Received: 11 October 2017 | Accepted: 28 November 2017

DOI: 10.1111/2041-210X.12957

FORUM

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Optimizing ensembles of small models for predicting the distribution of species with few occurrences

Frank T. Breiner[1,2] | Michael P. Nobis[3] | Ariel Bergamini[2] | Antoine Guisan[1,4]

[1]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

[2]Biodiversity and Conservation Biology, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

[3]Landscape Dynamics, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

[4]Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland

Correspondence
Frank Breiner
Email: frank.breiner@wsl.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 1528661; Swiss National Science Foundation (NSF; grant Number 1528661 to AG)

Handling Editor: Nick Isaac

Abstract

1. Ensembles of Small Models (ESM) represent a novel strategy for species distribution modelling with few observations. ESMs are built by calibrating many small models and then averaging them into an ensemble model where the small models are weighted by their cross-validated scores of predictive performance. In a previous paper (Breiner, Guisan, Bergamini, & Nobis, *Methods in Ecology and Evolution*, **6**, 1210–1218, 2015), we reported two major findings. First, ESMs proved largely superior to standard models in terms of model performance and transferability. Second, ESMs including different modelling techniques did not clearly improve model performance compared to single-technique ESMs. However, ESMs often require a large computation effort, which can become problematic when modelling large numbers of species. Given the appealing new perspectives offered by ESMs, it is especially important to investigate if some techniques yield increased performance while saving computation time and thus could be predominantly used for building ESMs.
2. Here, we present results from a reanalysis of a subset of the data used in Breiner et al. (2015). More specifically, we ran ESMs: (1) fitted with 10 modelling techniques separately (in Breiner et al., 2015 we used only three techniques); and (2) using various parameter options for each modelling technique (i.e., model tuning).
3. We show that ESMs vary in model performance and computation time across techniques, and some techniques are advantageous in terms of optimizing model performance and computation time (i.e., GLM, CTA and ANN). Including one of these modelling techniques could thus optimize computation time compared to using more computing-intensive techniques like GBM. Next, we show that parameter tuning can improve performance and transferability of ESMs, but often at the cost of computation time. Parameter tuning could therefore be used when computing resources are not a limiting factor.
4. These findings help improve the applicability and performance of ESMs when applied to large numbers of species.

KEYWORDS
computation time, endangered species, ensemble prediction, model performance, model tuning, species distribution modelling

© 2018 The Authors. Methods in Ecology and Evolution © 2018 British Ecological Society

1 | INTRODUCTION

The use of Ensembles of Small Models (ESMs) is a novel strategy proposed for predicting the distribution of rare species (Lomba et al., 2010), now available in the ECOSPAT R package (Broennimann et al., 2015; Di Cola et al., 2017). In our recent article (Breiner, Guisan, Bergamini, & Nobis, 2015), we showed that the predictive power of ESMs was significantly better than that of their standard model counterparts (i.e., using BIOMOD2 in a standard model, Thuiller, Lafourcade, Engler, & Araújo, 2009). The ESM strategy involves building many models with small subsets of predictors—two as implemented here (i.e., bivariate models). Each of these small models is then evaluated according to a cross-validated model performance measure. In the last step, all small models are averaged into a weighted ensemble, where weights are based on the small model performances, to produce the final ESM (Figure 1; for a more detailed description see Breiner et al., 2015). The great strength of ESMs is that model overfitting is avoided in each small model, where the ratio of the number of predictors to the number of observations remains appropriate even for species with few occurrences, while all important predictor variables can be considered within the final model. As a main result, we showed that the lower the sample size of a species the greater the gain in ESM performance compared to standard models. However, we also showed that building ESMs with an ensemble of different techniques (GLM, gradient boosting machine [GBM] and maximum entropy [Maxent])—thus a double ensemble—did not improve performance compared to single-technique ESMs. It therefore seemed sufficient to fit ESMs with only one technique. However, only three techniques were tested in Breiner et al. (2015), and many others are available which potentially perform better than those tested so far. Furthermore, computation time varied considerably among the three techniques used. It may thus be critical to select not only the best performing but also the fastest technique. Finding the optimal balance between high model performance and short computation time is especially important if ESMs are to be applied to a large number of species, e.g., when adding rare species to stacked models of species assemblages to account for rare species when modelling communities (e.g., D'Amen, Pradervand, & Guisan, 2015), or when using models to support real-time conservation practices for rare species (Guisan et al., 2013; Platts et al., 2014).

Here, we reanalyse the data from Breiner et al. (2015) by fitting ESMs with a large range of modelling techniques to assess how ESMs could be further improved, considering in particular the relationship between model performance and computation time. More specifically, we tested whether models are improved by: (1) comparing a larger range of modelling techniques to build bivariate models; and (2) optimizing tuning parameters in the bivariate models and allowing more complex relationships among parameters.

2 | MODELLING TECHNIQUES APPLIED WITH ESMS AND STANDARD MODELS

We explored 10 modelling techniques available in the BIOMOD2 R package for ensemble modelling (Thuiller et al., 2009). In addition to GLM, GBM and Maxent (Max.P as implemented by Phillips, Anderson, & Schapire, 2006) that were applied in Breiner et al. (2015), we additionally used multivariate adaptive regression splines (MARS; Friedman, 1991), generalized additive models (GAM; Hastie & Tibshirani, 1986), random forests (RF; Breiman, 2001), classification and regression trees (CTA; Breiman, Friedman, Olshen, & Stone, 1984), artificial neural networks (ANN; Ripley, 1996), surface range envelopes (SRE; also known as Bioclim; Busby, 1991; Booth, Nix, Busby, & Hutchinson, 2014) and

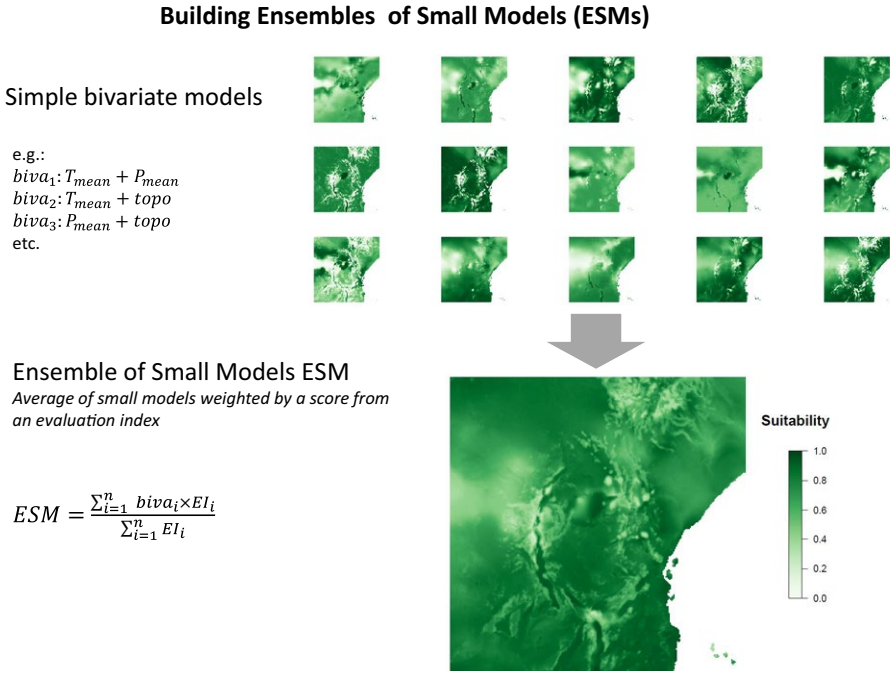

FIGURE 1 Framework to build an Ensemble of Small Models (ESM) to predict a species' distribution. First, many models with small subsets of predictors, in our case bivariate models (biva), are calibrated. Each of these small models is then evaluated according to a cross-validated model evaluation index (EI). To produce the final ESM, all small models are averaged into an ensemble weighted by the specific model EI. T_{mean} (T = temperature), P_{mean} (P = precipitation), topo (topographical index) are examples of environmental predictors used for modelling

TABLE 1 Model ranks according to different evaluation measures where the best performing model has rank one (where tfa: Transferability assessment; cv: cross-validation; ho: hold-hout). ESMs with best parameters and standard Maxent models (Max.P.s) are used in the table. Note that two distinct implementations of Maxent were used: Max.P (Phillips) and Max.T (Tsuruoka; more details in the main text)

	MEDIAN	AUC tfa	Boyce tfa	TSS tfa	Kappa tfa	AUC cv	Boyce cv	TSS cv	Kappa cv	AUC ho	Boyce ho	TSS ho	Kappa ho
Max.P.s	5	5	8	5	2	7	10	7	2	3	7	3	5
ANN	2.5	2	2	4	6	2	4	2	3	4	1	5	2
GLM	2.5	3	4	1	1	1	5	1	7	2	4	2	3
Max.P	3.5	1	1	2	5	5	3	6	8	1	3	4	6
CTA	4.5	4	6	3	4	3	7	3	5	5	8	6	4
GBM	5.5	6	3	6	7	4	1	4	4	7	5	9	9
GAM	6	9	5	10	10	6	2	5	6	6	2	7	7
Max.T	8	7	7	8	9	8	6	8	9	8	9	11	10
MARS	9.5	8	11	7	8	9	9	9	10	10	10	10	11
SRE	9.5	10	10	9	11	11	8	10	11	9	6	8	8
RF	10.5	11	9	11	3	10	11	11	1	NA	NA	NA	NA

a distinct implementation of Maxent (Max.T; Jurka & Tsuruoka, 2013; Tsuruoka, Tsujii, & Ananiadou, 2009). All the modelling techniques tested here are already implemented in the existing ESM functions of the ecospat R-package (Broennimann et al., 2015; Di Cola et al., 2017).

Additionally, we compared ESMs with standard Maxent models (Max.P.s; where s stands for standard). We selected Max.P.s because Maxent is often reported as performing well with low sample sizes (Pearson, Raxworthy, Nakamura, & Peterson, 2007; Wisz et al., 2008). For Max.P.s, we used a betamultiplier value of 2.5 to reduce the complexity of calibrated models, as recommended for modelling rare species (i.e., optimal values set between 1.5 and 4; Elith, Kearney, & Phillips, 2010; Moreno-Amat et al., 2015; Muscarella et al., 2014; Shcheglovitova & Anderson, 2013). This differed from the methods in Breiner et al. (2015), where we used standard parameters for Maxent (betamultiplier = 1). All models were computed on a high-performance cluster (www.vital-it.ch) equipped with 2.3 to 3 GHz processors (variations in processor speed should be averaged out given the large number of repetitions) and using up to 50 processors to run jobs in parallel. We provided 2 GB of memory to run each modelling algorithm, except for GBMs were available memory size needed to be doubled to 4 GB.

3 | MODEL BUILDING PROCESS

We used two subsets of the data from Breiner et al. (2015). The first subset contained all 34 vascular plant species from Switzerland with a sample size of 26 to 50 occurrences. We split the data four times geographically (across cardinal directions) so that 50% of the data were used for calibrating and 50% for evaluating the models independently (i.e., a transferability assessment: $W_{train} - E_{test}$; $E_{train} - W_{test}$; $S_{train} - N_{test}$; $N_{train} - S_{test}$), using AUC (Swets, 1988), the continuous Boyce index (Hirzel, Le Lay, Helfer, Randin, & Guisan, 2006), maxTSS and maxKappa (see Guisan, Thuiller, & Zimmermann, 2017 for details) as evaluation scores.

Bivariate models, i.e., all possible bivariate combinations of the predictor set, were then calibrated for each species and for each of the four training subsets. We first varied the model parameters and complexity separately for each bivariate model per technique and species and evaluated them using 10-fold cross-validation. For each bivariate model, the parameter setting yielding the highest 10-fold cross-validated AUC value was kept. These "tuned" bivariate models were then averaged, weighted by their mean cross-validated evaluation score, to form a final "tuned" ESM (ESM_{tuned}) for each of the four regions per species. For model tuning, we used the *BIOMOD_tuning* function, which is a wrapper function in the biomod2 R-package. It uses tuning functions from the caret R-package (to tune GBM, ANN, GAM, MARS, GLM and CTA; Kuhn, 2008; Kuhn & Johnson, 2013), enmeval R-package (to tune Max.P; Muscarella et al., 2014) and the maxent R-package (to tune Max.T; Jurka & Tsuruoka, 2013; Tsuruoka et al., 2009). In doing so, we aimed at calibrating bivariate models that better fit species' response curves, thereby increasing their predictive performance. All parameters, and their ranges, used to tune the bivariate models are presented in Table S1. Model tuning was used for

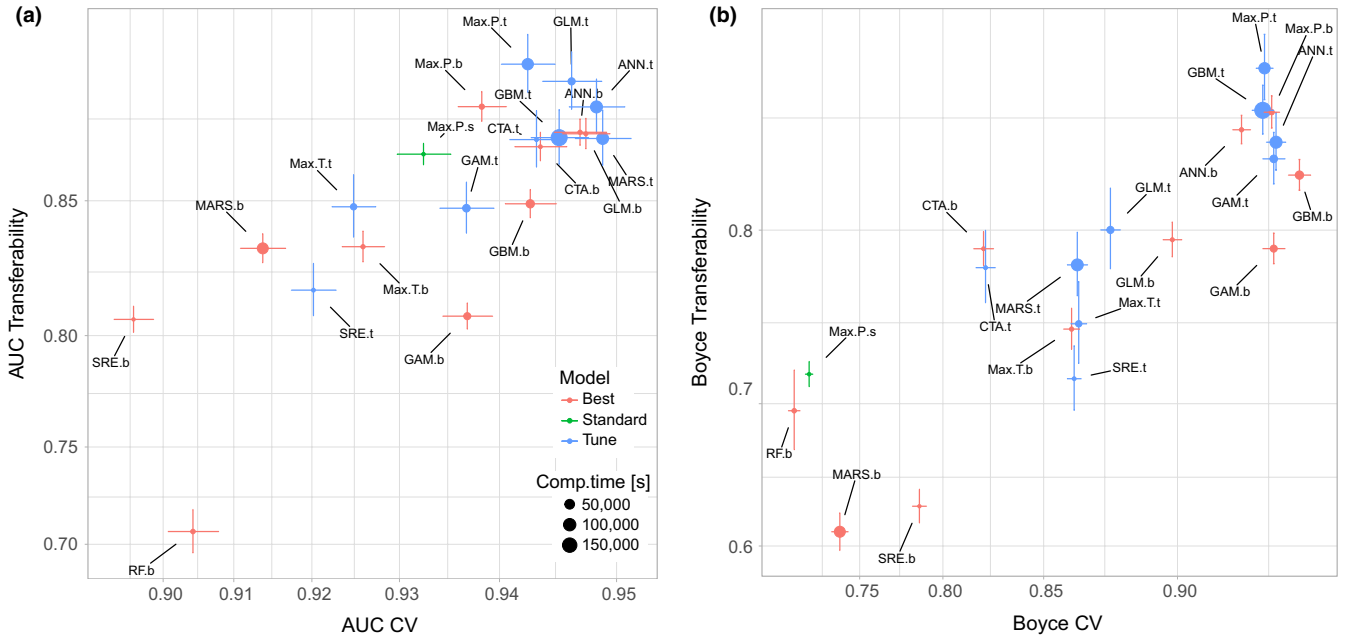

FIGURE 2 Evaluation scores (left: AUC; right: Boyce index) of ESMs for different techniques, based on 10-fold cross-validation (*x*-axes) and transferability assessment (*y*-axes) for tuned ESM models (small t after techniques' name) and ESMs with best-tuning parameters (small b after techniques' name), as well as for standard Maxent models (Max.P.s). The computation time of each model is indicated by the size of the point

all techniques except for RF because in this case the *mtry* parameter value could not exceed 2 for bivariate models.

Next, we refitted an ESM for each technique, with the parameters performing, on average, best with bivariate model tuning over all species (ESM_{best}). All parameter values used for ESM_{best} are presented in Table S1 and will be made available in the next ESM functions in the ECOSPAT package (Di Cola et al., 2017). ESM_{tuned} and ESM_{best} were built, using Somers' D, a rescaled version of the AUC (between 0 and 1), and the Boyce index to weight the bivariate models.

Each final ESM and the standard SDM Max.P.s were then evaluated using a procedure similar to that applied in Breiner et al. (2015), i.e., a transferability assessment with four, instead of only one, 50:50 splits for evaluating the models independently as well as an internal 10-fold cross-validation procedure on the training data instead of 50-fold split-sampling which increases computation time considerably.

For the second subset, we used the five most common species (sample sizes: 140, 128, 123, 115 and 112) in the data of Breiner et al. (2015). To simulate strongly undersampled species, we sampled this data randomly to 20, 15 and 10 occurrences for calibrating the models. Hold-out data for evaluation ranged thus from 92 to 130. The sampling procedure was repeated five times. For building ESMs with the second subset, we only used the best performing parameters without tuning and built ESMs using Somers' D only as the weighting score.

4 | TRADE-OFF BETWEEN MODEL PERFORMANCE AND COMPUTATION TIME

In agreement with our initial study, Max.P and GLM were among the best performing techniques in ESMs (Table 1, Figure 2). This was also true for the simulated undersampled species (Figures S1 and S2). ANN was also among the best performing techniques, but in contrast to our previous findings, GBMs showed only intermediate performance. The algorithms performing worst with ESMs were RF, SRE, MARS and Max.T. Standard Maxent models (Max.P.s) were on average better than those techniques (Figure 2). The tuning of bivariate models generally improved the performance of ESMs (Figure 2; Figures S3 and S4). However, many algorithms require intensive computation times, e.g., model tuning of GBMs, GAM, MARS and Maxent (Figure 2; Figure S4). SRE_{best} and standard Maxent (Max.P.s) were the fastest techniques to compute but model performance was rather low for SRE and intermediate for Max.P.s (Table 1).

We identified four modelling algorithms which were especially advantageous in terms of optimizing the trade-off between ESM performance and computation time for both subsets: ANN, GLM, Max.P and CTA. The trade-off between model performance and computation time was generally better for ESMs where the bivariate models were not optimized by model tuning (Figure 2; Figure S4): these models performed only slightly less well than the tuned models, but computation time was considerably smaller. We, therefore, promote using model tuning for bivariate models, at least when computation time is not a limitation. We further recommend using the parameters which turned out to perform best (Table S1) and the techniques with an optimal trade-off mentioned above if computing time is an issue. However, similar to other studies (Qiao, Soberón, & Peterson, 2015), we could not find a single best modelling technique (Table 1) and there are also no tuning parameters for the single techniques existing which are always best (Table S1) because many factors, e.g., sampling design, a missing covariate, spatial autocorrelation or species characteristics, affect model predictions (Thibaud, Petitpierre, Broennimann, Davison,

& Guisan, 2014). Hence, it remains difficult to give a definitive conclusion about which model technique should be selected, as the suitable level of performance and the limitation of computation time must be defined by the goals of the application.

With improving technologies, computation time will likely further decrease. However, as the fastest and the slowest algorithm took on average 75 s (SRE_{best}) and 54 hr (GBM_{tuned}) respectively to run, such differences might hardly be compensated so quickly by increasing technologies (Figure S4).

While it is straightforward to compare model performance and transferability between the different techniques, it is not that easy to explain the observed differences. Although a detailed understanding of the methodological causes behind these ESM performance differences between modelling techniques would require more detailed investigations, some preliminary interpretations can still be drawn. ESMs built with RF produced the worst results. Interestingly, ESMs built with RF also performed worse compared to ESMs built with CTA, possibly because, in contrast to CTA, RF uses a large number of bootstrap samples of the data to fit an average model, and is thus itself an ensemble of submodels. This might lead to severe overfitting of the single trees which are used to build the ensemble RF model (i.e., a "forest of trees models") because, within these, the ratio of the number of predictors to the number of observations is low. SRE was also performing poorly with ESMs, as shown before in SDMs (Guisan et al., 2007), likely because SRE follows a too simplistic approach (i.e., rectangular view of species' niches) and is not able to project species occurrences beyond the range of model calibration. Our results, therefore, advise against building ESMs with RF and SRE.

5 | LIMITATIONS AND PERSPECTIVES

In this study, we used rare and undersampled species with 10–25 occurrences. Using ESMs with even lower numbers of occurrences (<10) was not tested, but would be useful to further evaluate the usability of ESMs to very rare species.

We applied internal 10-fold cross-validation, a transferability assessment to calibrate and predict the models in four different regions within the study area, as well as an approach with hold-out data for testing the effect of strong undersampling. However, there are still other options available to evaluate models, like using a checkerboard stratification (Muscarella et al., 2014), which might be an appropriate addition to the stratifications selected here.

We evaluated ESMs using four different evaluation metrics, namely AUC, the continuous Boyce index, maximum TSS and maximum Kappa, but other evaluation metrics like the partial AUC (Peterson, Papeş, & Soberón, 2008; Robin et al., 2011) could also be useful for further evaluating the performance of ESMs, comparing them with each other and with standard models. The implementation of more evaluation metrics, like partial AUC, into existing modelling software like BIOMOD2, could be valuable future developments.

A useful next extension to ESMs would be the visualization of predicted responses of species distribution to environmental variables (Elith, Ferrier, Huettmann, & Leathwick, 2005), which would complement current evaluation practices (e.g., helping to understand cases where overfitting occurs) and support a better ecological understanding of a less known species.

We used the standard Maxent models by applying a betamultiplier value of 2.5 because higher values were performing best with rare species according to recent literature (Elith et al., 2010; Moreno-Amat et al., 2015; Muscarella et al., 2014; Shcheglovitova & Anderson, 2013), but it might be worth to compare ESMs also to tuned standard Maxent models in the future. Here, we used AUC to tune predictive performance of SDMs but it might be also very useful to test model tuning using other metrics, like the Boyce index or partial AUC.

Another option worth to be tested for tuning parameters, is the use of information criteria like AICc and BIC. Information criteria were even outperforming AUC-based parameter selection for finding the optimal model complexity for Maxent models in a previous study (Warren & Seifert, 2010). While information criteria can easily be applied for many techniques (like GLMs and Maxent), implementing them to other machine-learning algorithms (like RF, ANN and GBM) is not an easy task and was outperformed by cross-validation in other studies (Hauenstein, Dormann, & Wood, 2016).

Unfortunately, all these metrics are not implemented yet in the CARET package, which was used to tune most of the modelling techniques, and building tuning functions from scratch is a difficult and time intensive task.

6 | CONCLUSIONS

Our results allow proposing new recommendations on how to build optimal ESMs for modelling species with low sample sizes. The selection of the most useful algorithm depends on the computation resources available and on all the factors increasing computation time (e.g., the number of species to be modelled or whether the models should be projected to a large number of cells in the study area). It is also important to consider whether one wants models to perform well within a given study area or to transfer well to other areas. When the overall aim is to produce models with a high predictive performance, as measured by cross-validation, the best choices are ESMs with ANN_{tuned} and GBM. When the overall aim is to build robust models which are transferable to other regions, Max.P with tuned feature types would be the optimal solution. However, when computation time is particularly limited, e.g., when large numbers of species have to be modelled, then ESMs based on ANN, GLM and CTA are optimal. The remaining algorithms—SRE, MAX.T, GAM and MARS - performed less well and/or had a longer computing time than the previously mentioned modelling techniques.

ACKNOWLEDGEMENTS

We thank Melissa Dawes for revising the English of this article, the Swiss Federal Office of the Environment (FOEN) for financial support

to an initial phase of the project, the Swiss Floristic Data Center (Info Flora) for providing the species data and N.E. Zimmermann, K. Ecker and A. Lehmann for providing the environmental predictor layers for modelling. The computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics. We thank A. Townsend Peterson and two anonymous reviewers for useful comments on the manuscript. FB obtain partial support from the Swiss National Science Foundation (NSF; grant nr 1528661 to AG).

AUTHORS' CONTRIBUTIONS

All authors conceived the ideas for this manuscript. F.T.B. executed the analyses and led the writing, with contributions from all co-authors. A.B., M.P.N. and A.G. developed the larger framework and obtained the funding for the project.

DATA ACCESSIBILITY

Data on species occurrences (presence-only) were provided by the Swiss national data centre Info Flora (www.infoflora.ch) and are archived there. The functions used to calibrate ensembles of small models are available in the ecospat R-package (https://cran.r-project.org/package=ecospat).

ORCID

Frank T. Breiner http://orcid.org/0000-0003-4465-1684

REFERENCES

Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). Bioclim: The first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions*, *20*, 1–9. https://doi.org/10.1111/ddi.12144

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth and Brooks/Cole.

Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, *6*, 1210–1218. https://doi.org/10.1111/2041-210X.12403

Broennimann, O., Petitpierre, B., Randin, C.F., Engler, R., Breiner, F.T., D'Amen, M., … Guisan, A. (2015). Ecospat: Spatial ecology miscellaneous methods. R package version 1.0. http://CRAN.R-project.org/package=ecospat.

Busby, J. R. (1991). BIOCLIM - a bioclimate analysis and prediction system. In C. R. Margules, & M. P. Austin (Eds.), *Nature conservation: Cost effective biological surveys and data analysis* (pp. 64–68). Melbourne, Vic.: CSIRO.

D'Amen, M., Pradervand, J. N., & Guisan, A. (2015). Predicting richness and composition in mountain insect communities at high resolution: A new test of the SESAM framework. *Global Ecology and Biogeography*, *24*, 1443–1453. https://doi.org/10.1111/geb.12357

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., … Guisan, A. (2017). Ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, *40*, 774–787. https://doi.org/10.1111/ecog.02671

Elith, J., Ferrier, S., Huettmann, F., & Leathwick, J. R. (2005). The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. *Ecological Modelling*, *186*, 280–289. https://doi.org/10.1016/j.ecolmodel.2004.12.007

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, *1*, 330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.x

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, *19*, 1–67. https://doi.org/10.1214/aos/1176347963

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781139028271

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*, 1424–1435. https://doi.org/10.1111/ele.12189

Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Phillips, S., & Peterson, A. T. (2007). What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs*, *77*, 615–630. https://doi.org/10.1890/06-1060.1

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, *1*, 297–310. https://doi.org/10.1214/ss/1177013604

Hauenstein, S., Dormann, C.F., & Wood, S.N. (2016). Computing AIC for black-box models using generalized degrees of freedom: A comparison with cross-validation. *Communications in Statistics - Simulation and Computation*, https://doi.org/10.1080/03610918.2017.1315728

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C. F., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, *199*, 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

Jurka, T.P., & Tsuruoka, Y. (2013). Maxent: Low-memory multinomial logistic regression with support for text classification. R package version 1.3.3.1. https://CRAN.R-project.org/package=maxent.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*, 1–26.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York, NY: Springer. https://doi.org/10.1007/978-1-4614-6849-3

Lomba, A., Pellissier, L., Randin, C. F., 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. *Biological Conservation*, *143*, 2647–2657. https://doi.org/10.1016/j.biocon.2010.07.007

Moreno-Amat, E., Mateo, R. G., Nieto-Lugilde, D., Morueta-Holme, N., Svenning, J.-C., & García-Amorena, I. (2015). Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data. *Ecological Modelling*, *312*, 308–317. https://doi.org/10.1016/j.ecolmodel.2015.05.035

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. *Methods in Ecology and Evolution*, *5*, 1198–1205. https://doi.org/10.1111/2041-210X.12261

Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Peterson, A. T. (2007). Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *Journal of Biogeography*, *34*, 102–117.

Peterson, A. T., Papeş, M., & Soberón, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, *213*, 63–72. https://doi.org/10.1016/j.ecolmodel.2007.11.008

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Platts, P. J., Garcia, R. A., Hof, C., Foden, W., Hansen, L. A., Rahbek, C., & Burgess, N. D. (2014). Conservation implications of omitting narrow-ranging taxa from species distribution models, now and in the future. *Diversity and Distributions*, *20*, 1307–1320. https://doi.org/10.1111/ddi.12244

Qiao, H., Soberón, J., & Peterson, A. T. (2015). No silver bullets in correlative ecological niche modelling: Insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and Evolution*, *6*, 1126–1136. https://doi.org/10.1111/2041-210X.12397

Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511812651

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 77. https://doi.org/10.1186/1471-2105-12-77

Shcheglovitova, M., & Anderson, R. P. (2013). Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample sizes. *Ecological Modelling*, *269*, 9–17. https://doi.org/10.1016/j.ecolmodel.2013.08.011

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285–1293. https://doi.org/10.1126/science.3287615

Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A. C., & Guisan, A. (2014). Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution*, *5*, 947–955. https://doi.org/10.1111/2041-210X.12203

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD: A platform for ensemble forecasting of species distributions. *Ecography*, *32*, 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Tsuruoka, Y., Tsujii, J., & Ananiadou, S. (2009). Stochastic gradient descent training for L1-regularized log-linear models with cumulative penalty. Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pp. 477–485.

Warren, D. L., & Seifert, S. N. (2010). Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications*, *21*, 335–342.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, *14*, 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Breiner FT, Nobis MP, Bergamini A, Guisan A. Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Methods Ecol Evol*. 2018;9:802–808. https://doi.org/10.1111/2041-210X.12957