

# Not as good as they seem: the importance of concepts in species distribution modelling

Alberto Jiménez-Valverde<sup>1\*</sup>, Jorge M. Lobo<sup>1</sup> and Joaquín Hortal<sup>2</sup>

<sup>1</sup>Museo Nacional de Ciencias Naturales, Departamento Biodiversidad y Biología Evolutiva, C/José Gutiérrez Abascal 2, E-28006, Madrid, Spain, <sup>2</sup>NERC Centre for Population Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK

## ABSTRACT

**Aim** Nowadays, large amounts of species distribution data and software for implementing different species distribution modelling methods are freely available through the internet. As a result, methodological works that analyse the relative performance of modelling techniques, as well as those that study which species characteristics affect their performance, are necessary. We discuss three important topics that must be kept in mind when modelling species distributions, namely (i) the distinction between potential and realized distribution, (ii) the effect of the relative occurrence area of the species on the results of the evaluation of model performance, and (iii) the general inaccuracy of the predictions of the realized distribution provided by species distribution modelling methods.

**Location** Unspecific.

**Methods** Using some recent papers as a basis, we illustrate the three issues mentioned above and discuss the negative implications of neglecting them.

**Results** Considering a potential-realized distribution gradient, different modelling methods may be arranged along this gradient according to their ability to model any concept. Complex techniques may be more suitable to model the realized distribution than simple ones, which may be more appropriate to estimate the potential distribution. Comparisons among techniques must consider this scenario. The relative occurrence area of the species conditions the results of the evaluation scores, implying that models of rare species will unavoidably yield higher discrimination values. Moreover, discrimination values that are usually reported in the literature may imply considerable over or underestimations of the distribution of the species.

**Main conclusions** It is extremely important to establish a solid conceptual and methodological framework on which the emergent field of species distribution modelling can stand and develop.

## Keywords

Accuracy, error, extent, potential distribution, realized distribution, relative occurrence area, species distribution models, validation.

\*Correspondence: Alberto Jiménez-Valverde, Departamento Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), c/José Gutiérrez Abascal 2, 28006 Madrid, Spain. E-mail: mcnej651@mncn.csic.es

## INTRODUCTION

Predictive models of species distributions are being increasingly used to address questions related to the ecology, biogeography, and conservation of species (see Peterson, 2007). On one hand, a huge amount of distributional data are being gathered in biodiversity data bases and made freely available through the internet (see Soberón & Peterson, 2004 and <http://www.gbif.org/>). On the other, a wide variety of modelling techniques specially designed to model species distributions are currently available (see Guisan & Thuiller, 2005), including a number of presence-

only freeware modelling tools (e.g. Bioclim and Domain, Busby, 1991; Carpenter *et al.*, 1993, respectively; <http://www.diva-gis.org/>; Biomapper, Hirzel *et al.*, 2002; <http://www2.unil.ch/biomapper/>; GARP, Stockwell & Peters, 1999; <http://www.nhm.ku.edu/desktopgarp/>; Maxent, Phillips *et al.*, 2006; <http://www.cs.princeton.edu/~schapire/maxent/>). Therefore, a good knowledge of the performance of these techniques becomes extremely important to help researchers and technicians to select the method most appropriate to their purposes.

In a recent contribution, Tsoar *et al.* (2007) compared the performance of six species distribution modelling methods that

require only presence data as input (i.e. profile techniques). The value of this study could be of particular importance, since usually the only reliable information on the distribution of organisms is about their recorded presence. Contrary to presence data, reliable absence data are rare and hard to obtain; confirming that a species is absent from a locality is a difficult task (Gu & Swihart, 2004) that becomes almost unaffordable in the case of the coarse resolution grid cells used in most studies. After their comparison, Tsoar *et al.* (2007) reached two main conclusions: (i) complex techniques (i.e. those that establish more flexible relationships between the dependent and independent variables) are better predictors than simple methods, and (ii) the distribution data of species with restricted niches are modelled with higher accuracy than that of generalist species. These two statements are in agreement with the existing literature. Tsoar *et al.*'s conclusion on the complexity of techniques resembles the insight provided by Elith *et al.* (2006), who also concluded that methods able to fit complex responses are preferred over simple techniques. Their conclusion concerning the traits of the modelled species coincides with several previous studies that found that predictions are usually more accurate for the species with the smaller range sizes and the higher habitat specificity (McPherson & Jetz, 2007 and references therein).

Here, we argue that the interpretation of the results found in these comparative studies can vary if some methodological and theoretical considerations are taken into account. We use the discussion on both topics mentioned above as a means to reflect on the theoretical concepts that underlie species distribution modelling methodologies. We provide alternative and reasonable interpretations of the above-mentioned results that outperform the most widely agreed explanations.

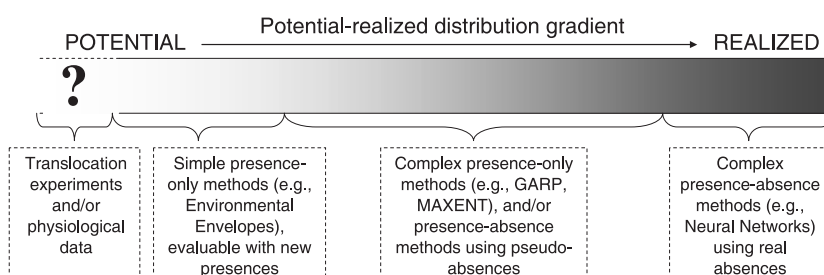
## ARE COMPLEX TECHNIQUES BETTER FOR THE PREDICTION OF SPECIES DISTRIBUTIONS THAN SIMPLE ONES?

In this paper, we deliberately avoid using the term *niche* to refer to species *distributions*. The concept of niche is often confused (Real & Levin, 1991; Colwell, 1992), and necessarily implies the understanding of the effects of biotic and abiotic factors on the fitness of organisms (Kearney, 2006). Many factors can result in the absence of a species from suitable habitats and/or its presence in unsuitable ones (Pulliam, 2000). Hence, the combination of statistical models with distribution data does not allow deriving the realized niche of the species, and even less their fundamental niches. Besides the current strong debate about the 'geographical'

definition of the niche (see Soberón & Peterson, 2005; Araújo & Guisan, 2006; Kearney, 2006; Peterson, 2007), Mike P. Austin wrote 'statistical models [...] can say little about the fundamental niche' (Austin, 2002; p. 104) and, in words of Jorge Soberón, modellers calculate 'abstract objects obviously *related* to niches' (Soberón, 2007; p. 1121). In this sense, correlative statistical models are able to project simulations of the distribution of species into the geographical space, but are not able to provide a description of species niches.

A good use of species distribution models requires a clear distinction of the differences between potential and realized distributions (see Soberón, 2007). While potential distribution refers to the places where a species could live, realized distribution does to the places where a species actually lives. Importantly, both concepts refer to a particular moment or a discrete period in time (usually, present time). Therefore, the places pertaining to the potential or realized distribution of a species vary with time. However, they do not vary in the same way. The potential distribution of a species varies geographically with the oscillation of climatic conditions, but is environmentally invariant. At the same time, the realized distribution of the same species will vary in both the geographical and the environmental spaces when subject to the same climatic variations. In other words, while it can be assumed that the potential response of a species to environmental gradients is constant under some conditions, its realized response is context dependent. Therefore, depending on the question asked, we will be interested in describing or modelling one characteristic of the species distribution or the other. Indeed, these two concepts would be better approached using different analytical frameworks (Soberón & Peterson, 2005; Jiménez-Valverde *et al.*, 2007; Lobo *et al.*, 2007; see Fig. 1). For the models to represent one of these two concepts or the other, they must be calibrated and validated using the appropriate data.

The kind of absence data used for the calibration of the models and the modelling technique used condition the characteristics of the distribution of the focal species that are described by model results (Fig. 1). Species distributions are not only constrained by abiotic (e.g. climate) factors. Rather, they are also shaped by biotic interactions; dispersal constraints; anthropogenic effects; stochastic events; and other historical, unique, and contingent factors (Pulliam, 2000; Soberón, 2007). These effects can only be accounted for using data on the absence of the species to restrict model predictions (Soberón & Peterson, 2005). Hence, the realized distribution of a species cannot be estimated without data on its absence from environmentally suitable localities (Lobo, 2008). In addition, predictors must not only



**Figure 1** A conceptual potential-realized distribution gradient showing the modelling techniques and the characteristics of the training (and evaluation) data that are more adequate to describe each portion of this gradient. The discontinuous lines in the extreme left of the gradient mean that the potential distribution is a hypothetical concept that is hard to describe without a high amount of evidence from different sources.

include environmental variables (scenopoetic variables *sensu* Soberón, 2007); it is necessary to incorporate other factors that might be restricting the distribution of the species (e.g. Lobo *et al.*, 2006). If, on the contrary, the goal is to estimate the potential distribution of a species, the absences caused by non-environmental factors must be avoided. Here, absence data must come from environmental conditions that are known to be unsuitable for the species (Chefaoui & Lobo, 2008). If information on absence due to environmental constraints is not available, two alternatives can be taken: (i) generate absences outside the environmental domain where the species is present and use them for model parameterization (see, for example, Jiménez-Valverde & Lobo, 2007a); and (ii) use profile techniques such as those evaluated by Tsoar *et al.* (2007) in order to estimate the location of climatically suitable places (see below and Fig. 1).

Validating these two distributional concepts is a different issue. Estimations of realized distributions must be evaluated using data of the realized distribution of the species, i.e. presences as well as true absences caused by either environmental (scenopoetic) or non-environmental factors. However, these true absences cannot be used to validate the estimations of potential distributions. Rather, identifying all the localities that host environmental conditions suitable for a species is impossible to achieve. Therefore, the potential distribution of a species is a hypothetical concept (see below) that could be partially evaluated using new presence information, or preferably with either physiological data (Kearny, 2006), translocation experiments, or additional evidence from species invasions (see Sax *et al.*, 2007). These alternative approaches, though, present their own difficulties.

Nevertheless, the distinction we make between potential and realized distribution, as well as between the techniques that are most appropriate to model one or the other, is not rigid. Rather, such distinction occurs along a continuous gradient where the position of each particular combination of data and modelling technique is uncertain (see Fig. 1). Also, as stated above, it is unlikely that any distributional data are able to reflect all the environmental potentiality of the species due to the influence of contingent factors. In addition, biodiversity inventories are often spatial and environmentally biased (Lobo *et al.*, 2007), providing a biased and incomplete picture of their responses to environmental gradients (Hortal *et al.*, 2008). Therefore, the potential

distribution should be considered a hypothetical extreme of the gradient described above, that can be approached only in an ideal scenario where the distribution of the species is fully in equilibrium with the environmental space defined by the scenopoetic variables.

In such context, any comparison on the performance of different presence-only modelling techniques must take into account that such techniques generally provide distributions close to the potential. Thus, if the outputs of these techniques are evaluated using presence and true absence data (i.e. data on the species' realized distributions), it can be erroneously concluded that the predictions from more complex techniques are more accurate than those from simpler ones (as, e.g., Elith *et al.*, 2006 or Tsoar *et al.*, 2007 conclude). We argue that this result comes from the nature of the evaluation data used and the nature of the modelling techniques evaluated rather than from the true accuracy of these techniques. Those techniques that are able to establish the more complex relationships between dependent and independent variables will overfit the presence data more strongly. Unavoidably, this will result in predicted extents of occurrence that are smaller than those suggested by simpler techniques. Due to this, a greater number of the true absences in the validation data will be predicted as absences by complex techniques than by the simple ones.

Regardless of any conceptual misunderstanding, species distribution models could provide good predictions if they fit the evaluation data tightly. Sometimes it could be possible to forecast a given part of the realized distribution of a species using methods that are more adequate to describe its potential distribution. In this case, it is necessary to be particularly demanding in the evaluation of the agreement between observations and predictions. However, the discrimination between 'good' and 'bad' models is based in subjective ranges of indices that measure only if the agreement between predicted and observed distribution is significantly higher than the expected by chance. For example, in the case of the kappa statistic, values equal or smaller than 0.6 are commonly thought to indicate reliable predictions (i.e. a good agreement between observed and predicted distributions; see, e.g. Elith *et al.*, 2006; Araújo & Luoto, 2007; or Tsoar *et al.*, 2007). However, a kappa value of 0.6 can be obtained with degrees of under- or overprediction of 40%, for a species that occupies half of the territory (Fig. 2a). In the case of a rare species occupying



**Figure 2** Predictive scenarios yielding a kappa score of 0.6. a1 and b1 represent the real distribution of two species (a and b), and a2, b2, and b3 the predicted distributions. The presence of the species is represented in black, and its absence in grey. In the first example, species a occupies half of the considered territory, and a kappa value of 0.6 can be obtained when the model overpredicts the presence of the species by 40%. Note that, in this case, the inverse is also true, i.e. if black was absence and grey presence, the model would underpredict the presence of the species by 40%. In the second example, species b occupies 5% of the considered territory, and a kappa value of 0.6 is obtained when the model overpredicts the presence of the species by 102% (b2; i.e. the predicted geographical range is duplicated with respect to the real range), or when the model underpredicts the presence of the species by a 44% (b3).

5% of the territory, a kappa value of 0.6 could mean an overprediction of 102% (i.e. the area of distribution is doubled) or an underprediction of 44% (i.e. nearly a half of the distribution of the species is not predicted) (Fig. 2b). Therefore, the adequacy of models with these kappa values is questionable for both basic and applied purposes. This is also a drawback of other commonly used agreement measures such as area under the receiver operating characteristic curve (AUC) (see Lobo, 2008).

In sum, the evaluation of model results is biased towards a better performance of complex techniques due to their potential to overfit models to the training data. Certainly, identifying those techniques that produce robust forecasts of the realized distribution of the species is a worthy and important task, even if the techniques used are conceptually more appropriate to represent the potential distribution. Indeed, the complex techniques usually thought to be the most effective (e.g. Elith *et al.*, 2006; Tsoar *et al.*, 2007) would be placed closer to the realized distribution end of the potential-realized adequacy gradient shown in Fig. 1. The conceptual framework discussed above should therefore be remembered when defining the objectives and interpreting the results of species predictive models. The performance of the models should be evaluated by examining errors of omission and commission separately (i.e. presence points predicted as absences and absence points predicted as presences, respectively), and by taking into account the ratio between the extent of occurrence and the whole extent of the region of study (the relative occurrence area, ROA; Lobo *et al.* 2008). This latter is an evaluation that studies comparing the performance of different species distribution modelling techniques do not report.

### ARE THE PREDICTIONS FOR SPECIALIST SPECIES MORE RELIABLE THAN FOR GENERALISTS?

Species with restricted environmental tolerances and/or distributions are usually reported to be well predicted (e.g. Tsoar *et al.*, 2007). Several biological explanations for this pattern have been proposed (see McPherson & Jetz, 2007 and references therein). Here, we argue that these good performances are usually the result of the properties of the data used for validation, due to the correlation between the ROA of the species in a given territory and the environmental tolerance of the focal species. The ROA is a function of the extent of the studied territory. Thus, the smaller the ROA, the greater the number of absences far from the environmental domain of the presences will be available and the better the models will describe distribution data. This phenomenon was described by Lobo *et al.* (2008) as being analogous to the artificial inflation of the explanatory capacity in a clinical study by selecting a control population that includes a number of people that are naturally resistant to the disease. Due to this, species with smaller ROA will show better performance in most validation metrics, a result that is merely an inevitable product of the data used for validation.

The immediate consequence is that the models developed for two species with different ROAs are not comparable (see Lobo

*et al.*, 2008). These models provide information about different processes, in the same way as the information on the same species provided by distribution models differs according to the scale (both of extent and resolution) at which they are performed. Indeed, the conclusion that the distributions of rare and specialist species are easier to model accurately than common and generalist ones (e.g. Segurado & Araújo, 2004; McPherson & Jetz, 2007; Tsoar *et al.*, 2007) might be a trivial result. Rather it is an artefact caused by the comparison of model performances for different species within the same extent, provided that rare/specialist and common/generalist gradients are extent-dependent concepts. In other words, a good validation result can be obtained simply by increasing the extent of analysis and thus decreasing the ROA of the species, independently of their actual range size.

Note here that ROA is not equivalent to the generally used term of prevalence; strictly speaking, prevalence is the ratio of the number of presences to the total number of data points used in model training (i.e. it is a property of the data sample; Jiménez-Valverde & Lobo, 2006). Thus, species with small ROAs can have high prevalence values, and vice versa. On the other hand, ROA is intimately related with marginality (i.e. the degree of departure of the conditions inhabited by the species from the mean environmental conditions of the studied region) due to the spatially autocorrelated structure of nature. However, marginality is a concept with a biological meaning that relates the extent of the studied area and the distribution data with the environmental variables used as predictors. The use of ROA instead of marginality or other related concepts designed with the purpose of providing biological explanations is preferable, because it highlights the artefactual effect of extent in model results.

### CONCLUDING REMARKS

Predictions of species distributions based on correlative models can help to understand the spatial patterns of biological diversity. The literature developing and comparing these modelling techniques increases steadily, as well as the number of studies applying these techniques. We believe that as the availability of distribution data and modelling software increases, so does the danger of developing and applying distribution models without a solid conceptual background. Therefore, the field of species distribution modelling needs a serious reflection about the conceptual basis that underlies species distribution models, as well as about the true meaning of their predictions. An increasing number of recently published studies are questioning and discussing important conceptual and methodological aspects of species distribution models (e.g. Soberón & Peterson, 2005; Araújo & Guisan, 2006; Jiménez-Valverde & Lobo, 2006, 2007b; Kearney, 2006; Real *et al.*, 2006; Austin, 2007; Soberón, 2007; Lobo *et al.*, 2008; Raes & ter Steege, 2008). Here, the different effects of the quality of predictors, quality of the distributional data, and of the adequacy of the species distribution modelling techniques can be analysed using virtual species, in order to avoid the effects of confusing factors (Austin *et al.*, 2007; Jiménez-Valverde & Lobo, 2007b; Meynard & Quinn, 2007).

To summarize, the design of future works evaluating, comparing, and applying species distribution modelling techniques should be rooted in a good understanding of their conceptual background. Indeed, the results of these works should be interpreted with caution and a critical eye. If species distribution models are to be a common-use tool for biodiversity research and conservation assessment, the foundations of their application must be much more solid than they are now.

## ACKNOWLEDGEMENTS

We thank some anonymous referees for their constructive and helpful review of a former version of this manuscript, and to Rich Grenyer for his advice and help with the English. AJ-V and JML were supported by a Fundación BBVA project and the MEC project (CGL2004-04309), and JH by the UK Natural Environment Research Council.

## REFERENCES

- Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.
- Araújo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743–753.
- Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.
- Austin, M. (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1–19.
- Busby, J.R. (1991) BIOCLIM – a bioclimate analysis and prediction system. *Nature conservation: cost effective biological surveys and data analysis* (ed. by C.R. Margules and M.P. Austin), pp. 64–68. CSIRO, Victoria, Australia.
- Carpenter, G., Gillison, A.N. & Winter, J. (1993) DOMAIN: a flexible modeling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, **2**, 667–680.
- Chefaoui, R. & Lobo, J.M. (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, **210**, 478–486.
- Colwell, R.K. (1992) The niche: environmental or population attribute? *Keywords in evolutionary biology* (ed. by E.F. Keller and E.A. Lloyd), pp. 241–248. Harvard University Press, Cambridge, MA.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Gu, W. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation*, **116**, 195–203.
- Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.
- Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecological niche factor analysis: how to compute habitat suitability maps without absence data? *Ecology*, **83**, 2027–2036.
- Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M. & Baselga, A. (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, doi: 10.1111/j.0030-1299.2008.16434.x.
- Jiménez-Valverde, A. & Lobo, J.M. (2006) The ghost of unbalanced species distribution data in geographic predictive models. *Diversity and Distributions*, **12**, 521–524.
- Jiménez-Valverde, A. & Lobo, J.M. (2007a) Potential distribution of the endangered spider *Macrothele calpeiana* (Walckenaer, 1805) (Araneae, Hexathelidae) and the impact of climate warming. *Acta Zoologica Sinica*, **53**, 865–876.
- Jiménez-Valverde, A. & Lobo, J.M. (2007b) Threshold criteria for conversion of probability of species presence to either- or presence-absence. *Acta Oecologica*, **31**, 361–369.
- Jiménez-Valverde, A., Ortuño, V.M. & Lobo, J.M. (2007) Exploring the distribution of *Stercorax* Ortuño, 1990 (Coleoptera, Carabidae) species in the Iberian Peninsula. *Journal of Biogeography*, **34**, 1426–1438.
- Kearny, M. (2006) Habitat, environment and niche: what are we modelling? *Oikos*, **115**, 186–191.
- Lobo, J.M. (2008) More complex distribution models or more representative data? *Biodiversity Informatics*, **5**, 14–19.
- Lobo, J.M., Verdú, J.R. & Numa, C. (2006) Environmental and geographical factors affecting the Iberian distribution of flightless *Jekelius* species (Coleoptera: Geotrupidae). *Diversity and Distributions*, **12**, 179–188.
- Lobo, J.M., Baselga, A., Hortal, J., Jiménez-Valverde, A. & Gómez, J.F. (2007) How does the knowledge about the spatial distribution of Iberian dung beetle species accumulate over time? *Diversity and Distributions*, **13**, 772–780.
- Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.
- McPherson, J.M. & Jetz, W. (2007) Effects of species' ecology on the accuracy of distribution models. *Ecography*, **30**, 135–151.
- Meynard, C.N. & Quinn, J.F. (2007) Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, **34**, 1455–1469.
- Peterson, A.T. (2007) Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics*, **3**, 59–72.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.

- Raes, N. & ter Steege, H. (2008) A null-model for significance testing of presence-only species distribution models. *Ecography*, **30**, 727–736.
- Real, L.A. & Levin, S.A. (1991) Theoretical advances: the role of theory in the rise of modern ecology. *Foundations of ecology: classic papers with commentaries* (ed. by L.A. Real and J.H. Brown), pp. 177–191. University of Chicago Press, Chicago, IL.
- Real, R., Barbosa, A.M. & Vargas, J.M. (2006) Obtaining environmental favourability functions from logistic regression. *Environmental and Ecological Statistics*, **13**, 237–245.
- Sax, D.F., Stachowicz, J.J., Brown, J.H., Bruno, J.F., Dawson, M.N., Gaines, S.D., Grosberg, R.K., Hastings, A., Holt, R.D., Mayfield, M.M., O'Connor, M.I. & Rice, W.R. (2007) Ecological and evolutionary insights from species invasions. *Trends in Ecology and Evolution*, **22**, 465–471.
- Segurado, P. & Araújo, M.B. (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**, 1555–1568.
- Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distribution of species. *Ecology Letters*, **10**, 1115–1123.
- Soberón, J. & Peterson, A.T. (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London B*, **359**, 689–698.
- Soberón, J. & Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1–10.
- Stockwell, D. & Peters, D. (1999) The GARP modeling system: problems and solutions to automated spatial prediction. *International Journal of Geographic Information Science*, **13**, 143–158.
- Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. (2007) A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, **13**, 397–405.

Editor: Dr Wilfred Thuiller