Evaluating presence-absence species distribution models

Robert Buchkowski

Introduction:

Species distribution models are one of the most common models used in biodiversity and ecological research. Consequently, there are an overwhelming number of pre-reviewed papers describing how models should be evaluated and best practices are changing rapidly. This makes it hard for non-specialists to know where to begin.

The goal here is to provide a short summary of key steps to take to evaluate a species distribution model that is built using a generalized linear or additive model (GLM or GAM) using presence-absence data. The assumption is that you have already completed some model selection process, confirmed that independent variables are not highly correlated, and now are trying to evaluation the result—although you may need to return to model selection if the model evaluation demonstrates that the model is poor.

Readers interested in the details can find them in the cited papers and many others.

Basics:

When an SDM is being used for interpolation or hypothesis generation, basic evaluation techniques are sufficient. The goal of these techniques is to confirm that model assumptions are met and that the final coefficients are predicting important effects consistent with our understand of the species natural history.

First, the model assumptions for GLMs and GAMs are well-documented and can be largely assessed by examining the residuals of the model. For example, do the residuals correlate with any explanatory variables included in the model? Are there trends in the model residuals relative to the fitted values? Are there spatial or temporal trends in the residuals? Residual plots, spatial variograms, and comparing models fit with different link functions (e.g., logit versus complementary log-log) should all be checked.

Second, the model coefficients should be examined to judge their biological relevance. This means that there are no coefficients of trivial magnitude (i.e., uninformative parameters; try plotting the model across the range in each independent variable) and that all the coefficients make biological sense. Ideally, all the coefficients should describe mechanistic relationships that are either documented or suspected but using some variables that contain indirect information (e.g., elevation as a surrogate for habitat or climate) may be ok if projection and prediction are not important.

Third, some model fit statistics should be used to evaluate the results. These are ones that you likely already applied during model selection, such as root mean squared error, log-likelihood, and relevant versions of R^2. Cut-off values for a 'good enough' model are subjective, so clear reporting and model comparisons are the best approach.

Finally, some form of cross-validation, where subsets of the data are withheld when fitting the model and then the model is used to predict those data, is standard in species distribution modelling. This cross-validation is often interpreted using statistics derived from the confusion matrix, such as true skill statistic, Kappa, etc. The area under the receiver operating characteristic curve (AUC), or a recently proposed variant called uniform AUC, is a controversial but very common metric to validate SDMs.

Advanced:

More rigorous evaluation techniques are necessary when an SDM is being used to predict outside the range of the data, such as in the case of an invasive species, range expansion, or into the future (i.e., projection). Two additional strategies can help confirm the reliability of an SDM model but go beyond the basics often used in published SDMs: validation using spatially and temporally independent data sets and comparison of coefficient values or effect sizes across model versions.

First, the 'gold standard' (*sensu* Araújo et al. 2019) for evaluating SDMs is to compare their predictions with independent data. Ideally, this would be data collected using different methods and different personnel. However, second best is to test the model performance on spatially or temporally biased subsets of the data to evaluate out-of-sample prediction. For example, companion studies recording species presence in separate protected areas could test model predictions from one area in the other or a model fit with data from one year can be used to test data collected in the next. Models tend to preform worse in these tests than they do when a random set of testing data is excluded because training and testing data are not necessarily comparable. However, a model that tests well out of sample is more valuable because it is more likely to capture true mechanisms.

A second technique is to compare coefficient values or effect sizes in multiple model versions. The simplest version of this is to run GLMs and GAMs with all possible coefficients and look for ones whose coefficient value remains consistent across model versions. This technique reduces the chances that non-mechanistic variables will be included in the model, but also increases the chance that mechanistic variables will be excluded because they correlate with other variables that are unrelated to a species distribution. Regardless, projecting species distribution with fewer mechanisms and higher uncertainty is arguably preferable to inaccurate projections with artificially high confidence intervals. The most comprehensive version of this technique would be to develop multiple models using different statistical approaches and compare the effect sizes of different variables. The output of this analysis could be ensemble modelling or model averaging or a qualitative discussion about which effects were consistent across model versions.

Some relevant literature:

Abdulwahab, U. A., Hammill, E., & Hawkins, C. P. (2022). Choice of climate data affects the performance and interpretation of species distribution models. *Ecological Modelling*, *471*, 110042. https://doi.org/10.1016/j.ecolmodel.2022.110042

Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858

Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, *33*(10), 1677–1688. https://doi.org/10.1111/j.1365-2699.2006.01584.x

Bryn, A., Bekkby, T., Rinde, E., Gundersen, H., & Halvorsen, R. (2021). Reliability in Distribution Modeling—A Synthesis and Step-by-Step Guidelines for Improved Practice. *Frontiers in Ecology and Evolution*, *9*, 658713. https://doi.org/10.3389/fevo.2021.658713

Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B., & Singer, A. (2012). Correlation and process in species distribution models: Bridging a dichotomy: Bridging the correlation-process dichotomy. *Journal of Biogeography*, *39*(12), 2119–2131. https://doi.org/10.1111/j.1365-2699.2011.02659.x

Hijmans, R. J. (2012). Cross-validation of species distribution models: Removing spatial sorting bias and calibration with a null model. *Ecology*, *93*(3), 679–688. https://doi.org/10.1890/11-0826.1

Jiménez-Valverde, A. (2022). The uniform AUC: Dealing with the representativeness effect in presence–absence models. *Methods in Ecology and Evolution*, *13*(6), 1224–1236. https://doi.org/10.1111/2041-210X.13826

Leroux, S. J. (2019). On the prevalence of uninformative parameters in statistical models applying model selection in applied ecology. *PLOS ONE*, *14*(2), e0206711. https://doi.org/10.1371/journal.pone.0206711

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*(2), 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x

Mouton, A. M., De Baets, B., & Goethals, P. L. M. (2010). Ecological relevance of performance criteria for species distribution models. *Ecological Modelling*, *221*(16), 1995–2002. https://doi.org/10.1016/j.ecolmodel.2010.04.017

Westwood, R., Westwood, A. R., Hooshmandi, M., Pearson, K., LaFrance, K., & Murray, C. (2020).
A field-validated species distribution model to support management of the critically endangered
Poweshiek skipperling (OARISMA POWESHIEK) butterfly in Canada. *Conservation Science and
Practice*, *2*(3). https://doi.org/10.1111/csp2.163

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X.,
Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T.,
Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., … Merow, C.
(2020). A standard protocol for reporting species distribution models. *Ecography*, *43*(9), 1261–
1277. https://doi.org/10.1111/ecog.04960

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models
and extensions in ecology with R*. Springer New York. https://doi.org/10.1007/978-0-387-87458-
6