workshop summary

The application of species distribution models in the megadiverse Neotropics poses a renewed set of research questions

Species distribution models: applications, challenges and perspectives – Belo Horizonte, Brazil, 29th–30th August 2011

The community of researchers and technicians interested in biogeography is large and growing in Brazil, with members coming from fields as diverse as ecology, evolution and conservation. The increment in the number of postgraduate programs in ecology and evolutionary biology is linked to many research questions about the causes and dynamics of species' ranges, as well as about their consequences for short-term evolutionary processes. As a result, the use (and abuse) of species distribution models (SDMs) as a tool in research and technical studies has grown rapidly in recent years. Systems that integrate biodiversity databases (e.g. SpeciesLink[1]) allow one to obtain distributional information about many species; and easy-to-use SDM software (e.g. MaxEnt[2], Phillips et al. 2004, or openModeller[3], Muñoz et al. 2011) is also available online. The easy availability of data and SDM tools provides a powerful means for answering questions about the geographic distribution of species. Potential users include non-specialist researchers, untrained postgraduate students and government technicians. But is this helping to develop a sound and solidly grounded knowledge of the distribution of Neotropical diversity?

In order to unite the Brazilian community of SDM users and provide them with a better understanding of the technique, the Postgraduate Course on Plant Biology of the Universidade Federal de Minas Gerais organized a workshop in Belo Horizonte last August. This workshop allowed more than 100 students and technicians to meet and discuss with researchers working with SDMs. The main conclusions of the meeting were that there is a growing interest in using the technique to study species' distributions and find unknown populations of rare species, and that there is a need for a code of good practice in both field surveys and SDM applications. These conclusions have been already discussed elsewhere (Kamino et al. 2011). Here we develop further one of the problems identified during the workshop: the lack of clear questions in many studies using SDMs; in other words, the mere application of species distribution modelling as a fashionable technique often 'justifies' a study.

Reviews summarizing the most important challenges for SDMs (e.g. Araújo and Guisan 2006, Zimmermann et al. 2010, Peterson et al. 2011) typically present how users view the field and what problems they perceive in each step in the modelling process. However, things move fast in this emerging field of research because, while the number of studies using SDMs increases steadily (see Fig. 1 in Lobo et al. 2010), the technique is also used to address completely new questions. Thus, the challenges to their application are themselves changing. We believe that the most important of these challenges are theoretically rather than methodologically grounded, although we recognize that both kinds of problems overlap to some extent. Here we outline the basic questions that we believe SDM users must take into account while studying current species' distributions.

Soberón (2007) provides perhaps the best starting point to understand the theoretical problems of SDMs (see also Colwell and Rangel 2009; Soberón 2010). The first problem is the definition of a clear research question. Here it is important to discriminate between theoretical questions such as "Why is this species here?" – which are more interesting in the long run – from those that are eminently practical such as "Where is this species?" – which unfortunately seem to be the most common. But even practical questions stimulated

1 http://www.splink.org.br; last accessed 10/11/2011
2 http://www.cs.princeton.edu/~schapire/maxent/; last accessed 10/11/2011
3 http://openmodeller.sourceforge.net/; last accessed 10/11/2011

by, for instance, the design of conservation programs need to be properly defined before SDMs are used, because they provide the basis for making adequate methodological choices.

Different questions may deserve different methods. One of the recurrent dilemmas for users is choosing the most appropriate modelling technique. However, when the research question is well defined, choosing one particular strategy comes more easily. The organization of SDM techniques in categories of complexity[4], together with acknowledging the advantages and drawbacks of the techniques in each category, may help to choose the best modelling strategy for each research question (Kamino et al. 2011). Since there are different products, they must have different uses. While simple models may help to find sites that are climatically similar to the known occurrences of the species, or be suited to describing its realized niche, more complex models may be best for describing the actual distribution of the species (see Jiménez-Valverde et al. 2008).

The definition of an adequate modelling strategy also raises several issues. Surprisingly, the first is that there may not be a minimum number of occurrences to be used in SDMs (other than they should be higher than zero). What is necessary, however, is that the question tackled is clearly understood and the relationship between the number of occurrences and the SDM technique used is known. Datasets with very few occurrences in combination with simple SDM methods may be useful for providing a description of the areas climatically similar to the recorded presences, allowing the planning of new field surveys (Siqueira et al. 2009) or a preliminary gap analysis to identify areas for conservation (Nóbrega and De Marco Jr 2011). However, larger datasets are needed for more complex methods and/or answering more sophisticated questions about the distribution of the species.

Determining how many predictors should be used to develop the models and how many parameters should be allowed while modelling deserves special attention. Many complex SDM techniques suffer from chronic overfitting – a disease that is easy to catch but hard to cure. Overfitted models may restrict too much the predicted distributions to the geographic and environmental domain of the observed occurrences, impeding identification of areas where the species could be potentially present, or the reliable representation of the realized response of the species to the environment (Lobo 2008).

Another theoretical problem of SDMs is that the current distributions of the species are often not in equilibrium with the environment (Araújo and Pearson 2005, Soberón 2007, 2010, Jiménez-Valverde et al. 2008). This may result from barriers to dispersal, but also from climate change, even in the absence of such barriers, if the dispersal ability of the species is not enough to track the changes in its potential geographic distribution. Similarly, recently originated species are expected to have restricted ranges and not be in equilibrium with the environment, if they lacked the time to disperse to their whole potential distribution. Such lack of equilibrium poses two key questions: how to restrict model predictions to the truly occupied areas (if that is the intention) and how to incorporate dispersal abilities into SDM strategies. The former can be tackled either by defining appropriate pseudo-absences to include in the calibration dataset (Lobo et al. 2010) or by restricting the extent of the study to the area that could have been colonized by the species (Anderson and Raza 2010). Incorporating dispersal abilities may be more difficult, but the incorporation of cellular automata or other spatially explicit methods (e.g. Brook et al. 2009) will generate a new generation of SDMs that may allow study of the spatiotemporal dynamics of species' ranges.

Finally, contrary to the general opinion of many users, the ability of species distribution models to measure the fundamental niche of a species is severely limited. First and foremost, the lack of equilibrium of the species' range with the

4 From simple (DOMAIN, BIOCLIM, Euclidean Environmental Distance) to moderate (Mahalanobis distance, GLM, GAM) and complex models (MaxEnt, Artificial Neural Networks, GARP) (T.F. Rangel unpublished, see also Jiménez-Valverde et al. 2008).

environment may interfere with any attempt to estimate its potential distribution (i.e., the fundamental niche in relation to climate). Gaining insights about the fundamental niche of a species based on environmental information and some recorded occurrences relies on prior knowledge of its requirements and having an adequate theoretical model for the niche itself. Additionally, many researchers tend to use variables only from widely available datasets (such as WorldClim[5]), instead of seeking other predictors that may be less easy – but yet possible – to obtain (Kamino et al. 2011, S. Amaral unpublished). We would like to emphasize that variables representing edaphic conditions are often not available, yet these could be important predictors for some species and are rarely considered in SDMs. These problems, in combination with the collinearity that is intrinsic to most spatially explicit variables, may be impeding the identification of predictors that are truly determining the geographic distribution of many species. It follows that it would be difficult to evaluate the relative importance of different environmental predictors in the absence of a good theoretical model and some knowledge on the physiology and ecology of the studied species. The interaction with students and technicians during the workshop showed us that a good understanding of the framework proposed by Soberón (2007, 2010) helps in clarifying the concepts and their meanings in this respect. But further work and a lot of didactics are needed to actually imprint the next generation of SDM users with a solidly rooted knowledge about the technique they use, the implications of the methodological choices they make and the power that these models actually have to solve theoretical and applied questions in ecology, evolution, biogeography and conservation.

Acknowledgements

We thank all attendees of the workshop for the discussions. The workshop was organized by LHYK and JRS and supported by UFMG (Pró-Reitoria de Pós-Graduação, Instituto Ciências Biológicas, Pós-Graduação em Biologia Vegetal, Genética, Zoologia, and Ecologia, Conservação e Manejo de Vida Silvestre), Conselho Regional de Biologia and Via-mundi. LHYK was financed by CAPES Grant, PDMJr and JRS by continuous CNPq productivity grants and JH by a CNPq Visiting Researcher Grant.

Luciana H. Y. Kamino[1], Paulo De Marco Jr[2], Thiago F. Rangel[2], Silvana Amaral[3], Marinez F. de Siqueira[4], Renato De Giovanni[5], João Renato Stehmann[1] and Joaquín Hortal[2,6]

[1]Depto. Botânica, Instituto Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. lucianakamino@gmail.com; http://www.icb.ufmg.br/pgbot/

[2]Depto. Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil

[3]Divisão de Processamento de Imagens, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil

[4]Unidade de Botânica Sistemática, Instituto de Pesquisa Jardim Botânico do Rio de Janeiro, Rio de Janeiro, Brazil

[5]Centro de Referência em Informação Ambiental, Campinas, Brazil

[6]Depto. Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), Madrid, Spain

References

Anderson, R.P. & Raza, A. (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. Journal of Biogeography, 37, 1378–1393.

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. Journal of Biogeography, 33, 1677–1688.

Araújo, M.B. & Pearson, R.G. (2005) Equilibrium of species' distributions with climate. Ecography, 28, 693–695.

Brook, B.W., Akçakaya, H.R., Keith, D.A., Mace, G.M., Pearson, R.G. & Araújo, M.B. (2009) Integrating bioclimate with population models to improve forecasts of species extinctions under climate change. Biology Letters, 5, 723–725.

Colwell, R.K. & Rangel, T.F. (2009) Hutchinson's duality: The once and future niche. Proceedings of the National Academy of Sciences USA, 106, 19651–19658.

5 http://www.worldclim.org/; last accessed 10/11/2011

ISSN 1948-6596

Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2008) Not as good as they seem: the importance of concepts in species distribution modelling. Diversity and Distributions, 14, 885–890.

Kamino, L.H.Y., Stehmann, J.R., Amaral, S., De Marco Jr, P., Rangel, T.F., de Siqueira, M.F., De Giovanni, R. & Hortal, J. (2011) Challenges and perspectives for species distribution modelling in the Neotropics. Biology Letters, in press. doi:10.1098/rsbl.2011.0942

Lobo, J.M. (2008) More complex distribution models or more representative data? Biodiversity Informatics, 5, 14–19.

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.

Muñoz, M.E.S., Giovanni, R., Siqueira, M.F., Sutton, T., Brewer, P., Pereira, R.S., Canhos, D.A.L. and Canhos, V.P. (2011) openModeller: a generic approach to species' potential distribution modelling. Geoinformatica, 15, 111–135.

Nóbrega, C.C. & De Marco Jr, P. (2011) Unprotecting the rare species: a niche-based gap analysis for odonates in a core Cerrado area. Diversity and Distributions, 17, 491–505.

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. Princeton University Press, Princeton.

Phillips, S.J., Dudík, M. and Schapire, R.E. (2004) A maximum entropy approach to species distribution modeling. In: Proceedings of the 21st International Conference on Machine Learning, ACM-Press, New York, pp. 655–662

Siqueira, M. F., Durigan, G., De Marco Jr, P. and Peterson, A. T. (2009) Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species. Journal for Nature Conservation, 17, 25–32.

Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters, 10, 1115–1123.

Soberón, J. (2010) Niche and area of distribution modeling: A population ecology perspective. Ecography, 33, 159–167.

Zimmermann, N.E., Edwards, T.C., Graham, C.H., Pearman, P.B. & Svenning, J.-C. (2010) New trends in species distribution modelling. Ecography, 33, 985–989.

Edited by Richard Field

