

# SMALL TO TRAIN, SMALL TO TEST: DEALING WITH LOW SAMPLE SIZE IN MODEL EVALUATION

Flavien Collart<sup>a</sup>, Antoine Guisan<sup>a,b</sup>

<sup>a</sup> *Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland*

<sup>b</sup> *Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland*

## KEYWORDS

Species distribution model ; Ecological niche ; Evaluation ; Sample size ; Rare species ; Conservation

## ABSTRACT

Sample size is a key issue in species distribution modelling. While many studies focused on the relevance of sample size for model calibration, the importance of the size of the dataset used for model evaluation has received much less attention. Here, we highlight two previously published approaches to address the problem, and which are relatively simple to implement: the pooling evaluation and the implementation of null models. We discuss the importance of these or other potential approaches that are critical for model evaluation in rare species, which represent the bulk of biodiversity, and for which accurate models are most necessary in a conservation context.

## Introduction

Species distribution models (SDM) - also called ecological niche models (ENM) or habitat suitability models (HSM) - are increasingly used in ecological studies, e.g. for testing biogeographic hypotheses, assessing global change impacts or planning conservation ([Araújo et al., 2019](#); [Guisan et al., 2013, 2017](#)). They are typically based on the ecological niche concept and allow predicting probabilities of occurrence or habitat suitability in a given geographical space and time period. SDMs can be affected by numerous factors, notably the algorithm used, data bias, or what is often shown as the major factor, sample size ([Fernandes et al., 2019](#); [Guisan et al., 2017](#); [Hernandez et al., 2006](#); [Jeliaskov et al., 2022](#); [Liu et al., 2018](#); [Stockwell and Peterson, 2002](#); [Valavi et al., 2022](#); [van Proosdij et al., 2016](#); [Wisz et al., 2008](#); [Zurell et al., 2020](#)). Sample size is especially a critical issue for rare species, which represent the bulk of biodiversity ([Lomba et al., 2010](#); [ter Steege et al., 2013](#)).

Modelling the distribution of rare species requires specific approaches ([Breiner et al., 2015, 2018](#); [Jeliaskov et al., 2022](#); [Lomba et al., 2010](#); [Mondanaro et al., 2023](#)). In particular, sample size limits the number of predictors that can be included in a model, with a general rule of thumb that the addition of a new predictor variable requires another 10 occurrences ([Harrell Jr. et al., 1996](#); [Peduzzi](#)

[et al., 1996](#)). Although models for rare species, which are more sensitive to climate change ([Vincent et al., 2020](#)) and prone to extinction ([Courchamp et al., 2006](#); [Enquist et al., 2019](#); [Isik, 2011](#)), they are crucially needed to assess threat levels and support conservation actions, such models are thus highly prone to overfitting issues ([Breiner et al., 2015](#); [Lomba et al., 2010](#)). To address this paradox, [Lomba et al. \(2010\)](#) proposed the “ensemble of small model” (ESM) approach, which consists of averaging models fitted with all possible combinations of pairs of predictors into a final ensemble. ESMs allow increasing the predictive power and avoiding overfitting (by reducing the number of predictors in each small model) while still allowing to consider all predictors deemed important for the species (i.e. defining its niche). Although other powerful methods exist to predict the distribution of species (see [Jeliaskov et al., 2022](#) for a review), ESMs, which significantly outperform standard SDMs also for more common species ([Breiner et al., 2015](#)), thus represent an elegant solution to the problem of fitting models with limited sample size. However, little attention has been paid to the problem of evaluating models with small sample size ([Bean et al., 2012](#); [Hallman and Robinson, 2020](#); [Jimenez-Valverde, 2020](#)). Here, we highlight this neglected issue and discuss two possible solutions to deal with it. We finally identify future refinements to these approaches, and potential tools associated.

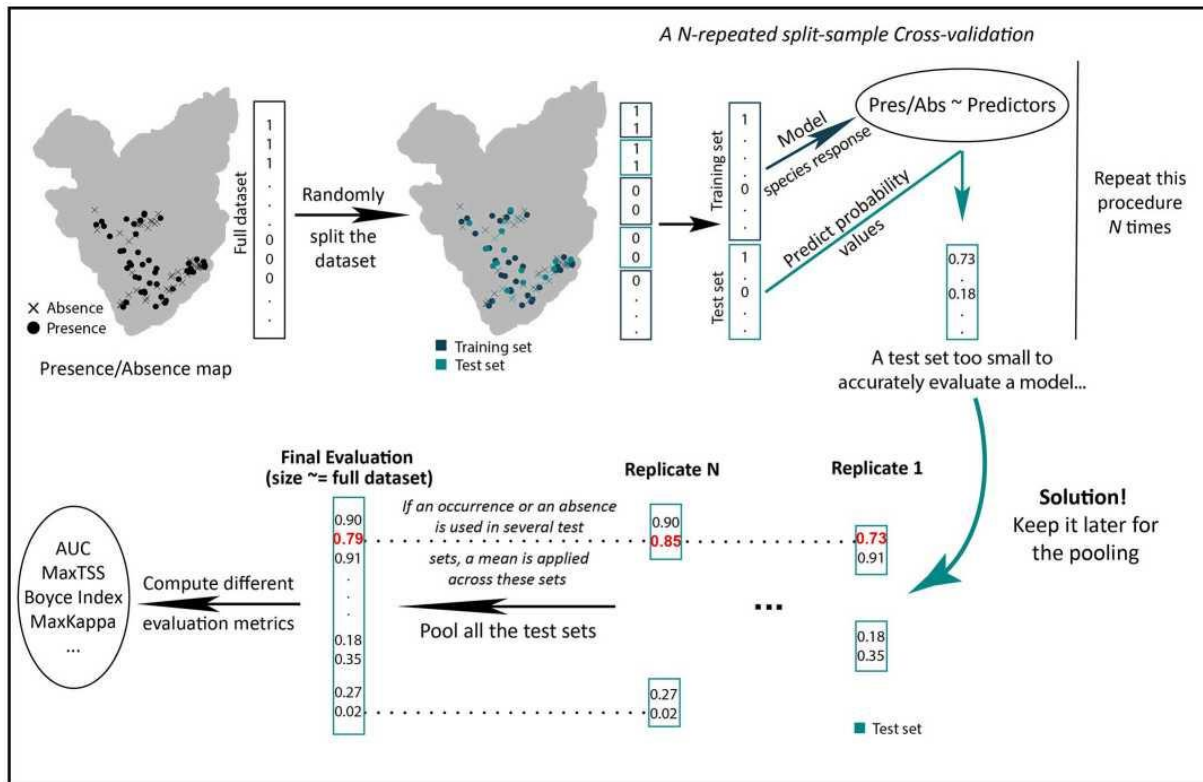
## The neglected one: Sample size effect in model evaluation. How to deal with low sample size?

The threshold used to binarize habitat suitability probabilities is impacted by the sample size which can likely influence the measurements of threshold-dependent evaluation metrics such as the True Skill Statistic or Cohen’s Kappa ([Bean et al., 2012](#)). In addition, [Hallman and Robinson \(2020\)](#) showed that the Partial Receiver Operating Characteristic (pROC), which is an alternative to the commonly-used nonthreshold-dependent evaluation metric ([Fourcade et al., 2018](#); [Guisan et al., 2017](#)), the Area Under the receiver operating Curve ([Fielding and Bell, 1997](#); [Peterson et al., 2008](#)), tends to be artificially inflated when small test sets are used. For presence-absence data, [Jimenez-Valverde \(2020\)](#) demonstrated through simulations and with the Area Under the ROC Curve (AUC) metric that a sample size lower than 10 presences and 10 absences can seriously hamper predictive power, leading them to recommend that test sets should include at least 15 occurrence and 15 absence points. Following these recommendations, the minimum initial sample size of a dataset should be greater or equal to 50 occurrences and 50 absences when implementing the most commonly cross-validation approach used for evaluating models from a single initial dataset, i.e., splitting it into a training set (typically 70% of the data) to calibrate the models and a test set (the remaining 30% of the data) to assess model performance ([Guisan et al., 2017](#)). To our knowledge, the minimal size of a test set has not yet been evaluated in the case of presence-only data, but since the latter carry less information than presence-absence data, the issue of the size of the test set will undoubtedly be even more critical than for presence-absence data. The size of the test set is therefore a crucial problem, especially for rare or under-represented species, and

further to conservation planning.

To address this issue, [Collart et al. \(2021\)](#) proposed a solution ([Fig. 1](#)), which we call hereafter “the pooling evaluation”. In the context of a standard repeated split-sample cross-validation approach (see above), a presence or an absence can be used or omitted several times in the evaluation process across replicates, resulting in some cases in having very few presences and highly unbalanced prevalence in some test sets. Furthermore, the sub-sampling of a dataset that is already, in the case of rare species, potentially small, leads to very few occurrences available for model evaluation in the test set. These issues hamper proper model evaluation, typically leading to a general inflation of the accuracy metrics ([Hallman and Robinson, 2020](#)). The idea of the proposed evaluation is to pool together all the left-out test sets to ‘reconstruct’ a single test set of sufficient sample size for proper evaluation. If an occurrence or an absence was, by chance, sampled in several test sets, the habitat suitability values at this point can simply be averaged across all test sets. At the end, one obtains a test set with roughly the same length as the original dataset, suitable to compute different evaluation metrics, such as the AUC, the True Skill Statistic and the Boyce Index. With this evaluation-by-pooling, one can thus palliate the small size of the test set, reducing the statistical accuracy of evaluation metrics and allowing to model species with a minimum of 10–15 occurrences without increasing computation time. This method is now available in the *ecospat* R package ([Broennimann et al., 2022](#)) for ESMs (*ecospat*.ESM.EnsembleEvaluation) but also for more conventional SDMs (*ecospat*.poolingEvaluation function).

**Fig. 1.** Procedure to evaluate species distribution models (SDMs) via the pooling evaluation (developed in [Collart et al., 2021](#)). A  $N$ -repeated split-sample cross-validation is first needed. The full dataset is thus randomly splitted into a training and a test set  $N$  times. The training set is used to infer the species ecological niche. This model is employed to afterwards make predictions on the test set. The  $N$  test sets are then pooled together to generate a single test set having roughly the same size as the full dataset. If an occurrence or an absence is used in several test sets (in bold red), a mean of probability values is applied. This resulting test set can be then used to compute different evaluation metrics such as the Area Under the Curve (AUC), the maximum value of True Skill Statistic (MaxTSS), the Boyce Index and/or the maximum value of kappa (MaxKappa). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



Another complementary and potentially powerful solution was proposed to evaluate models with small datasets (Bohl et al., 2019; Osborne et al., 2022; Raes and ter Steege, 2007; van Proosdij et al., 2016; Warren et al., 2021). It consists of running a series of null models, sampling a number of background points that is equal to the number of actual occurrences used for model calibration, and using these background points as ‘fake’ occurrences. The fitted null model is then evaluated with the same test set used to evaluate the empirical model (Bohl et al., 2019). This procedure is repeated several times (99 for van Proosdij et al., 2016; ~1000 for Raes and ter Steege, 2007 and Bohl et al., 2019) to obtain a null distribution of one or several evaluation metrics. The evaluation metrics of the initial models are finally compared with the evaluation metrics of the null models. However, because a fully random null model gives too optimistic evaluation of SDMs by not taking into account spatial autocorrelation nor sampling bias, it is recommended to generate bias-corrected null models based on target-group background sampling (Barber et al., 2022; Phillips et al., 2009; Ponder et al., 2001), i.e. randomly sampling background points in geographical (or environmental) space where the species occurs (van Proosdij et al., 2016). Null models can be run using, for example, the functions “ENMnulls” in the ENMeval R package (Kass et al., 2021) and nullRandom in dismo R package (Hijmans et al., 2022). When the study aim is to measure temporal changes of species distribution, Warren et al. (2021) proposed to project these null models onto different time periods, climate models and/or emission scenarios and computed the predicted changes in habitat suitability with the same method as with the empirical data. This null distribution of change can be afterwards compared with the empirical results, which allows to determine to what extent predictions are driven by our data. However, one drawback of null models is an increase of computation time, which

can be important when applied on ESMs as null models need to be implemented for each bivariate model.

If computation time is not a limitation, the most conservative option to deal with low sample size would be to combine the pooling evaluation and null models. By doing so, one can test if the generated SDMs for rare species (or with low sample size) are more accurate than expected by chance, while the pooling gives an approximation of how accurate models are. To apply this procedure, the idea would be to first generate several “null datasets” with fake occurrences randomly sampled from target-group background points and employ the pooling evaluation, thus performing a  $N$ -repeated split-sample cross-validation, pooling all the  $N$  test set and computing several evaluation metrics.

## What are the missing pieces?

Even though these approaches can palliate some aspects of SDM evaluation with low sample size, many other aspects still require further investigations. Remaining questions are for instance: (1) How do presence-absence evaluation metrics other than AUC, such as the True Skill Statistic (TSS) or Cohen’s kappa, and presence-only evaluation metrics, such as the Boyce Index, vary depending on sample size? Can we define a minimum sample size for presence-absence and presence-only datasets? (2) Regarding the previous questions, does sampling bias or spatial autocorrelation affect the minimum sample size for the evaluation?

Simulations by generating virtual species appear to be an appealing solution to answer these questions ([Meynard et al., 2019](#); [Meynard and Kaplan, 2013](#); [Zurell et al., 2010](#)). Virtual species can be defined as a simulated taxon for which its response to a particular environment is known and for which the sampling strategy can be totally controlled ([Meynard et al., 2019](#)). To understand the effect of sample size on model evaluation, several virtual species have to be generated, defining their responses to an environment and testing different niche properties (e.g. size). A sampling can then be applied defining where the species is present or not. In this step, one could create some noise, such as species misidentification (defining a species occurrence in a certain location where the species is actually not present) or adding species commission or omission errors (wrongly defining a species present or absent respectively; as in [Fernandes et al. \(2019\)](#)). These datasets can be split into training and test sets. Models can afterward be generated using the training sets and evaluated using the test sets. Then, one could decrease the sample size of the test sets to determine how the different metrics vary across the sample size. Virtual species can be generated via for example virtualspecies R package ([Leroy et al., 2016](#)). It is also possible applying the same procedure to test these questions using robust datasets based on real species (e.g., [Dubuis et al., 2011](#); [Elith et al., 2020](#); [Hallman and Robinson, 2020](#); [Valavi et al., 2022](#)).

## Concluding remarks

Although it is possible to accurately predict a species distribution with a low sample size, small test sets can provoke incorrect evaluations of these models. Here, we highlighted two solutions from recent papers to deal with a small number of occurrences in test sets. However, improving our understanding on how sample size affects the measurement of model performances is crucial to enhance future SDM studies, and will be particularly necessary to ensure robust SDM-based conservation planning, especially in the context of climate change. For now, with only limited knowledge on the impact of sample size on SDM evaluations, the next SDM studies should at least keep in mind that a small training set also means a small test set and acknowledge this potential issue, especially regarding rare or under-represented species.

## Funding

FC was funded by a grant from the Swiss National Science Foundation to AG (SNSF; Grant number 197777).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## References

- Araujo, M.B., Anderson, R.P., Marcia 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. *Sci. Adv.* 5 (1), eaat4858. <https://doi.org/10.1126/sciadv.aat4858>.
- Barber, R.A., Ball, S.G., Morris, R.K.A., Gilbert, F., 2022. Target-group backgrounds prove effective at correcting sampling bias in Maxent models. *Divers. Distrib.* 28 (1), 128–141. <https://doi.org/10.1111/ddi.13442>.

- Bean, W.T., Stafford, R., Brashares, J.S., 2012. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography* 35 (3), 250–258. <https://doi.org/10.1111/j.1600-0587.2011.06545.x>.
- Bohl, C.L., Kass, J.M., Anderson, R.P., 2019. A new null model approach to quantify performance and significance for ecological niche models of species distributions. *J. Biogeogr.* 46 (6), 1101–1111. <https://doi.org/10.1111/jbi.13573>.
- Breiner, F.T., Guisan, A., Bergamini, A., Nobis, M.P., 2015. Overcoming limitations of modelling rare species by using ensembles of small models. *Methods Ecol. Evol.* 6 (10), 1210–1218. <https://doi.org/10.1111/2041-210X.12403>.
- Breiner, F.T., Nobis, M.P., Bergamini, A., Guisan, A., 2018. Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Methods Ecol. Evol.* 9 (4), 802–808. <https://doi.org/10.1111/2041-210X.12957>.
- [Broennimann, O., Di Cola, V., Guisan, A., 2022. \*Ecospat: Spatial Ecology Miscellaneous Methods. \(R package version 3.4.\)\*](#)
- Collart, F., Hedenas, L., Broennimann, O., Guisan, A., Vanderpoorten, A., 2021. Intraspecific differentiation: Implications for niche and distribution modelling. *J. Biogeogr.* 48, 415–426. <https://doi.org/10.1111/jbi.14009>.
- Courchamp, F., Angulo, E., Rivalan, P., Hall, R.J., Signoret, L., Bull, L., Meinard, Y., 2006. Rarity value and species extinction: the anthropogenic Allee effect. *PLoS Biol.* 4 (12), e415 <https://doi.org/10.1371/journal.pbio.0040415>.
- Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P., Guisan, A., 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Divers. Distrib.* 17 (6), 1122–1131. <https://doi.org/10.1111/j.1472-4642.2011.00792.x>.
- Elith, J., Graham, C., Valavi, R., Abegg, M., Bruce, C., Ferrier, S., Ford, A., Guisan, A., Hijmans, R.J., Huettmann, F., Lohmann, L., Loiselle, B., Moritz, C., Overton, J., Peterson, A.T., Phillips, S., Richardson, K., Williams, S., Wiser, S.K., Zimmermann, N. E., 2020. Presence-only and presence-absence data for comparing species distribution modeling methods. *Biodivers. Inform.* 15 (2) <https://doi.org/10.17161/bi.v15i2.13384>. Article 2.
- Enquist, B.J., Feng, X., Boyle, B., Maitner, B., Newman, E.A., Jørgensen, P.M., Roehrdanz, P.R., Thiers, B.M., Burger, J.R., Corlett, R.T., Couvreur, T.L.P., Dauby, G., Donoghue, J.C., Foden, W., Lovett, J.C., Marquet, P.A., Merow, C., Midgley, G., Morueta-Holme, N., McGill, B.J., 2019. The commonness of rarity: global and future distribution of rarity across land plants. *Sci. Adv.* 5 (11), eaaz0414. <https://doi.org/10.1126/sciadv.aaz0414>.
- Fernandes, R.F., Scherrer, D., Guisan, A., 2019. Effects of simulated observation errors on the performance of species distribution models. *Divers. Distrib.* 25 (3), 400–413. <https://doi.org/10.1111/ddi.12868>.

- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24 (1), 38-49. <https://doi.org/10.1017/S0376892997000088>.
- Fourcade, Y., Besnard, A.G., Secondi, J., 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Glob. Ecol. Biogeogr.* 27 (2), 245-256. <https://doi.org/10.1111/geb.12684>.
- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Buckley, Y.M., 2013. Predicting species distributions for conservation decisions. *Ecol. Lett.* 16 (12), 1424-1435. <https://doi.org/10.1111/ele.12189>.
- Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. *Habitat Suitability and Distribution Models*. Cambridge University Press. <https://doi.org/10.1017/9781139028271>.
- Hallman, T.A., Robinson, W.D., 2020. Deciphering ecology from statistical artefacts: competing influence of sample size, prevalence and habitat specialization on species distribution models and how small evaluation datasets can inflate metrics of performance. *Divers. Distrib.* 26 (3), 315-328. <https://doi.org/10.1111/ddi.13030>.
- Harrell Jr., F.E., Lee, K.L., Mark, D.B., 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* 15 (4), 361-387. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960229\)15:4<361::AID-SIM168>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4).
- Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29 (5), 773-785. <https://doi.org/10.1111/j.0906-7590.2006.04700.x>.
- [Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2022. Dismo: Species Distribution Modeling. \(R package version 1.3-9\).](#)
- Isik, K., 2011. Rare and endemic species: why are they prone to extinction? *Turk. J. Bot.* 35 (4), 411-417. <https://doi.org/10.3906/bot-1012-90>.
- Jeliazkov, A., Gavish, Y., Marsh, C.J., Geschke, J., Brummitt, N., Rocchini, D., Haase, P., Kunin, W.E., Henle, K., 2022. Sampling and modelling rare species: conceptual guidelines for the neglected majority. *Glob. Chang. Biol.* 28 (12), 3754-3777. <https://doi.org/10.1111/gcb.16114>.
- Jimenez-Valverde, A., 2020. Sample size for the evaluation of presence-absence models. *Ecol. Indic.* 114, 106289 <https://doi.org/10.1016/j.ecolind.2020.106289>.
- Kass, J.M., Muscarella, R., Galante, P.J., Bohl, C.L., Pinilla-Buitrago, G.E., Boria, R.A., Soley-Guardia, M., Anderson, R.P., 2021. ENMeval 2.0: redesigned for customizable and reproducible modeling of species' niches and distributions. *Methods Ecol. Evol.* 12 (9), 1602-1608. <https://doi.org/10.1111/2041-210X.13628>.



- Leroy, B., Meynard, C.N., Bellard, C., Courchamp, F., 2016. Virtualspecies, an R package to generate virtual species distributions. *Ecography* 39 (6), 599-607. <https://doi.org/10.1111/ecog.01388>.
- Liu, C., Newell, G., White, M., 2018. The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. *Ecography*. <https://doi.org/10.1111/ecog.03188>.
- Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J., Guisan, A., 2010. Overcoming the rare species modelling paradox: a novel hierarchical framework applied to an Iberian endemic plant. *Biol. Conserv.* 143 (11), 2647-2657. <https://doi.org/10.1016/J.BIOCON.2010.07.007>.
- Meynard, C.N., Kaplan, D.M., 2013. Using virtual species to study species distributions and model performance. *J. Biogeogr.* 40 (1), 1-8. <https://doi.org/10.1111/jbi.12006>.
- Meynard, C.N., Leroy, B., Kaplan, D.M., 2019. Testing methods in species distribution modelling using virtual species: What have we learnt and what are we missing? *Ecography*. <https://doi.org/10.1111/ecog.04385> ecog.04385.
- Mondanaro, A., Di Febbraro, M., Castiglione, S., Melchionna, M., Serio, C., Girardi, G., Belfiore, A.M., Raia, P., 2023. ENphylo: a new method to model the distribution of extremely rare species. *Methods Ecol. Evol.* 14 (3), 911-922. <https://doi.org/10.1111/2041-210X.14066>.
- Osborne, O.G., Fell, H.G., Atkins, H., van Tol, J., Phillips, D., Herrera-Alsina, L., Mynard, P., Bocedi, G., Gubry-Rangin, C., Lancaster, L.T., Creer, S., Nangoy, M., Fahri, F., Lupiyaningdyah, P., Sudiana, I.M., Juliandi, B., Travis, J.M.J., Papadopoulos, A.S.T., Algar, A.C., 2022. Fauxcurrence: simulating multi-species occurrences for null models in species distribution modelling and biogeography. *Ecography* 2022 (7), e05880. <https://doi.org/10.1111/ecog.05880>.
- Peduzzi, P., Concato, J., Kemper, E., Holford, T.R., Feinstein, A.R., 1996. A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* 49 (12), 1373-1379. [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3).
- Peterson, A.T., Papes, M., Soberon, J., 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecol. Model.* 213 (1), 63-72. <https://doi.org/10.1016/j.ecolmodel.2007.11.008>.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19 (1), 181-197. <https://doi.org/10.1890/07-2153.1>.
- Ponder, W.F., Carter, G.A., Flemons, P., Chapman, R.R., 2001. Evaluation of museum collection data for use in biodiversity assessment. *Conserv. Biol.* 15 (3), 648-657. <https://doi.org/10.1046/j.1523-1739.2001.015003648.x>.
- Raes, N., ter Steege, H., 2007. A null-model for significance testing of presence-only species distribution models. *Ecography* 30 (5), 727-736. <https://doi.org/10.1111/j.2007.0906-7590.05041.x>.

- Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecol. Model.* 148 (1), 1-13. [https://doi.org/10.1016/S0304-3800\(01\)00388-X](https://doi.org/10.1016/S0304-3800(01)00388-X).
- ter Steege, H., Pitman, N.C.A., Sabatier, D., Baraloto, C., Salomao, R.P., Guevara, J.E., Phillips, O.L., Castilho, C.V., Magnusson, W.E., Molino, J.-F., Monteagudo, A., Nunez Vargas, P., Montero, J.C., Feldpausch, T.R., Coronado, E.N.H., Killeen, T.J., Mostacedo, B., Vasquez, R., Assis, R.L., Silman, M.R., 2013. Hyperdominance in the Amazonian tree Flora. *Science* 342 (6156), 1243092. <https://doi.org/10.1126/science.1243092>.
- Valavi, R., Guillera-Aroita, G., Lahoz-Monfort, J.J., Elith, J., 2022. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecol. Monogr.* 92 (1), e01486 <https://doi.org/10.1002/ecm.1486>.
- van Proosdij, A.S.J., Sosef, M.S.M., Wieringa, J.J., Raes, N., 2016. Minimum required number of specimen records to develop accurate species distribution models. *Ecography* 39 (6), 542-552. <https://doi.org/10.1111/ecog.01509>.
- Vincent, H., Bornand, C.N., Kempel, A., Fischer, M., 2020. Rare species perform worse than widespread species under changed climate. *Biol. Conserv.* 246, 108586 <https://doi.org/10.1016/j.biocon.2020.108586>.
- Warren, D.L., Dornburg, A., Zapfe, K., Iglesias, T.L., 2021. The effects of climate change on Australia's only endemic Pokemon: measuring bias in species distribution models. *Methods Ecol. Evol.* 12 (6), 985-995. <https://doi.org/10.1111/2041-210X.13591>.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Elith, J., Dudík, M., Ferrier, S., Huettmann, F., Leathwick, J.R., Lehmann, A., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. McC, Zimmermann, N.E., 2008. Effects of sample size on the performance of species distribution models. *Divers. Distrib.* 14 (5), 763-773. <https://doi.org/10.1111/j.1472-4642.2008.00482.x>.
- Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., Nehrbaas, N., Pagel, J., Reineking, B., Schroder, B., Grimm, V., 2010. The virtual ecologist approach: simulating data and observers. *Oikos* 119 (4), 622-635. <https://doi.org/10.1111/j.1600-0706.2009.18284.x>.
- Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., Guillera-Aroita, G., Guisan, A., Lahoz-Monfort, J.J., Leitao, P.J., Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schroeder, B., Serra-Diaz, J.M., Thuiller, W., Merow, C., 2020. A standard protocol for reporting species distribution models. *Ecography*. <https://doi.org/10.1111/ecog.04960> ecog.04960.