ISSN 1948-6596

update

Choosing the right path for species distribution modeling

Species distribution models (SDM) have been widely used to address the lack of knowledge about species' distributions (i.e. the so-called Wallacean shortfall; Lomolino 2004). However, SDM show great variability in their predictions due to the uncertainties that accumulate during the modeling process (Barry & Elith 2006). These uncertainties are generally separated in three major classes: data, model and predictions. Uncertainty in the data refers to variation arising from incomplete knowledge about the patterns and processes studied, bias in the sampling procedure, and quality and choice of covariates; it applies to data used to build and fit the models, as well as to assess their success. Model uncertainty is due to discrepancies in the assumptions and algorithms used to fit the data. Finally, given that the true distribution is unknown and the relationship between species' presence and the environment may differ in studied and unstudied geographic regions, we are uncertain whether a prediction is a perfect fit and even if it were, it may not be so at a different time or place. Currently, SDM applications lack approaches to manage such uncertainty, which could result in using them without considering the reliability of their outputs with the subsequent cost in the quality of the research based in these analyses. If, on the contrary, such uncertainty is acknowledged but is thought to be unmanageable, SDM would be flagged as unreliable for conservation planning or basic research despite their potential to provide good quality biogeographical data.

Beale and Lennon (2012) review the main sources of uncertainty associated with SDM and highlight research directions to improve SDM predictions. These authors categorize SDM along an axis ranging from those that are purely statistical and try to identify process from pattern (i.e. niche-based models) to those that identify directly the processes and mechanisms to then generate the distributional pattern (i.e. process-based models). For practical reasons the authors classified SDM into three types: niche-based distribution models (which estimate the niche from the species' geographic distribution and re-project it on a geographic space), demographic models (which correlate demographic parameters of the species with climate or weather to characterize its distribution) and process-based models (which identify the physiological responses of the species and use them to determine the geographical distribution).

The authors argue that, for all model types, the most critical sources of uncertainty are model uncertainty and prediction uncertainty. While the type of model is clearly relevant to niche-based models for its influence in niche identification, the type of model also is important for demographic models to identify the actual links between population growth and weather, and for process-based models to correctly estimate parameters. The uncertainties associated with predictions also affect all model types because of current deficiencies in the measures of the model fit. Beale and Lennon (2012) state that the performance of models can be improved by the incorporation of uncertainty in environmental covariates and by development of measures of model fit that take into account model complexity but are not affected by prevalence and spatial autocorrelation.

Specifically for niche-based SDMs, Beale and Lennon (2012) identify that the quality of distribution data presents particular challenges, since these data are not only used for model fit but also for model building. They argue, however, that suitable tools exist to assess the uncertainty at all steps of the modeling process of niche-based models (data quality, choice of covariates, modeling technique and evaluation). Once these uncertainties are estimated, it is possible to incorporate spatial error terms into the model building. This allows evaluation of the effects of uncertainties in the predictions, or even obtaining SDM results corrected by the underlying uncertainty.

Process-based models, however, require detailed species-specific information which often is unavailable, and present the additional difficulty of identifying the interaction between species and the environment and how the interaction could be codified to be used as a predictor. A plausible option for improving the predictions of these models would be to use trait-based analyses to model assemblages or major types of communities, instead of single species.

Finally, demographic models need accurate identification and measurement of the demographic processes that limit population growth. This requires extensive field work, thereby limiting their use to a few species and reduced geographical extents. Further, as demographic models are developed, uncertainty from the measures of model fit will have a more important role, thus again necessitating development of more adequate ways to assess goodness-of-fit. Beale and Lennon (2012) conclude that it would be desirable to combine different model types statistically, or to integrate models acting at different scales with the use of hierarchical frameworks (e.g., Anderson et al. 2009).

Beale and Lennon (2012) provide guidelines for the technical development of SDM in the near future. They identify topics that need further research—assessment of both predictive performance and the significance of model covariates, and development of hierarchical models—and usefully guess the characteristics of new statistical tools that will improve SDM predictions. The importance of Beale and Lennon's (2012) review may be that it highlights that improvement of SDM is only possible by the identification and quantification of the uncertainties associated with model predictions. This is of great relevance to decision makers, promoting a higher awareness of the real degree of confidence in predictions and thus the uncertainties that actually are involved in their management decisions. A greater knowledge of the uncertainties in SDM in turn may encourage development of new tools able to manage these errors and obtain better predictions.

Geiziane Tessarolo
Depto. Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brasil.
geites@gmail.com

References

Anderson, B.J., Akçakaya, H.R., Araújo, M.B., Fordham, D.A., Martinez-Meyer, E., Thuiller, W. & Brook, B.W. (2009) Dynamics of range margins for metapopulations under climate change. Proceedings of the Royal Society B, 276, 1415–1420.

Barry, S. & Elith, J. (2006) Error and uncertainty in habitat models. Journal of Applied Ecology, 43, 413–423.

Beale, C.M. & Lennon, J.J. (2012) Incorporating uncertainty in predictive species distribution modelling. Philosophical transactions of the Royal Society B, 367, 247–258.

Lomolino, M.V. (2004) Conservation Biogeography. In: Frontiers of Biogeography – new directions in the geography of nature (ed. by M.V. Lomolino and L.R. Heaney), pp. 293–296. Sinauer Associates, Inc., Sunderland, MA.

Edited by Núria Roura-Pascual