de Pelotas e seus arredores Fasc. 1–Lepidópteros. **Boletim da Escola de Agronomia Eliseu Maciel** **25**: 1–32.
- Biezanko, C. M. & F. D. Seta. 1939. **Catálogo dos insetos encontrados em Rio Grande e seus arredores**. Fasc. 1 – Lepidópteros. Pelotas. Universal. 15 p.
- Bini, L. M.; J. A. F. Diniz-Filho; T. F. L. V. B. Rangel; R. P. Bastos & M. P. Pinto. 2006. Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. **Diversity and Distributions** **12**: 475–482.
- Buisson, L.; W. Thuiller; N. Casajus; S. Lek & G. Grenouillet. 2010. Uncertainty in ensemble forecasting of species distribution. **Global Change Biology** **16**: 1145–1157.
- Coetzee, B. W. T.; M. P. Robertson; B. F. N. Erasmus; B. J. van Rensburg & W. Thuiller. 2009. Ensemble models predict Important Bird Areas in southern Africa will become less effective for conserving endemic birds under climate change. **Global Ecology and Biogeography** **18**: 701–710.
- Colwell, R. K. & T. F. L. V. B. Rangel. 2009. Hutchinson's duality: the once and future niche. **Proceedings of National Academy of Sciences USA** **106**: 19651–19658.
- De Marco, P. Jr.; J. A. F. Diniz-Filho & L. M. Bini. 2008. Spatial analysis improves species distribution modelling during range expansion. **Biology Letters** **4**: 577–580.
- Diniz-Filho, J. A. F.; L. M. Bini & B. A. Hawkins. 2003. Spatial autocorrelation and red herrings in geographical ecology. **Global Ecology and Biogeography** **12**, 53–64.
- Diniz-Filho, J. A. F.; L. M. Bini; T. F. L. V. B. Rangel; R. D. Loyola; C. Hof; D. Nogués-Bravo & M. B. Araújo. 2009a. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. **Ecography** **32**: 897–906.
- Diniz-Filho, J. A. F.; G. Oliveira; L. M. Bini; R. D. Loyola; J. C. Nabout & T. F. L. V. B. Rangel. 2009b. Conservation biogeography and climate change in the Brazilian Cerrado. **Natureza Conservação** **7**: 100–112.
- Diniz-Filho, J. A. F.; P. DeMarco & B. A. Hawkins. 2010a. Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. **Insect Conservation and Diversity** **3**: 172–179.
- Diniz-Filho, J. A. F.; J. C. Nabout; L. M. Bini; R. D. Loyola; T. F. Rangel; D. Nogue's-Bravo & M. B. Araújo. 2010b. Ensemble forecasting shifts in climatic suitable areas for *Tropidacris cristata* (Orthoptera: Acridoidea: Romaleidae). **Insect Conservation and Diversity** **3**: 213–221.
- Dormann, C. F.; O. Puschke; J. R. G. Marquez; S. Lautenbach & B. Schroder. 2008. Components of uncertainty on species distribution analysis: a case study of the great grey shrike. **Ecology** **89**: 3371–3386.
- Elith, J. & C. H. Graham. 2009. Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. **Ecography** **32**: 66–77.
- Elith, J.; C. H. Graham; R. P. Anderson; M. Dudik; S. Ferrier; A. Guisan; R. J. Hijmans; F. Huettmann; J. R. Leathwick; A. Lehmann; J. Li; L. G. Lohmann; B. A. Loiselle; G. Manion; C. Moritz; M. Nakamura; Y. Nakazawa; J. M. Overton; A. T. Peterson; S. J. Phillips; K. Richardson; R. Scachetti-Pereira; R. E. Schapire; J. Soberon; S. Williams; M. S. Wisz & N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. **Ecography** **29**: 129–151.
- Escalante, T.; M. Linaje; P. Illoldi-Rangel; M. Rivas; P. Estrada; F. Neira & J. J. Morrone. 2009. Ecological niche models and patterns of richness and endemism of the southern Andes genus *Eurymetopum* (Coleoptera, Cleridae). **Revista Brasileira de Entomologia** **53**: 379–385.
- Ferreira, P. S. F.; A. S. Paula & D. S. Martins. 1995. Análise faunística de Lepidoptera Arctiidae em área de reserva natural remanescente de floresta tropical em Viçosa, Minas Gerais. **Anais da Sociedade Entomológica do Brasil** **24**: 123–133.
- Ferro, V. G. & J. A. Teston. 2009. Composition of the Arctiidae species (Lepidoptera) in southern Brazil: relationship among vegetation types and among habitat spatial configuration. **Revista Brasileira de Entomologia** **53**: 278–286.
- Ferro, V. G.; P. R. Guimarães-Jr & J. R. Trigo. 2006. Why do larvae of *Utetheisa ornatrix* penetrate and feed in pods of *Crotalaria* species? Larval performance vs. chemical and physical constraints. **Entomologia Experimentalis et Applicata** **121**: 23–29.
- Guisan, A. & W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. **Ecology Letters** **8**: 993–1009.
- Hawkins, B. A.; R. Field; H. V. Cornell; D. J. Currie; J. F. Guégan; D. M. Kaufman; J. T. Kerr; G. G. Mittelbach; T. Oberdorff; E. M. O'Brien; E. E. Porter & J. R. G. Turner. 2003. Energy, water, and broad-scale geographic patterns of species richness. **Ecology** **84**: 3105–3117.
- Hortal, J.; A. Jiménez-Valverde; J. F. Gómez; J. M. Lobo & A. Baselga. 2008. Historical bias in biodiversity inventories affects the observed realized niche of the species. **Oikos** **117**: 847–858.
- Lobo, J. M.; A. Jimenez-Valverde & R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. **Global Ecology and Biogeography** **17**: 145–151.
- Marinoni, R. C. & R. R. C. Dutra. 1996. Levantamento da fauna entomológica do Estado do Paraná. II. Ctenuchidae (Lepidoptera). **Revista Brasileira de Zoologia** **13**: 435–461.
- Marmion, M.; M. Parviainen; M. Luoto; R. K. Heikkinen & W. Thuiller. 2009. Evaluation of consensus methods in predictive species distribution modeling. **Diversity and Distributions** **15**: 59–69.
- Mittelbach, G. G.; D. W. Schemske; H. V. Cornell; A. P. Allen; J. M. Brown; M. B. Bush; S. P. Harrison; A. H. Hurlbert; N. Knowlton; H. A. Lessios; C. M. McCain; A. R. McCune; L. A. McDade; M. A. McPeck; T. J. Near; T. D. Price; R. E. Ricklefs; K. Roy; D. F. Sax; D. Schluter; J. M. Sobel & M. Turelli. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. **Ecology Letters** **10**: 315–331.
- Nogués-Bravo, D. 2009. Predicting the past distribution of species climatic niches. **Global Ecology and Biogeography** **18**: 521–531.
- O'Haney, J. R. 2009. Neural Ensembles: a neural network based ensemble forecasting program for habitat and bioclimatic suitability analysis. **Ecography** **32**: 89–93.
- Pearman, P. B.; A. Guisan & O. Broennimann. 2008. Niche dynamics in space and time. **Trends in Ecology and Evolution** **23**: 149–158.
- Pearson, R. G. & T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelopes useful? **Global Ecology and Biogeography** **12**: 361–171.
- Pearson, R. G.; W. Thuiller; M. B. Araújo; E. Martinez-Meyer; L. Brotons; C. Mclean; L. J. Miles; P. Segurado; T. E. Dawson & D. C. Lees. 2006. Model-based uncertainty in species' range prediction. **Journal of Biogeography** **33**: 1704–1711.
- Pease, R. W. 1968. Evolution and hybridization in the *Utetheisa ornatrix* complex (Lepidoptera: Arctiidae): inter and intra population variation and its relation to hybridization. **Evolution** **22**: 719–735.
- Peterson, A. T.; M. Papes & M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. **Ecography** **30**: 550–560.
- Peterson, A. T.; M. Papes & J. Soberón. 2008. Rethinking receiver operating characteristic analysis: applications in ecological niche modeling. **Ecological Modeling** **213**: 63–72.
- Phillips, S. J. & M. Dudik. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. **Ecography** **31**: 161–175.
- Randin, C. F.; T. Dirnböck; S. Dullinger; N. E. Zimmermann; M. Zappa & A. Guisan. 2006. Are niche-based species distribution models transferable in space? **Journal of Biogeography** **33**: 1689–1703.
- Roura-Pascal, N.; L. Brotons; A. T. Peterson & W. Thuiller. 2009. Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. **Biological Invasions** **11**: 1017–1031.
- Segurado, P. & M. B. Araújo. 2004. An evaluation of methods for modelling species distributions. **Journal of Biogeography** **31**: 1555–1568.
- Signoretti, A. G. C.; D. E. Nava; J. M. S. Bento & J. R. P. Parra. 2008. Biology and thermal requirements of *Utetheisa ornatrix* (L.) (Lepidoptera: Arctiidae) reared on artificial diet. **Brazilian Archives of Biology and Technology** **51**: 647–653.
- Soberon, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. **Ecology Letters** **10**: 1115–1123.
- Teston, J. A. & E. Corseuil. 2004. Diversidade de Arctiinae (Lepidoptera, Arctiidae) capturados com armadilha luminosa, em seis comunidades no Rio Grande do Sul, Brasil. **Revista Brasileira de Entomologia** **48**:

- 77–90.
- Teston, J. A.; A. Specht; R. A. Di Mare & E. Corseuil. 2006. Arctiinae (Lepidoptera, Arctiidae) coletados em unidades de conservação estaduais do Rio Grande do Sul, Brasil. **Revista Brasileira de Entomologia** **50**: 280–286.
- Thuiller, W. 2003. BIOMOD: optimising predictions of species distributions and projecting potential future shifts under global change. **Global Change Biology** **9**: 1353–1362.
- Thuiller, W. 2004. Patterns & uncertainties of species' range shifts under climate change. **Global Change Biology** **10**: 2020–2027.
- Thuiller, W.; B. Lafourcade; R. Engler & M. B. Araújo. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. **Ecography** **32**: 369–373.
- Zikán, J. F. & W. Zikán. 1968. Inseto-fauna do Itatiaia e da Mantiqueira 3: Lepidoptera. **Pesquisa Agropecuária Brasileira** **3**: 45–109.