

Methods in Ecology and Evolution

JAMIE KASS (Orcid ID : 0000-0002-9432-895X)

DR ROBERT MUSCARELLA (Orcid ID : 0000-0003-3039-1076)

MR PETER J. GALANTE (Orcid ID : 0000-0002-7025-3551)

Article type : Application

Editor : Laura Graham

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

Authors:

Jamie M. Kass [1,2,3] (*corresponding author*): <https://orcid.org/0000-0002-9432-895X>

Robert Muscarella [4]: <https://orcid.org/0000-0003-3039-1076>

Peter J. Galante [5]: <https://orcid.org/0000-0002-7025-3551>

Corentin L. Bohl [6]: <https://orcid.org/0000-0003-1418-8430>

Gonzalo E. Pinilla-Buitrago [2,3]: <https://orcid.org/0000-0002-0065-945X>

Robert A. Boria [7]: <https://orcid.org/0000-0001-6813-4222>

Mariano Soley-Guardia [8]

Robert P. Anderson [2,3,9]: <https://orcid.org/0000-0002-7706-4649>

1. Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna-son, Kunigami-gun, Okinawa, 904-0495, Japan
email: jamie.m.kass@gmail.com
2. Ph.D. Program in Biology, The Graduate Center, City University of New York, New York, NY, 10016, USA

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/2041-210X.13628](https://doi.org/10.1111/2041-210X.13628)

This article is protected by copyright. All rights reserved

- Accepted Article
3. Department of Biology, City College of New York, City University of New York, New York, NY, 10031, USA
 4. Plant Ecology and Evolution, Evolutionary Biology Centre, Uppsala University, SE-75236, Uppsala, Sweden
 5. Center for Biodiversity and Conservation, American Museum of Natural History, New York, NY, 10024, USA
 6. Innland County Council, Hadeland Videregående Skole, Gran, Norway
 7. Quantitative and Systems Biology Graduate group, University of California-Merced, Merced, CA, 95343, USA
 8. Escuela de Biología, Universidad de Costa Rica, Ciudad Universitaria, 11501-2060 San Pedro, Costa Rica
 9. Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, New York, NY, 10024, USA

Headline: customizable and reproducible niche/range modeling

Abstract

1. 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.
2. 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).
3. `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.
4. This redesigned and expanded version can help promote progress in the field and improve the information available for decision-making.

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

表題：ENMeval 2.0: 再設計されたカスタマイズおよび再現可能な種のニッチ分布モデル

要旨

1) 種の分布予測を行う生態学的ニッチモデル (ecological niche models=ENMs) の過剰適合を避けるため、モデルの複雑さを最適化するための定量評価を行うことが一般的である。

ENMeval はモデルチューニングとも呼ばれるこの評価を、Maxent モデルに対して広く利用可能にする初の R パッケージである。またこのパッケージは、種の在データを分割する複数の手法・評価測定を提供する。

2) ユーザからのリクエストや本研究分野の最近の発展、ソフトウェアの互換性の問題などから、ENMeval の大幅な再開発と拡張につながる事となった。さらに ENMeval の使用トレンド (2015–2019) を調べるため、文献調査を行った。

3) ENMeval 2.0 には、他のアルゴリズムを追加する新たなオブジェクト指向ストラクチャ、アルゴリズムと評価指標のカスタマイズ、大量のメタデータ生成、統計学的有意性と効果量を測定するヌルモデルの実装、解析手法と視覚化の選択肢の充足といった拡張が組み込まれている。また、文献調査によりモデル評価とパラメータ設定に関する報告の不足、AICc のモデル選択に依存過剰であること、更に空間的交差検証の使用頻度が低い事が明らかとなった。我々はこれらの問題を ENMeval 2.0 がどのように対処できるかを議論する。

4) 再設計により拡大された ENMeval 2.0 は、本研究分野の発展を促進し、意思決定に利用可能な情報の充足に寄与する事が期待される。

キーワード：生態学的ニッチモデル、交差検証、メタデータ、モデル評価、モデルチューニング、ソフトウェア、空間的、種の分布モデル

Título: ENMeval 2.0: Rediseñado para el modelamiento personalizable y reproducible de nichos y distribuciones de especies

Resumen

1) Se ha vuelto un estándar la evaluación cuantitativa para optimizar la complejidad en los modelos de nicho ecológico (MNE), los cuales estiman la distribución geográfica potencial de las especies, con el propósito de evitar su sobreajuste. ENMeval fue el primer paquete de R que hizo que esta evaluación (denominada también ajuste del modelo) fuera accesible para el algoritmo Maxent. También permitió el uso de múltiples métodos para hacer particiones de los datos de presencia y de varias métricas para evaluar el desempeño de los modelos.

2) La versión 2.0 fue rediseñada, incluyendo nuevas funcionalidades, como respuesta a las solicitudes de usuarios y a los desarrollos recientes en el campo, asimismo para garantizar su compatibilidad con otros programas. Adicionalmente, realizamos una revisión de la literatura para investigar las tendencias en el uso de ENMeval entre 2015 y 2019.

3) ENMeval 2.0 tiene una nueva estructura orientada a objetos para agregar otros algoritmos, permite personalizar la parametrización de los algoritmos y las métricas de desempeño, genera metadatos extensos, implementa una aproximación de modelo nulo para cuantificar la significancia y el tamaño del efecto en métricas de desempeño, e incluye características que incrementan la variedad de opciones de análisis y de visualización. En nuestra revisión de la literatura, encontramos insuficiente reporte de las métricas de desempeño y la parametrización de los modelos, una gran dependencia a seleccionar modelos solamente con AICc y una poca utilización de la validación cruzada con particiones espaciales. Además, explicamos cómo ENMeval 2.0 puede ayudar a abordar los problemas encontrados en esta revisión.

4) Esta versión rediseñada y ampliada busca promover avances en el campo de modelamiento de nicho ecológico y mejorar la información disponible para la toma de decisiones.

Palabras claves: ajuste del modelo, espacial, evaluación del modelo, metadatos, modelo de distribución de especies, modelo de nicho ecológico, software, validación cruzada

ENMeval 2.0: restructurée pour une modélisation personnalisable et reproductible de la niche écologique et de la distribution des espèces

Titre: ENMeval 2.0: restructurée pour une modélisation personnalisable et reproductible de la niche écologique et de la distribution des espèces

Résumé

- 1) Les évaluations qualitatives pour optimiser la complexité sont devenues communes afin d'éviter le surajustement des modèles de niche écologique (ENM) qui estiment les distributions géographiques potentielles des espèces. ENMeval fut la première librairie R à rendre largement accessibles ce type d'évaluations (souvent appelées "model tuning" en anglais) pour l'algorithme Maxent. Elle proposait également de multiples méthodes pour partitionner les données d'occurrence, ainsi que des mesures de performance variées.
- 2) Les demandes des utilisateurs, les récentes avancées de cette discipline, et les besoins de compatibilité avec d'autres logiciels, ont conduit à une restructuration et une extension majeure de la librairie ENMeval. En outre, nous avons effectué une revue de littérature afin d'examiner ses tendances d'utilisation (2015–2019).
- 3) ENMeval 2.0 présente donc une nouvelle structure orientée objet pour incorporer d'autres algorithmes, permettre l'optimisation des paramètres des algorithmes et des mesures de performance, générer des métadonnées approfondies, intégrer une approche par modèle nul afin de quantifier la significativité statistique et la taille d'effet, ainsi que de nouvelles composantes pour élargir les possibilités d'analyse et de visualisation. Dans notre revue de littérature, nous avons relevé que les mesures de performance et de paramétrage des modèles sont insuffisamment rapportées, que l'AICc est massivement utilisée pour la sélection des modèles, et que la validation croisée spatialisée est peu utilisée; nous expliquons comment ENMeval 2.0 peut contribuer à résoudre ces problèmes.
- 4) Cette version restructurée et élargie peut favoriser des progrès dans cette discipline et améliorer l'information disponible pour la prise de décisions.

Mots-clés: ajustement des paramètres des modèles, évaluation des modèles, logiciel, métadonnées, modélisation de distribution des espèces, modélisation de la niche écologique, validation croisée spatialisée

Introduction

Models that estimate species' niches and geographic distributions (hereafter ecological niche models, or ENMs) based on occurrence data and environmental predictor variables are frequently employed to test biogeographical hypotheses and inform management and conservation (Peterson et al. 2011, Guisan et al. 2017). 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 Group 2020) to provide automated model evaluation and tuning for ENMs (using the algorithm `Maxent`; Phillips et al., 2006, Phillips et al., 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 (Fig. 1). Below, we describe the main features of `ENMeval` 2.0 (current version 1.9.0 [will be updated to 2.0 with acceptance]: <https://github.com/jamiemkass/ENMeval/>). Other important changes and enhancements are outlined in Table 1 and Fig. 1, as well as discussed in detail in Appendix 2, and highlighted through worked examples in an updated vignette. 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.

Key new features of `ENMeval` 2.0

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.

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 catalog 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 example:

<https://github.com/jamiemkass/ENMeval/blob/master/vignettes/ENMeval-2.0.0-Vignette.html.zip>).

Metadata objects can be saved as text files for sharing or use in supplemental materials.

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).

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: (1) if ENMeval was used for analysis or only cited for concepts, (2) which functionalities were used, (3) which method was used to partition data, (4) which performance metric(s) were used and reported (studies may report other metrics in addition to those used to select models), (5) whether the selected Maxent settings (i.e., regularization multiplier and feature class combinations) were indicated, and (6) 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 (Fig. 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) (Fig. 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 (Fig. 2b). A smaller 57% of the 141 provided the selected model settings, and only 24% characterized variation of model performance across the settings used.

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.

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 key aspects of uncertainty (as called for by Zurell et al. 2020 and Araújo et al. 2019) 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.

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.

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 et al. 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 artifactually 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).

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 innovation 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 apply emerging standards for reporting and assessing such models. In sum, we look forward to continued reciprocal advances in both methods and community standards.

Acknowledgments

JMK 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; RM by grant 2019-03758 from the Swedish Research Council (Vetenskapsrådet); and RAB by the Ford Foundation (Predoctoral Fellowship). JMK, GEPB, and RPA acknowledge support from the U.S. National Science Foundation (NSF; DBI-1661510) and National Aeronautics and Space Administration (80NSSC18K0406 to Mary E. Blair). We also thank Steven J. Phillips for development of the clamping function; the Maxent and *Wallace* Google Group users for invaluable input; and Mary E. Blair, the Anderson Lab group, and two anonymous reviewers for constructive comments on this manuscript. Finally, we thank Olivier Broennimann, Angela P. Cuervo-Robayo, Claudio Mota-Vargas, Octavio R. Rojas-Soto, Nao Takashina, and Mari Takenouchi for help with translating the abstract and title to French, Japanese, and Spanish.

Data availability

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://www.doi.org/10.5281/zenodo.4724213>; Kass et al. 2021). The vignette is available on Github (<https://github.com/jamiemkass/ENMeval/blob/master/vignettes/ENMeval-2.0.0-Vignette.html.zip>). 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>).

Authors' contributions

JMK envisioned the redesign and content of the updated package, in consultation with RPA, RM, CLB, and GEP-B. JMK programmed the package with input from GEP-B, RM, and CLB. JMK and PJG wrote the vignette with input from RPA and GEP-B. JMK, RAB, GEP-B, CLB, and MS-G conducted package testing, including running unit tests and comparing with prior worked analyses. RM designed the literature review with feedback from all authors and conducted it with JMK, PJG, GEP-B, and MS-G. JMK drafted the manuscript with major contributions by RPA, MS-G, CLB, and GEP-B. All authors contributed to and approved the final manuscript.

References

1. Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., & O'Hara, R. B. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5: eaat4858.
2. 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.
3. 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.
4. Buisson, L., Thuiller, W., Casajus, N., Lek, S., & Grenouillet, G. (2010). Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, 16: 1145–1157.
5. 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.
6. Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1: 330-342.
7. 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.
8. 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.
9. 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.
10. Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: with applications in R*. Cambridge University Press.
11. 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.

- Accepted Article
12. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY.
 13. Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2020). *dismo: Species Distribution Modeling*. R package version 1.3-3. <https://CRAN.R-project.org/package=dismo>
 14. 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.
 15. 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.
 16. 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.
 17. 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. <http://doi.org/10.5281/zenodo.4724213>
 18. 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.
 19. 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.
 20. 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.

21. 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.
22. Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190: 231–259.
23. 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.
24. Radosavljevic, A. & Anderson, R. P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, 41: 629–643.
25. Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guisera-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.
26. 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.
27. Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guisera-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.
28. 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.
29. Veloz, S. D. (2009). Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, 36: 2290–2299.
30. 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.

31. Wickham H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
32. 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.
33. Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillerá-Arroita, G., Guisan, A. & Lahoz-Monfort, J. J. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43: 1261–1277.

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 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 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.

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 tan, 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 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).

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.



