

ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions

Jamie M. Kass^{1,2,3}  | Robert Muscarella⁴  | Peter J. Galante⁵  | Corentin L. Bohl⁶  |
 Gonzalo E. Pinilla-Buitrago^{2,3}  | Robert A. Boria⁷  | Mariano Soley-Guardia⁸ |
 Robert P. Anderson^{2,3,9} 

¹Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan; ²PhD Program in Biology, The Graduate Center, City University of New York, New York, NY, USA; ³Department of Biology, City College of New York, City University of New York, New York, NY, USA; ⁴Plant Ecology and Evolution, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden; ⁵Center for Biodiversity and Conservation, American Museum of Natural History, New York, NY, USA; ⁶Innland County Council, Hadeland Videregående Skole, Gran, Norway; ⁷Quantitative and Systems Biology Graduate group, University of California-Merced, Merced, CA, USA; ⁸Escuela de Biología, Universidad de Costa Rica, and Ciudad Universitaria, San Pedro, Costa Rica and ⁹Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, New York, NY, USA

Correspondence

Jamie M. Kass
 Email: jamie.m.kass@gmail.com

Funding information

Japan Society for the Promotion of Science; National Aeronautics and Space Administration, Grant/Award Number: 80NSSC18K0406; National Science Foundation, Division of Biological Infrastructure, Grant/Award Number: DBI-1661510; Ford Foundation; Vetenskapsrådet, Grant/Award Number: 2019-03758

Handling Editor: Laura Graham

Abstract

- Quantitative evaluations to optimize complexity have become standard for avoiding overfitting of ecological niche models (ENMs) that estimate species' potential geographic distributions. ENMeval was the first R package to make such evaluations (often termed model tuning) widely accessible for the Maxent algorithm. It also provided multiple methods for partitioning occurrence data and reported various performance metrics.
- Requests by users, recent developments in the field, and needs for software compatibility led to a major redesign and expansion. We additionally conducted a literature review to investigate trends in ENMeval use (2015–2019).
- ENMeval 2.0 has a new object-oriented structure for adding other algorithms, enables customizing algorithmic settings and performance metrics, generates extensive metadata, implements a null-model approach to quantify significance and effect sizes, and includes features to increase the breadth of analyses and visualizations. In our literature review, we found insufficient reporting of model performance and parameterization, heavy reliance on model selection with AICc and low utilization of spatial cross-validation; we explain how ENMeval 2.0 can help address these issues.
- This redesigned and expanded version can promote progress in the field and improve the information available for decision-making.

KEY WORDS

cross-validation, ecological niche model, metadata, model evaluation, model tuning, software, spatial, species distribution model

1 | INTRODUCTION

Models that estimate species' niches and geographic distributions (hereafter ecological niche models [ENMs]) based on occurrence

data and environmental predictor variables are frequently employed to test biogeographical hypotheses and inform management and conservation (Guisan et al., 2017; Peterson et al., 2011). Many algorithms exist, but machine learning approaches have gained

popularity due to their ability to fit responses of varied complexity with high predictive performance (Elith & Graham, 2009). Although choices related to complexity can have great impacts on models (Hallgren et al., 2019), default settings have often been used. This commonly leads to overfit models that reflect noise in the data (Elith & Graham, 2009; Merow et al., 2014) and have poor transferability when applied to novel conditions (Guevara et al., 2018). Quantitative evaluation constitutes a key element of modeling, and assessing different levels of complexity to select optimal models (often termed model tuning) can help determine the level of complexity warranted (Radosavljevic & Anderson, 2014).

ENMeval (Muscarella et al., 2014) was the first package in the R programming language (R Core Team, 2021) to provide automated model evaluation and tuning for ENMs (using the algorithm Maxent; Phillips et al., 2006, 2017). Since its release, ENMeval has been frequently cited in a variety of theoretical and applied studies (from March 2021: Web of Science 'highly cited paper' [499 citations], Google Scholar [702 citations], over 71,000 CRAN downloads). It is now the basis for model-tuning analyses in the *Wallace* ecological modeling application (Kass et al., 2018) and has a vibrant user community via the Maxent and *Wallace* Google Groups.

The popularity of ENMeval is likely due to three interrelated characteristics. First, it provides a simple workflow that: (a) partitions occurrence and background data into subsets for training and validation (Hastie et al., 2009), (b) builds models with different algorithmic settings and (c) evaluates their performance. Second, it

offers different partitioning methods, including spatial options then new to the field. Third, to aid in model selection, it calculates various performance metrics appropriate for such ENMs made with different settings but identical data (Peterson et al., 2011). Other R packages provide functionality for building and evaluating ENMs, but few provide multiple partitioning tools, have a focus on model tuning and evaluation or report a broad spectrum of appropriate performance metrics (Appendix 1).

Recent developments and requests by users guided the redesign of the package. We aimed to: make the package more generalizable to other algorithms, implement emerging metadata standards to improve reproducibility (Merow et al., 2019), include new model evaluation methods (Bohl et al., 2019) and address user requests for compatibility with different data types. This required large structural changes and resulted in a complete redesign of the package, while keeping the same simple workflow (Figure 1). Below, we describe the main features of ENMeval 2.0 (<https://jamiemkass.github.io/ENMeval/>). Other important changes and enhancements are outlined in Table 1 and Figure 1, as well as discussed in detail in Appendix 2, and highlighted through worked examples in an updated vignette (<https://jamiemkass.github.io/ENMeval/articles/ENMeval-2.0.0-vignette.html>). We also report the results of a literature review to uncover trends in ENMeval use, and assess how the new package can address several shortcomings in the field. We conclude with how this package can help move ENM research forward and finally outline its future development.

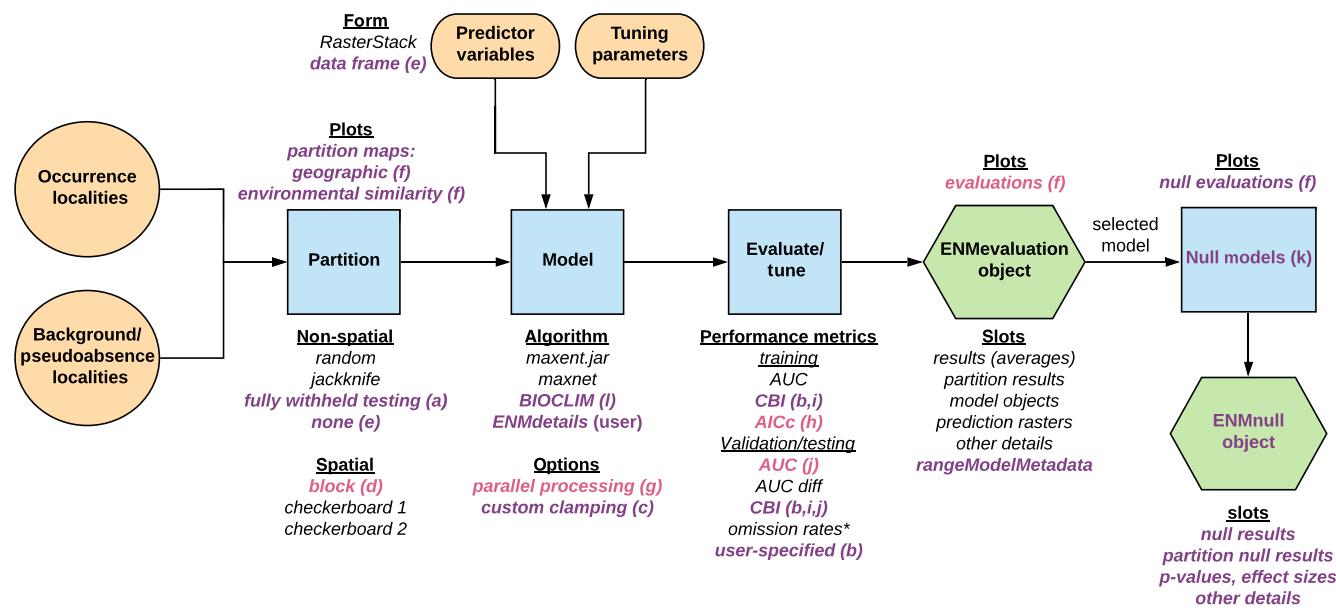


FIGURE 1 Schematic of the ENMeval 2.0 workflow, placing advances in the context of steps included in the earlier version. Purple text indicates new functionalities, and pink denotes extended ones; letters in parentheses refer to sections of Appendix 2 that explain each in detail. Shapes illustrating the workflow correspond to the following elements: input data appear in orange, operations in blue and outputs in green. The main analysis steps include partitioning data, building models and evaluating the models/tuning their settings—model selection by the researcher is based on the quantitative measures of performance calculated by ENMeval. Because of computational intensity, the null-model functionalities occur separately and are envisioned to be performed for the selected model. *The two omission rates provided automatically for validation/testing data correspond to the following thresholds determined based on training data: minimum training presence and 10% omission of training presences (Muscarella et al., 2014)

TABLE 1 New enhancements in ENMeval 2.0 (detailed in Appendix 2) in addition to the ‘key features’ described in the text. Note that in ENMeval 2.0, these are now available for use by any algorithm specified

Feature	Description
Fully withheld test-data evaluations	Model evaluation on data never used for training; in comparison, cross-validation in ENMeval evaluates models on successively withheld partitions (i.e., validation data) that are ultimately used in building the final model
New and customizable performance metrics	Continuous Boyce Index (Hirzel et al., 2006) for training/validation data and ability to use R functions specifying custom metrics
Customizable clamping function	Independent clamping of upper and lower tails of individual predictor variables; selection can be based, for example, on whether the marginal response curve is increasing or decreasing at the point of truncation (Guevara et al., 2018)
Custom spatial-block partition orientation	Flexibility to change longitude–latitude orientation of spatial block partition
Model tuning without raster data	Model tuning using data tables specifying predictor variable information for occurrence and background data
Expanded plotting functions	New ggplot2 (Wickham, 2016) functions to map data partitions, map environmental similarity (including MESS; Elith et al., 2010) among partitions to determine the degree of extrapolation during cross-validation and plot results of null models and model-tuning exercises
Extensive package unit tests	New automated unit tests ensure that key functions work with new package development

2 | KEY NEW FEATURES OF ENMEVAL 2.0

2.1 | Integrating additional algorithms and customizing their settings

ENMeval 2.0 has a new structure enabling modeling with any algorithm given user-determined specifications. As there can be considerable differences in the implementations of algorithms, the package now uses an object-oriented approach to flexibly specify algorithms and their critical settings. The customizable `ENMdetails` object, shareable as an R script file, contains information regarding how to run the model, extract the number of model coefficients and make predictions. Several other R packages have features for implementing algorithms with custom R objects (Appendix 1), but `ENMdetails` objects are simpler and designed to facilitate model evaluation exercises. The algorithms Maxent (Phillips et al., 2017), via `maxnet` (R package) and `maxent.jar` (Java) and BIOCLIM (with R package `dismo`; Hijmans et al., 2020) are now implemented as example `ENMdetails` objects to illustrate the capacity for expansion (details for BIOCLIM implementation in Appendix 2). These examples serve as templates for developing new `ENMdetails` objects for other algorithms. Going forward, we plan to coordinate with experts in other algorithms to add them using best practices for presence–background data.

2.2 | Automatic generation of metadata

The package now produces metadata for the analyses run. The *Range Model Metadata Standards* framework (Merow et al., 2019) provides a comprehensive catalogue of different aspects of ENM analyses. This framework is the basis for a recently proposed standardized protocol for reporting on ENMs (Zurell et al., 2020), following comprehensive reproducibility standards (Feng et al., 2019) and facilitating the assessment of model adequacy for applied uses

(Araújo et al., 2019). ENMeval 2.0 uses the `rangeModelMetadata` package (Merow et al., 2019) to input metadata categories crucial for ENM reporting (see Tables S1 and S2 for relationships with Feng et al., 2019 and Zurell et al., 2020). Metadata corresponding to non-automated analysis steps (i.e., data collection, model selection) should be added independently (see the updated vignette for an example). Metadata objects can be saved as text files for sharing or for use in supplemental materials.

2.3 | Null models: Significance and effect size for performance metrics

ENMeval 2.0 automates a recently developed null-model approach to quantify significance and effect sizes of performance metrics. Null models can account for features of the system—such as spatial autocorrelation or unequal environmental representation—that often lead to incorrect statistical conclusions, especially with background or pseudo-absence data (Bohl et al., 2019). Without needing to quantify these features explicitly, the approach implemented here evaluates both empirical and null models with the same withheld data (instead of random data), and has been shown to produce more statistically reliable results than earlier ones (Bohl et al., 2019; Kass et al., 2020).

3 | LITERATURE REVIEW

We conducted a literature review of articles citing ENMeval to understand how the package has been used and inform recommendations for the use of version 2.0. We searched Web of Science (accessed June 2019) for articles that cited the original publication from October 2015 to May 2019. The search returned 214 studies, and of these we reviewed a total of 182 (Appendix 3; some were

omitted due to access constraints). For the main text, and supplemental materials when necessary, we recorded: (a) if ENMeval was used for analysis or only cited for concepts, (b) which functionalities were used, (c) which method was used to partition data, (d) which performance metric(s) were used and reported (studies may report other metrics in addition to those used to select models), (e) whether the selected Maxent settings (i.e., regularization multiplier and feature class combinations) were indicated and (f) whether variation in model performance across settings was given.

Of the 182 studies reviewed, 78% ($n = 141$) used ENMeval in the analysis. Of these, the majority (93%) used the package to evaluate model performance or select among candidate models (i.e., tune to identify 'optimal' settings). Most (82%) indicated the method used to partition data (Figure 2a). A majority of the 141 studies (90%) reported at least one performance metric to select approximate 'optimal' settings, and only 18% used more than one. Most of these studies employed AICc, and the second most common metric was

validation AUC (called AUC_{TEST} in previous ENMeval versions; Figure 2b,c). However, 82% of the 141 gave the actual values of the chosen metrics: the most reported was validation AUC, followed by AICc and then validation omission rates (Figure 2b). A smaller 57% of the 141 provided the selected model settings, and only 24% characterized variation of model performance across the settings used.

4 | DISCUSSION

Although most studies citing ENMeval over the past few years indeed used the package to evaluate model performance and identify optimal levels of complexity, key methodological and reporting issues remain. We first discuss how each can impede reproducible science and lead to incorrect conclusions, then explain how new features in ENMeval 2.0 can help address them. Finally, we outline ideas for future work on this package and the broader field.

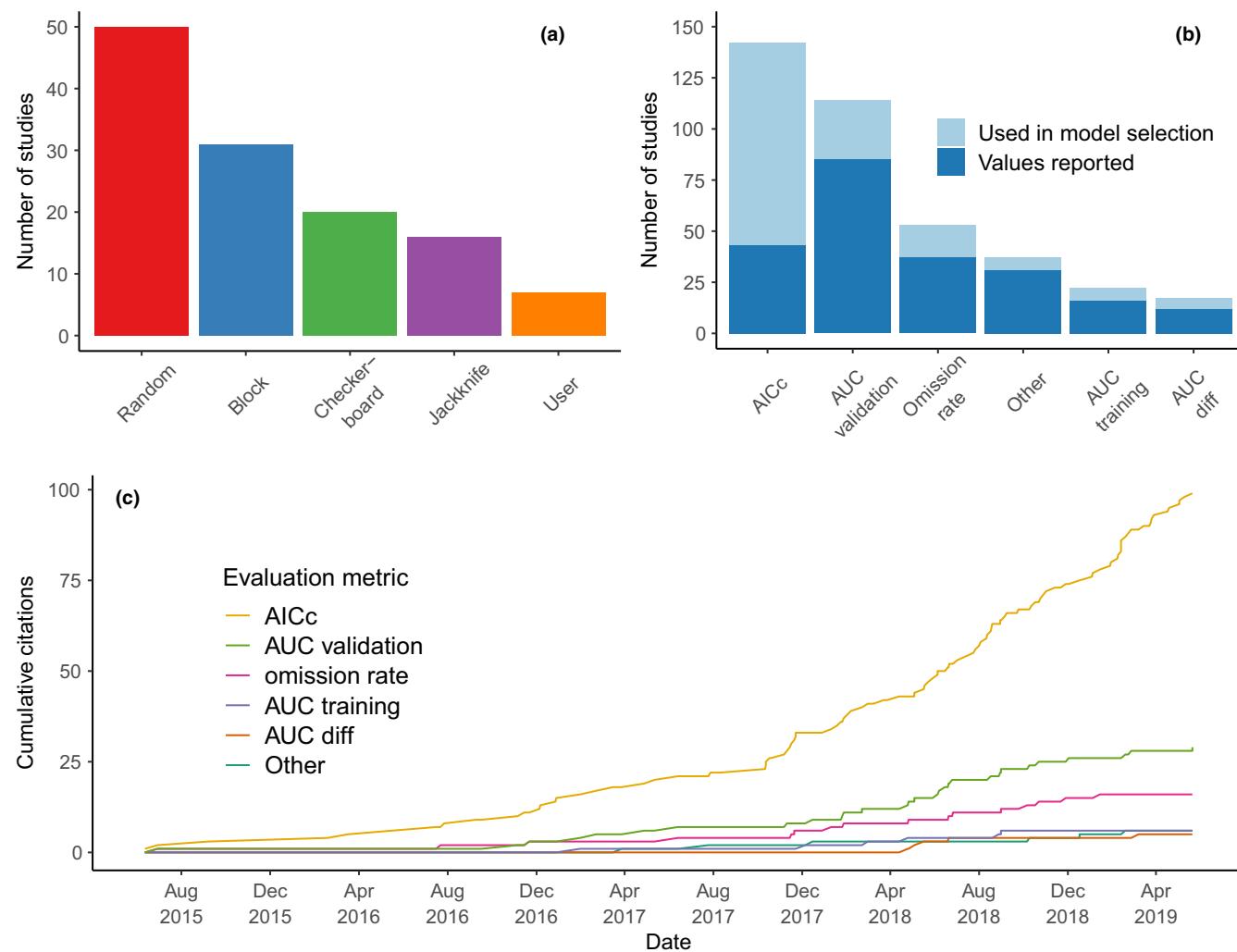


FIGURE 2 Results of the literature review for studies that used ENMeval for analysis (October 2015–May 2019). (a) The number of studies that used particular ENMeval partitioning methods for cross-validation (out of 115 reporting, 8 reported >1 partition method). (b) The number of studies that: (1) used each evaluation metric for model selection (out of 127 reporting, 26 reported >1 metric) and (2) gave values for that metric (out of 116 reporting, 73 reported >1 metric). (c) Cumulative citations for the analyzed period that used each evaluation metric for model selection (AUC diff = AUC training minus AUC validation)

4.1 | Reporting on model performance and parameterization

Although most studies reported how data were partitioned for evaluation, the details of the tuning analysis were often minimal, presenting challenges for reproducibility (Zurell et al., 2020). Most studies reported only one performance metric, and just over half provided optimal model settings. In their literature review of the field overall, Feng et al. (2019) found that less than half of surveyed studies gave details on the data evaluation method or the model settings. In comparison, our results show that studies using ENMeval more consistently reported key information regarding the evaluations. Importantly, ENMeval 2.0 has default metadata generation to help users with reporting. In addition, users can characterize the key aspects of uncertainty (as called for by Araújo et al., 2019 and Zurell et al., 2020) by comparing the predictions of models showing similar performance, as well as by making ensemble predictions for different settings of the same algorithm (Boria et al., 2017) or for different algorithms (Buisson et al., 2010) using **ENMdetails** objects.

4.2 | Overreliance of model selection with AICc

The most common model-selection method across the surveyed studies was AICc, which is calculated on the full dataset, rather than on validation or fully withheld testing data (Table 1). Such metrics do not consider the results of model validation or testing, and thus neither provide a true evaluation of the model nor allow assessment of its transferability (Roberts et al., 2017; Velasco & González-Salazar, 2019). In particular for studies that transfer to new areas or time periods, model selection using *spatial* cross-validation (see below) or fully withheld test data from another region should result in models with better transferability (Soley-Guardia et al., 2019). When users prefer to select models with AICc, which comes with a statistical caveat for Maxent (Warren & Seifert, 2011), ENMeval 2.0 can additionally quantify how well such models predict withheld data compared with null models (Kass et al., 2020). Lastly, AICc can also be considered along with other performance metrics (Galante et al., 2018), and ENMeval 2.0 both provides more than the earlier version and allows users to add others.

4.3 | Underutilization of spatial cross-validation

Only about half of the surveyed studies chose spatial partitions over random ones, although the former is recommended for most ENM applications with sufficient data (Veloz, 2009). Random partitions typically lead to low independence between training and validation datasets, which can be particularly problematic when biological sampling bias exists (Radosavljevic & Anderson, 2014), and in such situations results in artificially high performance (Roberts et al., 2017). Spatial cross-validation (i.e., 'block' and 'checkerboard' partitions in ENMeval, see Table 1 and Appendix 2 for new options) increases

the independence of the validation data and tends to force models to extrapolate more environmentally, thus evaluating transferability to new conditions (Soley-Guardia et al., 2019). Importantly, spatial partitioning also can be performed using tools from other packages, such as blockCV (Valavi et al., 2019), and then used in ENMeval for modeling (see updated vignette). Further, the new plots showing Multivariate Environmental Similarity Surfaces (MESS; Table 1; Elith et al., 2010) for partitions should help inform users about the magnitude of environmental extrapolation introduced for different partitioning schema, and the new clamping function provides more flexible control of how models extrapolate (Table 1).

4.4 | Potential uses and future development

ENMeval 2.0 enables a range of potential basic and applied studies, which should both promote progress in the field and improve the information available for decision-making. For example, broader model-selection strategies informed by different performance metrics and evaluation approaches can be compared to produce species' range estimates with uncertainty characterized, which is particularly important for conservation or management scenarios (Wright et al., 2015). Additionally, how the choice of partition schema affects the level of extrapolation in model evaluations can be better explored.

Finally, it is also important to note that feedback from users of the R-based GUI software *Wallace* for modeling niches and distributions (Kass et al., 2018) inspired many of the enhancements that led to ENMeval 2.0, especially functionalities to use algorithms other than Maxent and to yield standardized metadata. The tight interactions between these two packages and their active user communities should help to drive innovations in both of them that advance the field. In particular, we anticipate coordinating with experts to implement presence-background ENMs with other algorithms in the ENMeval 2.0 framework and to applying emerging standards for reporting and assessing such models. In sum, we look forward to continued reciprocal advances in both methods and community standards.

ACKNOWLEDGEMENTS

J.M.K. was supported by the Japan Society for the Promotion of Science Postdoctoral Fellowships for Foreign Researchers Program and the Okinawa Institute of Science and Technology Graduate University; R.M. was supported by grant 2019-03758 from the Swedish Research Council (Vetenskapsrådet); and R.A.B. was supported by the Ford Foundation (Predoctoral Fellowship). J.M.K., G.E.P.-B. and R.P.A. acknowledge support from the US National Science Foundation (NSF; DBI-1661510) and National Aeronautics and Space Administration (80NSSC18K0406 to Mary E. Blair). The authors thank Steven J. Phillips for development of the clamping function, the Maxent and *Wallace* Google Group users for invaluable input, Mary E. Blair, the Anderson Lab group, and two anonymous reviewers for constructive comments on this manuscript. Finally, they thank Olivier Broennimann, Angela P. Cuervo-Robayo, Claudio Mota-Vargas, Octavio R. Rojas-Soto, Nao

Takashina, and Mari Takenouchi for helping with translating the abstract and title to French, Japanese and Spanish.

AUTHORS' CONTRIBUTIONS

J.M.K. envisioned the redesign and content of the updated package, in consultation with R.P.A., R.M., C.L.B. and G.E.P.-B. J.M.K. programmed the package with input from G.E.P.-B., R.M. and C.L.B. J.M.K. and P.J.G. wrote the vignette with input from R.P.A. and G.E.P.-B. J.M.K., R.A.B., G.E.P.-B., C.L.B. and M.S.-G. conducted package testing, including running unit tests and comparing with prior worked analyses. R.M. designed the literature review with feedback from all the authors and conducted it with J.M.K., P.J.G., G.E.P.-B. and M.S.-G. J.M.K. drafted the manuscript with major contributions by R.P.A., M.S.-G., C.L.B. and G.E.P.-B. All the authors contributed to and approved the final manuscript.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13628>.

DATA AVAILABILITY STATEMENT

The R package code is available on Github (<https://github.com/jamiemkass/ENMeval>) and CRAN (<https://CRAN.R-project.org/package=ENMeval>), and is additionally archived on Zenodo (<https://doi.org/10.5281/zenodo.4768605>; Kass et al., 2021). The vignette is available on the package website (<https://jamiemkass.github.io/ENMeval/articles/ENMeval-2.0.0-vignette.html>). The publication dates, citations and DOIs of the papers used in the literature review are archived on figshare (<https://www.doi.org/10.6084/m9.figshare.14464743>).

ORCID

Jamie M. Kass  <https://orcid.org/0000-0002-9432-895X>
 Robert Muscarella  <https://orcid.org/0000-0003-3039-1076>
 Peter J. Galante  <https://orcid.org/0000-0002-7025-3551>
 Corentin L. Bohl  <https://orcid.org/0000-0003-1418-8430>
 Gonzalo E. Pinilla-Buitrago  <https://orcid.org/0000-0002-0065-945X>
 Robert A. Boria  <https://orcid.org/0000-0001-6813-4222>
 Robert P. Anderson  <https://orcid.org/0000-0002-7706-4649>

REFERENCES

Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naiimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5, eaat4858.

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. *Journal of Biogeography*, 46, 1101–1111.

Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2017). A single-algorithm ensemble approach to estimating suitability and uncertainty: Cross-time projections for four Malagasy tenrecs. *Diversity & Distributions*, 23, 196–208.

Buisson, L., Thuiller, W., Casajus, N., Lek, S., & Grenouillet, G. (2010). Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, 16, 1145–1157.

Elith, J., & Graham, C. H. (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, 32, 66–77.

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1, 330–342.

Feng, X., Park, D. S., Walker, C., Peterson, A. T., Merow, C., & Papeş, M. (2019). A checklist for maximizing reproducibility of ecological niche models. *Nature Ecology and Evolution*, 3, 1382–1395.

Galante, P. J., Alade, B., Muscarella, R., Jansa, S. A., Goodman, S. M., & Anderson, R. P. (2018). The challenge of modeling niches and distributions for data-poor species: A comprehensive approach to model complexity. *Ecography*, 41, 726–736.

Guevara, L., Gerstner, B. E., Kass, J. M., & Anderson, R. P. (2018). Toward ecologically realistic predictions of species distributions: A cross-time example from tropical montane cloud forests. *Global Change Biology*, 24, 1511–1522.

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: With applications in R*. Cambridge University Press.

Hallgren, W., Santana, F., Low-Choy, S., Zhao, Y., & Mackey, B. (2019). Species distribution models can be highly sensitive to algorithm configuration. *Ecological Modelling*, 408, 108719.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2020). *dismo: Species distribution modeling*. R package version 1.3-3. Retrieved from <https://CRAN.R-project.org/package=dismo>

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199, 142–152.

Kass, J. M., Anderson, R. P., Espinosa-Lucas, A., Juárez-Jaimes, V., Martínez-Salas, E., Botello, F., Tavera, G., Flores-Martínez, J. J., & Sánchez-Cordero, V. (2020). Biotic predictors with phenological information improve range estimates for migrating monarch butterflies in Mexico. *Ecography*, 43, 341–352.

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, April 28). *jamiemkass/ENMeval: ENMeval v2.0.0 (Version v2.0.0)*. Zenodo. <https://doi.org/10.5281/zenodo.4768605>

Kass, J. M., Vilela, B., Aiello-Lammens, M. E., Muscarella, R., Merow, C., & Anderson, R. P. (2018). *Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion*. *Methods in Ecology and Evolution*, 9, 1151–1156.

Merow, C., Maitner, B. S., Owens, H. L., Kass, J. M., Enquist, B. J., Jetz, W., & Guralnick, R. (2019). Species' range model metadata standards: RMMS. *Global Ecology and Biogeography*, 28, 1912–1924.

Merow, C., Smith, M. J., Edwards Jr., T. C., Guisan, A., McMahon, S. M., Normand, S., Thuiller, W., Wüest, R. O., Zimmermann, N. E., & Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography*, 37, 1267–1281.

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, 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.

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 (MPB-49)*. Princeton University Press.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, 40, 887–893.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography*, 41, 629–643.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., & Warton, D. I. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913–929.

Soley-Guardia, M., Carnaval, A. C., & Anderson, R. P. (2019). Sufficient versus optimal climatic stability during the Late Quaternary: Using environmental quality to guide phylogeographic inferences in a Neotropical montane system. *Journal of Mammalogy*, 100, 1783–1807.

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10, 225–232.

Velasco, J. A., & González-Salazar, C. (2019). Akaike information criterion should not be a 'test' of geographical prediction accuracy in ecological niche modelling. *Ecological Informatics*, 51, 25–32.

Veloz, S. D. (2009). Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, 36, 2290–2299.

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications*, 21, 335–342.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. Retrieved from <https://ggplot2.tidyverse.org>

Wright, A. N., Hijmans, R. J., Schwartz, M. W., & Shaffer, H. B. (2015). Multiple sources of uncertainty affect metrics for ranking conservation risk under climate change. *Diversity and Distributions*, 21(1), 111–122.

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. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43, 1261–1277.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Kass JM, Muscarella R, Galante PJ, et al. ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. *Methods Ecol Evol*. 2021;12:1602–1608. <https://doi.org/10.1111/2041-210X.13628>