BIODIVERSITY LETTER

New measures for assessing model equilibrium and prediction mismatch in species distribution models

A. Márcia Barbosa[1,2]*, Raimundo Real[3], A.-Román Muñoz[4,5] and Jennifer A. Brown[6]

[1]*'Rui Nabeiro – Delta Cafés' Biodiversity Chair, Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), University of Évora, 7004-516 Évora, Portugal,* [2]*Division of Biology, Imperial College London, Silwood Park Campus, Ascot (Berkshire) SL5 7PY, UK,* [3]*Biogeography, Diversity and Conservation Lab, Department of Animal Biology, Faculty of Sciences, University of Málaga, 29071, Málaga, Spain,* [4]*Fundación MIGRES, N-340, Km. 96 Huerta Grande, Pelayo, E-11390, Algeciras, Spain,* [5]*Área de Didáctica de las Ciencias Experimentales, Departamento de Didáctica de la Matemática, de las Ciencias Sociales y de las Ciencias Experimentales, Faculty of Education Sciences, University of Malaga, E-29071, Malaga, Spain,* [6]*Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand*

Correspondence: A. Márcia Barbosa, Cátedra 'Rui Nabeiro – Delta Cafés' Biodiversidade, Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), Universidade de Évora, 7004-516 Évora, Portugal.
E-mail: barbosa@uevora.pt

ABSTRACT

Models based on species distributions are widely used and serve important purposes in ecology, biogeography and conservation. Their continuous predictions of environmental suitability are commonly converted into a binary classification of predicted (or potential) presences and absences, whose accuracy is then evaluated through a number of measures that have been the subject of recent reviews. We propose four additional measures that analyse observation-prediction mismatch from a different angle – namely, from the perspective of the predicted rather than the observed area – and add to the existing toolset of model evaluation methods. We explain how these measures can complete the view provided by the existing measures, allowing further insights into distribution model predictions. We also describe how they can be particularly useful when using models to forecast the spread of diseases or of invasive species and to predict modifications in species' distributions under climate and land-use change.

Keywords

Commission, distribution equilibrium, model evaluation, omission, over-prediction, prediction error, predictive performance, under-prediction.

INTRODUCTION

Models based on species distributions are increasingly used in ecology, conservation and management, serving a number of important purposes (see e.g. Jiménez-Valverde & Lobo, 2007 for a brief review). The predictions from such models, usually continuous values of environmental suitability or similar, are often converted into a binary classification of presence or absence, determined by a threshold above which the model is considered to predict the species to be present (Jiménez-Valverde & Lobo, 2007; Nenzén & Araújo, 2011). After this binary conversion, a confusion matrix (Fig. 1) can be generated from the numbers of observed and predicted presences and absences (e.g. Fielding & Bell, 1997; Manel et al., 2001; Anderson et al., 2003). From this matrix, several measures can be calculated to evaluate the capacity of a model to correctly classify presences and absences, including

measures of match and of mismatch between predictions and observations; such measures have been recently reviewed (Liu et al., 2009, 2011). Among the measures of mismatch are the omission and commission rates (Anderson et al., 2003), also known as false-negative and false-positive rates (Fielding & Bell, 1997; Liu et al., 2009, 2011): omission refers to species' presences that are missed by the model (i.e. classified as absences), and commission refers to the presences that are predicted outside the area where the species was observed (i.e. absences classified as presences).

We would first like to point out that, although these measures (especially omission) are commonly referred to as errors (e.g. Guisan & Zimmermann, 2000; Teixeira et al., 2001; Anderson et al., 2003; Bulluck et al., 2006; Elith et al., 2006; Liu et al., 2011; Nenzén & Araújo, 2011; Peterson et al., 2011), neither omission nor commission are necessarily shortcomings of a model. Models are meant to infer,

Figure 1 Confusion matrix showing match (white background) and mismatch (grey background) between observed and predicted presences and absences of a modelled species' distribution. Encircled are the elements used to calculate the omission and commission rates (dashed lines; Anderson *et al.*, 2003) and the proposed under- and over-prediction rates (solid lines).

from the recorded distribution, the environmentally suitable areas for the species. As we detail below, a species may be absent from suitable areas, or present in less adequate areas, without this meaning that the model has made a mistake (see also Sillero *et al.*, 2010).

Omission (presences not predicted by the model), while being more likely to reflect prediction error than commission, may also result from errors of identification or georeferencing of particular species records, as no data set can be deemed completely error free. Omissions may also reveal areas where a species is present under suboptimal conditions (e.g. sinks in the source–sink theory; Pulliam, 1988) due to spatially contagious processes such as dispersal or immigration. In the case of generalist or widespread species, it is common to observe presences in regions below the putative presence–absence (or suitable–unsuitable) threshold, as well as absences above this threshold, because generalists can usually tolerate a wider range of environmental conditions, and effective thresholds are difficult to define.

Commission (presences predicted outside the observed occurrence area) can point to areas where the modelled species occurs but has not been detected or sufficiently surveyed (again, no data set is guaranteed to be complete and error free). Commissions may also represent suitable areas to where the species has not managed to disperse (due to physical barriers, insufficient dispersal ability or lack of time), or where it has become temporarily extinct due to recent disturbance events (e.g. suitable unoccupied patches in metapopulation theory; Levins, 1969); or areas that are suitable on the basis of the environmental variables that were included in the model, but that are unsuitable on the basis of other factors such as biotic interactions (Anderson *et al.*, 2003; Real *et al.*, 2009; Barbosa *et al.*, 2009, 2010).

Hence, rather than a drawback, model misclassifications can allow the extraction of ecological and evolutionary inferences by comparison of the observed and the predicted (potential) distributions of species (Anderson *et al.*, 2003). As such, omission and commission should generally be referred to as rates rather than errors; this may also help in distinguishing error associated with the accuracy of the field data.

That said, additional informative measures can be calculated, regarding under- or over-predicted presences and absences, that are not included in the published reviews on the evaluation measures of binary-converted models (Fielding & Bell, 1997; Liu *et al.*, 2009, 2011). We present four new measures that can be added to the existing suite of model evaluation metrics and provide useful insights into the potential or predicted distributions of species.

RATIONALE AND CALCULATION

The omission and commission rates are calculated in relation to the observed data: omission is the proportion of predicted absences in the recorded presence area, and commission is the proportion of predicted presences in the observed (or assumed) absence area (Fielding & Bell, 1997; Anderson *et al.*, 2003; Liu *et al.*, 2009, 2011). In other words, omission and commission measure how many of the observations are incorrectly classified by the model. Omission is calculated based on the number of observed presences, and commission is calculated based on the number of observed/assumed absences (Fig. 1).

However, this procedure may pose some problems. Firstly, the omission and commission rates are the complements of model sensitivity and specificity (i.e. the proportions of correctly classified presences and absences, respectively), which are widely used in species distribution modelling. Hence, if we have sensitivity $[Se = a/(a + c)]$, the omission rate is redundant $(Om = c/(a + c) = 1 - Se)$, and the same goes for specificity $[Sp = d/(b + d)]$ and the commission rate $(Co = b/(b + d) = 1 - Sp$; see Fig. 1 for the meanings of a, b, c and d).

Secondly, calculating omission and commission in this manner can sometimes lead to unrealistic assessments of model fit. For example, for a species with a restricted distribution within the studied territory, even a model that predicts more than twice the number of recorded presences may exhibit a low commission rate, given the high number of (assumed) absences relative to which this rate is calculated (e.g. Teixeira *et al.*, 2001).

Thirdly, as a result of the frequent and generally recommended procedure of optimizing the binary conversion threshold to maximize both sensitivity and specificity (or to minimize the difference between the two), generally with a preference towards sensitivity (Manel *et al.*, 2001; Jiménez-Valverde & Lobo, 2007), sensitivity and specificity often show similar values, with sensitivity being slightly higher. Therefore, their complements omission and commission also take similar values, with commission being generally slightly higher. It is thus difficult to gauge, from omission and commission, whether a model mainly tends to either under- or over-predict a species' distribution.

We propose two additional measures, the under-prediction and over-prediction rates (UPR and OPR, respectively), that approach the problem from a different angle and are calculated relative to the predicted rather than the observed data:

under-prediction refers to the proportion of observed presences in the predicted absence area, and over-prediction refers to the proportion of observed/assumed absences in the predicted presence area:

$$\text{Under} - \text{Prediction Rate(UPR)} = \frac{\text{unsuitable \& occupied}}{\text{unsuitable}}$$
$$= \frac{c}{c+d}$$

$$\text{Over} - \text{Prediction Rate(OPR)} = \frac{\text{suitable \& unoccupied}}{\text{suitable}}$$
$$= \frac{b}{a+b}$$

where a, b, c and d are the elements of the confusion matrix (Fig. 1). In other words, these rates measure the proportion of predictions that are not matched by observations, rather than the proportion of observations that are not correctly predicted. Under-prediction is calculated based on the number of predicted presences, while over-prediction is calculated based on the number of predicted absences (Fig. 1). The under-prediction rate assesses the probability that the species occurs at a place where the model predicts it to be absent; the over-prediction rate assesses the probability that the species is not found at a place where the model predicts it to occur. These measures, which were not included in previous reviews of model evaluation statistics (Fielding & Bell, 1997; Liu *et al.*, 2009, 2011), provide additional information on observation/prediction mismatch, over and above the customary measures of sensitivity and specificity (and omission and commission).

The under- and over-prediction rates are the complements of the negative and positive predictive power (NPP and PPP; Fielding & Bell, 1997), also called negative and positive predictive value (NPV and PPV; Liu *et al.*, 2009, 2011), respectively. However, although NPP and PPP are relatively popular in fields such as medical diagnostics, they are seldom used in species distribution modelling (Liu *et al.*, 2009). This could be because NPP and PPP are measures of goodness of fit, for which distribution modellers tend to prefer sensitivity and specificity. Distribution modellers are, however, interested in counterbalancing sensitivity and specificity with measures of disagreement between predictions and observations. While omission and commission are not suitable in this case, given that they do not add any information to sensitivity and specificity, the under- and over-prediction rates are useful to assess lack of model fit while completing the view provided by sensitivity and specificity.

Two further measures can be calculated from elements of the confusion matrix (Fig. 1) and added to the existing model evaluation toolset: the potential presence increment (PPI), that is, the proportional increase (positive values) or decrease (negative values) in the number of potential (predicted) relative to observed presences (see also Muñoz & Real, 2006; who calculated a similar measure based on the ratio of predicted to observed presences); and the potential absence increment (PAI), that is, the proportional increase (positive values) or decrease (negative values) in the number of potential relative to observed/assumed absences (a, b, c and d are the elements of the confusion matrix, Fig. 1):

$$\text{Potential Presence Increment(PPI)} = \frac{\text{suitable}}{\text{occupied}} - 1$$
$$= \frac{a+b}{a+c} - 1$$

$$\text{Potential Absence Increment(PAI)} = \frac{\text{unsuitable}}{\text{unoccupied}} - 1$$
$$= \frac{c+d}{b+d} - 1$$

A PPI or PAI of zero would mean no difference between the total number (irrespective of the location) of observed and predicted presences or absences, respectively; a positive or negative value would measure how much the potential occurrence (or the potential non-occurrence) area exceeds the actually occupied (or the actually unoccupied) area. Depending on the ecological and biogeographical characteristics of the species under analysis, these measures may be useful when predicting the spread of diseases or invasive species, the potential habitat to be occupied by species colonizing new areas or the evolution of species' distributions under climate and land-use change scenarios.

CASE STUDIES AND POTENTIAL APPLICATIONS

We illustrate the use of these measures on the Iberian mole (*Talpa occidentalis*), an insectivorous mammal endemic to the Iberian Peninsula (SW Europe), whose distribution in Spain was modelled previously (Ribas *et al.*, 2006; Fig. 2). More details on the data and modelling method are provided in Appendix S1 (see Supporting Information), where we also describe a series of additional case studies on species with varying range sizes (restricted to widespread) and biogeographical characteristics (native, invasive, metapopulational).

For the Iberian mole, omission and commission (like their complements sensitivity and specificity) become balanced near the 0.5 favourability threshold, and their values, which are calculated relative to the observed occurrence area, denote high model accuracy. However, from the perspective of the predicted occurrence area, over-prediction is substantial at the same threshold, with 66% of the predicted occurrence area not being actually occupied. The potential presence increment is also relatively high at this threshold, as the model predicts more than twice the observed occurrence area. Equilibrium between observed and predicted occupancy is not attained until the 0.72 threshold, where the potential increments in presences and absences approach zero (Fig. 2).

Further insights arise from analysing species with varying prevalence or relative occurrence area (see Appendix S1). While omission and commission (following sensitivity and specificity) had similar values for medium thresholds within every model, the under- and over-prediction rates were often
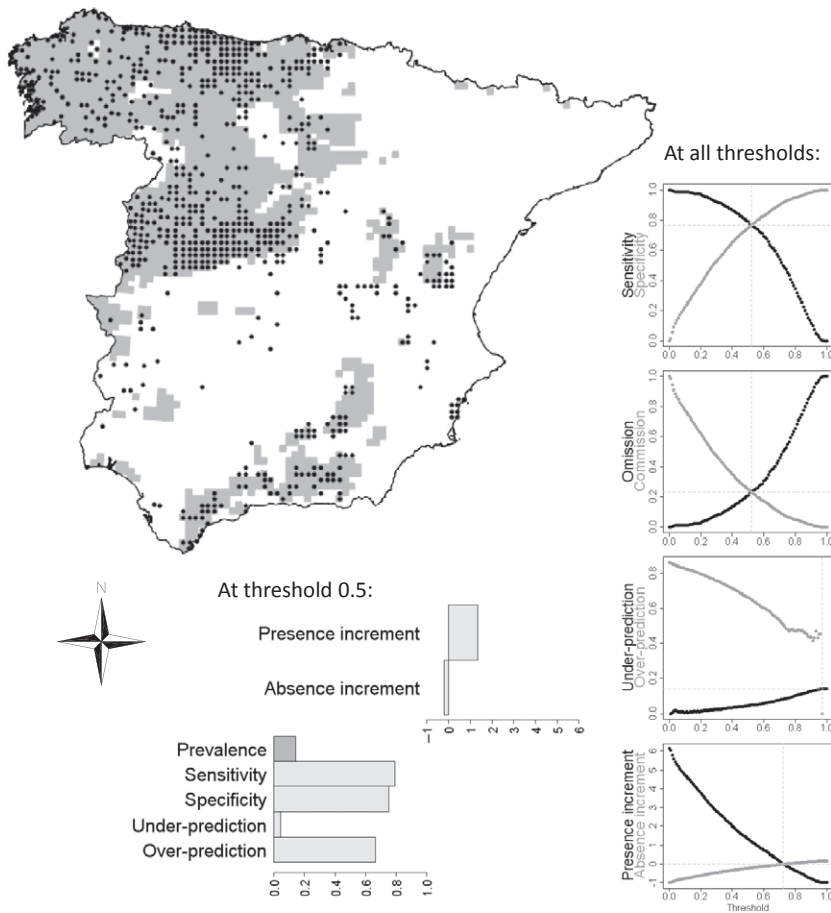
Figure 2 Observed (black; Palomo & Gisbert, 2002) and potential presences (grey; Ribas *et al.*, 2006) of Iberian mole *Talpa occidentalis* on UTM 10 × 10-km grid cells of mainland Spain; species prevalence and model evaluation measures at the 0.5 environmental favourability threshold; and variation of these measures along the entire range of possible thresholds.

visibly different from each other. Moreover, over-prediction was higher than under-prediction for some species and lower than under-prediction for others (Figs S1 and S2) and, except for the most widespread species, this occurred along most of the range of possible thresholds separating predicted presences from predicted absences (Figs S3 and S4).

For restricted-range species, although commission rates were low (following the high specificity), over-prediction rates were substantial, reflecting the fact that a high proportion of the predicted favourable areas are not actually occupied (Figs S1 and S3). The most widespread species, on the other hand, have relatively high rates of under-prediction (unfavourable localities that are actually occupied), despite the substantially lower omission (Figs S2 and S4). Equilibrium between potential and occupied area (i.e. null presence and absence increment) is achieved at very high favourability thresholds for restricted species (Fig. S3). This reflects specialists with low-entropy distributions, requiring excellent environmental thresholds to occupy the whole suitable area; under those thresholds, there are always more favourable than actually occupied sites. As species prevalence increases, this equilibrium threshold decreases, approaching the sensitivity–specificity balance threshold. Middle favourability thresholds thus provide equilibrium between potential and observed distributions for these widespread species, reflecting less environmentally demanding occurrence patterns (Fig.

S4). This information is not provided by the omission–commission (nor by the sensitivity–specificity) plots. The proposed measures thus allow further insights into the models' tendency for either under- or over-predicting species' occurrence areas, from a novel point of view, independently of the information provided by omission and commission (or their complements sensitivity and specificity).

The measures presented here, as well as their variation across the range of decision thresholds (Fig. 2), can be easily calculated for analogous data sets (binary observations versus continuous predictions) with the *modEvA* package for R (Barbosa *et al.*, 2013), which is currently in beta version. Until a stable version is officially released, the package (along with a set of simple instructions for users inexperienced with R) is available upon request to the authors.

CONCLUDING COMMENTS

Model performance measures such as sensitivity and specificity should be complemented with assessments of prediction mismatch, which does not necessarily indicate model failure and is useful for understanding species' distributions, their equilibrium with the environment, or their potential for change. While omission and commission do not add any information to the widely used sensitivity and specificity, the proposed under- and over-prediction rates analyse the problem from a different

perspective, by assessing prediction mismatch over the potential rather than the observed occurrence area. They thus allow more complete assessments of model classification performance. In species distribution modelling, where sensitivity and specificity tend to be optimized, under-prediction and over-prediction can be particularly useful to assess misclassification rates without repeating information. The potential increments in presences and absences, in addition, measure the equilibrium between the observed and the potential area of occupancy – that is, between the model and the species' distribution.

There is an increasing use of models for forecasting modifications in species' distributions under climate and land-use changes, especially to inform conservation planning for threatened species, as well as a growing interest in predicting and monitoring the spread of diseases and invasive species. We expect these measures to be particularly useful in such studies, as they assess how the potential distribution compares with the observed one and may thus provide clues on how a species' range may be expected to expand or contract.

ACKNOWLEDGEMENTS

REFERENCES

Anderson, R.P., Lew, D., Peterson, A.T.T., Anderson, R.P., Lew, D. & Peterson, A.T. (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, **162**, 211–232.

Barbosa, A.M., Brown, J.A., & Real, R. (2013) modEvA – an R package for model evaluation and analysis. Beta version, available upon request.

Barbosa, A.M., Real, R. & Vargas, J.M. (2009) Transferability of environmental favourability models in geographic space: the case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. *Ecological Modelling*, **220**, 747–754.

Barbosa, A.M., Real, R. & Vargas, J.M. (2010) Use of coarse-resolution models of species' distributions to guide local conservation inferences. *Conservation Biology*, **24**, 1378–1387.

Bulluck, L., Fleishman, E., Betrus, C. & Blair, R. (2006) Spatial and temporal variations in species occurrence rate affect the accuracy of occurrence models. *Global Ecology and Biogeography*, **15**, 27–38.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Jiménez-Valverde, A. & Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta Oecologica*, **31**, 361–369.

Levins, R. (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, **15**, 4.

Liu, C., White, M. & Newell, G. (2009) Measuring the accuracy of species distribution models: a review. *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Cairns, Australia 13-17 July 2009*, (ed. by R.S. Anderssen, R.D. Braddock and L.T.H. Newham), pp. 4241–4247. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, Cairns, Australia.

Liu, C., White, M. & Newell, G. (2011) Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography*, **34**, 232–243.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921–931.

Muñoz, A.-R. & Real, R. (2006) Assessing the potential range expansion of the exotic monk parakeet in Spain. *Diversity and Distributions*, **12**, 656–665.

Nenzén, H.K. & Araújo, M.B. (2011) Choice of threshold alters projections of species range shifts under climate change. *Ecological Modelling*, **222**, 3346–3354.

Palomo, L.J. & Gisbert, J. (2002) *Atlas de los mamíferos terrestres de España*. Dirección General de Conservación de la Naturaleza-SECEM-SECEMU, Madrid.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological Niches and Geographic Distributions*. University Press, Princeton and Oxford.

Pulliam, H.R. (1988) Sources, sinks, and population regulation. *The American Naturalist*, **132**, 652–661.

Real, R., Barbosa, A.M., Rodríguez, A., García, F.J., Vargas, J.M., Palomo, L.J. & Delibes, M. (2009) Conservation biogeography of ecologically interacting species: the case of the Iberian lynx and the European rabbit. *Diversity and Distributions*, **15**, 390–400.

Ribas, A., Barbosa, A.M., Casanova, J.C., Real, R., Feliu, C. & Vargas, J.M. (2006) Geographical patterns of the species richness of helminth parasites of moles (Talpa spp.) in Spain: separating the effect of sampling effort from

those of other conditioning factors. *Vie et Milieu*, **56**, 1–8.

Sillero, N., Barbosa, A.M., Martínez-Freiría, F. & Real, R. (2010) Los modelos de nicho ecológico en la herpetología ibérica: pasado, presente y futuro. *Boletín de la Asociación Herpetológica Española*, **21**, 2–24.

Teixeira, J., Ferrand, N. & Arntzen, J.W. (2001) Biogeography of the golden-striped salamander Chioglossa lusitanica: a field survey and spatial modelling approach. *Ecography*, **24**, 618–624.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Case studies on species with varying prevalence.

Figure S1 Observed (black) and potential (grey) presences of restricted-range (including an invasive) species on UTM 10×10-km grid cells of mainland Spain, their prevalence and model evaluation measures.

Figure S2 Observed (black) and potential (grey) presences of relatively widespread species on UTM 10×10-km grid cells of mainland Spain, their prevalence and model evaluation measures.

Figure S3 Variation of the four analysed evaluation measure pairs along the range of possible thresholds (at 0.01 intervals) for the four restricted-range species analysed.

Figure S4 Variation of the four analysed evaluation measure pairs along the range of possible thresholds (at 0.01 intervals) for the more widespread species analysed.

BIOSKETCH

A. Márcia Barbosa is a postdoctoral fellow jointly hosted by CIBIO – University of Évora (Portugal) and Imperial College London (UK) – and was a visitor to the University of Canterbury (New Zealand) while preparing this article. Her research interests include biogeography, macroecology, distribution modelling, comparative phylogeography, biodiversity patterns and conservation. The team have a common interest in the analysis of biogeographical patterns and its applications to conservation and management.

Author contributions: A.M.B. and R.R. conceived the ideas. A.M.B. gathered and analysed the data, programmed the R functions and led the writing. A.R.M., J.A.B. and R.R. provided ideas for additional analyses and improved writing, interpretation and presentation.

Editor: Mark Robertson