Title

thesis abstract: Applying species distribution modeling for the conservation of Iberian protected invertebrates

Permalink

https://escholarship.org/uc/item/3mx5p9pz

Author

Chefaoui, Rosa María

Applying species distribution modeling for the conservation of Iberian protected invertebrates

Rosa María Chefaoui

PhD Thesis, Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, c/ José Gutiérrez Abascal 2, 28006 Madrid, Spain.
e-mail: rosa.chef@gmail.com; http://www.biogeografia.org/

Abstract. This article outlines the approaches to modeling the distribution of threatened invertebrates using data from atlases, museums and databases. Species Distribution Models (SDMs) are useful for estimating species' ranges, identifying suitable habitats, and identifying the primary factors affecting species' distributions. The study tackles the strategies used to obtain SDMs without reliable absence data while exploring their applications for conservation. I examine the conservation status of *Copris* species and *Graellsia isabelae* by delimiting their populations and exploring the effectiveness of protected areas. I show that the method of pseudo-absence selection strongly determines the model obtained, generating different model predictions along the gradient between potential and realized distributions. After assessing the effects of species' traits and data characteristics on accuracy, I found that species are modeled more accurately when sample sizes are larger, no matter the technique used.

Keywords: Environmental niche modeling, Iberian Peninsula, invertebrates, predictive accuracy, species distribution models

The rapid disappearance of habitats and species starkly contrasts the need to conserve biodiversity against our inability to inventory and protect all species individually. Knowledge about biodiversity remains insufficient because many species are still not described (the "Linnean Shortfall"; Brown and Lomolino 1998) and the distributions of described species often are inadequately defined (the "Wallacean Shortfall"; Lomolino 2004). It is therefore essential to identify threatened species and describe their distributions using approaches that overcome the time and budget constraints of systematic conservation planning.

Araújo et al. (2007) demonstrated the need for additional protected areas for the effective conservation of the diversity of plants and vertebrates in the Iberian Peninsula. Preliminary data suggest that the existing network of reserves also would be ineffective in representing invertebrate species (Verdú and Galante 2009). Unfortunately, the conservation of invertebrates faces serious challenges due to their high diversity, complex life cycles and difficult taxonomy, among other factors (see New 1998).

Geographic Information Systems (GIS) significantly advanced the conservation of endangered species because they allow us to delimit species' potential distributions (e.g. Hortal et al. 2005), to control their populations (e.g. Davies et al. 2005), to analyze their niche (Peterson et al. 2002), design networks of protected areas (e.g. Pearce and Boyce 2006), and to forecast the future (e.g. Hill et al. 2002). Together, the databases taken from atlases, museums and herbaria have emerged as a valuable source of species' occurrence records (e.g. Elith and Leathwick 2007). Unfortunately, these data from heterogeneous sources may contain errors or have been obtained using a biased sampling procedure (Hortal et al. 2007, 2008, Newbold 2010). Besides, they do not usually provide reliable absences needed to perform consistent predictive models (Anderson et al. 2003, Lobo et al. 2007), so alternatives have been sought generating models based only on presences (Hirzel et al. 2002, Pearce and Boyce 2006), sometimes employing pseudo-absences obtained in different ways (Zaniewski et al. 2002, Engler et al. 2004,

Lobo et al. 2006, 2010).

For my doctoral thesis, I evaluated the utility of SDMs for the conservation of threatened invertebrates in the Iberian Peninsula (Chefaoui 2010). The majority of the species studied here have been designated by the European Union as species of "community interest" requiring protection and conservation (Habitats Directive). I used presence-only data on Iberian threatened invertebrates obtained from museums, atlases and databases. I applied presence-only methods such as ENFA (Ecological Niche Factor Analysis) and MDE (Multi-Dimensional Niche Envelope), in addition to other methods that require presences and absences (here, pseudo-absences): GAM (Generalized Additive Models), GLM (Generalized Linear Models) and NNET (Neural Networks Models). I approached methodological issues concerning the difficulties associated with predicting the distribution of species when reliable absence data are not available, and explored the possibilities of SDMs as a tool for conservation of endangered and threatened Iberian invertebrates. In this respect, I explored the applications of SDM to estimate species ranges, identify suitable habitats and the primary factors affecting species' distribution in order to assess the conservation status of threatened invertebrates.

Dung beetle populations, which are in decline in the Iberian Peninsula, play a critical ecological role in extensive pasture ecosystems by recycling organic matter. We delimited the potential distribution of the two species of *Copris* (Coleoptera, Scarabaeidae) that inhabit the Iberian Peninsula using ENFA (Chefaoui et al. 2005). ENFA is a presence-only method that compares the environmental values of the localities where the species has been observed with respect to the environmental values of the territory studied (Hirzel et al. 2002). We explored the environmental niche occupied by each species in a small region, the Community of Madrid (CM), to restrict the role of dispersal constraints discriminating possible areas of co-occurrence and identifying the specific environmental characteristics of each species. We identified that solar radiation and the presence of calcareous soils are critical to the

presence of *Copris hispanus*, while *Copris lunaris* requires siliceous soils and high rainfall. Both *Copris* species are distributed along a geographic and environmental gradient from the Tajo basin (warmer, dryer, with strong annual weather variations) where only *C. hispanus* is found, towards the mountain slopes of the Sistema Central (colder, higher rainfall) where *C. lunaris* predominates. The environmental niches of both species are distributed along a Dry-Mediterranean to Wet-Alpine axis, and overlap in areas of moderate temperatures and precipitations in the north of CM.

We also studied the degree of protection of key populations of *C. hispanus* and *C. lunaris*, making a proposal to improve their conservation. To evaluate the conservation status of *Copris* species, we took into account the size of protected sites as well as the values of habitat suitability in each protected natural site and Natura 2000 network. We found that *Copris* species were poorly conserved in the previous protected sites network: for *C. hispanus* only two protected sites measured around 30 km^2, and for *C. lunaris* a single area measured 183 km^2. However, protection provided by Sites of Community Importance (SCIs) seems to improve the general conservation status of these species in CM because the area and connectivity of protected sites have been increased substantially.

Chefaoui and Lobo (2008) assessed the effects of pseudo-absences on model performance when reliable absence data are not available. We compared seven procedures to generate pseudo-absence data to be used in GLM-logistic regressed models. These pseudo-absences were selected randomly or by means of presence-only methods (ENFA and MDE) to model the distribution of a threatened endemic Iberian moth species (*Graellsia isabelae*). Our purpose was to show the possibility of achieving different forecasted distributions depending on the method and the threshold used to select these pseudo-absences.

The results showed that the pseudo-absence selection method greatly influenced the percentage of explained variability, the scores of the accuracy measures and, most importantly, the

predicted range size. As we extracted pseudo-absences from environmental regions further from the optimum established by presence data, the models obtained better accuracy scores, and over-prediction increased. Conversely, the profile techniques that generated wider unsuitable areas, produced functions with lower percentages of explained deviance and poorer accuracy scores, but more restricted predictive distribution maps, similar to the observed distribution. The random selection of pseudo-absences generated the most constrained predictive distribution map.

Based on results of the aforementioned work, we identified the environmental variables most relevant for explaining the distribution of *Graellsia isabelae* and assessed this species' conservation status (Chefaoui and Lobo 2007). We modeled the potential distribution of the insect by performing GLM with pseudo-absence data selected from an ENFA model. We found that the best predictor variables were summer precipitation (ranging from 1250 mm to 3250 mm), aridity, and mean elevation. This species prefers habitats with mid-range mountain conditions. With respect to host plants, the presence of *G. isabelae* was associated mainly with *Pinus sylvestris* and *P. nigra*.

Moreover, we found 8 areas exclusively in the eastern Iberian territory, and a larger unoccupied habitat in the western Iberian Peninsula, indicating that this species is probably not in equilibrium with its environment because of historical factors (Chefaoui and Lobo 2007). We suggested that the current distribution of the species was associated with the dynamism of its host plants during glacial periods of the Holocene, when the forests of *Pinus sylvestris* decreased strongly in the northwestern part of the peninsula. After analyzing the possibility of connectivity and fragmentation of the eight populations delimited as well as the degree of protection of *G. isabelae* on the SCIs, we found that the SCIs under protection did not seem sufficient to maintain current populations. Moreover, our study rejected the idea that the species was expanding its range due to reforestation. Because the conservation of *G. isabelae* depends on the forests of *Pinus sylves-*

tris and *P. nigra* located both inside and near to SCIs, we suggested that the reintroduction of the species in these habitats could improve its conservation.

To understand the limitations and possibilities of SDM techniques, we evaluated the effects of species' traits and data characteristics on the accuracy of SDMs for red-listed invertebrates (Chefaoui et al. 2011). We applied three SDM techniques (GAM, GLM and NNET) using pseudo-absences to model the distribution of 20 threatened Iberian invertebrates. We correlated the accuracy of the obtained models with several data characteristics and species' ecological traits. We examined two data characteristics, the amount of data (N) and the relative occurrence area (ROA), and both significantly affected the accuracy of the models. Greater AUC values and higher sensitivity scores were obtained from samples for which there were more than 200 records. In general, species whose distributions were most accurately modelled were those with a greater sample size or smaller ROA. In addition, species related to habitats that are problematic to detect using GIS data, such as riparian or humid areas, seemed to be more difficult to predict.

Summary

The performance of SDMs depends on the type of data and the characteristics of the species. Presence-only methods (ENFA and MDE) achieved worse validation results and overpredicted more than techniques using pseudo-absences. Nevertheless, presence-only methods can be very useful for obtaining pseudo-absences and discovering the environmental response of species. The method of pseudo-absence selection strongly determined the predicted range size, generating different model predictions along the gradient between potential and realized distributions. There is an added difficulty in obtaining predictions that closely approximate the realized distribution of species under non-equilibrium conditions, because both presence and absence data may be possible under similar environmental conditions. Irrespective of the approach used, species' distributions are modelled more accurately when sam-

ple sizes are larger. Species in habitats that are difficult to detect using GIS data, such as riparian species, thus may tend to be more difficult than most to predict.

Availability of thesis

Printed and PDF copies are available in the Science Faculty Library, Universidad Autónoma de Madrid (http://biblioteca.uam.es/ciencias/). A PDF copy is also available at request from the author.

Acknowledgements

I would like to thank my two supervisors, Jorge M. Lobo and Joaquín Hortal for their support and encouragement.

References

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.

Araújo, M.B., Lobo, J.M. & Moreno, J.C. (2007) The effectiveness of Iberian protected areas in conserving terrestrial biodiversity. Conservation Biology, 21, 1423–1432.

Brown, J.H. & Lomolino, M.V. (1998) Biogeography, 2nd edn. Sinauer Press, Sunderland, Massachusetts.

Chefaoui, R.M., Hortal, J., & Lobo, J.M. (2005) Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian Copris species. Biological Conservation, 122, 327–338.

Chefaoui, R.M. & Lobo, J.M. (2007) Assessing the conservation status of an Iberian moth using pseudo-absences. The Journal of Wildlife Management, 8, 2507–2516.

Chefaoui, R.M. & Lobo, J.M. (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. Ecological Modelling, 210, 478–486.

Chefaoui, R.M. (2010) Modelos predictivos aplicados a la conservación de invertebrados protegidos ibero-baleares. Ph.D. Thesis. Universidad Autónoma de Madrid, Departamento de Biología, Facultad de Ciencias, 196 pp.

Chefaoui, R.M., Lobo, J.M. & Hortal J. (2011) Effects of species' traits and data characteristics on distribution models of threatened invertebrates. Animal Biodiversity and Conservation, 34, (in press).

Davies, Z.G., Wilson, R.J., Brereton, T.M. & Thomas, C.D. (2005) The re-expansion and improving status of the silver-spotted skipper butterfly (Hesperia comma) in Britain: a metapopulation success story. Biological Conservation, 124, 189–198.

Elith, J. & Leathwick, J.R. (2007) Predicting species' distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. Diversity and Distributions, 13, 165–175.

Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology, 41, 263–274.

Hill, J.K., Thomas, C.D., Fox, R., Telfer, M.G., Willis, S.G., Asher, J. & Huntley, B. (2002) Responses of butterflies to twentieth century climate warming: implications for future ranges. Proceedings of the Royal Society, 269, 2163–2171.

Hirzel, A., 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., Borges, P.A.V., Dinis, F., et al. (2005) Using ATLANTIS – Tierra 2.0 and GIS environmental information to predict the spatial distribution and habitat suitability of endemic species. Direção Regional de Ambiente and Universidade dos Açores, Horta, Angra do Heroísmo and Ponta Delgada, Horta, Faial.

Hortal, J., Lobo, J.M., & Jiménez-Valverde, A. (2007) Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife (Canary Islands). Conservation Biology, 21, 853–863.

Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M., & Baselga, A. (2008) Historical bias in biodiversity inventories affects the observed realized niche of the species. Oikos, 117, 847–858.

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., & Hortal, J. (2010) The uncertain nature of absences and their importance in species distribution modelling. Ecography, 33, 103–114.

Lomolino, M.V. (2004) Conservation biogeography. Frontiers of Biogeography: new directions in the geography of nature (ed. by M. V. Lomolino and L. R. Heaney), pp. 293–296. Sinauer Associates, Sunderland, Massachusetts.
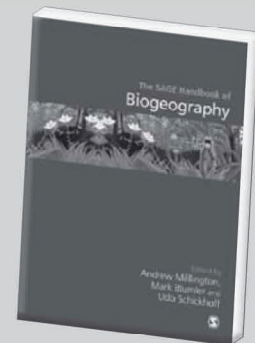
New, T.R. (1998) Invertebrate surveys for conservation. Oxford University Press, New York.

Newbold, T. (2010) Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Progress in Physical Geography, 34, 3–22.

Pearce, J. & Boyce, M.S. (2006) Modelling distribution and abundance with presence-only data. Journal of Applied Ecology, 43, 405–412.

Peterson, A.T., Ball, L.G. & Cohoon, K.P. (2002) Predicting distributions of Mexican birds using ecological niche modelling methods. Ibis, 144, E27–E32.

Verdú, J.R. & Galante, E., eds. (2009) Atlas de los Invertebrados Amenazados de España (Especies en peligro crítico y en peligro). Dirección General para la Biodiversidad, Ministerio de Medio Ambiente, Madrid, 340 pp.

Zaniewski, A.E., Lehmann, A. & Overton, J.M. (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling, 157, 261–280.

Edited by Richard Pearson